Inverse Problems **15** (1999) 1439–1454. Printed in the UK PII: S0266-5611(99)01573-7

A Bayesian test for the appropriateness of a model in the biomagnetic inverse problem

R Hasson† and S J Swithenby‡

† Applied Mathematics Department, The Open University, Milton Keynes MK7 6AA, UK
‡ Physics Department, The Open University, Milton Keynes MK7 6AA, UK

Received 4 February 1999, in final form 5 July 1999

Abstract. This paper extends previous work on the Bayesian foundations of the biomagnetic inverse problem. It derives the *a posteriori* source current probability distribution given a prior source current probability distribution, a source space weight function and a data set. This calculation enables the construction of a Bayesian test for the appropriateness of any *a priori* choice of source distribution including the optimal distribution associated with any specific model. In this way the adequacy of a model may be tested.

The procedure is as follows. A model for the sources is chosen, e.g. a single equivalent current dipole. The method then produces a map in source space of the probability that the data imply a significant difference between the real distribution and that associated with the model. The procedure is illustrated using both simulated data generated by a multi-dipolar source set and the results of a study of early latency processing of images of human faces.

1. Introduction

The magnetoencephalographic (MEG) inverse problem (a special case of the biomagnetic inverse problem) is the process of identifying the source current distribution inside the brain that gives rise to a given set of magnetic field measurements outside the head. The problem is difficult because the detectors are limited in number and are sensitive to widespread source currents, and because of the existence of silent and near-silent sources. 'Silent sources' are configurations of current density inside the brain which give zero magnetic field outside the head (e.g. all radial sources are silent when the head is modelled as a conducting sphere). It follows that the general biomagnetic inverse problem is ill-posed and under-determined.

The most common way of reducing the problem and rendering it tractable is by characterizing the source in terms of a limited number of effective current dipoles. Such source descriptions provide links with the dominant functional architecture model of the brain in which processing is described in terms of localized activity with interactions between essentially separate regions. Multiple dipole models have enjoyed considerable success in the analysis of sensory and motor cortex (e.g. [1–4]).

Growing evidence for the existence of more diffuse brain networks have led to an interest in distributed source algorithms. Several have been proposed [5–10]. These algorithms have been designed to cope with the non-uniqueness of the problem, primarily by restriction of the source space and by regularization. Each algorithm leads to a unique solution (from the infinite number available) through its particular choice of source basis, weight functions, noise assumptions, and, in many cases, cost function. There has been an extended debate about the accuracy and value of these methods. This proceeds at two levels; the technical ability

0266-5611/99/061439+16$30.00 © 1999 IOP Publishing Ltd

of the various algorithms to recover a simulated source distribution (often quoted in terms of one or more source current dipoles), and the electrophysiological appropriateness of the type of source structure favoured by particular algorithm parameters. So, for example, the simple minimum norm solution [6] which tends to produce a grossly smeared and superficial source distribution may be compared with the LORETA solution [9] which favours smooth but regionally confined current distributions. The issues have been fully debated in recent conferences [11, 12].

The many-to-one nature of the mapping of sources to magnetic fields suggests that a probabilistic approach to reconstructing the sources from the magnetic field could be used [13, 14]. A Bayesian probabilistic approach for reconstructing distributed sources was developed formally by Clarke [15, 16]. More recently, Baillet *et al* have described an iterative approach which combines both spatial and temporal constraints within a unified Bayesian framework designed to allow the estimation of source activity that varies rapidly with position, e.g. across a sulcus [17]. Schmidt *et al* have developed a probabilistic algorithm in which a bridge is made between distributed and local source solutions through the use of a regional descriptor within the source representation [18]. In this case, a Monte Carlo method is used in the absence of an analytic solution for the expectation value of the source current.

Here, we are proposing an alternative Bayesian approach. It includes the explicit assumption of both a prior source current probability distribution and a source space weight function, and allows the calculation of the *a posteriori* source current probability distribution. The derivation is detailed in section 3. The inclusion of the prior probability provides a means by which the adequacy of a prior source distribution can be tested against the data to reveal those parts of the source distribution that are statistically valid. If this prior source distribution is chosen to be the optimum solution with a given source model (e.g. a single equivalent current dipole), then the method allows the testing of the appropriateness of that model. A straightforward extension of this idea is the direct comparison of two data sets to reveal where, within a given model, there are significant differences in their associated source distributions.

In section 4, both simulated and real data are used to illustrate various uses of the technique. We show how the method may be used to identify the number of dipole sources needed to adequately model a data set. This is done in a simulation where the sources are localized and known and in an analysis of real data where the source distribution is a matter of current debate. In addition, we demonstrate the method's applicability to the identification of significant differences in the brain's responses to different experimental conditions. Such analysis underpins a great deal of neuroscience research.

2. Specification of the problem

The arrangement of sources and detectors for the biomagnetic inverse problem is shown in figure 1. The sources giving rise to the measurements are assumed to be restricted within a source space Q, which may be smaller than the whole head volume (e.g. if the sources are assumed to be cortical). The current density within the volume Q is assumed to belong to the space of square-integrable vector fields on Q, which we call J.

The measurement process typically gives successive sets of data ($\sim$100 channels) every millisecond. In this paper the data for each time instant are processed independently, and the data from a single time instant are collected into a vector $m \in \mathbb{R}^N$. If $\vec{j}(\vec{r}) \in J$ then the measurement process can be represented by a functional $z : J \to \mathbb{R}^N$. A subscript notation will be used to identify the sensor, i.e. z_i is the ideal reading from the ith sensor. So the basic

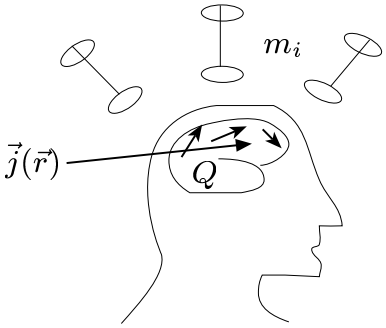


Figure 1. A schematic experimental geometry. The diagram shows a current distribution $\vec{j}(\vec{r})$ confined within the source space Q (a subregion of the brain volume). Also shown are stylized measurement channels which give a data set $\{m_i : i = 1, \ldots, N\}$. The measurement channels shown represent first-order axial gradiometers which consist of two counter-wound sensing coils. These are linked by a flux transformer to a SQUID detector.

equation is

$$m_i = z_i(\vec{j}) + e_i \tag{1}$$

where the e_i are the measurement errors, which are assumed to be normally distributed with zero mean and covariance matrix $\alpha^2 D$ (where α is the standard deviation of the errors and D is a symmetric, positive-definite matrix).

To compute the functional z on a computer (i.e. to solve the forward problem) requires a volume conductor model of the head. In this paper the precise model used is irrelevant, so the final results will be written in terms of z. This is done via the Green functions $\vec{L}_i$ for the problem, which are defined by

$$z_i(\vec{j}) = \int_Q \vec{L}_i(\vec{r}) \cdot \vec{j}(\vec{r})\, \mathrm{d}\vec{r}. \tag{2}$$

Stated simply, the inverse problem is to estimate $\vec{j}(\vec{r})$ given the data vector m. The given data are not enough to determine $\vec{j}(\vec{r})$ uniquely for several reasons. Firstly, the data are limited in number (a typical instrument has $\sim$100 channels). Secondly, the data are noisy (there are several noise sources, e.g. thermal noise in the SQUIDs themselves and external magnetic fields). Thirdly, and more crucially, even if the magnetic field was measured with zero noise on a surface completely surrounding the head then $\vec{j}(\vec{r})$ would still not be uniquely determined [19–21]. This is most easily explained in terms of 'silent sources' (i.e. source distributions which give zero magnetic field outside the head). In the special case of a homogeneous spherical conductor, Fokas *et al* [20] derive a unique decomposition of the source current distribution into three contributions, one of which is uniquely determined whilst the other two are arbitrary (and so characterize the silent sources).

The approach adopted here starts from the same point as used in Clarke [16], a statement of Bayes's theorem:

$$\mathcal{P}(\vec{j} \in A | m \in B) = \frac{\mathcal{P}(m \in B | \vec{j} \in A)\mathcal{P}(\vec{j} \in A)}{\mathcal{P}(m \in B)} \tag{3}$$

where A is a set of currents and B is a set of measurements. This equation reads, the *a posteriori* probability of a current set A after the measurement B is proportional to the probability of producing the measurement B given that the current is in the set A times the *a priori* probability of the current set A. ($\mathcal{P}(m \in B)$ is a constant for any measurement set B.)

In this paper we will work with the probability density functions of our distributions and will simplify the analysis by shrinking the measurement set B to a single point $\{m\}$. This ignores the finiteness of the precision of the measurements. Equation (3) then becomes

$$\rho_m(\vec{j}) \propto \rho(\vec{j}) \times \epsilon(m - z(\vec{j})) \tag{4}$$

where ρ is the *a priori* distribution, ρ_m is the *a posteriori* distribution and ϵ is the error distribution. Throughout the paper, probability density functions will only be determined up to a constant. The constant of proportionality is found by requiring that the probability is normalized to one. In this paper both ρ and ϵ are assumed to be Gaussian and then ρ_m is calculated to be Gaussian. The advantage of assuming that both ρ and ϵ are Gaussian is that an explicit expression for the *a posteriori* probability density function is derived which avoids the use of Monte Carlo techniques to evaluate the integrals involved. This ensures that the method is quick to compute and makes it suitable for MEG analysis where large numbers of data are usually involved.

An error probability density function ϵ consistent with the Gaussian assumption may be written as

$$\epsilon(e) \propto \exp\left\{-\frac{1}{2\alpha^2} e^{\mathrm{T}} D^{-1} e\right\}. \tag{5}$$

This generally valid expression will be retained throughout the derivation in this paper. In practice, the noise covariance matrix $\alpha^2 D$ may be difficult to estimate and, for simplicity, the simple form $D = I$ will be used in the later illustrations.

3. Derivation

In this section the *a posteriori* probability distribution is derived under specific assumptions. An explicit formula for a Bayes estimator of the expectation of the *a posteriori* probability distribution is derived. Also derived is an explicit formula for a Bayes estimator of the variance of the *a posteriori* probability distribution which can be thought of as assigning a variance at each point in source space Q and each direction. The starting point for our method is to define an inner product on J. Several choices can be made, but in this paper we use

$$\langle \vec{j}_1, \vec{j}_2 \rangle = \int_Q \frac{\vec{j}_1(\vec{r}) \cdot \vec{j}_2(\vec{r})}{\omega(\vec{r})} \, \mathrm{d}\vec{r} \tag{6}$$

where $\omega(\vec{r})$ is a weighting distribution defined on the source space Q. This provides a method of inputting prior information of the location of sources (e.g., gained from MRI images) into the algorithm.

Clarke [16] assumed that the prior current density was identically zero. Here that restriction is avoided and an arbitrary prior current $\vec{j}^{\,\mathrm{prior}}$ will be introduced as a parameter of the method. The *a priori* probability distribution on J is defined using $\vec{j}^{\,\mathrm{prior}}$ and the inner product

$$\rho(\vec{j}) \propto \exp\left\{-\frac{1}{2\beta^2} \langle \vec{j} - \vec{j}^{\,\mathrm{prior}}, \vec{j} - \vec{j}^{\,\mathrm{prior}} \rangle\right\} \tag{7}$$

where β is the assumed *a priori* standard deviation.

To proceed further a basis is needed for J. A 'natural' choice is a basis that is related to the measurement functional z. So an obvious candidate is a basis derived from the adjoint map to the measurement map from J to $\mathbb{R}^N$. This gives a map $\vec{z} : \mathbb{R}^N \to J$ (since J is self-dual) defined by

$$\langle \vec{z}(a), \vec{j} \rangle = a^{\mathrm{T}} z(\vec{j}). \tag{8}$$

Explicitly, for our choice of inner product on J, this gives the set of linearly independent distributions $\{\omega(\vec{r})\vec{L}_i(\vec{r})\}$. This set is extended into a basis of J that includes the silent sources by adding vectors $\{\vec{\mathcal{L}}_i\}$ which are chosen to be orthogonal to the $\{\omega\vec{L}_i\}$, i.e. $\langle \omega\vec{L}_i, \vec{\mathcal{L}}_j \rangle = 0 \, \forall i, j$.

Since $\{\omega(\vec{r})\vec{L}_i\} \bigcup \{\vec{\mathcal{L}}_i\}$ is a basis of J a general current density $\vec{j}(\vec{r})$ can be written in terms of this basis as

$$\vec{j}(\vec{r}) = \sum_{i=1}^{N} a_i \omega(\vec{r}) \vec{L}_i(\vec{r}) + \sum_i b_i \vec{\mathcal{L}}_i(\vec{r}). \tag{9}$$

To simplify the notation the components of currents are written in column vector notation:

$$\vec{j} = \begin{pmatrix} a \\ b \end{pmatrix}, \qquad \vec{j}^{\text{prior}} = \begin{pmatrix} a^{\text{prior}} \\ b^{\text{prior}} \end{pmatrix}. \tag{10}$$

A simple computation gives

$$\langle \vec{j} - \vec{j}^{\text{prior}}, \vec{j} - \vec{j}^{\text{prior}} \rangle = \begin{pmatrix} a - a^{\text{prior}} \\ b - b^{\text{prior}} \end{pmatrix}^{\text{T}} \begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix} \begin{pmatrix} a - a^{\text{prior}} \\ b - b^{\text{prior}} \end{pmatrix} \tag{11}$$

where $P_{ij} = \langle \omega \vec{L}_i, \omega \vec{L}_j \rangle$ and $Q_{ij} = \langle \vec{\mathcal{L}}_i, \vec{\mathcal{L}}_j \rangle$ since by construction $\langle \omega \vec{L}_i, \vec{\mathcal{L}}_j \rangle = 0$.

Now the two *a priori* probability density functions (equations (7) and (5)) may be combined with Bayes's theorem (equation (4)) to obtain the *a posteriori* probability density and the task is to manipulate the equation so that it is explicitly in the form of a Gaussian distribution. As a first step the exponentials are combined to give

$$\rho_m(\vec{j}) \propto \exp\left\{ -\frac{1}{2\alpha^2}\left[\begin{pmatrix} a - a^{\text{prior}} \\ b - b^{\text{prior}} \end{pmatrix}^{\text{T}} \begin{pmatrix} \zeta P & 0 \\ 0 & \zeta Q \end{pmatrix} \begin{pmatrix} a - a^{\text{prior}} \\ b - b^{\text{prior}} \end{pmatrix} + (m - Pa)^{\text{T}} D^{-1} (m - Pa) \right] \right\} \tag{12}$$

where we have introduced the notation $\zeta = \alpha^2/\beta^2$ and the $z(\vec{j})$ in equation (4) has been written as Pa. Next, the terms involving operators on a are simplified by completing the square (all constant terms can be absorbed into the normalization constant)

$$\rho_m(\vec{j}) \propto \exp\left\{ -\frac{1}{2\alpha^2} \begin{pmatrix} a - a' \\ b - b^{\text{prior}} \end{pmatrix}^{\text{T}} \begin{pmatrix} PD^{-1}P + \zeta P & 0 \\ 0 & \zeta Q \end{pmatrix} \begin{pmatrix} a - a' \\ b - b^{\text{prior}} \end{pmatrix} \right\} \tag{13}$$

where a' is defined so as to make the linear terms in a agree between equations (12) and (13), i.e.

$$a' = (\zeta D + P)^{-1}(m + \zeta D a^{\text{prior}}). \tag{14}$$

Equation (13) can now be written explicitly in the form of a multivariate Gaussian distribution:

$$\rho_m(\vec{j}) \propto \exp\left\{ -\frac{1}{2\alpha^2} \begin{pmatrix} a - a' \\ b - b^{\text{prior}} \end{pmatrix}^{\text{T}} \begin{pmatrix} (PD^{-1}P + \zeta P)^{-1} & 0 \\ 0 & Q^{-1}/\zeta \end{pmatrix}^{-1} \begin{pmatrix} a - a' \\ b - b^{\text{prior}} \end{pmatrix} \right\}. \tag{15}$$

From equation (15) it can be seen that the mean value of b is b^{prior} which accords with intuition because there is no change from our prior knowledge. This is because by construction all the information that the experiment provides is orthogonal to the $\vec{\mathcal{L}}_i$.

In order to produce images of the *a posteriori* probability distribution, it is necessary to reduce the dimensionality of the solution and compute the spatial variation of a scalar quantity that is a useful and well-defined descriptor of some aspect of the *a posteriori* probability distribution. This scalar statistic may be defined through a 'test current' $\vec{t} = \chi_{V_k}(\vec{r})\hat{e}_\alpha$ where $\chi_{V_k}(\vec{r})$ is the characteristic function of a voxel in the brain and $\hat{e}_\alpha$ is a unit vector. This is another departure from Clarke [16] in which a delta function test current is assumed. His choice would cause problems because the inner product $\langle \vec{t}, \vec{t} \rangle$, which is needed below (see equation (37)), is undefined for a delta function. The scalar statistic, λ, is defined to be the inner product of

current with our 'test current', i.e. $\lambda = \langle \vec{t}, \vec{j} \rangle$. The *a posteriori* probability distribution defined for current densities $\vec{j}$ now induces a corresponding probability distribution on the statistic λ. An image is produced by scanning the voxel V_k over the source space, Q.

The probability distribution of the statistic, λ, will now be determined. First, the coefficients of the basis elements required to construct $\vec{t}$ are identified:

$$\lambda = \langle \vec{t}, \vec{j} \rangle \tag{16}$$

$$= \left\langle \vec{t}, \sum_i a_i \omega(\vec{r}) \vec{L}_i(\vec{r}) + \sum_i b_i \vec{\mathcal{L}}_i(\vec{r}) \right\rangle \tag{17}$$

$$= \sum_i a_i \langle \vec{t}, \omega(\vec{r}) \vec{L}_i(\vec{r}) \rangle + \sum_i b_i \langle \vec{t}, \vec{\mathcal{L}}_i(\vec{r}) \rangle \tag{18}$$

$$= u^{\mathrm{T}} a + v^{\mathrm{T}} b, \text{ say.} \tag{19}$$

Equation (15) is projected onto this particular linear combination of coordinates to find the probability density of λ:

$$\rho_m(\lambda) \propto \exp\left\{ -\frac{1}{2\alpha^2} \frac{(\lambda - u^{\mathrm{T}} a' - v^{\mathrm{T}} b^{\mathrm{prior}})^2}{u^{\mathrm{T}} (P D^{-1} P + \zeta P)^{-1} u + v^{\mathrm{T}} Q^{-1} v / \zeta} \right\}. \tag{20}$$

The mean of λ can be identified from equation (20) by inspection to be $u^{\mathrm{T}} a' + v^{\mathrm{T}} b^{\mathrm{prior}}$. This expression cannot be used directly because the term $v^{\mathrm{T}} b^{\mathrm{prior}}$ cannot be computed directly since the basis fields $\vec{\mathcal{L}}_i$ have not been defined explicitly. This problem may be overcome by expanding $\langle \vec{t}, \vec{j}^{\mathrm{prior}} \rangle$:

$$\langle \vec{t}, \vec{j}^{\mathrm{prior}} \rangle = \left\langle \vec{t}, \sum_i a_i^{\mathrm{prior}} \omega(\vec{r}) \vec{L}_i(\vec{r}) + \sum_i b_i^{\mathrm{prior}} \vec{\mathcal{L}}_i(\vec{r}) \right\rangle \tag{21}$$

$$= \sum_i a_i^{\mathrm{prior}} \langle \vec{t}, \omega(\vec{r}) \vec{L}_i(\vec{r}) \rangle + \sum_i b_i^{\mathrm{prior}} \langle \vec{t}, \vec{\mathcal{L}}_i(\vec{r}) \rangle \tag{22}$$

$$= u^{\mathrm{T}} a^{\mathrm{prior}} + v^{\mathrm{T}} b^{\mathrm{prior}}. \tag{23}$$

The expression for the mean of λ can now be rewritten using only references to known vector fields as

$$\text{mean of } \lambda = u^{\mathrm{T}} a' + \langle \vec{t}, \vec{j}^{\mathrm{prior}} \rangle - u^{\mathrm{T}} a^{\mathrm{prior}} \tag{24}$$

$$= u^{\mathrm{T}} (\zeta D + P)^{-1} (m + \zeta D a^{\mathrm{prior}}) + \langle \vec{t}, \vec{j}^{\mathrm{prior}} \rangle - u^{\mathrm{T}} a^{\mathrm{prior}} \tag{25}$$

$$= u^{\mathrm{T}} (\zeta D + P)^{-1} (m - P a^{\mathrm{prior}}) + \langle \vec{t}, \vec{j}^{\mathrm{prior}} \rangle. \tag{26}$$

This equation explicitly writes the expectation value of the statistic $\lambda = \langle \vec{t}, \vec{j} \rangle$ as a sum of two terms. The second term is the statistic for the prior current, and so the first term can be identified as the correction to the prior suggested by the measurements; i.e., the first term shows the difference between the expectation of $\langle \vec{t}, \vec{j} \rangle$ before and after the experiment was made. This is the first central result of this paper and it is worth stating explicitly:

$$\text{change in expectation as a result of experiment} = u^{\mathrm{T}} (\zeta D + P)^{-1} (m - z(\vec{j}^{\mathrm{prior}})). \tag{27}$$

Using equation (20) it is possible go further than this and determine the statistical significance of the statistic. This is because the variance of the variable λ can also be read off from equation (20) as

$$\text{variance of } \lambda = \alpha^2 \left[u^{\mathrm{T}} (P D^{-1} P + \zeta P)^{-1} u + \frac{1}{\zeta} v^{\mathrm{T}} Q^{-1} v \right]. \tag{28}$$

In order to derive an expression in the form of computable matrices, the term $v^{\mathrm{T}} Q^{-1} v$ must be rewritten. To do so, $\vec{t}$ is written as a linear combination of basis elements:

$$\vec{t}(\vec{r}) = \sum_i x_i \omega(\vec{r}) \vec{L}_i(\vec{r}) + \sum_i y_i \vec{\mathcal{L}}_i(\vec{r}). \tag{29}$$

Of course, x and y are related to u and v. In fact,

$$u_i = \langle \vec{t}, \omega(\vec{r})\vec{L}_i(\vec{r})\rangle \tag{30}$$

$$= \left\langle \sum_j x_j \omega(\vec{r})\vec{L}_j(\vec{r}) + \sum_j y_j \vec{\mathcal{L}}_j(\vec{r}), \omega(\vec{r})\vec{L}_i(\vec{r})\right\rangle \tag{31}$$

$$= \sum_j x_j \langle \omega(\vec{r})\vec{L}_j(\vec{r}), \omega(\vec{r})\vec{L}_i(\vec{r})\rangle \tag{32}$$

$$= \sum_j x_j P_{ij}. \tag{33}$$

Similarly, $v = Qy$. Using these relationships, the inner product of $\vec{t}$ with itself can be computed:

$$\langle \vec{t}, \vec{t}\rangle = \left\langle \sum_i x_i \omega(\vec{r})\vec{L}_i(\vec{r}) + \sum_i y_i \vec{\mathcal{L}}_i(\vec{r}), \sum_j x_j \omega(\vec{r})\vec{L}_j(\vec{r}) + \sum_j y_j \vec{\mathcal{L}}_j(\vec{r})\right\rangle \tag{34}$$

$$= x^{\mathrm{T}} P x + y^{\mathrm{T}} Q y \tag{35}$$

$$= u^{\mathrm{T}} P^{-1} u + v^{\mathrm{T}} Q^{-1} v. \tag{36}$$

Equations (28) and (36) can now be combined to generate the following formula for the variance:

$$\text{variance of } \lambda = \alpha^2 \left[u^{\mathrm{T}} (P D^{-1} P + \zeta P)^{-1} u + \frac{1}{\zeta}(\langle \vec{t}, \vec{t}\rangle - u^{\mathrm{T}} P^{-1} u)\right]. \tag{37}$$

This equation is a generalization of the results of Clarke [16], some consequences of which were explored in [22, 23].

It may be helpful to relate features of this equation to the measurement system. The second term in equation (37) is multiplied by $\alpha^2/\zeta = \beta^2$ and so is independent of the assumed noise levels in the detectors. It represents a variance derived from the finite number of measurements and the geometry of the experiment. Since it is proportional to β^2, it can be interpreted in terms of the truncation error becoming less and less important as the certainty of the prior distribution $\vec{j}^{\text{prior}}$ increases.

The first term in equation (37) is proportional to α^2. It shows how the noise in data is reflected into source space. The unregularized form of this term was derived previously by Ioannides *et al* in [24] using an *ad hoc* argument. Ioannides *et al* obtained the regularized form by replacing occurrences of P^{-1} in the unregularized form by $(P + \zeta I)^{-1}$. In the notation used here this gives a first term as follows:

$$\text{first term} = u^{\mathrm{T}} (P + \zeta I)^{-1} (P + \zeta I)^{-1} u. \tag{38}$$

This does not agree with equation (37) which has the form (in the case of the measurement error being uncorrelated, i.e. $D = I$)

$$\text{first term} = u^{\mathrm{T}} P^{-1} (P + \zeta I)^{-1} u. \tag{39}$$

The behaviour as ζ tends to infinity is that the variance tends to zero. This is reasonable since, for fixed experimental noise levels (i.e. fixed α), ζ tending to infinity corresponds to β tending to zero, which in turn corresponds to greater and greater certainty that the prior is correct. When β is zero the *a priori* current distribution is known with absolute certainty. This is consistent with the above analysis which indicates that, in this case, the *a posteriori* current density is certain to be equal to the *a priori* current density.

Note that, when computing the term $u^{\mathrm{T}} P^{-1} u$ in equation (37), any reasonable algorithm (e.g. Choleski's algorithm) for computing $P^{-1}u$ can be used even though the matrix P is

ill-conditioned. This is because the large residual vector which results in this calculation is annihilated by the inner product with u. So the computation of the whole term is well conditioned.

In order to apply the method a way of determining the parameters appearing in equations (26) and (37), namely α, ζ and D, is required. The covariance matrix D and the parameter α may be estimated from analysis of the instrumental output during a period where there are no active sources. In an evoked response experiment, it is usually assumed that the pre-stimulus interval satisfies this condition. If sources are present, this will be apparent from the covariance matrix and will lead to an overestimate of α and a consequent underestimate of the significance of the result. Alternatively, it is possible to estimate α from measured instrumental noise alone. This tends to underestimate α and overestimate significance.

The parameter ζ appears in equation (26) in exactly the same way that the regularization parameter appears when the Tikhonov regularization technique is applied to this problem. Our proposal is to treat ζ as a regularization parameter and use one of the methods used in Tikhonov regularization to fix ζ. In the illustrations in this paper the L-curve [25] method is used.

4. Applications

The main analytical results of this paper (equations (26), (27) and (37)) provide the means of solving the MEG inverse problem with specific assumptions and of assessing the robustness of the solution. In this section, this approach will be illustrated through three studies: a simulation of a few-dipole source set; an analysis of the appropriate dipole model for data from a real experiment on face processing; and a comparison of responses to different visual stimuli from the same real experiment.

All the illustrations are based on the same experimental arrangement and the same instrument, the Neuromag-122^{TM} [26]. This is a helmet MEG system that contains 61 pairs of first-order gradiometers ($\partial B_r/\partial\theta$, $\partial B_r/\partial\phi$ in spherical polar coordinates), covering the head (figure 2). The outputs of each pair of detectors are closely related to the dominant tangential ionic current flow in the region underlying the relevant sensors. Also shown in figure 2 is the assumed source space, a two-dimensional spherical shell of radius 0.08 m covering a 2 rad by 2 rad solid angle over posterior regions of the cortex.

The source configuration for the first simulation study was three dipoles distributed in an isosceles triangle configuration (coordinates (0, −0.08 m, −0.02 m), (−0.05 m, −0.06 m, 0.02 m) and (0.05 m, −0.06 m, 0.02 m) with orientations (1, 0, 1), (1, 0, −1), (1, 0, 1), respectively). The positions of the dipoles were chosen so as to represent approximately a central primary and two bilateral secondary source areas. The precise locations are not exactly on the source space shell but are displaced by 3, 1, and 1 mm, respectively, from the shell. The reason for this displacement is to simulate typical misalignment errors that occur when obtaining a source space from co-registered MRI images. The dipole locations and activation curves are shown in figure 3. The central source is activated first, followed by synchronous activation of the lateral sources. The forward problem is solved using a homogeneous sphere conductor model centred at the origin. Gaussian noise has been added to the computed dipole signal so that the integrated noise power is equal to 50% of the integrated signal power. Examples of simulated data are shown in figures 3(*c*)–(*e*).

The simplest approach to the inverse problem is to use equation (26) with a zero prior current distribution. This simplification results in the same formulae as the probabilistic algorithm that has been used for several years [15, 27]. The resulting expectation of the *a posteriori* current distribution is shown in figure 4(*a*). In this illustration we have used a value

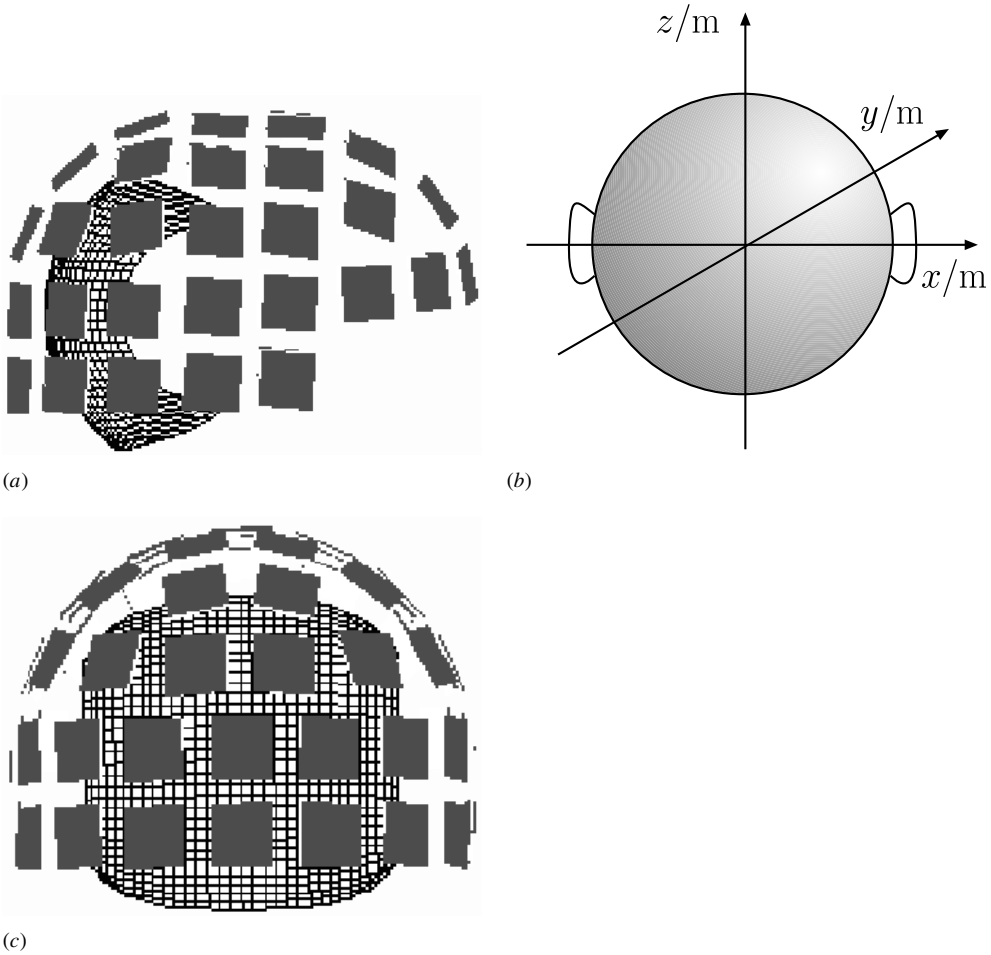


Figure 2. The experimental geometry. (*a*), (*c*) The 61 measurement sites in the helmet arrangement. Each square represents a pair of orthogonal detectors. Also shown is the source space, consisting of a discretized square mesh covering a part of a spherical shell. Note that the elements of the mesh are of equal area. (*b*) The coordinate system with the x-axis along the line joining the pre-auricular points and the y-axis joining the inion and nasion.

of $\alpha = 5.0$ computed directly from the added Gaussian noise. The optimal regularization parameter was determined by the L-curve method [25] to be $51 \times \mathrm{trace}(P)/N$.

However, using our analysis, it is possible to employ an approach which goes further in comparing different source descriptions. Current distributions estimated using the dipole model (either a single or a few dipoles) can be used as prior current distributions in our analysis. It is then possible to identify the appropriateness of the model by computing from equation (27) the change in the expectation associated with including the measurement information without constraining the final solution to a dipolar form. Because the statistic provides spatial information it indicates directly those areas where the dipole model solution has been modified. From the variance associated with the *a posteriori* current (computed from equation (37)) we can plot the significance of the change in expectation at each point in source space.

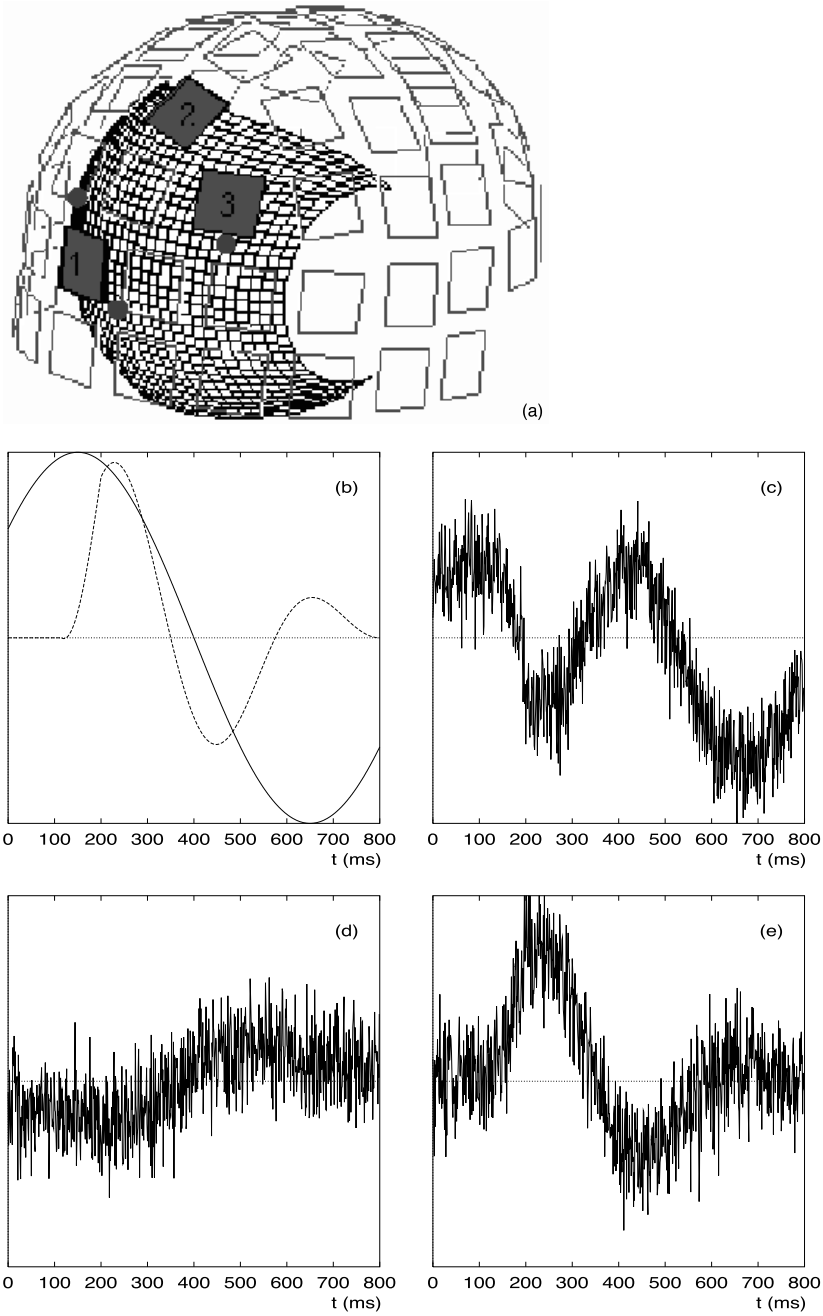


Figure 3. (*a*) An isometric projection of the experimental geometry. The dots on the source space are the feet of the perpendiculars from the three dipole sources used in this simulation. The activation curves (dipole moment versus time in milliseconds) for these three dipoles are shown in (*b*). The longer period curve corresponds to the dominant (central) source. The other curve corresponds to the synchronous and equally activated lateral sources. The highlighted detector sites in (*a*) are those for which data are shown in (*c*) channel 1, (*d*) channel 2 and (*e*) channel 3. The *y* scale is in arbitrary units and is the same for graphs (*c*)–(*e*). Note that it is only in the case of channel 3 that the signal clearly follows the activation curve for the closest dipole.

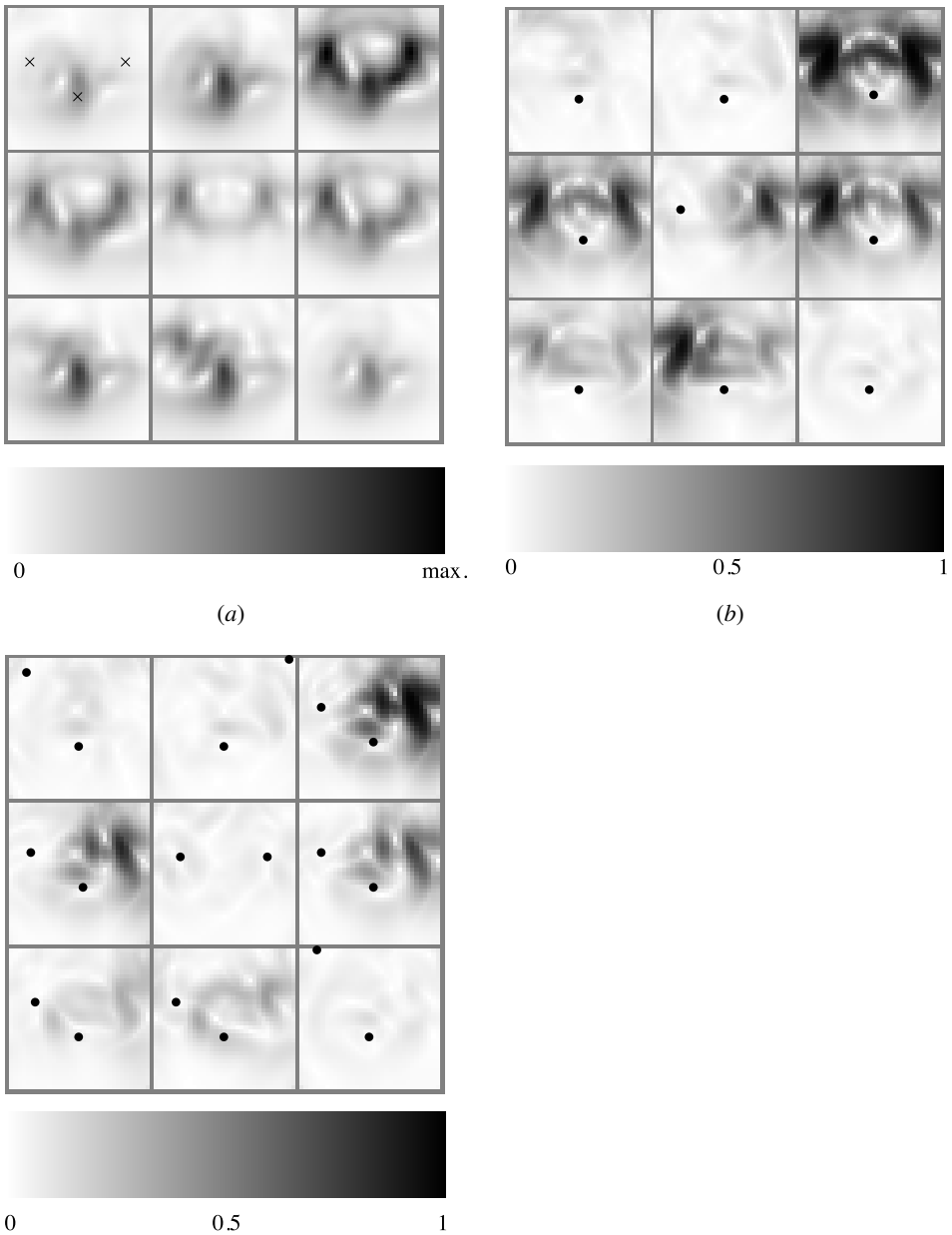


Figure 4. (*a*) The current distribution that is the expectation of the *a posteriori* probability distribution with a zero *a priori* current distribution using a linear greyscale normalized with black equal to the maximum expectation. Each frame is a two-dimensional representation of the shell source space pictured in figure 2. The nine frames are snapshots at time 0, 100, 200 ms, etc (see figure 3). They should be read from left to right and top to bottom. The crosses on the first frame of this sequence show the positions of the three source dipoles in this representation of the source space. (*b*) The significance of the differences between a single moving dipole model and computed data. The black dots denote the projected position of the fitted dipole. (*c*) As (*b*) but with a two moving dipole model.

To illustrate this technique, two prior models of the source current for these data are postulated—a single moving dipole model and a two moving dipole model. The optimal solutions have been found through exhaustive search of the discretized source space by least squares minimization of the fit to the data set. The positions of the fitted dipoles for nine time slices are shown in figure 4(*b*). They may be compared with the nearest points on the source space to the actual dipole positions. Using each model in turn as a prior current and equations (27) and (37), the statistical significance of the differences between the *a posteriori* and the *a priori* current distributions are calculated (figure 4(*b*)).

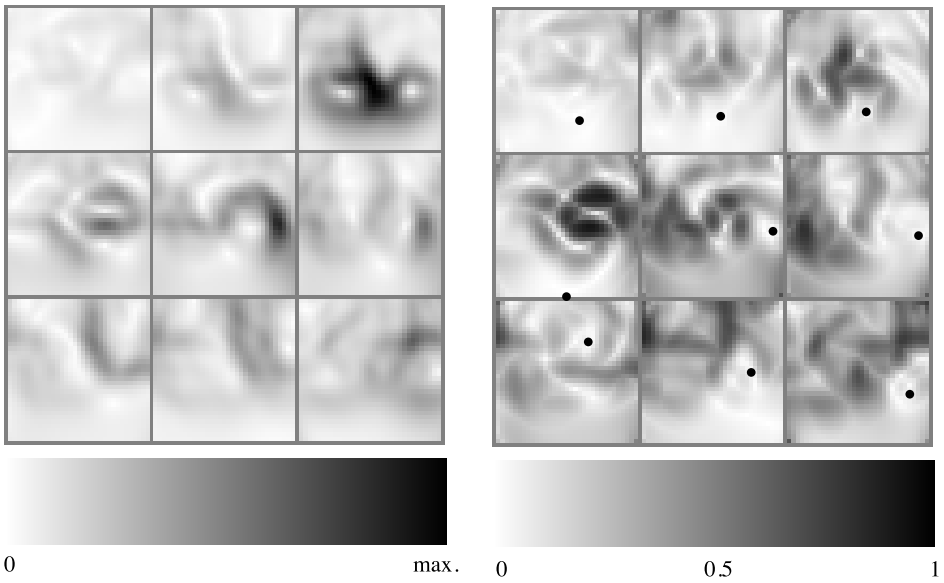
It can be seen that a single dipole is a good model for the first two time instants and also for the ninth. This is not surprising as only one dipole is active at these times. The significance plots for the other times suggest systematic discrepancies between model and data. The restricted localization of these significant differences points to additional localized sources that have been omitted from the model. The next step is the two-dipole model (figure 4(*c*)). The two-dipole significant difference diagram suggests that this model is adequate for all but three time instants. Comparison with the activation curves identifies these as times when all three dipoles are active.

The second illustration uses data from an evoked response study of face-processing using the same experimental system [28]. Human subjects were presented briefly with photographs of human faces and control objects (e.g. animals) and their neural responses were recorded as a function of time after the stimulus. It is known that the early response to face images involves widespread activity in the posterior brain but there is limited evidence for the precise distribution and timecourse of the neuronal sources. One suggestion is that there are three major areas of activity: in the occipital cortex and both the right and left ventral occipito-temporal cortex [29,30]. Strong occipital activity (starting about 100 ms after the stimulus) is expected to lead to concurrent activity in the two other regions with a stronger response in the right hemisphere [28]. The hypothesized arrangement is therefore similar in geometry to the simulated measurement already discussed.

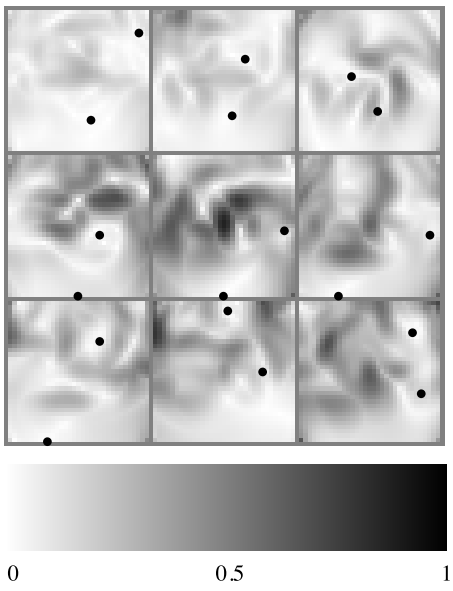
An important added complication in considering this real data is the choice of α. Inspection of the full ensemble of signals across all channels in the pre-stimlus interval reveals a close correspondence with Gaussian assumptions in spite of significant cross-correlation between some of the channels. In subsequent figures we have opted to use the value of $\alpha = 6.4$ given by the standard deviation of the observed signal histogram. Given the significant cross-correlations in the signals the computed significances should be regarded as representing lower bounds. The regularization parameter ζ chosen by using the L-curve method [25] was found to be $20 \times \text{trace}(P)/N$.

Figure 5 shows the same set of outputs that were presented for the simulated system. In this case, the fitted dipoles may be thought of as representing a discrete limited region of source current. Figure 5(*a*) suggests early central activity (frame 3) followed by less prominent localized activity on the right (frame 5). These source regions are reflected in the one-dipole solutions (figure 5(*b*)). However, comparison with figure 4(*b*) shows that the accuracy of the single-dipole fit is less than for the simulated data even though the noise levels are comparable. This may suggest that there are other active sources present. There is no evidence that these are recovered by the two-dipole model (figure 5(*c*)) as there is little improvement in the significant difference maps generated using a two-dipole model as the prior (see for example the strong similarity between frame 5 in figures 5(*b*) and (*c*)). It would be reasonable to infer that the additional sources are diffuse.

Figure 5 shows the same set of outputs that were presented for the simulated system. In this case, a fitted dipole may be thought of as representing a discrete limited region of source current. Figure 5(*a*) suggests early central activity (frame 3) followed by less prominent



(*a*) (*b*)



(*c*)

Figure 5. (*a*) The expectation of the *a posteriori* probability distribution for the face response with a zero *a priori* current distribution. The nine frames represent equal steps in time from 70 to 241 ms after the stimulus. (*b*) The significant differences between a single moving dipole model and the computed data. The black dots denote the projected positions of a single dipole. (*c*) As (*b*) but with a two moving dipole model.

localized activity on the right (frame 5). These source regions are reflected in the one-dipole solutions (figure 5(*b*)) in which the fitted dipoles are placed in a central position in the first three frames and in a right lateral position in frames five and six. Comparison with figure 4(*b*) shows that the accuracy of the single-dipole fit is less than for the simulated data even though the noise levels are comparable. This may suggest that there are other active sources present. However, figure 5(*c*) demonstrates that a two moving dipole model offers only limited advantages. The significance of the differences is reduced only slightly from those in figure 5(*b*), in marked contrast to the reduction seen in figure 4. It would be reasonable to infer that the additional sources are diffuse.

The third illustration relates to the main thrust of the face-processing study by Swithenby *et al* [28], which was to identify statistically significant evidence for differences between the responses to faces and other complex visual stimuli. In the initial analysis the strength of evoked activity within a certain region and latency span was parametrized in terms of the signal power integrated over a group of channels and a specified latency span. These calculations revealed that the brain activity in the right occipito-temporal region following face presentation is significantly different ($p = 0.05$) from activity following non-face images during the latency span 110–170 ms. No other consistent and statistically significant differences were found. This data-space analysis, though useful, was complicated by the need to survey the large number of possible choices of channel group and latency range.

The Bayesian framework developed here provides an alternative direct means of directly comparing responses to two stimuli. One data set is used as the prior current distribution with the hypothesis that there are the same underlying generators for the two stimuli. Comparing the prior with the other data set it is possible to identify those regions in source space where there is a statistically significant difference between the two source structures. We have carried out this calculation for the face and control stimuli as a function of position and latency for a simple two-dimensional source space consisting of a part spherical shell whose radius is similar to that of the cortical surface (figure 6).

The evidence for statistically significant differences in the right occipito-temporal region at about 155 ms is clear. However, there is no evidence for differences in earlier latencies, in particular with respect to the early source shown in figure 5(*a*).

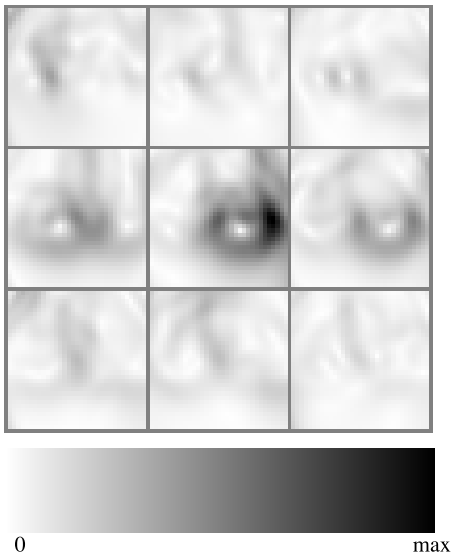


Figure 6. The current density which is the expectation of the *a posteriori* probability distribution for the face response with the *a priori* current distribution derived from the control experiment with pictures of animals. The nine frames represent equal steps in time from 70 to 241 ms after the stimulus.

5. Discussion

The illustrations provided above offer ways of exploiting the new Bayesian results that have been derived in this paper. Dipole predictions have been systematically examined using a measure that goes beyond a simple scalar goodness of fit measure (i.e. percentage of variance accounted for) to a statistically valid map of fitting significance. This addresses a long-standing issue: the unreliability of the goodness of fit measure as a reliable test of the appropriateness of a given model in explaining a given set of data [31]. The spatially discriminated Bayesian approach gives a test that is more reliable when there is fundamental concern about the appropriateness of a model. The second example shows how these ideas may be applied to real data to assess the complexity of the dipole model that a given data set can sustain. In similar fashion, the third example illustrates how the Bayesian framework allows the direct comparison of two data sets in order to identify quantitatively the regions of statistically significant source activities.

These ideas may be extended further within MEG (and EEG) data analysis. An obvious extension is to perform dipole analysis as a precursor to a distributed Bayesian analysis. This may provide a way of not only refining information about the depth of source activity but also of assessing the reliability of depth estimates. Other possibilities include exploration of the dynamics. For example, times at which there are statistically significant changes in the data could be identified by using a source distribution estimated from each time as the prior for an analysis of the data from the next sampling instant.

In summary, the analysis presented here comprises a Bayesian estimator of a source current distribution in the biomagnetic inverse problem. This is generalizable to other systems and may be used as the engine for tests of significant difference in source models and data.

References

[1] Vanni S, Rockstroh B and Hari R 1996 Cortical sources of human short-latency somatosensory evoked fields to median and ulnar nerve stimuli *Brain Res.* **737** 25–33

[2] Buchner H, Adams L, Müller A, Ludwig I, Knepper A, Thron A, Niemann K and Scherg M 1995 Somatotopy of human hand somatosensory-evoked potentials and 3D-NMR tomography *Evoked Potentials—Electroencephalogr. Clin. Neurophysiol.* **96** 121–34

[3] Mauguière F, Merlet I, Forss N, Vanni S, Jousmäki V, Adeleine P and Hari R 1997 Activation of a distributed somatosensory cortical network in the human brain: a dipole modelling study of magnetic fields evoked by median nerve stimulation: 1. Location and activation timing of SEF sources *Evoked Potentials—Electroencephalogr. Clin. Neurophysiol.* **104** 281–9

[4] Hoshiyama M, Kakigi R, Koyama S, Watanabe S and Shimojo 1997 Activity in posterior parietal cortex following somatosensory stimulation in man: magnetoencephalographic study using spatio-temporal source analysis *Brain Topography* **10** 23–30

[5] Hämäläinen M S and Ilmoniemi R J 1984 Interpreting measured magnetic fields of the brain: estimates of current distributions *Technical Report* TKK-F-A559, Helsinki University of Technology

[6] Hämäläinen M S and Ilmoniemi R J 1994 Interpreting magnetic fields of the brain: minimum norm estimates *Med. Biol. Engng Comput.* **32** 35–42

[7] Ioannides A A, Bolton J P R, Hasson R and Clarke C J S 1989 Localized and distributed source solutions for the biomagnetic inverse problem II *Advances in Biomagnetism* ed S J Williamson *et al* (New York: Plenum) pp 591–5

[8] Wang J Z, Williamson S J and Kaufman L 1992 Magnetic source images determined by a lead-field analysis—the unique minimum-norm least-squares estimation *IEEE Trans. Biomed. Engng* **39** 665–75

[9] Pascual-Marqui R D 1994 Low resolution brain electromagnetic tomography *Brain Topography* **7** 180

[10] Gorodnitsky I F, George J S and Rao B D 1995 Neuromagnetic source imaging with focuss: a recursive weighted minimum norm algorithm *Electroencephalogr. Clin. Neurophysiol.* **95** 231–51

[11] Skrandies W (ed) 1995 Extended discussion of LORETA *International Society for Brain Electromagnetic Topography Newsletter* 6, ISSN 0947-5133

[12] Wood C (ed) 1999 Workshop on the MEG inverse problem *Advances in Biomagnetism Research: Biomag96* ed C Aine *et al* (New York: Springer) at press
[13] Hari R and Ilmoniemi R J 1986 Cerebral magnetic-fields 1986 *CRC Crit. Rev. Biomed. Engng* **14** 93–126
[14] Hämäläinen M S, Haario H and Lehtinen M S 1988 Inferences about sources of neuromagnetic fields using Bayesian parameter estimation *Preprint* TKK-F-A620, Helsinki University of Technology
[15] Clarke C J S 1991 Probabilistic modelling of continuous current sources *Biomagnetic Localization and 3D Modelling* ed J Nenonen, H-M Rajala and T Katila (Helsinki: Helsinki University of Technology, Department of Technical Physics) pp 117–25
[16] Clarke C J S 1994 Error estimates in the biomagnetic inverse problem *Inverse Problems* **10** 77–86
[17] Baillet S and Garnero L 1997 A Bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem *IEEE Trans. Biomed. Engng* **44** 374–85
[18] Schmidt D M, George J S and Wood C C 1999 Bayesian inference applied to the electromagnetic inverse problem *Human Brain Mapping* **7** 195–212
[19] Helmholtz H 1853 Ueber einige Gesetze der Vertheilung elektrischer Strome in korperlichen Leiten, mit Anwendung auf die thierischen elektrischen Versuche *Ann. Phys. Chem.* **89** 211–33
Helmholtz H 1853 Ueber einige Gesetze der Vertheilung elektrischer Strome in korperlichen Leiten, mit Anwendung auf die thierischen elektrischen Versuche *Ann. Phys. Chem.* **89** 353–77
[20] Fokas A S, Gelfand I M and Kurylev Y 1996 Inversion method for magnetoencephalography *Inverse Problems* **12** L9–11
[21] He S and Romanov V G 1998 Identification of dipole sources in a bounded domain for Maxwell's equations *Wave Motion* **28** 25–40
[22] Hasson R and Swithenby S J 1993 A quantitative analysis of the EEG and MEG inverse problems *Biomagnetism: Fundamental Research and Clinical Applications* ed C Baumgartner *et al* (Amsterdam: Elsevier) pp 455–7
[23] Hasson R and Swithenby S J 1999 Aspects of non-uniqueness: how one source affects the reconstruction of another *Advances in Biomagnetism Research: Biomag96* ed C Aine *et al* (New York: Springer) at press
[24] Ioannides A A, Bolton J P R and Clarke C J S 1990 Continuous probabilistic solutions to the biomagnetic inverse problem *Inverse Problems* **6** 523–42
[25] Hansen P C 1994 Regularization tools *Numer. Algorithms* **6** 1–35
[26] Knuutila J E T, Ahonen A I, Hämäläinen M S, Kajola M J, Laine P P, Lounasmaa O V, Parkkonen L T, Simola J T A and Tesche C D 1993 A 122-channel whole cortex squid system for measuring the brains magnetic-fields *IEEE Trans. Magn.* **29** 3315–20
[27] Clarke C J S and Janday B S 1989 The solution of the biomagnetic inverse problem by maximum statistical entropy *Inverse Problems* **5** 483–500
[28] Swithenby S J, Bailey A J, Bräutigam S, Josephs O E, Jousmäki V and Tesche C D 1998 Neural processing of human faces: a magnetoencephalographic study *Exp. Brain Res.* **118** 501–10
[29] Lu S T, Hämäläinen M S, Hari R, Ilmoniemi R J, Lounasmaa O V, Sams M and Vilkman V 1991 Seeing faces activates 3 separate areas outside the occipital visual cortex in man *Neurosci.* **43** 287–90
[30] Halgren E, Raij T, Marinkovic K, Jousmäki V and Hari R 1995 Magnetic fields evoked by faces in the human brain: 1. Topography and equivalent dipole locations *Soc. Neurosci.* **21** 562
[31] Janday B S and Swithenby S J 1989 The use of the symmetric sphere model in magnetoencephalographic analysis *Advances in Biomagnetism Functional Localization: A Challenge for Biomagnetism* ed S N Erne and G L Romani (London: World Scientific) pp 153–60