

# Bayesian M/EEG source reconstruction with spatio-temporal priors

Nelson J. Trujillo-Barreto,<sup>a,\*</sup> Eduardo Aubert-Vázquez,<sup>a</sup> and William D. Penny<sup>b</sup>

<sup>a</sup>Brain Dynamics Department, Cuban Neuroscience Centre, P.O. Box 6412/6414, Ave. 25, Esq. 158, No. 15202, Cubanacán, Playa, Havana, Cuba

<sup>b</sup>Wellcome Department of Imaging Neuroscience, UCL, London, UK

Received 9 March 2007; revised 30 June 2007; accepted 27 July 2007

Available online 22 August 2007

This article proposes a Bayesian spatio-temporal model for source reconstruction of M/EEG data. The usual two-level probabilistic model implicit in most distributed source solutions is extended by adding a third level which describes the temporal evolution of neuronal current sources using time-domain General Linear Models (GLMs). These comprise a set of temporal basis functions which are used to describe event-related M/EEG responses. This places M/EEG analysis in a statistical framework that is very similar to that used for PET and fMRI. The experimental design can be coded in a design matrix, effects of interest characterized using contrasts and inferences made using posterior probability maps. Importantly, as is the case for single-subject fMRI analysis, trials are treated as fixed effects and the approach takes into account between-trial variance, allowing valid inferences to be made on single-subject data. The proposed probabilistic model is efficiently inverted by using the Variational Bayes framework under a convenient mean-field approximation (VB-GLM). The new method is tested with biophysically realistic simulated data and the results are compared to those obtained with traditional spatial approaches like the popular Low Resolution Electromagnetic Tomography (LORETA) and minimum variance Beamformer. Finally, the VB-GLM approach is used to analyze an EEG data set from a face processing experiment.

© 2007 Elsevier Inc. All rights reserved.

**Keywords:** M/EEG source localization; Spatio-temporal priors; GLM; Bayesian models; Variational Bayes; Ensemble learning

## Introduction

This article describes a model-based spatio-temporal deconvolution method for M/EEG source reconstruction. The underlying forward or “generative” model incorporates two mappings. The first specifies a time-domain General Linear Model (GLM) at each point in source space. This relates effects of interest at each generator to source activity at that generator. This is identical to the “mass-univariate” approach that is widely used in the analysis of fMRI

(Frackowiak et al., 2003). Additionally, effects of interest are constrained to be similar at nearby generators through use of a spatial prior. The second mapping relates source activity to sensor activity at each time point using the usual spatial-domain lead-field matrix.

There are two potential benefits of the approach. First, as we will show, the use of temporal (as well as spatial) priors can result in more accurate source reconstructions. This may allow signals to be found that cannot otherwise be detected. Second, it provides an analysis framework for M/EEG that is very similar to that used in PET and fMRI. The experimental design can be coded in a design matrix, the model fitted to data, and various effects of interest can be characterized using “contrasts” (Frackowiak et al., 2003). These effects can then be tested for statistically using posterior probability maps (PPMs), as described in Friston and Penny (2003). Importantly, the model does not need to be refitted to test for multiple experimental effects that are potentially present in any single data set. Source parameters are estimated once only using a spatio-temporal deconvolution rather than separately for each temporal component of interest.

The new method is to be contrasted with approaches which follow a single-pass serial processing strategy in which either (i) spatial processing first proceeds to create estimates at each source location and then temporal models are applied at these “virtual depth electrodes” (Darvas et al., 2004; Kiebel and Friston, 2004; Brookes et al., 2004), or (ii) time-series methods are applied in sensor space to identify components of interest using, e.g., time windowing (Rugg and Coles, 1995) or time–frequency estimation (Durka et al., 2005), and source reconstructions are then based on these components. The algorithm we propose comprises a multiple-pass strategy in which temporal and spatial parameter estimates are improved iteratively to provide an optimized and mutually constrained solution.

The new algorithm is similar to existing distributed source solutions in employing spatial priors but differs from the standard generative models implicit in source reconstruction by having an additional level that embodies temporal priors. The spatial prior we use is the spatial Laplacian employed in, for example, Low Resolution Electromagnetic Tomography (LORETA) (Pascual-Marqui et al., 1994). This uses an  $L^2$ -norm, which embodies a belief that sources are diffuse and highly distributed. These are to be

\* Corresponding author. Fax: +53 7 208 6707.

E-mail address: trujillo@cneuro.edu.cu (N.J. Trujillo-Barreto).

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

contrasted with priors based on  $L^1$ -norms (Fuchs et al., 1999),  $L^p$ -norms (Auranen et al., 2005), Variable Resolution Electromagnetic Tomography (VARETA) (Valdés-Sosa et al., 2000), models with multiple priors (Mattout et al., 2006) or models employing Bayesian Model Averaging (BMA) (Trujillo-Barreto et al., 2004), which can accommodate more focal sources. In this paper, we use a single Laplacian spatial prior, as it is the simplest available and because we want to focus on the benefit of using temporal priors in addition to spatial priors. In the future, we envisage augmenting the approach with more flexible spatial priors.

The use of both spatial and temporal constraints is not unique within the source reconstruction community. Indeed, there have been a number of approaches that also make use of temporal priors. Baillet and Garnero (1997), in addition to considering edge-preserving spatial priors, have proposed temporal priors that penalize quadratic differences between neighboring time points. Schmidt et al. (2000) have extended their dipole-like modelling approach using a temporal correlation prior which encourages activity at neighboring latencies to be correlated. Similarly, Daunizeau et al. (2005) propose magnitude priors and temporal smoothness priors based on second derivatives. Galka et al. (2004) have proposed a spatio-temporal Kalman filtering approach which is implemented using linear autoregressive models with neighborhood relations. This work has been extended by Yamashita et al. (2004), who have developed a “Dynamic LORETA” algorithm in which the Kalman filtering step is approximated using a recursive penalized least squares solution. The algorithm is, however, computationally costly, taking several hours to estimate sources in even low-resolution source spaces. Compared to these approaches, our algorithm perhaps embodies stronger dynamic constraints. However, the computational simplicity of fitting GLMs, allied to the efficiency of our inference procedure, results in a relatively fast algorithm. Moreover, the GLM can accommodate damped sinusoidal and wavelet approaches that are ideal for modelling the transient and nonstationary responses in M/EEG.

The manuscript is organized as follows. In the Theory section, we describe the model and relate it to the existing literature on distributed solutions. The success of the approach rests on our ability to characterize neuronal responses, and task-related differences in them, using GLMs. We describe how this can be implemented for the analysis of evoked responses and show how the model can be inverted to produce source estimates using Variational Bayes (VB). In the Results section, the framework is applied to simulated data and data from an EEG study of face processing.

## Methods

### Notation

Bold and regular lowercase variable names denote vectors and scalars, respectively. Bold uppercase names denote matrices with dimensions denoted by regular uppercase names. By convention, all vectors are assumed to be column vectors, whether it corresponds to a row or a column of a matrix will be denoted by using a dot (“.”) as a subscript indicating the non-singleton dimension. That is,  $\mathbf{x}_i$ . ( $\mathbf{x}_i$ ) is a column vector containing the elements of the  $i$ th row (column) of matrix  $\mathbf{X}$ . In what follows,  $N(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$  denotes a multivariate normal density over  $\mathbf{x}$ , having mean  $\boldsymbol{\mu}$  and covariance  $\boldsymbol{\Sigma}$ . The precision of a Gaussian variate is the inverse (co)variance. A gamma density over the scalar random variable  $x$  is written as  $Ga(x; a, b)$ . We also use  $\|\mathbf{x}\|^2 = \mathbf{x}^T \mathbf{x}$ , denote

the trace operator as  $tr(\mathbf{X})$ , use  $\text{diag}(\mathbf{x})$  to denote a diagonal matrix with diagonal entries given by the vector  $\mathbf{x}$  and the symbol  $\otimes$  for Kronecker’s product.

### Probabilistic generative model

The aim of the M/EEG inverse problem (or source reconstruction) is to estimate the primary current density (PCD)  $\mathbf{J}$  from M/EEG measurements  $\mathbf{Y}$ . If we have  $m=1, \dots, M$  sensors,  $g=1, \dots, G$  generators,  $r=1, \dots, R$  trials (repetitions) and  $t=1, \dots, T$  time bins, then  $\mathbf{J}$  and  $\mathbf{Y}$  are multivariate time series of dimensions  $G \times RT$  and  $M \times RT$ , respectively. In order to keep notation simple, we will first describe a single-trial model ( $R=1$ ) and then will generalize to the multiple-trial case.

The applications in this paper use a cortical source space in which PCD orientations are constrained to be perpendicular to gray/white matter interface. Each entry in  $\mathbf{J}$  therefore corresponds to the scalar value (magnitude and sign) of the PCD vector at particular locations and time points. This is related to sensor measurements by solving the forward problem (FP) of the M/EEG that uses Maxwell’s equations governing electromagnetic fields (Baillet et al., 2001).

Because measurements always have attached uncertainties, it is natural to take a probabilistic approach. In this case, we are not interested in a particular solution, but in the ensemble of possible solutions. That is, one always starts with a probability distribution representing a priori information, and the use of observations *narrows* this distribution. The solution of the inverse problem is not a particular model but the (posterior) probability distribution over the model space.

Most established distributed source reconstruction or “imaging” methods (Darvas et al., 2004) implicitly rely on the following hierarchical model.

$$\begin{aligned} \bar{\mathbf{Y}} &= \mathbf{KJ} + \mathbf{E} \\ \mathbf{J} &= \mathbf{Z} \end{aligned} \quad (1)$$

in which random fluctuations  $\mathbf{E}$  correspond to sensor noise and source activity  $\mathbf{J}$  is generated by random innovations  $\mathbf{Z}$ . Here we have assumed that the signal at the sensors has been averaged over trials to give the ERP  $\bar{\mathbf{Y}}$ . This corresponds to the two-level probabilistic generative model (PGM)

$$\begin{aligned} p(\bar{\mathbf{y}}_t | \mathbf{J}, \boldsymbol{\Omega}) &= \prod_{t=1}^T N(\bar{\mathbf{y}}_t; \mathbf{Kj}_t, \boldsymbol{\Omega}^{-1}) \\ p(\mathbf{J} | \alpha) &= \prod_{t=1}^T N(\mathbf{j}_t; \mathbf{0}, \alpha^{-1} \mathbf{D}^{-1}) \end{aligned} \quad (2)$$

also shown schematically in Fig. 1, where  $\mathbf{j}_t$  and  $\bar{\mathbf{y}}_t$  are the source and sensor column vectors at time  $t$  and  $\boldsymbol{\Omega}^{-1}$  is the sensor noise covariance. The matrix  $\mathbf{D}$  reflects the choice of spatial prior and  $\alpha$  is a spatial precision variable.

Our approach is then based on the following three-level model

$$\begin{aligned} \mathbf{Y} &= \mathbf{KJ} + \mathbf{E} \\ \mathbf{J}^T &= \mathbf{XW} + \mathbf{Z} \\ \mathbf{W} &= \mathbf{R} \end{aligned} \quad (3)$$

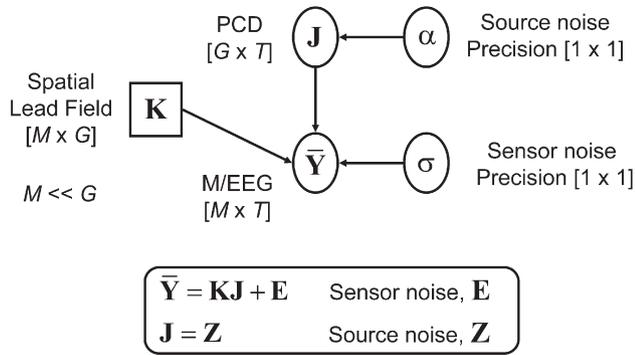


Fig. 1. Graphical representation of the probabilistic generative model implicit in most distributed source reconstruction methods. It comprises two levels. The second level specifies a spatial prior on the primary current density and the first level incorporates the observation equation given by the EEG forward model.

Here we have random innovations  $\mathbf{Z}$  which are “temporal errors,” i.e., lack of fit of the temporal model, and  $\mathbf{R}$  which are “spatial errors,” i.e., lack of fit of a spatial model. In this case, the spatial models are simply zero mean Gaussians with covariances  $\alpha_k^{-1}\mathbf{D}_k^{-1}$ . That is, before observing data we believe that  $\mathbf{W}=\mathbf{R}$ , i.e., that the regression coefficients are given by a random variation with zero mean and spatial regularity  $\alpha_k^{-1}\mathbf{D}_k^{-1}$ . This belief will be updated after observing our M/EEG data. We can regard  $\mathbf{X}\mathbf{W}$  as an empirical prior on the expectation of source activity.

The above equations can be re-expressed as the probabilistic generative model

$$p(\mathbf{Y}|\mathbf{J},\Omega) = \prod_{t=1}^T N(\mathbf{y}_t; \mathbf{K}\mathbf{j}_t, \Omega^{-1}) \quad (4)$$

$$p(\mathbf{J}|\mathbf{W},\Lambda) = \prod_{t=1}^T N(\mathbf{j}_t^T; \mathbf{x}_t, \mathbf{W}, \Lambda^{-1}) \quad (5)$$

$$p(\mathbf{W}|\alpha) = \prod_{k=1}^K N(\mathbf{w}_k^T; \mathbf{0}, \alpha_k^{-1}\mathbf{D}_k^{-1}) \quad (6)$$

The first level, Eq. (4), is identical to the standard model. In the second level, however, source activity at each generator is constrained using a  $T \times K$  matrix of temporal basis functions,  $\mathbf{X}$ . The PGM is shown schematically in Fig. 2.

The precision of the source noise is given by  $\Lambda$ . In this paper,  $\Lambda = \text{diag}(\lambda)$ , where the diagonal element  $\lambda_g$  is the noise precision at the  $g$ th generator. That is, event-related source activity is described by the time-domain GLM and remaining source activity will correspond to unmodelled responses. The quantity  $\Lambda^{-1}$  can therefore be thought of as the variance of spontaneous and/or induced activity in source space. The regression coefficients  $\mathbf{W}$  determine the weighting of the temporal basis functions.

The third level of the model is a spatial prior that reflects our prior uncertainty about  $\mathbf{W}$ . The  $k$ th row of  $\mathbf{W}$ ,  $\mathbf{w}_k$ , is a map of regression coefficients in source space. It provides a generator-specific weighting of the  $k$ -th column of the design matrix, i.e., of the  $k$ -th putative experimental effect or temporal basis function. Each regression coefficient map is constrained by setting  $\mathbf{D}_k$  to correspond to the usual  $L^2$ -norm spatial prior. The spatial prior that is usually on the PCD now appears at a superordinate level. Different choices of  $\mathbf{D}_k$  result in different weights and different

neighborhood relations. This lends the model a higher degree of flexibility by allowing the different effects to be assigned different spatial priors.

The applications in this paper use  $\mathbf{D}_k = \mathbf{D} = \mathbf{L}^T\mathbf{L}$ , where  $\mathbf{L}$  is a discrete surface Laplacian as defined by Huiskamp (1991), which implements second-order differences on geodesic distances. The parameter  $\alpha_k$  then controls the spatial smoothness of the  $k$ th map  $\mathbf{w}_k$ . This is important because it allows different response components to have different spatial characteristics, e.g., response components with longer time scales may be more spatially diffuse (Buzsaki and Draguhn, 2004).

The first level of the model assumes that there is Gaussian sensor noise  $\mathbf{e}_t$ , with zero mean and covariance  $\Omega^{-1}$ . This covariance can be estimated from prestimulus or baseline periods when such data are available (Sahani and Nagarajan, 2004). Alternatively, we assume that  $\Omega = \text{diag}(\sigma)$  where the  $m$ th element of  $\sigma$  is the noise precision on the  $m$ th sensor, and provide a scheme for estimating  $\sigma_m$ , should this be necessary. For this, we also place conjugate Gamma priors on the precision variables  $\sigma$ ,  $\lambda$  and  $\alpha$

$$p(\sigma) = \prod_{m=1}^M \text{Ga}(\sigma_m; b_{\sigma_m}, c_{\sigma_m})$$

$$p(\lambda) = \prod_{g=1}^G \text{Ga}(\lambda_g; b_{\lambda_g}, c_{\lambda_g})$$

$$p(\alpha) = \prod_{k=1}^K \text{Ga}(\alpha_k; b_{\alpha_k}, c_{\alpha_k}) \quad (7)$$

This allows the inclusion of further prior information into the source localization. For example, instead of using baseline periods to estimate a full covariance matrix  $\Omega^{-1}$ , we could use these data to estimate the noise variance at each sensor. This information could then be used to set  $b_{\sigma_m}$  and  $c_{\sigma_m}$ , allowing noise estimates during periods of interest to be constrained softly by those from baseline

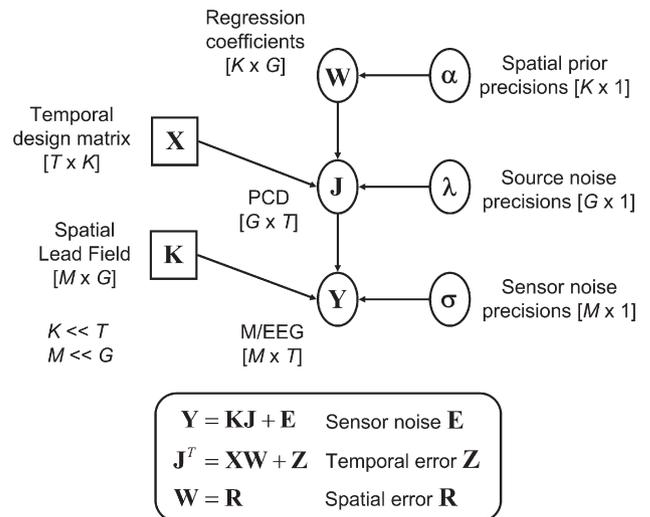


Fig. 2. Graphical representation of the proposed probabilistic generative model. It comprises three levels. The third level specifies a spatial prior on the regression coefficients of the temporal GLM proposed for the primary current density. This temporal model is specified in the second level, whereas the first level encodes the observation equation given by the EEG forward model.

periods. Similarly, we may wish to enforce stronger or weaker spatial regularization on  $w_k$  by setting  $b_{\alpha_k}$  and  $c_{\alpha_k}$  appropriately. The applications in this paper, however, use uninformative gamma priors by setting all scale and shape parameters in (7) to 1000 and 0.001 (mean 1 and variance 1000), respectively. This means that  $\sigma$ ,  $\lambda$  and  $\alpha$  will be estimated solely from the data  $\mathbf{Y}$ .

In summary, the addition of the superordinate level to our generative model induces a partitioning of source activity into signal and noise. This empirical Bayes perspective means that the conditional estimates of source activity  $\mathbf{J}$  are subject to “bottom-up” constraints provided by the data, and “top-down” predictions from the third level of our model. We will use this heuristic later to understand the update equations used to estimate source activity.

### Temporal models

The usefulness of the present spatio-temporal approach rests on our ability to characterize neuronal responses using GLMs. Fortunately, there is a large literature that suggests this is possible. The type of temporal model necessary will depend on the M/EEG response one is interested in. These components could be (i) single trials, (ii) evoked components (steady-state or ERPs; Rugg and Coles, 1995) or (iii) induced components (Tallon-Baudry et al., 1996). In this paper we focus on single trials and ERPs leaving steady-state and induced components the subject of future publications.

The basis functions will form columns in the GLM design matrix,  $\mathbf{X}$  (see Eq. (3) and Fig. 2). Basis functions could be derived from damped sinusoids (Demiralp et al., 1998) or principal components (Trejo and Shensa, 1999; Friston et al., 2006) but in this paper we use a wavelet representation. That is, given an M/EEG signal,  $f$

$$f = \sum_{k=1}^K w_k \mathbf{x}_k \quad (8)$$

where  $\mathbf{x}_k$  are wavelet basis functions and  $w_k$  are wavelet coefficients. Wavelets are derived by translating and dilating a mother wavelet and provide a tiling of time–frequency space that gives a balance between time and frequency resolution. The  $Q$ -factor of a filter or basis function is defined as the central frequency to bandwidth ratio. Wavelet bases are chosen to provide constant  $Q$  (Unser and Aldroubi, 1996). This makes them good models of nonstationary signals, such as ERPs and induced EEG components (Tallon-Baudry et al., 1996). If  $K=T$ , then the mapping  $f \rightarrow w$  is referred to as a wavelet transform, and for  $K > T$  we have an overcomplete basis set. More typically, we have  $K \leq T$ .

In the ERP literature, the particular subset of basis functions used is chosen according to the type of ERP component one wishes to model. Popular choices are wavelets based on B-splines (Unser and Aldroubi, 1996). In statistics, however, it is well known that an appropriate subset of basis functions can be automatically selected using a procedure known as “wavelet shrinkage” or “wavelet denoising.” This relies on the property that natural signals such as images, speech or neuronal activity can be represented using a sparse code comprising just a few large wavelet coefficients. Gaussian noise signals, however, produce Gaussian noise in wavelet space. This comprises a full set of wavelet coefficients whose size depends on the noise variance. By “shrinking” these noise coefficients to zero using a thresholding procedure (Donoho and Johnstone, 1994; Clyde et al., 1998), and transforming back

into signal space, one can denoise data. This amounts to defining a temporal model. We will use this approach for the empirical work reported in this paper. We also note that it is possible to incorporate the wavelet shrinkage methods into the probabilistic generative model by modifying Eq. (7). This has been implemented for models of fMRI data with spatial wavelet priors (Flandin and Penny, 2007). In this paper, however, wavelet shrinkage is implemented outside of the model by using a standard thresholding procedure (Donoho and Johnstone, 1994).

### Multiple-trials model

When considering  $R$  independent trials or repetitions, our PGM can be written as

$$p(\tilde{\mathbf{Y}}|\tilde{\mathbf{J}}, \Omega) = \prod_{r=1}^R \prod_{t=1}^T N(\mathbf{y}_{\cdot tr}; \mathbf{K}\mathbf{j}_{tr}, \Omega^{-1})$$

$$p(\tilde{\mathbf{J}}|\mathbf{W}, \Lambda) = \prod_{r=1}^R \prod_{t=1}^T N(\mathbf{j}_{tr}^T; \mathbf{x}_{rt} \cdot \mathbf{W}, \Lambda^{-1})$$

$$p(\mathbf{W}|\alpha) = \prod_{k=1}^K N(\mathbf{w}_k^T; \mathbf{0}, \alpha_k^{-1} \mathbf{D}_k^{-1}) \quad (9)$$

where  $\tilde{\mathbf{Y}}$  ( $M \times RT$ ) and  $\tilde{\mathbf{J}}$  ( $G \times RT$ ) are multivariate time series obtained by concatenating all trials of the measured ERP and the estimated PCD, respectively; and  $\mathbf{x}_{rt}$  is the  $K \times 1$  vector of regressors for the  $t$ th time bin and the  $r$ th trial. Here we have assumed that the effect of interest  $\mathbf{W}$  is the same in all trials. This treats trials as fixed rather than random effects, as is the case for standard analyses of single-subject fMRI data (Frackowiak et al., 2003). Thus, the multiple-trial design matrix  $\tilde{\mathbf{X}}$  in this case is constructed by block repeating our design matrix for a single trial

$$\tilde{\mathbf{X}} = \mathbf{1}_R \otimes \mathbf{X} \quad (10)$$

where  $\mathbf{1}_R$  denotes a column vector of ones with length  $R$ . Note that the multiple-trial PGM and the corresponding hierarchical model have the same form as for the single-trial case, if we just use  $\mathbf{X} = \tilde{\mathbf{X}}$  in Eq. (3), and take the index  $t$  to run over time and across trials in Eqs. (4) and (5) ( $t=1, \dots, RT$ ). Thus, for simplicity, we will keep the same notation as before. In summary, multiple trials are treated by forming concatenated data and design matrices.

### Bayesian inference

To make inferences about the sources underlying M/EEG, we need to invert our PGM to produce the posterior density  $p(\mathbf{J}|\mathbf{Y})$ . This is straightforward in principle and can be achieved using standard Bayesian methods (Gelman et al., 1995). For example, one could use Markov Chain Monte Carlo (MCMC) to produce samples from the posterior. This has been implemented efficiently for dipole-like inverse solutions (Schmidt et al., 1999) in which sources are parameterized as spheres of unknown number, extent and location. It is, however, computationally demanding for distributed source solutions, taking several hours for source spaces comprising  $G > 1000$  generators (Auranen et al., 2005). In this work we adopt the computationally efficient approximate inference

framework called Variational Bayes (VB) (Lappalainen and Miskin, 2000; Beal, 2003; Friston et al., 2007).

#### Variational Bayes framework

This is a recent development from the machine learning community and is based on the Variational Free Energy method of Feynman and Bogoliubov. The central quantity of interest is the posterior distribution  $p(\boldsymbol{\theta}|\mathbf{Y})$ . This implies estimation of both the parameters and the uncertainties associated with their estimation. Given a PGM of the data, the log-evidence or marginal likelihood can be written as

$$\begin{aligned} \log p(\mathbf{Y}) &= \int q(\boldsymbol{\theta}) \log p(\mathbf{Y}) d\boldsymbol{\theta} = \int q(\boldsymbol{\theta}) \log \left[ \frac{p(\mathbf{Y}, \boldsymbol{\theta}) q(\boldsymbol{\theta})}{q(\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{Y})} \right] d\boldsymbol{\theta} \\ &= F + KL[q(\boldsymbol{\theta})||p(\boldsymbol{\theta}|\mathbf{Y})] \end{aligned} \quad (11)$$

here,  $q(\boldsymbol{\theta})$  is the approximate posterior. We have

$$F = \int q(\boldsymbol{\theta}) \log \frac{p(\mathbf{Y}, \boldsymbol{\theta})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} \quad (12)$$

which is known (to physicists) as the negative variational free energy and

$$KL[q(\boldsymbol{\theta})||p(\boldsymbol{\theta}|\mathbf{Y})] = \int q(\boldsymbol{\theta}) \log \frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta}|\mathbf{Y})} d\boldsymbol{\theta} \quad (13)$$

is the KL divergence (Cover and Thomas, 1991) between the approximate posterior  $q(\boldsymbol{\theta})$  and the true posterior  $p(\boldsymbol{\theta}|\mathbf{Y})$ .

The aim of VB learning is to maximize  $F$  and so make the approximate posterior as close as possible to the true posterior. One generic procedure for ensuring that the integrals in  $F$  are tractable is to assume that the approximating density factorizes over groups of parameters (mean-field approximation)

$$q(\boldsymbol{\theta}) = \prod_i q(\boldsymbol{\theta}_i) \quad (14)$$

where  $\boldsymbol{\theta}_i$  is the  $i$ th group of parameters.

#### Approximate posteriors

For our source reconstruction model we assume the following factorization of the approximate posterior

$$q(\mathbf{J}, \mathbf{W}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\sigma}) = q(\mathbf{J})q(\mathbf{W})q(\boldsymbol{\alpha})q(\boldsymbol{\sigma})q(\boldsymbol{\lambda}) \quad (15)$$

We also assume that the approximate posterior for the regression coefficients factorizes over generators

$$q(\mathbf{W}) = \prod_{g=1}^G q(\mathbf{w}_g) \quad (16)$$

This approximation was used in the spatio-temporal model for fMRI described in Penny et al. (2005). Because of the spatial prior (Eq. (6)), the regression coefficients in the true posterior  $p(\mathbf{W}|\mathbf{Y})$  will clearly be correlated. Our perspective, however, is that this is too computationally burdensome for current personal computers to take account of. Moreover, as we shall see, updates for our approximate factorized densities  $q(\mathbf{w}_g)$  do encourage the approximate posterior means to be similar at nearby generators, thereby achieving the desired effect of the prior.

Now that we have defined the probabilistic model and our factorization of the approximate posterior, we can use VB to derive expressions for each component of the approximate posterior. We do not present details of these derivations in this paper. Similar derivations have been published elsewhere (Penny et al., 2005). The following sections describe each distribution and the updates of its sufficient statistics required to maximize the lower bound on the model evidence,  $F$ .

#### Primary current density

Updates for the sources are given by

$$q(\mathbf{J}) = \prod_{t=1}^T q(\mathbf{j}_t) \quad (17)$$

$$q(\mathbf{j}_t) = N(\mathbf{j}_t; \hat{\mathbf{j}}_t, \hat{\boldsymbol{\Sigma}}_t) \quad (18)$$

$$\hat{\boldsymbol{\Sigma}}_t = (\mathbf{K}^T \hat{\boldsymbol{\Omega}} \mathbf{K} + \hat{\boldsymbol{\Lambda}})^{-1} \quad (19)$$

$$\hat{\mathbf{j}}_t = \hat{\boldsymbol{\Sigma}}_t (\mathbf{K}^T \hat{\boldsymbol{\Omega}} \mathbf{y}_t + \hat{\boldsymbol{\Lambda}} \hat{\mathbf{W}}^T \mathbf{x}_t^T) \quad (20)$$

where  $\mathbf{j}_t$  is the  $t$ th column of  $\mathbf{J}$  and  $\hat{\boldsymbol{\Omega}}$ ,  $\hat{\boldsymbol{\Lambda}}$  and  $\hat{\mathbf{W}}$  are estimated parameters defined in the following sections. We have not assumed that  $q(\mathbf{J})$  factorizes over time, but this ‘falls out’ of the equations, primarily because we have assumed that the additive noise  $\mathbf{E}$  factorizes over time (i.e., IID observation noise). Given that the source covariance matrix does not change with time, Eq. (20) can be rewritten in a more compact form

$$\hat{\mathbf{J}} = \hat{\boldsymbol{\Sigma}}_t (\mathbf{K}^T \hat{\boldsymbol{\Omega}} \mathbf{Y} + \hat{\boldsymbol{\Lambda}} \hat{\mathbf{W}}^T \mathbf{X}^T) \quad (21)$$

This expression shows that our source estimates are the result of a spatio-temporal deconvolution. The spatial contribution to the estimate is  $\mathbf{K}^T \mathbf{Y}$  and the temporal contribution is  $\hat{\mathbf{W}}^T \mathbf{X}^T$ . From the perspective of the hierarchical model, shown in Fig. 3, these are the ‘‘bottom-up’’ and ‘‘top-down’’ predictions. Importantly, each

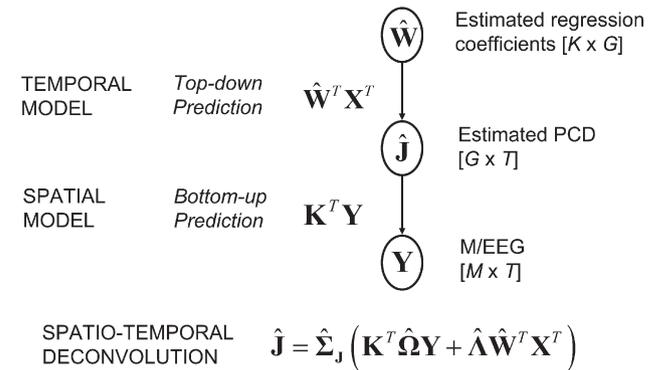


Fig. 3. Hierarchical model representing the spatio-temporal deconvolution embodied by the VB estimator of the PCD. The estimated PCD receives contributions from two terms: (i) a ‘‘top-down’’ prediction from the temporal GLM and (ii) a ‘‘bottom-up’’ prediction from the spatial lead-field model, both weighted by their respective precisions.

prediction is weighted by its relative precision. Moreover, the parameters controlling the relative precisions  $\hat{\Omega}$  and  $\hat{\Lambda}$  are estimated from the data. This means that our source estimates derive from an automatically regularized spatio-temporal deconvolution. This property is shared by the spatio-temporal model for fMRI, described in Penny et al. (2005).

An important characteristic of the source update in Eq. (21) is that the source estimate at each time point depends on the posterior covariance matrix between different sources,  $\hat{\Sigma}_{\mathbf{J}}$ . This allows the deconvolution algorithm to accommodate correlation between sources and should be contrasted with the minimum variance beamformer estimate (Darvas et al., 2004)

$$\hat{\mathbf{j}}_{g\cdot}^T = \mathbf{b}_{\cdot g}^T \mathbf{Y} \quad (22)$$

where the  $1 \times T$  time series at generator  $g$  is given by projecting M/EEG data  $\mathbf{Y}$  onto the spatial filter  $\mathbf{b}_{\cdot g}^T$ . This projection is repeated separately for all source locations  $g$ . The spatial filter, which is derived by assuming that different sources are uncorrelated, is given by

$$\mathbf{b}_{\cdot g} = (\mathbf{k}_{\cdot g}^T \Omega \mathbf{k}_{\cdot g})^{-1} \Omega \mathbf{k}_{\cdot g} \quad (23)$$

As a result, the beamformer is unable to localize correlated activity (Sahani and Nagarajan, 2004). We will return to this important issue in the Results section.

We end this section by noting that statistical inferences about current sources are more robust than point predictions. This property has been used to great effect with Pseudo-z beamformer statistics (Robinson and Vrba, 1999), sLORETA (Pascual-Marqui, 2002) and VARETA (Bosch-Bayard et al., 2001) source reconstructions, which divide current source estimates by their standard deviations. This approach can be adopted in the current framework as the standard deviations are readily computed from the diagonal elements of  $\hat{\Sigma}_{\mathbf{J}}$ . Moreover, we can threshold these statistic images to create posterior probability maps (PPMs), as introduced by Friston and Penny (2003).

#### Regression coefficients

Updates for the regression coefficients are given by

$$q(\mathbf{w}_{\cdot g}) = N(\mathbf{w}_{\cdot g}; \hat{\mathbf{w}}_{\cdot g}, \hat{\Sigma}_{\mathbf{w}_{\cdot g}})$$

$$\hat{\Sigma}_{\mathbf{w}_{\cdot g}} = (\hat{\lambda}_g \mathbf{X}^T \mathbf{X} + \text{diag}(\mathbf{d}_{gg}) \text{diag}(\hat{\boldsymbol{\alpha}}))^{-1}$$

$$\hat{\mathbf{w}}_{\cdot g} = \hat{\Sigma}_{\mathbf{w}_{\cdot g}} (\hat{\lambda}_g \mathbf{X}^T \hat{\mathbf{j}}_{g\cdot} + \text{diag}(\hat{\boldsymbol{\alpha}}) \mathbf{r}_g) \quad (24)$$

where  $\hat{\boldsymbol{\alpha}}$  is the estimated parameter defined later on,  $\mathbf{d}_{ijk}^T$  is a  $K \times 1$  vector containing the  $(i, j)$ th element of all the  $\mathbf{D}_k$  matrices and  $\mathbf{r}_g$  is the weighted sum of neighboring regression coefficient estimators and is given by

$$\mathbf{r}_g = \sum_{g'=1, g' \neq g}^G \text{diag}(\mathbf{d}_{gg'}) \hat{\mathbf{w}}_{\cdot g'} \quad (25)$$

The update for  $\hat{\mathbf{w}}_{\cdot g}$  in Eq. (24) therefore indicates that the regression coefficient estimates at a given generator regress toward those at nearby generators. This is the desired effect of the spatial

prior and it is preserved despite the factorization in the approximate posterior. This equation can again be thought of in terms of the hierarchical model where the regression coefficient estimate is a combination of a “bottom-up” prediction from the level below,  $\mathbf{X}^T \hat{\mathbf{j}}_{g\cdot}$ , and a “top-down” prediction from the prior,  $\mathbf{r}_g$ . Again, each contribution is weighted by its relative precision.

The update for the covariance in Eq. (24) shows that the only off-diagonal contributions are due to the design matrix. If the temporal basis functions are therefore chosen to be orthogonal then this posterior covariance will be diagonal, thus making a potentially large computational saving. One benefit of the proposed framework, however, is that non-orthogonal bases can be accommodated. This may allow for a more natural and compact description of the data.

#### Precision of temporal models

Updates for the precision of the temporal model are given by

$$q(\lambda) = \prod_{g=1}^G Ga(\lambda_g; \hat{b}_{\lambda_g}, \hat{c}_{\lambda_g})$$

$$\frac{1}{\hat{b}_{\lambda_g}} = \frac{1}{b_{\lambda_g}} + \frac{1}{2} \sum_{t=1}^T \left[ (\hat{j}_{gt} - \hat{\mathbf{w}}_{\cdot g}^T \mathbf{x}_t)^2 + (\hat{\Sigma}_{\mathbf{J}})_{gg} + \mathbf{x}_t^T \hat{\Sigma}_{\mathbf{w}_{\cdot g}} \mathbf{x}_t \right]$$

$$\hat{c}_{\lambda_g} = \frac{1}{2} T + c_{\lambda_g}$$

$$\hat{\lambda}_g = \hat{b}_{\lambda_g} \hat{c}_{\lambda_g} \quad (26)$$

Where  $(\hat{\Sigma}_{\mathbf{J}})_{gg}$  is the  $g$ th diagonal element of  $\hat{\Sigma}_{\mathbf{J}}$ . In the context of ERP analysis, these expressions amount to an estimate of the variance of spontaneous and/or induced activity at generator  $g$ ,  $\hat{\lambda}_g^{-1}$  given by the squared error between the evoked component estimate,  $\hat{\mathbf{w}}_{\cdot g}^T \mathbf{x}_t$ , and source estimate,  $\hat{j}_{gt}$  at the given generator, averaged over time, and the other approximate posteriors.

#### Precision of forward model

Updates for the precision of the sensor noise are given by

$$q(\boldsymbol{\sigma}) = \prod_{m=1}^M q(\sigma_m)$$

$$q(\sigma_m) = Ga(\hat{b}_{\sigma_m}, \hat{c}_{\sigma_m})$$

$$\frac{1}{\hat{b}_{\sigma_m}} = \frac{1}{b_{\sigma_m}} + \frac{1}{2} \sum_{t=1}^T (y_{mt} - \mathbf{k}_m^T \hat{\mathbf{j}}_t)^2 + \frac{1}{2} \mathbf{k}_m^T \hat{\Sigma}_{\mathbf{J}} \mathbf{k}_m$$

$$\hat{c}_{\sigma_m} = \frac{T}{2} + c_{\sigma_m}$$

$$\hat{\sigma}_m = \hat{b}_{\sigma_m} \hat{c}_{\sigma_m} \quad (27)$$

These expressions amount to an estimate of observation noise variance at the  $m$ th sensor,  $\hat{\sigma}_m^{-1}$ , given by the squared error between the forward model and sensor data, averaged over time and the other approximate posteriors.

### Precision of spatial prior

Updates for the precision of the spatial prior are given by

$$q(\boldsymbol{\alpha}) = \prod_{k=1}^K q(\alpha_k)$$

$$q(\alpha_k) = \text{Ga}(\hat{b}_{\alpha_k}, \hat{c}_{\alpha_k})$$

$$\frac{1}{\hat{b}_{\alpha_k}} = \frac{1}{b_{\alpha_k}} + \left\| \mathbf{D}_k \hat{\mathbf{w}}_k^T \right\|^2 + \sum_{g=1}^G d_{ggk} (\hat{\boldsymbol{\Sigma}}_{\mathbf{w}_g})_{kk}$$

$$\hat{c}_{\alpha_k} = \frac{G}{2} + c_{\alpha_k}$$

$$\hat{\alpha}_k = \hat{b}_{\alpha_k} \hat{c}_{\alpha_k} \quad (28)$$

where  $(\hat{\boldsymbol{\Sigma}}_{\mathbf{w}_g})_{kk}$  is the  $k$ th diagonal element of  $\hat{\boldsymbol{\Sigma}}_{\mathbf{w}_g}$ . These expressions amount to an estimate of the “spatial noise variance”  $\hat{\alpha}_k^{-1}$ , given by the discrepancy between neighboring regression coefficients, averaged over space and the other approximate posteriors.

To summarize, our source reconstruction model is fitted to data by iteratively applying the update equations until the change in the negative free energy  $F$ , is less than some user-specified tolerance. This procedure is summarized in the pseudo-code in Fig. 4 and will lead to a local maximum of  $F$  (for a general discussion of convergence issues, see Friston et al. (2007)).

This amounts to a process in which sensor data is spatially deconvolved, time-series models are fitted in source space, and then the precisions (accuracy) of the temporal and spatial models are estimated. This process is then iterated and results in a spatio-temporal deconvolution in which all aspects of the model are optimized to maximize a lower bound on the model evidence. The algorithm can be efficiently implemented as described in Appendix A.

## Results

This section describes the application of the approach presented here to (i) biophysically realistic simulated data and (ii) EEG from a face-processing experiment. In all cases, we used the same sensor and source spaces. The sensor space was defined using  $M=128$  electrodes from the BioSemi ActiveTwo System. The source space then consisted of a mesh of nodes (generators) corresponding to the vertices of the triangles obtained by tessellation of the gray/white matter interface of the realistic digital brain phantom developed at the Montreal Neurological Institute (MNI) (Collins et al., 1998). The tessellation comprised 12,000 triangles and  $G=6004$  vertices.

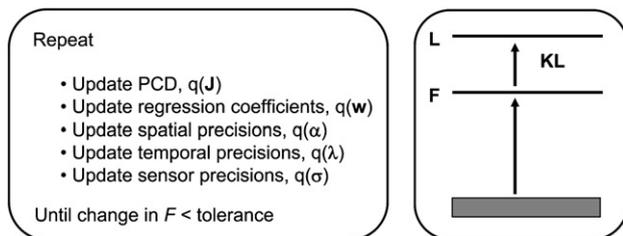


Fig. 4. Pseudo-code for the Variational Bayes algorithm. Iterative update of the approximate posterior components results in increasing the lower bound,  $F$ , on the model evidence.

We used the three concentric sphere model to calculate the electric lead field (Rush and Driscoll, 1969). The centre and radius of the spheres were fitted to the scalp, skull and cerebral tissue of the same brain. In what follows we refer to the spatio-temporal approach as “VB-GLM.”

### ERP simulation

We used our generative model to simulate ERP-like activity by using the waveforms and spatial profiles shown in Fig. 5. As can be seen, the two waveforms are temporally correlated ( $\text{Corr}=0.86$ ) with main peaks that mimic an ERP component at about  $t=200$  ms post-stimulus. These waveforms were derived from a neural mass model describing activity in a distributed network of cortical areas (David and Friston, 2003), which lends these simulations a degree of biological plausibility. The two spatial profiles in turn consisted of Gaussian blobs with identical maximum amplitudes of 10, and identical full width at half maximums (FWHM) of 20 mm. The spatial extent of the activated areas was constrained by taking a geodesic neighborhood of 3 nodes around the centre of each Gaussian and setting to zero the activity outside.

These temporal and spatial profiles were then used respectively as design matrix and regression coefficients to generate data from our model. Ten trials of sensor data were generated using signal-to-noise ratios (SNR) of 10 and 40 at the sensor and source levels, respectively. Here, we defined SNR as the ratio of the signal standard deviation to noise standard deviation. Signal epochs of 512 ms were then produced with a sampling period of 4 ms, giving a total of 5120 ms of EEG (1280 time bins).

We estimated the sources underlying the sample ERP (i) with an overspecified temporal model that incorporated two spurious regressors in addition to the ones used to generate the data and (ii) with a temporal model that consisted of Battle-Lemarie wavelets obtained by application of the wavelet shrinkage algorithm to the first eigenvector of the simulated sensor data.

### Overspecified temporal model

The four regressors that form the overspecified temporal model for a single trial are shown in Fig. 6A. Note that Regressors 1 and 2 contain the source waveforms that were used to generate the data. These four regressors were then concatenated to form the design matrix that models the 10 simulated trials (Fig. 6B). It is important to note that this design matrix is not orthogonal because the four regressors are temporally correlated.

The model was then fitted to the data using VB-GLM. As shown in Fig. 7, the true effects (regression coefficients) are accurately recovered, whereas the spurious regression coefficients are shrunk toward zero. The shrinking effect is evident when looking at the estimated spatial precision,  $\hat{\alpha}_k$ , for each regression coefficient shown in the upper panel of Fig. 7. These determine how precisely the effects are constrained around zero. A large precisions implies a strong shrinkage toward zero. As can be seen, the precisions corresponding to the spurious regressors are four orders of magnitude greater than the precisions corresponding to the true regressors. These results are a consequence of the spatial prior and the iterative spatio-temporal deconvolution, and demonstrate that source reconstruction with temporal priors is robust to model mis-specification. This also shows that VB-GLM, in contrast to, for instance, traditional beamforming approaches (see section “VB-GLM vs. minimum variance beamformer”), is capable of localizing temporally correlated sources.

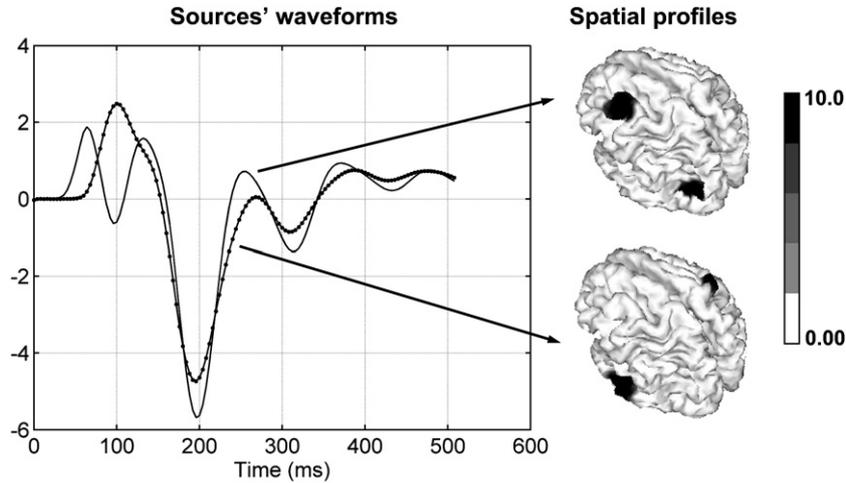


Fig. 5. Temporal (left) and spatial (right) profiles of the biophysically realistic sources used for generating simulated data. The two waveforms incorporate a negative component at  $t \approx 200$  ms.

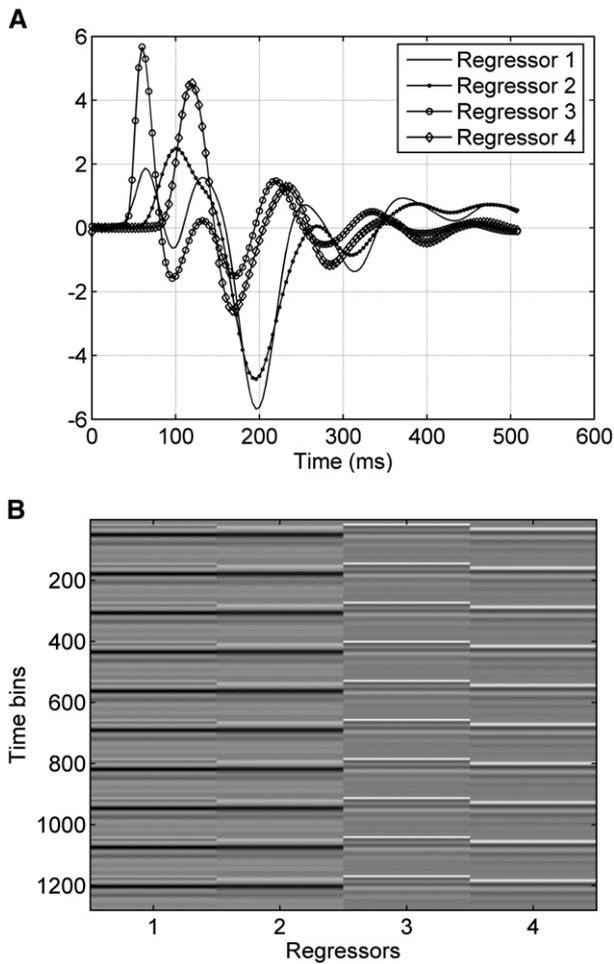


Fig. 6. Definition of the overspecified temporal model used for source reconstruction of simulated data. (A) Four regressors used for constructing the design matrix of the temporal GLM of the PCD. Regressors 1 and 2 are the same as used for generating the data, while 3 and 4 are spurious regressors. (B) Image of the overspecified multiple-trial design matrix. Note that all regressors are temporally correlated.

Additionally, note that the reconstructed effect sizes are diminished with respect to the true ones. This could be due to the increasing source amplitude bias with depth that is inherent to all distributed inverse solutions (Trujillo-Barreto et al., 2004). The bias is less in the case of Regression Coefficient 1 because the left-frontal and left-occipital sources are closer to the sensors than the left-temporal and right-postcentral sources in Regression Coefficient 2. These depth effects can be overcome with sparse priors (Trujillo-Barreto et al., 2004). It is possible, however, that this bias is not only a depth effect but it could also reflect the well-known overconfidence problem of mean-field approximations. In our case, this is most likely the consequence of assuming that the approximate posterior for the regression coefficients factorizes over generators, resulting in posterior certainties that are slightly too high.

*Wavelet temporal model*

The simple example we have just described, although useful for illustrative purposes, is of limited use because, in practice, we can never know which are the “true” regressors.

We now describe how the temporal model can be constructed generically using the data at hand. We first partition the sensor data into two halves. The first five trials were used to fit the temporal model, while the remaining five trials were used for source reconstruction. We then extracted the first eigenvector of the ERP calculated from the first five trials using a singular value decomposition (SVD) and fitted a Battle-Lemarie wavelet model to this time series. The upper panel of Fig. 8 shows the corresponding time series estimate. This employed  $K=33$  basis functions, as determined by application of the wavelet shrinkage algorithm (Donoho and Johnstone, 1994; Clyde et al., 1998). The corresponding single-trial design matrix of our temporal model is shown in the lower panel of Fig. 8. This matrix was extended to the multiple-trial case using Eq. (10), and then used for source reconstruction.

Because it is impractical to present the source estimates for all time instants, we will show results for a single time point. Fig. 9 shows the true and estimated PCD, averaged over trials and normalized to the respective maximum absolute values, for  $t=200$  ms. Note that, at this latency, both spatial profiles are

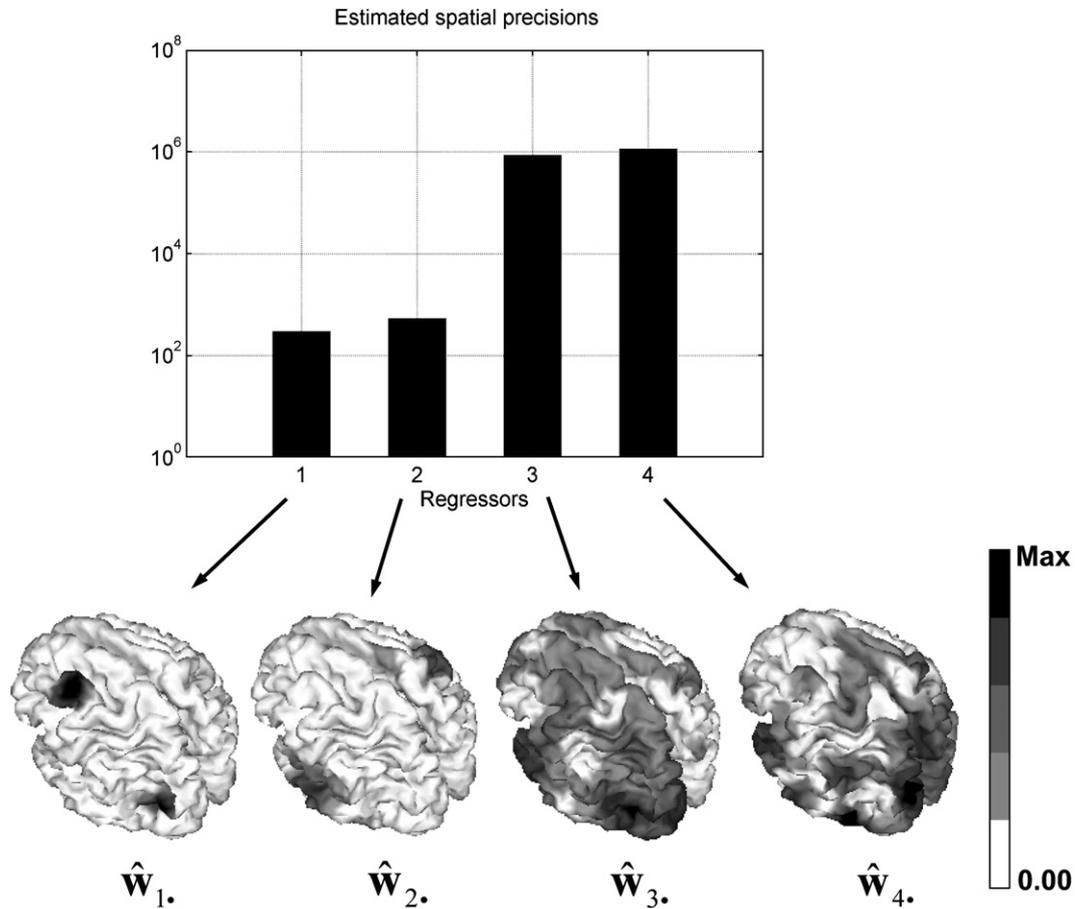


Fig. 7. Results of the VB-GLM approach with an overspecified temporal model. Upper panel: Estimated spatial precisions,  $\alpha_k$ , for each regression coefficient. Lower panels: Estimated regression coefficients  $\hat{\mathbf{w}}_k$ . Coefficients 1 and 2 correspond to the regressors used to generate the simulated data and are correctly reconstructed (compare to Fig. 5). Coefficients 3 and 4 corresponding to the spurious regressors, are shrunk toward zero. The maximum of the scale has the following values (from left to right): 5.72, 5.72, 0.03, and 0.02. Note that the maximum of the scale for Regression Coefficient 2 has been set to the maximum of Regression Coefficient 1 for comparison.

activated and correspond to negative peaks of activity. This pattern is satisfactorily recovered by VB-GLM.

#### *Spatial vs. spatio-temporal approach*

In order to quantitatively assess the effect of the temporal prior, we compared the results of our full VB-GLM approach to a limited version where the temporal prior is not used. This is achieved by using  $\mathbf{X}=\mathbf{I}_T$  and  $\mathbf{Z}=\mathbf{0}_{M \times T}$ . In this case, the three-level PGM depicted in Fig. 2 reduces to the two-level one in Fig. 1, which underlies the majority of the “instantaneous” approaches reported in the literature. In our case, given the spatial Laplacian prior that we have assumed for the regression coefficients, this reduced model can be considered to be equivalent to an “instantaneous” LORETA solution (Pascual-Marqui et al., 1994).

We used two measures for comparison. First, we calculated the receiver operating characteristic (ROC) for the two approaches. This is a plot of the sensitivity versus 1 minus the specificity, and was generated by declaring a generator to be active if the effect size was larger than some arbitrary threshold. Although ROC curves have been extensively applied to evaluate the detection accuracy of diagnostic imaging techniques, they do not provide an explicit assessment of localization accuracy. Therefore, we also calculated

the distance-based localization receiver operating characteristic (DL-ROC) (Biscay-Lirio et al., 1992). This curve describes the variation of localization error over the range of arbitrary thresholds used (see Appendix A).

The results for  $t=200$  ms are shown in Fig. 10 (right panel). As can be seen, the VB-GLM outperforms the instantaneous approach in both detection and localization accuracy. The results also indicate that increased sensitivity can be achieved while maintaining high specificity. Additionally, the source reconstruction obtained with VB-GLM is less blurred and contains less ghosting (the curse of traditional linear inverse solutions).

Another perspective on these simulations is given by the temporal evolution of the activity for the generator of maximum amplitude at  $t=200$  ms. This is shown in Fig. 11. We see that the estimated VB-GLM time course is much smoother than with the instantaneous approach. This is clearly a consequence of the temporal prior used.

#### *VB-GLM vs. minimum variance beamformer*

We have demonstrated that VB-GLM is capable of recovering highly correlated sources, which has been reported not to be the case for the minimum variance beamformer (MV-BF) (Sahani and

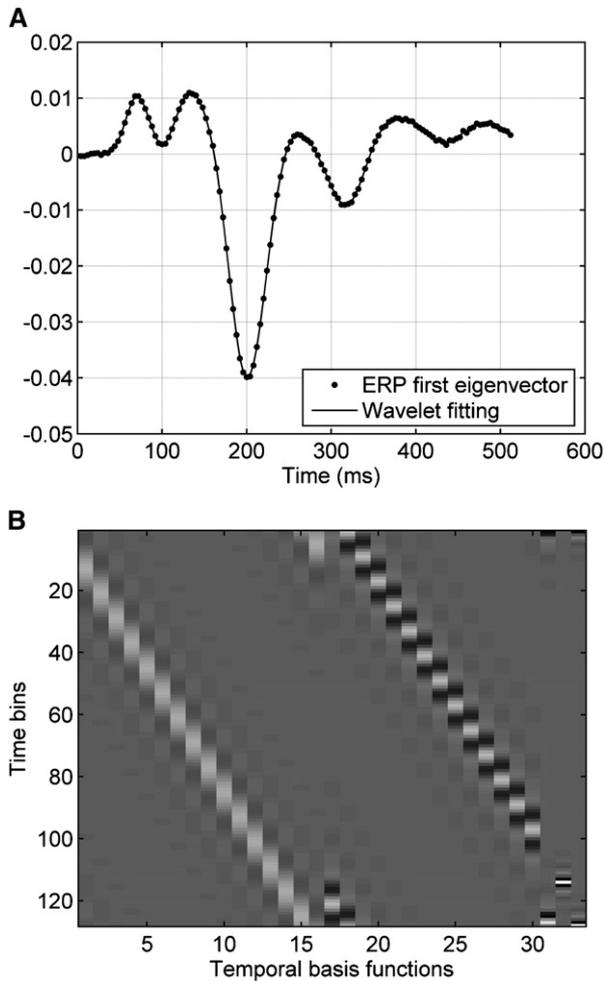


Fig. 8. Definition of the wavelet temporal model used for source reconstruction of simulated data. (A) Fitting of the data first eigenvector with  $K=33$  Battle-Lemarie wavelets as calculated using wavelet shrinkage. (B) Image of the corresponding single-trial design matrix. The leftmost columns contain lower frequencies with progressively higher frequencies to the right. This design matrix is replicated over trials using Eq. (10).

Nagarajan, 2004). This motivated an empirical comparison between the two approaches. For this, Eqs. (22) and (23) for the MV-BF were implemented and applied to the simulated data described previously. Additionally, data from a simulation comprising a single left-occipital source were analyzed with both methods. The time series for this single source simulation was given by Regressor 1 in Fig. 6. The rest of the simulation settings were kept as before. The initial MV-BF estimates were found to contain erroneously large values near the centre of the sphere used for the forward calculation (Sekihara et al., 2001). This is because the norm of the lead field becomes very small in that region. To avoid these artefacts, a normalized lead-field matrix was used in Eq. (23) (Van Veen et al., 1997; Robinson and Vrba, 1999).

The true and estimated PCDs at  $t=200$  ms for the two simulated data sets and for both the VB-GLM and MV-BF approaches are shown in Fig. 12. All maps have been normalized to their respective maximum absolute values. As can be seen, VB-GLM outperforms MV-BF in the two cases. In the single-source case, the MV-BF reconstruction, although giving activity in the area of the true activation, it is significantly more blurred than the VB-GLM

solution. Moreover, MV-BF is unable to recover multiple correlated sources, as expected.

A quantitative comparison was also carried out by calculating the ROC and DL-ROC curves for all cases. The results are shown in Fig. 13. In all cases, VB-GLM showed higher sensitivity for any level of specificity. This is critical in the case of multiple correlated sources, for which MV-BF performed very poorly. Note also that even in the single-source case, where MV-BF showed its best detection accuracy (ROC), its localization accuracy (DL-ROC) was very low.

#### Face ERPs

This section presents an analysis of a face processing ERP data set from Henson et al. (2003). Details of the experimental paradigm as well as the full data set can be found at [www.fil.ion.ucl.ac.uk/spm/data/mmfaces.html](http://www.fil.ion.ucl.ac.uk/spm/data/mmfaces.html).

#### Experimental paradigm

The experiment involved randomized presentation of 86 faces and 86 scrambled faces, as described in Fig. 14. Half of the faces are familiar and half unfamiliar, creating three event-types (conditions) in total, although only the basic contrast of faces vs. scrambled faces is described here. The faces condition in this case was obtained by collapsing over familiarity.

The scrambled faces were created by 2-D Fourier transformation, random phase permutation, inverse transformation and outline-masking of each face. Thus, faces and scrambled faces are closely matched for low-level visual properties such as spatial frequency power density. The subject had to judge the left–right symmetry of each stimulus (face and scrambled) around an imaginary vertical line through the centre of the image. Faces were presented for 600 ms, every 3600 ms.

The EEG data were acquired on a 128-channel BioSemi ActiveTwo system (see Fig. 5), sampled at 1024 Hz, plus electrodes on the left earlobe, right earlobe, and two each to measure HEOG and VEOG. The data were referenced to the average of left and right earlobe electrodes and epoched from  $-200$  ms to  $+600$  ms. These epochs were then detrended and examined for artefacts, defined as time points that exceeded an absolute threshold of  $120 \mu\text{V}$  (mainly in the VEOG). A total of 29 of the 172 trials were rejected.

#### Data analysis

The epochs were averaged according to the two trial types faces (F) and scrambled faces (S) to produce condition specific ERPs, for visualization purposes. The first clear difference F–S was maximal around 170 ms, appearing as an enhancement of a negative component (peak N170) at occipito-temporal channels, or enhancement of a positive peak near Cz (e.g., channel C1). These effects are shown as a differential topography and as time series in Fig. 15.

The source reconstruction method (VB-GLM) was then applied to the single-trial (unaveraged) data. Before applying the model, the data were first down-sampled by a factor of 4, and the 128 samples following stimulus onset were extracted. These steps were taken as we used WaveLab to generate the wavelet bases (for the GLM) which uses a pyramidal algorithm to compute coefficients, thus requiring the number of samples to be a power of two.

We then extracted the first eigenvector of the ERP for each condition using SVD and fitted Battle-Lemarie wavelet models to

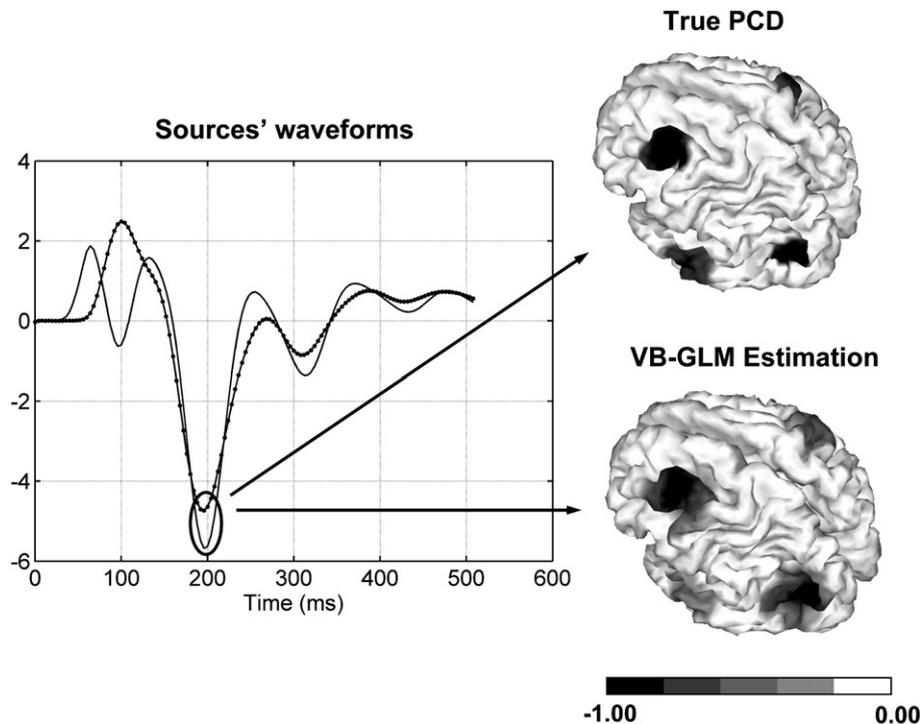


Fig. 9. True and VB-GLM source estimates for the simulated data, based on the wavelet temporal model. The left panel indicates the time point of interest ( $t=200$  ms). At this latency, the two simulated sources show simultaneous negative peaks. The right panel shows the true PCD at  $t=200$  ms, as well as the corresponding VB-GLM estimate, averaged across trials. The two maps have been normalized to their respective maximum absolute values.

these time series. Fig. 16 shows the corresponding time series estimates. We employed  $K=23$  and  $K=30$  basis functions for conditions F and S, respectively, as determined by wavelet shrinkage (Donoho and Johnstone, 1994). These functions were then used to construct the single-trial design matrices for the two conditions, each comprising the Battle-Lemarie basis (see Fig. 16). These matrices were then repeated for all trials to produce the multiple-trial condition-specific design matrices of dimensions  $11,008 \times 23$  and  $11,008 \times 30$  (see Eq. (10)). Finally, the full design matrix for all trials and for the two conditions was constructed as a block diagonal matrix, where each block contained the multiple-trials design matrix for each condition. This matrix, which fully integrates the experimental design, is of dimension  $22,016 \times 53$ .

All trials for faces and scrambled faces were then concatenated to form a vector of 22,016 elements at each electrode. The sensor data matrix was then of dimension  $128 \times 22,016$ . The source space used was the same as for the simulations.

We then applied the source reconstruction algorithm and obtained a solution after 6 min of processing. The estimated PCDs averaged across trials for conditions F and S at  $t=170$  ms, are shown in Fig. 17. The two solutions have been normalized to the maximum of the solution for condition F. As can be seen, the spatial distribution of the sources in both cases show bilateral activity in the fusiform area, with the cluster of maximum activation in the right hemisphere. The temporal waveforms corresponding to the generators with maximum activity in the two conditions are also shown in Fig. 17 (right panel). As expected, maximum differences between conditions are obtained for  $t=170$  ms.

The overall effect of faces was obtained by applying the appropriate contrast to the fitted source reconstruction and is shown more clearly in Fig. 18. The image in the upper panel shows differences between conditions at each generator, normalized to the

maximum positive difference. The overall activation pattern shows a number of clusters of positive and negative differences. By convention we have constrained the PCD to be perpendicular to the cortical surface and directed outward. Then positive differences can be interpreted as an increased outward or a decreased inward PCD, while negative differences can be associated with decreased outward or an increased inward PCD.

In order to better characterize the effect of faces, the positive and negative differences were normalized to their respective maximum values and thresholded at 30% and 80%. The results are shown in the lower panels of Fig. 18. At 30%, four main clusters appear at (i) right fusiform, (ii) left fusiform (iii) right temporal, and (iv) anterior frontal regions. With an 80% threshold, only the right fusiform and right temporal activations are present. These activations are consistent with previous fMRI and MEG analyses (Henson et al., 2003) and the classical “core model” for face recognition and perception (Haxby et al., 2002; Gobbini and Haxby, 2007).

## Discussion

This paper has described a model-based spatio-temporal deconvolution approach to source reconstruction. Sources are reconstructed by inverting a forward model comprising a temporal process as well as a spatial process. This approach relies on the fact that EEG and MEG signals are extended in time as well as space.

It rests on the notion that MEG and EEG reflect the neuronal activity of a spatially distributed dynamical system. Depending on the nature of the experimental task, this activity can be highly localized or highly distributed and the dynamics can be more, or less, complex. At one extreme, listening for example to simple auditory stimuli produces brain activations that are highly localized

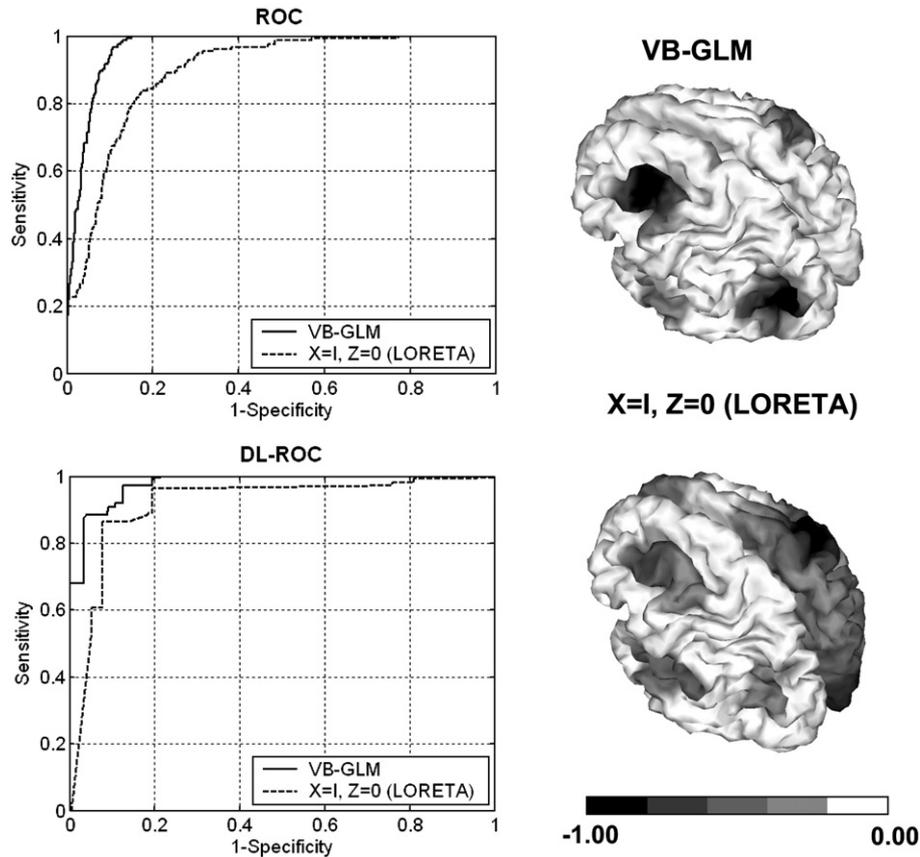


Fig. 10. Spatial effects: VB-GLM vs. instantaneous solution. The left panels show ROC and DL-ROC curves corresponding to VB-GLM and the instantaneous solution for  $t=200$  ms. In this case, the instantaneous solution was obtained by reducing the full VB-GLM approach to the traditional two-level model of Fig. 1 ( $X=I, Z=0$ ). The right panels show the corresponding PCD estimate averaged across trials. The two maps have been normalized to their respective maximum absolute values.

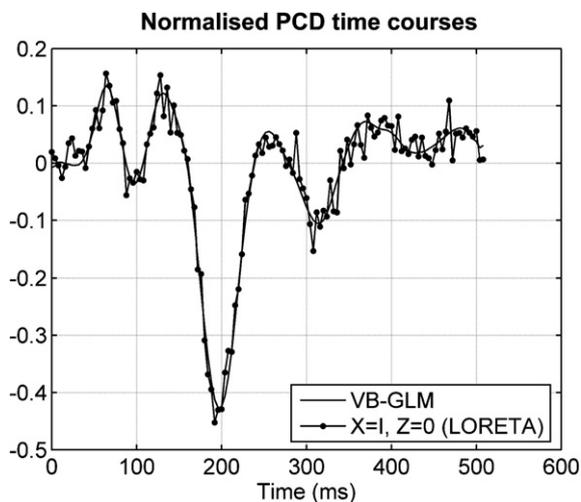


Fig. 11. Temporal effects: VB-GLM vs. instantaneous solution. The figure shows the time course of the VB-GLM and instantaneous estimations for the generator with maximum activation at  $t=200$  ms. The two time courses have been normalized to the maximum absolute value of the true activity. The use of temporal priors leads to smoother estimated time courses.

in time and space. This activity is well described by a single dipole located in brainstem and reflecting a single burst of neuronal activity at, e.g.,  $t=20$  ms post-stimulus. More complicated tasks, such as oddball paradigms, elicit spatially distributed responses and more complicated dynamics that can appear in the ERP as damped sinusoidal responses. In this paper we have taken the view that by explicitly modelling these dynamics one can obtain better source reconstructions.

Compared to previous spatio-temporal models (Baillet and Garnero, 1997; Schmidt et al., 2000; Galka et al., 2004; Yamashita et al., 2004; Daunizeau et al., 2005), our algorithm perhaps embodies stronger dynamic constraints. But the computational simplicity of fitting GLMs, allied to the efficiency of variational inference, results in a relatively fast algorithm. Also, the GLM can accommodate damped sinusoidal and wavelet approaches that are ideal for modelling transient and nonstationary responses.

The dynamic constraints implicit in our model help to regularize the solution. Indeed, with  $M$  sensors,  $G$  sources,  $T$  time points and  $K$  temporal, if  $K < MT/G$  the inverse problem is no longer underdetermined. In practice, however, spatial regularization will still be required to improve estimation accuracy.

The method proposed in the present paper embodies well-known phenomenological descriptions of evoked responses. A similar method has recently been proposed in Friston et al. (2006), but the approaches differ in a number of respects. First, in Friston et al. (2006), scalp data  $Y$  are (effectively) projected onto a temporal basis

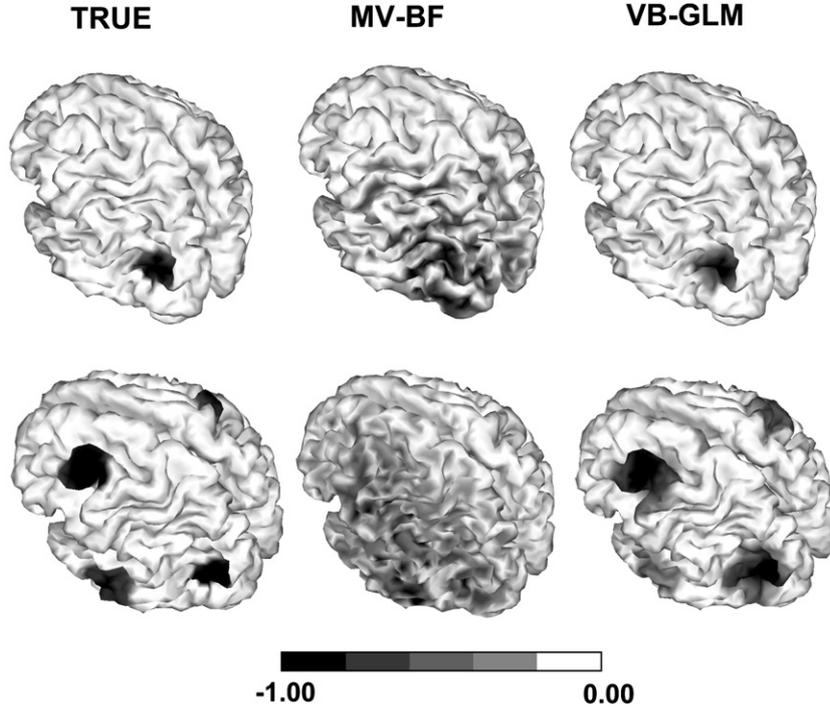


Fig. 12. Qualitative comparison between VB-GLM and MV-BF methods. The figure shows the true PCD spatial distribution for the single source (upper row) and the multiple correlated sources (lower row) at  $t=200$  ms, as well as the corresponding VB-GLM and MV-BF source reconstructions. All maps have been normalized to their respective maximum absolute values.

set  $\mathbf{X}$  and source reconstructions are made in this reduced space. This results in a computationally efficient procedure based on restricted maximum likelihood (ReML), but one in which the between-trial variance is not taken into account. This will result in inferences about  $\mathbf{W}$  and  $\mathbf{J}$  which are overconfident. If one is interested in population inferences based on summary statistics (i.e.,  $\hat{\mathbf{W}}$ ) from a group of subjects, then this does not matter. If, however, one wishes to make within-subject inferences, then VB-GLM is the preferred approach. Second, in Friston et al. (2006), the model has been augmented to account for trial-specific responses. This treats each trial as a “random effect” and provides a method for making inferences about induced responses. The algorithm described in this paper, however, is restricted to treating trials as fixed effects. This mirrors standard first-level analyses of fMRI in which multiple trials are treated by forming concatenated data and design matrices.

### Acknowledgments

Will Penny is supported by the Wellcome Trust and Nelson J. Trujillo-Barreto is supported by a Science Link programme from the British Council. The authors would also like to thank Rik Henson for providing the EEG data, and Karl Friston for discussing similarities and differences between the algorithm in this paper and the approach described in Friston et al. (2006).

## Appendix A

### A.1. Implementation details

A practical difficulty with the update equations for the PCD is that the covariance matrix  $\hat{\Sigma}_{\mathbf{J}}$  is of dimension  $G \times G$ . Even

low-resolution source grids typically contain  $G > 1000$  elements. This therefore presents a problem. A solution is found, however, with use of a singular value decomposition (SVD). First, we define a modified lead-field matrix  $\bar{\mathbf{K}} = \hat{\Omega}^{1/2} \mathbf{K} \hat{\Lambda}^{-1/2}$  and compute its SVD

$$\bar{\mathbf{K}} = \mathbf{U} \mathbf{S} \mathbf{V}^T = \mathbf{U} \bar{\mathbf{V}} \quad (\text{A1})$$

where  $\bar{\mathbf{V}}$  is an  $M \times G$  matrix, the same dimension as the lead field  $\mathbf{K}$ . It can then be shown using the matrix inversion lemma (Golub and Van Loan, 1996) that

$$\hat{\Sigma}_{\mathbf{J}} = \hat{\Lambda}^{-1/2} (\mathbf{I}_G - \mathbf{P}) \hat{\Lambda}^{-1/2} \quad (\text{A2})$$

$$\mathbf{P} = \bar{\mathbf{V}}^T (\mathbf{I}_M + \mathbf{S} \mathbf{S}^T)^{-1} \bar{\mathbf{V}}$$

which is simple to implement computationally, as it only requires inversion and square root of diagonal matrices.

Source estimates can be computed as shown in Eq. (20). In principle, this means the estimated sources over all time points and source locations are given by

$$\hat{\mathbf{J}} = \hat{\Sigma}_{\mathbf{J}} \mathbf{K}^T \hat{\Omega} \mathbf{Y} + \hat{\Sigma}_{\mathbf{J}} \hat{\Lambda} \hat{\mathbf{W}}^T \mathbf{X}^T \quad (\text{A3})$$

In practice, however, it is inefficient to work with such a large matrix during estimation. We therefore do not implement Eqs. (19) and (20) but, instead, work in the reduced space  $\hat{\mathbf{J}}_{\mathbf{x}} = \hat{\mathbf{J}} \mathbf{X}$  which are

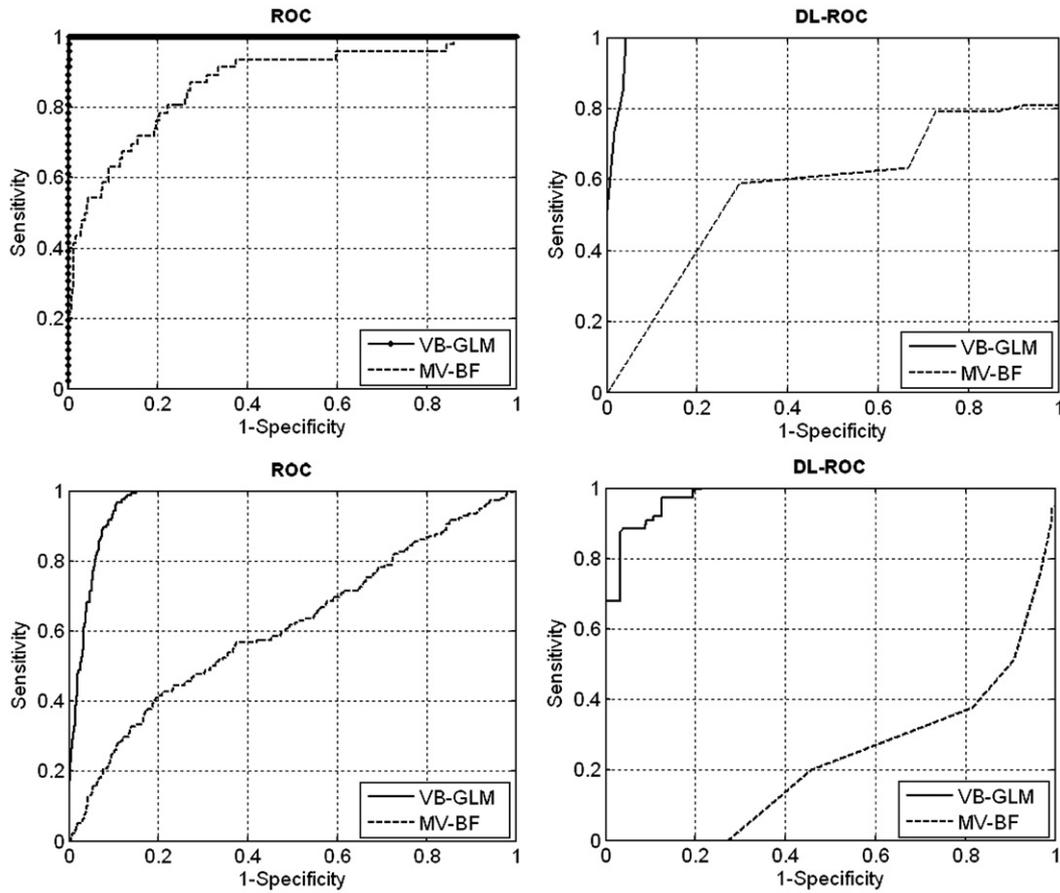


Fig. 13. Quantitative comparison between VB-GLM and MV-BF methods. The figure shows the ROC and DL-ROC curves corresponding to VB-GLM and MV-BF methods for the single-source (upper row) and the multiple correlated sources (lower row) at  $t=200$  ms.

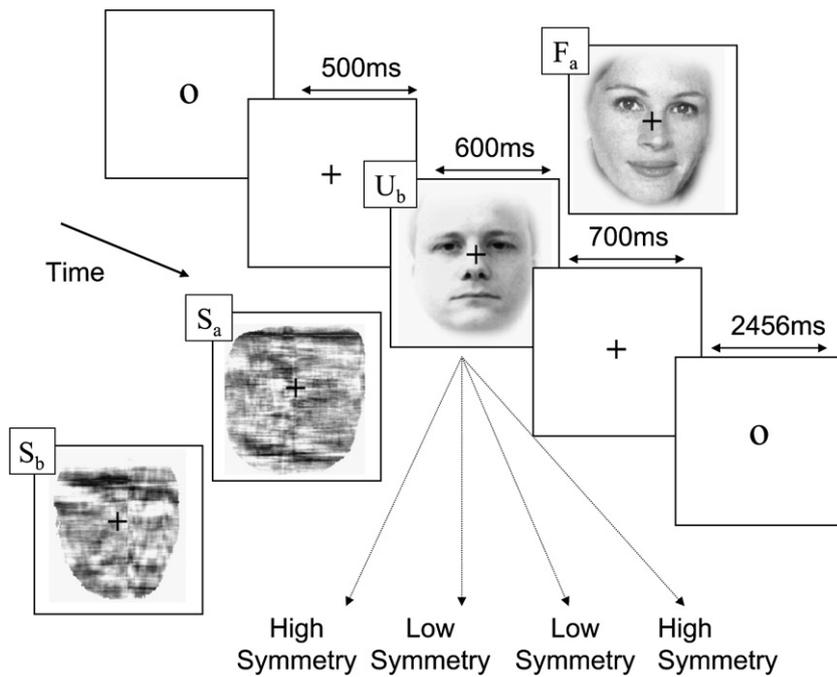


Fig. 14. Description of the face processing experiment. The experiment involved randomized presentation of 86 faces and 86 scrambled faces. Half of the faces were familiar and half unfamiliar, creating three event-types (conditions) in total. The subject had to judge the left–right symmetry of each stimulus (face and scrambled) around an imaginary vertical line through the centre of the image. Faces were presented for 600 ms, every 3600 ms.

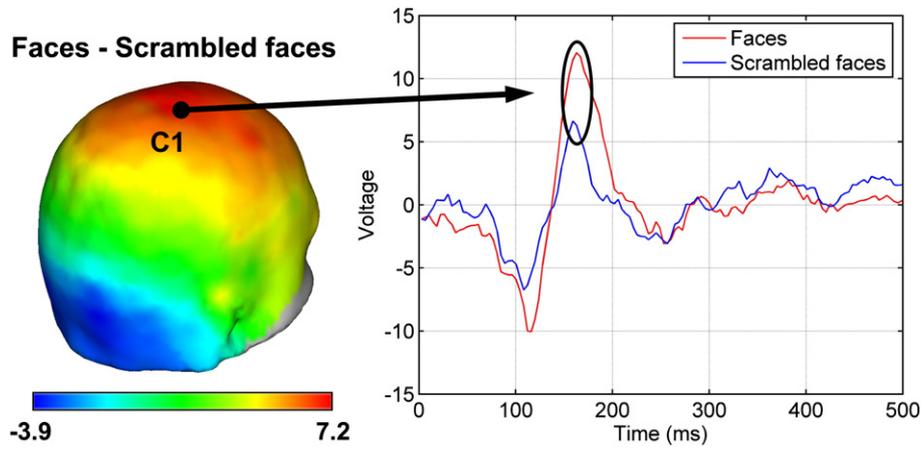


Fig. 15. Face processing ERP. Left panel shows the topographic map of the ERP data at  $t=170$  ms (N170). At this latency, the difference of faces–scrambled faces is maximum. The time courses for the two conditions at electrode C1 are shown in the right panel.

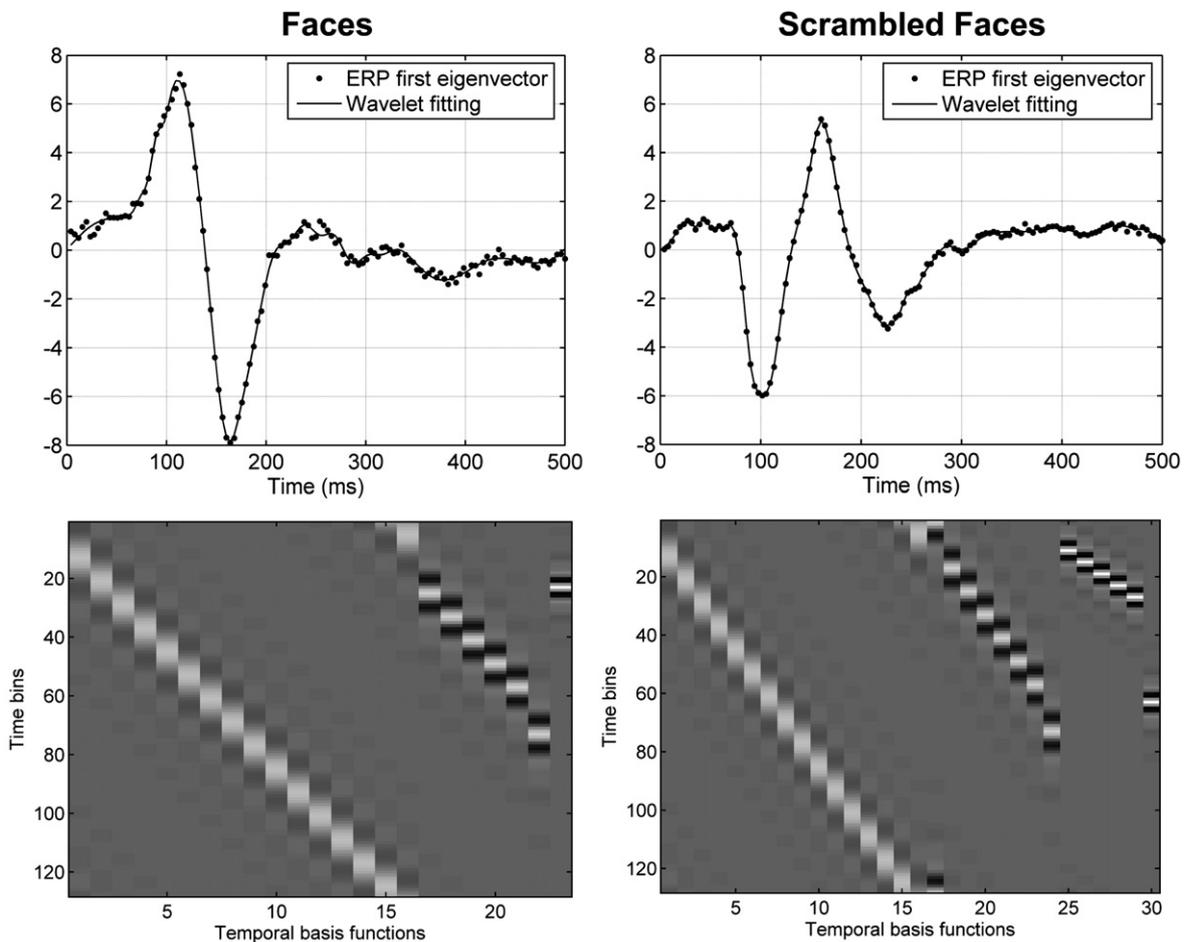


Fig. 16. Definition of the wavelet temporal model used for source reconstruction of the face processing data set. Upper panels: Fitting of the data first eigenvector for conditions faces and scrambled faces, respectively, with  $K=23$  and  $K=30$  Battle-Lemarie wavelets as calculated by wavelet shrinkage. Lower panels: Images of the design matrices (for a single trial) for the two conditions.

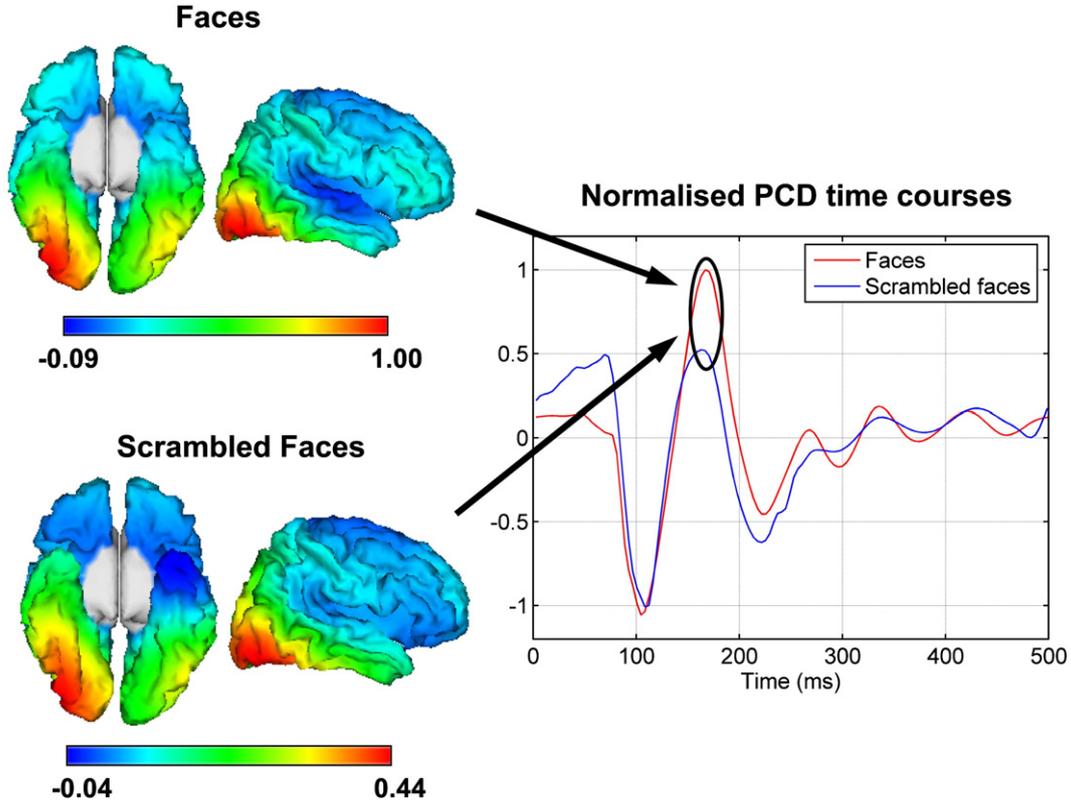


Fig. 17. Left panels show the VB-GLM source estimates for the two conditions at  $t = 170$  ms, averaged across trials. The two images have been normalized with respect to the maximum activity in the faces condition. The right panel shows the normalized average of the source time courses for the generator with maximum activity in the two conditions.

the sources projected onto the design matrix. These projected source estimates are given by

$$\begin{aligned} \hat{\mathbf{J}}_X &= \hat{\mathbf{J}}\mathbf{X} \\ &= \hat{\Sigma}_J \mathbf{K}^T \hat{\Omega} \mathbf{Y} \mathbf{X} + \hat{\Sigma}_J \hat{\Lambda} \hat{\mathbf{W}}^T \mathbf{X}^T \mathbf{X} \\ &= \mathbf{A}_{K\Omega} \mathbf{Y} \mathbf{X} + \mathbf{A}_{\Lambda W} \mathbf{X}^T \mathbf{X} \end{aligned} \quad (\text{A4})$$

where  $\mathbf{Y} \mathbf{X}$  and  $\mathbf{X}^T \mathbf{X}$  can be pre-computed and the intermediate quantities are given by

$$\begin{aligned} \mathbf{A}_{K\Omega} &= \hat{\Sigma}_J \mathbf{K}^T \hat{\Omega} \\ &= (\hat{\Lambda}^{-1} \mathbf{K}^T - \hat{\mathbf{P}} \hat{\Lambda}^{-1/2} \mathbf{K}^T) \hat{\Omega} \\ \mathbf{A}_{\Lambda W} &= \hat{\Sigma}_J \hat{\Lambda} \hat{\mathbf{W}}^T \\ &= (\hat{\mathbf{W}}^T - \hat{\Lambda}^{-1/2} \hat{\mathbf{P}} \hat{\Lambda}^{1/2} \hat{\mathbf{W}}^T) \end{aligned} \quad (\text{A5})$$

Because these matrices are only of dimension  $G \times M$  and  $G \times K$ , respectively,  $\hat{\mathbf{J}}_X$  can be efficiently computed. The term  $\mathbf{X}^T \hat{\mathbf{j}}_g$  in Eq. (24) is then given by the  $g$ th row of  $\hat{\mathbf{J}}_X$ .

The intermediate quantities can also be used to compute model predictions as

$$\begin{aligned} \hat{\mathbf{Y}} &= \mathbf{K} \hat{\mathbf{J}} \\ &= \mathbf{K} \mathbf{A}_{K\Omega} \mathbf{Y} + \mathbf{K} \mathbf{A}_{\Lambda W} \mathbf{X}^T \end{aligned} \quad (\text{A6})$$

The entry  $(m, t)$  in  $\hat{\mathbf{Y}}$  then corresponds to the  $\mathbf{k}_m^T \hat{\mathbf{j}}_t$  term in Eq. (27). Other computational savings are as follows. For Eq. (27), we use the result

$$\mathbf{k}_m^T \hat{\Sigma}_J \mathbf{k}_m = \frac{1}{\hat{\sigma}_m} \sum_{m'=1}^M \frac{s_{m'm'}^2 u_{mm'}^2}{(s_{m'm'}^2 + 1)} \quad (\text{A7})$$

where  $s_{ij}$  and  $u_{ij}$  are the  $(i, j)$ th entries in  $S$  and  $U$ , respectively. For Eq. (26) we use the result

$$\{\hat{\Sigma}_J\}_{gg} = \frac{1}{\hat{\lambda}_g} \left( 1 - \hat{\lambda}_g \sum_{g'=1}^M \frac{s_{g'g'}^2 v_{gg'}^2}{s_{g'g'}^2 + 1} \right) \quad (\text{A8})$$

where  $v_{ij}$  is the  $(i, j)$ th entry in  $\mathbf{V}$ .

#### A.2. Distance based localization receiver operating characteristic (DL-ROC)

Consider a continuous image  $I$  within which a source  $S$  is required to be localized and let  $D$  be the detection region defined by the classifier. In our case, the classifier was defined by declaring a generator to be active (included in the source) if the estimated PCD at the given generator exceeded a specified threshold. The region  $D$  then defines the labelling of each generator  $X$  by the classifier as included ( $X \in D$ ) or not included ( $X \in \bar{D}$ ) in the source.

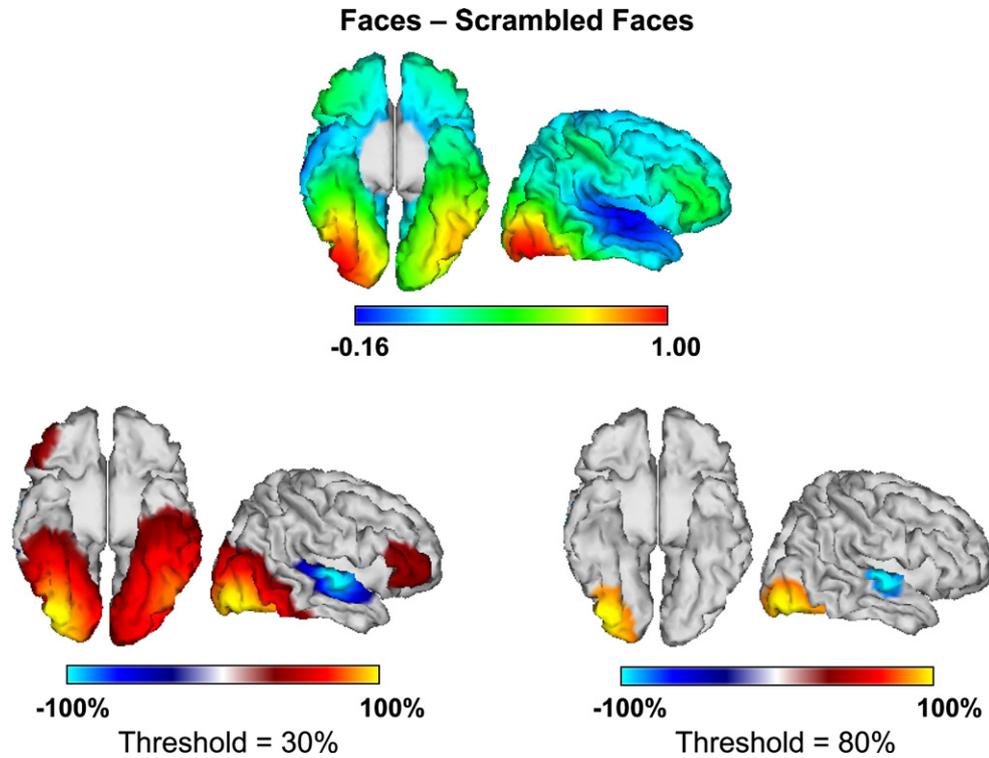


Fig. 18. Overall effect of faces. The upper panel shows the normalized differences of faces minus scrambled faces at  $t = 170$  ms. In the lower panels, positive and negative differences have been normalized to the maximum positive and negative activities, respectively. The images show the sources that survive a 30% (lower left) and an 80% (lower right) threshold.

Now denote by  $A \setminus B$ ,  $A \cup B$  and  $A \cap B$  the difference, union and intersection operations on sets  $A$ ,  $B$  of generators in  $I$ . The sets  $D \setminus S$ ,  $S \setminus D$ ,  $D \cap S$  and  $(I \cap D) \cup (I \cap S)$  are formed by the generators with false positive (FP), false negative (FN), true positive (TP) and true negative (TN) classifications, respectively. Thus, the regions  $D \setminus S$  and  $S \setminus D$  contain all the incorrectly classified generators.

Let  $D_2(X, A)$  be the distance of a generator  $X$  from a (non-empty) region  $A$  defined as:

$$D_2(X, A) = \inf_{Y \in A} d_2(X, Y) \quad (\text{A9})$$

where  $d_2$  is the geodesic distance between generators, and “inf” denotes the greatest lower bound of a set of numbers, or infimum. And let  $d_I$  be the diameter of  $I$ , i.e.,  $d_I = \sup d_2(X, Y)$ , where “sup” denotes the least upper bound, or supremum, with respect to all the pixels  $X$  and  $Y$  of  $I$ . We can then calculate the re-scaled distances  $d(X, Y) = d_2(X, Y)/d_I$  and  $d(X, A) = d_2(X, A)/d_I$ , for any generators  $X$ ,  $Y$  and any region  $A$ . Based on this, supremum measures of false positive (FPLE) and false negative localization error (FNLE), for specified regions  $D$  and  $S$  are then defined as:

$$\text{FPLE}(D, S) = \sup_{X \in D} d(X, S)$$

$$\text{FNLE}(D, S) = \sup_{X \in S} d(X, D) \quad (\text{A10})$$

Average and integral measures of FPLE and FNLE can also be defined (Biscay-Lirio et al., 1992).

For a given source  $S$ , the detection region  $D$  and therefore the measures  $\text{FPLE}(D, S)$  and  $\text{FNLE}(D, S)$  depend on the classifier’s

decision threshold  $C$ , which determines the level of certainty used by the classifier to consider a generator  $X$  as belonging to the source (i.e.,  $X \in D$ ). Then, by analogy with conventional ROC methodology, the variation of the measures of localization error over the range of the classifier’s decision thresholds can be described with the curve  $(\text{FPLE}(C), 1 - \text{FNLE}(C))$ . This is called the distance-based localization receiver operating characteristic (DL-ROC).

## References

- Auranen, T., Nummenmaa, A., Hammalain, M., Jaaskelainen, I., Lampinen, J., Vehtari, A., Sams, M., 2005. Bayesian analysis of the neuromagnetic inverse problem with lp norm priors. *NeuroImage* 26 (3), 870–884.
- Baillet, S., Garnero, L., 1997. A Bayesian approach to introducing anatomofunctional priors in the EEG/MEG inverse problem. *IEEE Trans. Biomed. Eng.* 374–385.
- Baillet, S., Mosher, J.C., Leahy, R.M., November 2001. Electromagnetic brain mapping. *IEEE Signal Process. Mag.* 14–30.
- Beal, M., 2003. Variational algorithms for approximate Bayesian inference. PhD thesis, Gatsby Computational Neuroscience Unit, University College London.
- Biscay-Lirio, R., Galán-García, L., Valdés-Sosa, P., Virués-Alba, T., Neira-Blaquier, L., Rojas-Viggo, J., 1992. Localization error in biomedical imaging. *Comput. Biol. Med.* 22 (4), 277–286.
- Bosch-Bayard, J., Valdés-Sosa, P., Virués-Alba, E., Aubert-Vázquez, E., John, R., Harmony, T., Riera-Díaz, J., Trujillo-Barreto, N., 2001. 3D statistical parametric mapping of variable resolution electromagnetic tomography (VARETA). *Clin. Electroencephalogr.* 32 (2), 47–66.
- Brookes, M., Gibson, A., Hall, S., Furlong, P., Barnes, G., Hillebrand, A., Singh, K., Halliday, I., Francis, S., Morris, P., 2004. A general linear model for MEG beamformer imaging. *NeuroImage* 23 (3), 936–946.

- Buzsaki, G., Draguhn, A., 2004. Neuronal oscillations in cortical networks. *Science* 304, 1926–1929.
- Clyde, M., Parmigiani, G., Vidakovic, B., 1998. Multiple shrinkage and subset selection in wavelets. *Biometrika* 85, 391–402.
- Collins, D.L., Zijdenbos, A.P., Kollokian, V., Sled, J.G., Kabani, N.J., Holmes, Colin J., Evans, Alan C., 1998. Design and construction of a realistic digital brain phantom. *IEEE Trans. Med. Imag.* 17 (3), 463–468.
- Cover, T.M., Thomas, J.A., 1991. *Elements of Information Theory*. John Wiley, p. 22.
- Darvas, F., Pantazis, D., Kucukaltun Yildirim, E., Leahy, R., 2004. Mapping human brain function with MEG and EEG: methods and validation. *NeuroImage* 25, 383–394.
- Daunizeau, J., Mattout, J., Clonda, D., Goulard, B., Benali, H., Lina, J.M., 2005. Bayesian spatio-temporal approach for EEG source reconstruction: conciliating ECD and distributed models. *IEEE Trans. Biomed. Eng.* 53, 503–516.
- David, O., Friston, K.J., 2003. A neural mass model for MEG/EEG: coupling and neuronal dynamics. *NeuroImage* 20 (3), 1743–1755.
- Demiralp, T., Ademoglu, A., I Stefanopoulos, Y., Gulcur, H.O., 1998. Analysis of event-related potentials (ERP) by damped sinusoids. *Biol. Cybern.* 78, 487–493.
- Donoho, D.L., Johnstone, I.M., 1994. Ideal spatial adaptation by wavelet shrinkage. *Biometrika* 81, 425–455.
- Durka, P.J., Martínez-Montes, E., Valdés-Sosa, P., Blinowska, J., 2005. Multichannel matching pursuit and EEG inverse solutions. *J. Neurosci. Methods* 148, 49–59.
- Frackowiak, R.S.J., Friston, K.J., Frith, C., Dolan, R., Price, C.J., Zeki, S., Ashburner, J., Penny, W.D., 2003. *Human Brain Function*, 2nd edition. Academic Press.
- Friston, K.J., Penny, W.D., 2003. Posterior probability maps and SPMs. *NeuroImage* 19 (3), 1240–1249.
- Friston, K., Henson, R., Phillips, C., Mattout, J., 2006. Bayesian estimation of evoked and induced responses. *Hum. Brain Mapp.* 27, 722–735.
- Friston, K.J., Mattout, J., Trujillo-Barreto, N.J., Ashburner, J., Penny, W., 2007. Variational free energy and the Laplace approximation. *NeuroImage* 34 (1), 220–234.
- Fuchs, M., Wagner, M., Kohler, T., Wischman, H.A., 1999. Linear and nonlinear current density reconstructions. *J. Clin. Neurophysiol.* 16 (3), 267–295.
- Galka, A., Yamashita, O., Ozaki, T., Biscay, R., Valdés-Sosa, P., 2004. A solution to the dynamical inverse problem of EEG generation using spatiotemporal Kalman filtering. *NeuroImage* 23 (2), 435–453.
- Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 1995. *Bayesian Data Analysis*. Chapman and Hall, Boca Raton.
- Gobbini, M.I., Haxby, J.V., 2007. Neural systems for recognition of familiar faces. *Neuropsychologia* 45 (1), 32–41.
- Golub, G.H., Van Loan, C.F., 1996. *Matrix Computations*, 3rd edition. John Hopkins Univ. Press.
- Haxby, J.V., Hoffman, E.A., Gobbini, M.I., 2002. Human neural systems for face recognition and social communication. *Biol. Psychiatry* 51, 59–67.
- Henson, R.N.A., Goshen-Gottstein, Y., Ganel, T., Otten, L.J., Quayle, A., Rugg, M.D., 2003. Electrophysiological and hemodynamic correlates of face perception, recognition and priming. *Cereb. Cortex* 13, 793–805.
- Huiskamp, G., 1991. Difference formulas for the surface Laplacian on a triangulated surface. *J. Comput. Phys.* 95, 477–496.
- Kiebel, S.J., Friston, K.J., 2004. Statistical parametric mapping for event-related potentials: II. A hierarchical temporal model. *NeuroImage* 22 (2), 503–520.
- Lappalainen, H., Miskin, J.W., 2000. Ensemble learning. In: Girolami, M. (Ed.), *Advances in Independent Component Analysis*. Springer-Verlag.
- Mattout, J., Phillips, C., Penny, W.D., Rugg, M., Friston, K.J., 2006. MEG source localisation under multiple constraints: an extended Bayesian framework. *NeuroImage* 30 (3), 753.
- Pascual-Marqui, R., 2002. Standardised low resolution electromagnetic tomography (sLORETA): technical details. *Methods Find. Exp. Clin. Pharmacol.* 24, 5–12.
- Pascual-Marqui, R.D., Michel, C.M., Lehman, D., 1994. Low resolution electromagnetic tomography: a new method for localizing electrical activity of the brain. *Int. J. Psychophysiol.* 18, 49–65.
- Flandin, G., Penny, W.D., 2007. Bayesian fMRI data analysis with sparse spatial basis function priors. *NeuroImage* 34 (3), 1108–1125.
- Penny, W.D., Trujillo-Barreto, N.J., Friston, K.J., 2005. Bayesian fMRI time series analysis with spatial priors. *NeuroImage* 24 (2), 350–362.
- Robinson, S., Vrba, J., 1999. *Functional neuroimaging by synthetic aperture magnetometry (SAM)*. Recent Advances in Biomagnetism. Tohoku Univ. Press, Sendai, Japan.
- Rugg, M.D., Coles, M.G.H., 1995. *Electrophysiology of Mind: Event-related Potentials and Cognition*. Oxford Univ. Press.
- Rush, S., Driscoll, D., 1969. EEG electrode sensitivity—an application of reciprocity. *IEEE Trans. Biomed. Eng.* 16 (1), 15–22.
- Sahani, M., Nagarajan, S.S., 2004. Reconstructing MEG sources with unknown correlations. In: Saul, L., Thrun, S., Schoelkopf, B. (Eds.), *Advances in Neural Information Processing Systems*, vol. 16. MIT, Cambridge, MA.
- Schmidt, D.M., George, J.S., Wood, C.C., 1999. Bayesian inference applied to the electromagnetic inverse problem. *Hum. Brain Mapp.* 7, 195–212.
- Schmidt, D.M., Ranken, D.M., George, J.S., Wood, C.C., 2000. Spatial-temporal Bayesian inference for MEG/EEG. 12th International Conference on Biomagnetism, Helsinki, Finland, August.
- Sekihara, K., Nagarajan, S.S., Poeppel, D., Marantz, A., Miyashita, Y., 2001. Reconstructing spatio-temporal activities of neural sources using an MEG vector beamformer technique. *IEEE Trans. Biomed. Eng.* 48, 760–771.
- Tallon-Baudry, C., Bertrand, O., Delpuech, C., Pernier, J., 1996. Stimulus specificity of phase-locked and non phase-locked 40 Hz visual responses in human. *J. Neurosci.* 16 (13), 4240–4249.
- Trejo, L., Shensa, M.J., 1999. Feature extraction of event-related potentials using wavelets: an application to human performance monitoring. *Brain Lang.* 66, 89–107.
- Trujillo-Barreto, N.J., Aubert-Vázquez, E., Valdés-Sosa, P.A., 2004. “Bayesian model averaging in EEG/MEG imaging”. *NeuroImage* 21 (4), 1300–1319.
- Unser, M., Aldroubi, A., 1996. A review of wavelets in biomedical applications. *Proc. IEEE* 84, 626–638.
- Valdés-Sosa, P., Marti, F., Garcia, F., Casanova, R., 2000. Variable resolution electric-magnetic tomography. In: Aine, C.J., Okada, Y., Stroink, G., Switheyby, S.J., Wood, C.C. (Eds.), *Biomag 96’*: Proceedings of the Tenth International Conference on Biomagnetism, vol. II. Springer-Verlag, New York, pp. 373–376.
- Van Veen, B.D., van Drongelen, W., Yuchtman, M., Suzuki, A., 1997. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans. Biomed. Eng.* 44, 867–880.
- Yamashita, O., Galka, A., Ozaki, T., Biscay, R., Valdés-Sosa, P., 2004. Recursive penalised least squares solution for dynamical inverse problems of EEG generation. *Hum. Brain Mapp.* (21), 221–235.