

Bayesian Spatio-Temporal Approach for EEG Source Reconstruction: Conciliating ECD and Distributed Models

Jean Daunizeau*, Jérémie Mattout, Diego Clonda, Bernard Goulard, Habib Benali, and Jean-Marc Lina

Abstract—Characterizing the cortical activity sources of electroencephalography (EEG)/magnetoencephalography data is a critical issue since it requires solving an ill-posed inverse problem that does not admit a unique solution. Two main different and complementary source models have emerged: equivalent current dipoles (ECD) and distributed linear (DL) models. While ECD models remain highly popular since they provide an easy way to interpret the solutions, DL models (also referred to as imaging techniques) are known to be more realistic and flexible. In this paper, we show how those two representations of the brain electromagnetic activity can be cast into a common general framework yielding an optimal description and estimation of the EEG sources. From this extended source mixing model, we derive a hybrid approach whose key aspect is the separation between temporal and spatial characteristics of brain activity, which allows to dramatically reduce the number of DL model parameters. Furthermore, the spatial profile of the sources, as a temporal invariant map, is estimated using the entire time window data, allowing to significantly enhance the information available about the spatial aspect of the EEG inverse problem. A Bayesian framework is introduced to incorporate distinct temporal and spatial constraints on the solution and to estimate both parameters and hyperparameters of the model. Using simulated EEG data, the proposed inverse approach is evaluated and compared with standard distributed methods using both classical criteria and ROC curves.

Index Terms—Bayesian inference, distributed model, ECD, EEG, hybrid, inverse problem, spatio-temporal.

I. INTRODUCTION

SINCE it gives access to the localization and dynamics of cerebral activity sources, neuroimaging has now become a major tool for the investigation of cognitive processes. In contrast to positron emission tomography (PET) and functional

magnetic resonance imaging (fMRI) which measure cerebral metabolic and vascular variations resulting from changes in neuronal activity, electroencephalography (EEG) and magnetoencephalography (MEG) are direct physical measurements of neuronal currents [1]. Moreover, they are the only modalities capable of resolving temporal patterns of neuronal activity in the millisecond range [2], [3]. However, knowing the scalp electric (respectively, magnetic) field topologies does not allow to estimate in a unequivocal manner the underlying current generators. The so-called EEG/MEG inverse problem is indeed known as being mathematically ill-posed: it has no unique solution in the most general unconstrained case.

Two types of approaches have been proposed to address this issue [4]–[6]. Both of them rely on the physical concept of current dipole, which is a plausible bioelectric source model under suitable conditions of temporal synchronization of local neuronal populations firing [7]–[9]. The first approaches, called “dipole fit” methods, consist of estimating the location and orientation of a few number (up to about five) user-defined equivalent current dipoles (ECDs) whose contributions fit the data best [10]–[12]. Each ECD aims at modeling the activity of a whole cortical area and its single temporal dynamic is evaluated together with the other parameters by means of a nonlinear optimization process. This hypothesis of a single local time course is both convenient and fairly realistic since closed neuronal populations are expected to be strongly coherent in time. However, this constraint has a spatial drawback. The ECD position may well approximate the center of mass of the cortical patch, yet it is not able to describe its extent. Indeed, in case of a highly extended cortical area, even the optimal ECD localization may be meaningless. Nevertheless, despite the lack of physical relevance of these dipoles, ECD methods are widely used in the neuroscience research community. The second and more recent so-called “imaging” approaches consist of estimating the amplitude of a predefined dense set of dipoles, typically distributed all over the cortical sheet [13]. Each dipole unit models a neuronal macrocolumn, i.e., a surface element of a putative source. Given this model, estimating the amplitude of each dipole becomes a linear inverse problem, and should allow to localize the activated areas as well as to quantify their spatial extent. The major drawback of such methods comes from the huge number of parameters to be estimated (about hundred times the amount of available data). Thus, the problem remains under-determined and regularization is required to constrain the solution space by some *a priori* information. Most existing inverse techniques hence differ by the nature of

Manuscript received October 26, 2004; revised June 11, 2005. The work of J. Daunizeau was supported by the Association pour la Recherche contre le Cancer. Asterisk indicates corresponding author.

*J. Daunizeau is with UMR 678 Inserm/UPMC, Paris, France and also with the Centre de Recherches Mathématiques, Montréal H3C 3J7 QC, Canada (e-mail: jean.daunizeau@imed.jussieu.fr).

J. Mattout is with the Wellcome Department of Imaging Neuroscience, WC1N 3BG London, U.K. (e-mail: jmmattout@fil.ion.ucl.ac.uk).

D. Clonda and B. Goulard are with the Centre de Recherches Mathématiques, Montréal H3C 3J7QC, Canada (e-mail: clonda@crm.umontreal.ca; goulard@crm.umontreal.ca).

H. Benali is with UMR 678 Inserm/UPMC, Paris 75013, France and also with the Centre de Recherches Mathématiques, Montréal H3C 3J7 QC, Canada (e-mail: habib.benali@imed.jussieu.fr).

J.-M. Lina is with UMR 678 Inserm/UPMC, Paris 75013, France and also with the Centre de Recherches Mathématiques, Montréal H3C 3J7 QC, Canada. He is also with the Ecole de Technologie Supérieure, Montréal H3C 3J7 QC, Canada (e-mail: jmlina@ele.etsmtl.ca).

Digital Object Identifier 10.1109/TBME.2005.869791

the spatial prior they introduce [14]–[25], pointing out that the choice of the regularization scheme also constitutes a crucial issue. However, very few developments have been focusing on temporal constraints. In [17], [26], temporal smoothness has been enforced through a highly nonlinear optimization procedure, but without explicit control of the regularization process. More recently, the temporal issue has been addressed in [27], applying a simplified Kalman filtering, but the proposed algorithm is highly sensitive to its (spatial) initialization.

A hybrid approach has already been proposed in [28], [29]. After a first step of multiple ECD fit, the method consists of a “remapping” process that solves a local inverse problem on each distributed source patch corresponding to one of the ECDs estimated previously. This method aims at taking advantage of both the parsimonious nature of ECD and the anatomophysiological realism of distributed linear (DL) models. Besides, the scalp measurements themselves are only used to initialize the method (the ECD fit step), which is a highly critical step. Moreover, the approach does not deal with any temporal issue, yielding sub-optimal estimates of the activity sources.

In this paper, we propose a hybrid approach that associates a set of ECDs with functionally and anatomically meaningful clusters made of some elementary dipoles of the distributed source model. Following the ECD model, which associates a regional source to a single time course, we then constrain each distributed dipole of a given cluster to have a common temporal dynamic. As a consequence, space (\mathbf{x}) and time (t) variables can be dissociated and the current density (described as a spatio-temporal field $f(\mathbf{x}, t)$) can be rewritten as the product of a spatial and a temporal field: $f(\mathbf{x}, t) = f_{\mathbf{x}}(\mathbf{x}) \cdot f_t(t)$. This simplification of the original DL model leads to a significant diminution of the number of model parameters. The time courses of the active areas as well as their spatial activation profile, which is time invariant, are estimated based on the whole data window. The subsequent Bayesian spatio-temporal approach (BASTA) framework described afterwards enables us to introduce both spatial and temporal specific prior informations in order to provide a realistic estimate of the bioelectromagnetic activity sources.

This paper is organized as follows. The extended source mixing model and the associated spatiotemporal decomposition are described in Section II. In Section III, we detail the proposed Bayesian estimation framework. In Section IV, we describe both the numerical simulations and the evaluation metrics used to compare the proposed approach to three classical inverse estimators [the minimum norm method (MN), the weighted minimum norm method (WMN), and the low-resolution electromagnetic tomography algorithm (LORETA)]. The simulation results are then presented in Section V. Finally, results are discussed together with the proposed methodology in Section VI.

NOTATIONS

In the following, x denotes a real number, \mathbf{x} a vector, and \mathbf{X} a matrix. \mathbf{X}^T , \mathbf{X}_i , X_{ij} and $\text{tr}(\mathbf{X})$ indicate the transposed matrix of \mathbf{X} , the i th vector column of \mathbf{X} , the scalar element of the i th column and j th row of \mathbf{X} and the trace of \mathbf{X} , respectively. \mathbf{I}_n , $\mathbf{e}_{i,n}$, $\mathbf{1}_n$, and $\mathbf{0}_n$ stand for the $n \times n$ identity matrix, the i th

column of the $n \times n$ identity matrix, the $n \times 1$ vector of ones, and the $n \times 1$ null vector, respectively. For any $n \times 1$ vector \mathbf{x} , $\text{diag}(\mathbf{x})$ denotes the $n \times n$ diagonal matrix whose diagonal is \mathbf{x} . \otimes refers to the Kronecker product and ∇ to the nabla operator. “ \propto ” relates two expressions that are proportional. For two variables x and y , $x|y$ stands for “ x given y ” and $p(x)$ for the probability of x . $\mathcal{N}(\mathbf{m}, \mathbf{V})$ is the Gaussian probability density function (pdf) with mean \mathbf{m} and covariance matrix \mathbf{V} ; $G(a, b)$ is the Gamma pdf with shape parameter a and inverse scale parameter b ; $\delta[C(X)]$ is the Dirac delta distribution for the random variable X ($\delta[C(X)]$ is identically null when $C(X) \neq 0$). Given an $a \times b$ matrix \mathbf{X} , let us also define an $ab \times 1$ vector-valued function denoted by $\text{vec}(\mathbf{X})$ such that

$$\text{vec}(\mathbf{X}) = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_b \end{pmatrix}.$$

II. BAYESIAN SPATIO-TEMPORAL APPROACH: THE EXTENDED SOURCE MIXING MODEL

A. Spatio-Temporal Separability

On one hand, solving the EEG/MEG inverse problem within the DL framework amounts to find a unique solution to the following linear system:

$$\mathbf{M} = \mathbf{G}\mathbf{J} + \mathbf{F} \quad (1)$$

where \mathbf{M} indicates the $p \times t$ matrix of measurements ($p \sim 10^2$: number of sensors, t : number of time samples), \mathbf{F} is an additive measurement noise, \mathbf{J} is the $n \times t$ matrix of the unknown time courses of dipoles ($n \sim 10^4$: number of elementary distributed dipoles), and \mathbf{G} is the $p \times n$ gain matrix (forward operator) associated with the position and orientation of the dipoles.

\mathbf{G} is obtained by solving the so-called forward problem [30], [31] for a given set of dipoles with fixed position and orientation (distributed perpendicularly to the cortical surface). Each column \mathbf{G}_j of \mathbf{G} indicates the putative contribution of dipole j to the scalp data (its so-called forward field).

On the other hand, the ECD model assumes that the observed data can be explained by a set of q active extended cortical areas, each of them being modeled by a unique ECD such that

$$\mathbf{M} = \mathbf{A}\mathbf{X} + \mathbf{E} \quad (2)$$

where \mathbf{A} is a $p \times q$ unknown mixing matrix containing the unknown forward fields of the q ECD’s, \mathbf{X} is an also unknown $q \times t$ matrix containing the ECD time courses and \mathbf{E} is an additive noise. By definition, \mathbf{A} remains the same throughout time: it should express the proper time invariant signatures of the extended sources on the scalp. This is the key point to ensure proper estimation and interpretation of the corresponding ECDs.

The hybrid model first consists of linking the ECD (1) and the DL (2) models by simply expressing each ECD forward field \mathbf{A}_i as a weighted sum of each contribution of the elementary dipoles of the DL model

$$\mathbf{A}_i = \sum_{j=1}^n K_{ji} \mathbf{G}_j$$

where K_{ji} is the relative contribution of dipole j to the i th extended source. This can be rewritten, in matrix formulation, as the kernel convolution

$$\mathbf{A} = \mathbf{G}\mathbf{K}. \quad (3)$$

Equation (3) explicitly links the ECD and the DL models. It states that the measurements can be equivalently explained by these two different configurations of current source. Neglecting modeling errors (assuming $\mathbf{F} = \mathbf{E}$), (3) yields a spatio-temporal decomposition of the distributed sources such that

$$\mathbf{J} = \mathbf{K}\mathbf{X} \quad (4)$$

where \mathbf{X} models the temporal characteristics (or variables) of the brain electric activity, while the spatial kernel \mathbf{K} represents its cortical spatial signature.

Based on this spatio-temporal decomposition, our proposed extended source mixing model further constrains the shape of the spatial kernel \mathbf{K} as follows. Assuming a spatial parcelling of the cortical surface into q anatomically and functionally meaningful clusters [32], [33], we associate each parcel $P_i (i = 1, \dots, q)$ with one ECD. To do so, we restrain the support of the i th ECD cortical signature \mathbf{K}_i to the elementary dipoles belonging to the cortical patch P_i , as follows:

$$\mathbf{A}_i = \sum_{j \in P_i} w_j \mathbf{G}_j. \quad (5)$$

\mathbf{K} is, thus, now defined such that

$$\begin{cases} K_{ji} = w_j, & \text{if } j \in P_i \\ K_{ji} = 0, & \text{otherwise} \end{cases}.$$

or

$$\mathbf{K} = \mathbf{W}\mathbf{C} \quad (6)$$

where \mathbf{W} is the $n \times n$ diagonal matrix (whose diagonal is the $n \times 1$ vector \mathbf{w} containing the weights w_j , i.e., $\mathbf{W} = \text{diag}(\mathbf{w})$) and \mathbf{C} is the $n \times q$ matrix describing the cortex parcelling

$$\begin{cases} C_{ji} = 1, & \text{if } j \in P_i \\ C_{ji} = 0, & \text{otherwise} \end{cases}. \quad (7)$$

Equation (3) and (6) are then to be unified as

$$\mathbf{A} = \mathbf{G}\mathbf{W}\mathbf{C}. \quad (8)$$

Equation (8) constitutes the ground of our extended source mixing model. Together with (1) and (2), it reconciliates the ECD and DL models by describing the data as follows:

$$\mathbf{M} = \mathbf{G}\mathbf{W}\mathbf{C}\mathbf{X} + \mathbf{E} \quad (9)$$

where \mathbf{G} (respectively, \mathbf{C}) relates to the known forward operator of the DL model (respectively, a fixed data-driven cortex parcelling) and \mathbf{w} and \mathbf{X} respectively, indicate the spatial and temporal characteristics of the current source densities that remain to be estimated. This leads to the following expression for the parameters of the distributed sources

$$\mathbf{J} = \mathbf{W}\mathbf{C}\mathbf{X}. \quad (10)$$

\mathbf{C} is interpreted as a biresolution matrix, linking the local scale (n elementary dipoles) to the regional scale (q clusters). The vector \mathbf{w} hence expresses the relative current intensity distribution within regions. It describes the spatial profile of each active extended cortical source and is, by definition, time invariant. Therefore, the entire time window data can be used to estimate both the temporal invariant \mathbf{w} and the time course matrix \mathbf{X} .

B. Standard Models as Particular Cases of the Extended Source Mixing Model

It is important to emphasize the way the standard ECD and DL models relate to this proposed general and hybrid framework. We show how the two standard models can be seen as two particular and extreme cases of the extended source mixing model.

The standard ECD model can be derived from the extended source mixing model by assuming the following:

- 1) the very few number q of active sources (ECDs) is fixed *a priori*;
- 2) $n = q$ and $\mathbf{W} = \mathbf{I}_q$: the extent of the source is not modeled.

Similarly, the extended source mixing model appears equivalent to the standard distributed model when the following hold:

- 1) $q = n$ and $\mathbf{C} = \mathbf{I}_n$: each elementary dipole has its own time course;
- 2) the product $\mathbf{J} = \mathbf{K}\mathbf{X} = \mathbf{W}\mathbf{X}$ is directly estimated, without any explicit separation of space and time variables.

C. Identifiability of the Model

As any decomposition technique (e.g., “blind deconvolution”), our model has a hidden characteristic that requires to be highlighted. Indeed, since both \mathbf{w} and \mathbf{X} are unknown, the extended source mixing model described by (9) suffers from nonidentifiability of its parameters. In other words, the same prediction of the measurements may be obtained by either of the two pairs (\mathbf{X}, \mathbf{A}) and $(\tilde{\mathbf{X}}, \tilde{\mathbf{A}})$, that would be linked as follows:

$$E[\mathbf{M}] = \mathbf{A}\mathbf{X} = \mathbf{A}\mathbf{V}^{-1}\mathbf{V}\mathbf{X} = \tilde{\mathbf{A}}\tilde{\mathbf{X}}$$

where \mathbf{V} is a $q \times q$ matrix. When \mathbf{V} is any invertible matrix (e.g. permutation matrix), this type of nonidentifiability is often referred to as “factor indeterminacy” in the context of factor analysis [34]. In our case, since \mathbf{A} is of the form given by (8), \mathbf{V} is restricted to any nonzero $q \times q$ diagonal matrix. As a consequence, we have to choose the power balance between the \mathbf{w} map and the ECDs time courses matrix \mathbf{X} . Among all possibilities, a convenient solution consists of enforcing the normalization of one of these two terms in the estimation procedure, allowing the other one to express the relative power between the extended sources (see Section III).

D. Model Insights

In order to introduce spatial and temporal specific constraints in the estimation, we need to rewrite the previous extended source mixing model in a more practical way. Applying the “vec” operator to the left and right sides of (9) yields two equivalent formulations of the extended source mixing model

$$\text{vec}(\mathbf{M}) = (\mathbf{I}_t \otimes \mathbf{A})\text{vec}(\mathbf{X}) + \text{vec}(\mathbf{E}) \quad (11)$$

$$= \mathbf{D}\mathbf{w} + \text{vec}(\mathbf{E}) \quad (12)$$

where \mathbf{D} is the $pt \times n$ forward operator of \mathbf{w} whose j th column is given by

$$\mathbf{D}_j = [\mathbf{CX}]_j^T \otimes \mathbf{G}_j \quad (13)$$

where $[\mathbf{CX}]_j^T$ stands here for the transposed j th row of the matrix \mathbf{CX} and corresponds to the time course of the j th elementary dipole. Note that we benefited from the diagonal nature of \mathbf{W} in order to derive (12). Equations (11) and (12) emphasize the temporal and spatial aspects of the model, respectively. Either (11) or (12) will be considered for estimating parameters of the extended source mixing model.

First, note that depending on the length of the data window, the ratio between the size of available data and the number of parameters to be estimated changes. Indeed, the estimation of the whole set of parameters of the extended source mixing model becomes overdetermined (which is also a feature of ECD models) as soon as $t \geq (n/(p-q))$. However, this does not state that this system of equations is well-posed. For instance, the estimation of \mathbf{w} given the time courses matrix \mathbf{X} is well-posed if and only if

$$\begin{cases} t \geq \frac{n}{p} \\ \mathbf{D} \text{ is full rank.} \end{cases} \quad (14)$$

The first statement is conditional on the length of the data window. The second condition depends on the putative linear relationship between various columns of \mathbf{D} , each of them being associated with a particular elementary dipole of the DL model. The linear independence between columns of \mathbf{D} that correspond to two different patches can be brought by a time course contrast between the corresponding ECDs. Nevertheless, within a patch, the term $[\mathbf{CX}]_j^T$ is constant. Therefore, the conditioning of \mathbf{D} directly relies upon the linear independence of the forward fields \mathbf{G}_j of dipoles belonging to the same cluster, which cannot be assured in general. In brief, the temporal behavior of the active sources has consequences on the efficiency of the estimation of \mathbf{w} .

III. THE BAYESIAN SPATIO-TEMPORAL APPROACH: ESTIMATION FRAMEWORK

As mentioned above, the issue of estimating the parameters of the extended source mixing model is not-in general-well-posed, and hence requires prior information. Therefore, we derive a Bayesian statistical inference framework through the use of a dedicated hierarchical model [35], [36]. This is afforded by specifying a prior pdf to \mathbf{X} and \mathbf{w} that contains a hyperparameter vector, whose own prior pdf is fixed. Having dealt with the issue of identifiability of the parameters, we may reasonably assume that the mass of the joint posterior pdf is concentrated

about a single dominant mode. Then, the primary objective of Bayesian probabilistic inference may be focused on characterizing the location and the extent of the highest posterior density region.

We propose to learn about the parameters at both levels of inference by maximizing their joint posterior pdf. Assuming that \mathbf{X} and \mathbf{w} are *a priori* independent (which means that, in the absence of data, one cannot predict anything about \mathbf{X} given \mathbf{w} and reciprocally), the chain rule gives

$$p(\mathbf{w}, \mathbf{X}, \alpha, \beta, \gamma | \mathbf{M}) \propto p(\mathbf{M} | \mathbf{w}, \mathbf{X}, \alpha) \cdot p(\mathbf{X} | \beta) \cdot p(\mathbf{w} | \gamma) \cdot p(\alpha) \cdot p(\beta) \cdot p(\gamma) \quad (15)$$

where the following hold.

- $p(\mathbf{w}, \mathbf{X}, \alpha, \beta, \gamma | \mathbf{M})$ is the joint posterior pdf of \mathbf{w} , \mathbf{X} and a set of mutually independent hyperparameters (α, β, γ) . The maximum *a posteriori* (MAP) estimate of \mathbf{w} , \mathbf{X} , α , β and γ maximizes this quantity.
- $p(\mathbf{M} | \mathbf{w}, \mathbf{X}, \alpha)$ is the likelihood of the data given the model parameters (\mathbf{w} and \mathbf{X}) and hyperparameter α . It is derived from the assumption of independent identically distributed (i.i.d.), 0-mean Gaussian noise for \mathbf{E}

$$p(\mathbf{M} | \mathbf{w}, \mathbf{X}, \alpha) \propto \alpha^{\frac{pt}{2}} \exp \left(-\frac{\alpha}{2} \sum_{j=1}^t \|\mathbf{M}_j - \mathbf{GWCX}_j\|^2 \right) \quad (16)$$

with α being the inverse of the noise variance $\sigma_{\mathbf{E}}^2$.

- $p(\mathbf{X} | \beta)$ (respectively, $p(\mathbf{w} | \gamma)$) is the prior pdf on \mathbf{X} (respectively, \mathbf{w}) conditional on hyperparameter β (respectively, γ) (see Sections III-A and III-B).
- $p(\alpha)$, $p(\beta)$ and $p(\gamma)$ refer to a level of higher inference, and embody the priors on the hyperparameters themselves (see Section III-C).

According to (15), the MAP estimate of \mathbf{w} , \mathbf{X} , α , β and γ can be derived by the iterative maximization of the conditional pdf of each single variable (see Appendix I for convergence proof). Then, the i th iteration of the proposed algorithm can be summarized as follows:

$$\begin{cases} \mathbf{X}_{[i]} = \underset{\mathbf{X}}{\text{argmax}} [p(\mathbf{M} | \mathbf{w}_{[i-1]}, \mathbf{X}, \alpha_{[i-1]}) \cdot p(\mathbf{X} | \beta_{[i-1]})] \\ \mathbf{w}_{[i]} = \underset{\mathbf{w}}{\text{argmax}} [p(\mathbf{M} | \mathbf{w}, \mathbf{X}_{[i]}, \alpha_{[i-1]}) \cdot p(\mathbf{w} | \gamma_{[i-1]})] \\ \alpha_{[i]} = \underset{\alpha}{\text{argmax}} [p(\mathbf{M}, \mathbf{X}_{[i]}, \mathbf{w}_{[i]}, \alpha) \cdot p(\alpha)] \\ \beta_{[i]} = \underset{\beta}{\text{argmax}} [p(\mathbf{X}_{[i]} | \beta) \cdot p(\beta)] \\ \gamma_{[i]} = \underset{\gamma}{\text{argmax}} [p(\mathbf{w}_{[i]} | \gamma) \cdot p(\gamma)] \end{cases} \quad (17)$$

where $\mathbf{Y}_{[i]}$ denotes the estimate of variable \mathbf{Y} at the i th iteration and assuming a known initial set $(\mathbf{X}_{[0]}, \mathbf{w}_{[0]}, \alpha_{[0]}, \beta_{[0]}, \gamma_{[0]})$ (see Section III-D).

A. Temporal Prior

Once the temporal and spatial characteristics of the sources have been separated, we can introduce specific temporal priors using explicit constraints on matrix \mathbf{X} . A first and strong constraint lies in the above-mentioned normalization, which enables us to then estimate a unique set of parameters \mathbf{X} and \mathbf{w} (see Section II-C). Let us define the $q \times qt$ operator \mathbf{N} such

that $\mathbf{Nvec}(\mathbf{X})$ is a $q \times 1$ vector whose elements are the square L_2 -norm of the time course of each ECD

$$\mathbf{Nvec}(\mathbf{X}) = \begin{pmatrix} \sum_{j=1}^t X_{1j}^2 \\ \vdots \\ \sum_{j=1}^t X_{ij}^2 \\ \vdots \\ \sum_{j=1}^t X_{qj}^2 \end{pmatrix}.$$

Hence the normalization constraint is written as

$$\mathbf{Nvec}(\mathbf{X}) = \mathbf{1}_q. \quad (18)$$

Note that \mathbf{N} depends on \mathbf{X} . The i th row of \mathbf{N} can be expressed as

$$\mathbf{N}_i^\top = \mathbf{X}_i^\top \otimes \mathbf{e}_{q,i}^\top \quad (19)$$

where \mathbf{X}_i^\top denotes the i th row of the matrix \mathbf{X} .

In addition to that strong constraint, we may also introduce soft temporal constraints through the prior pdf $p(\mathbf{X}|\beta)$. Since EEG is known to provide an over-sampled measure of the bioelectric activity [7], we here propose a temporal smoothness prior. This is performed by enforcing the second derivative of the time courses of the ECD to be minimal. Precisely, let us denote by \mathbf{T} the discrete second derivative operator of $\text{vec}(\mathbf{X})$ such that $\mathbf{Tvec}(\mathbf{X})$ is the $qt \times 1$ vector containing the second temporal derivative of $\text{vec}(\mathbf{X})$, as follows:

$$\begin{cases} \mathbf{T}_{ii} = -2 \\ \mathbf{T}_{ij} = 1, & \text{if } j = i \pm (q+1) \\ \mathbf{T}_{ij} = 0, & \text{otherwise.} \end{cases} \quad (20)$$

Our assumed prior is such that $\mathbf{Tvec}(\mathbf{X})$ behaves as an i.i.d, 0-mean Gaussian random variable

$$\mathbf{Tvec}(\mathbf{X}) \sim \mathcal{N}\left(\mathbf{0}_{qt}, \frac{1}{\beta} \mathbf{I}_{qt}\right) \quad (21)$$

where β is the inverse variance of the temporal second derivative of \mathbf{X} .

Taking the constraints for both the normalization and the temporal smoothness into account yields the following prior law on \mathbf{X}

$$p(\mathbf{X}|\beta) \propto \beta^{\frac{qt}{2}} \exp\left(-\frac{\beta}{2} \|\mathbf{Tvec}(\mathbf{X})\|^2\right) \delta[\mathbf{Nvec}(\mathbf{X}) - \mathbf{1}_q] \quad (22)$$

where \mathbf{N} is defined by (19). Then the conditional posterior pdf of \mathbf{X} at iteration i is derived from (16) and (17). It is expressed as

$$\begin{aligned} & p(\mathbf{X}|\mathbf{M}, \mathbf{w}_{[i-1]}, \alpha_{[i-1]}, \beta_{[i-1]}) \\ & \propto p(\mathbf{M}|\mathbf{w}_{[i-1]}, \alpha_{[i-1]}, \mathbf{X}) \cdot p(\mathbf{X}|\beta_{[i-1]}) \propto \alpha_{[i-1]}^{\frac{qt}{2}} \beta_{[i-1]}^{\frac{qt}{2}} \\ & \times \exp\left(-\frac{\alpha_{[i-1]}}{2} \sum_{j=1}^t \|\mathbf{M}_j - \mathbf{G}\mathbf{W}_{[i-1]}\mathbf{C}\mathbf{X}_j\|^2\right) \\ & \quad - \frac{\beta_{[i-1]}}{2} \|\mathbf{Tvec}(\mathbf{X})\|^2 \delta[\mathbf{Nvec}(\mathbf{X}) - \mathbf{1}_q]. \end{aligned} \quad (23)$$

Due to the nonlinearity of the normalization constraint, it is not possible to derive an analytic expression of the condi-

tional MAP estimate of \mathbf{X} . We hence use a Newton optimization scheme [37] involving a vector series $\mathbf{X}_{[i,k]}$ that converges toward the normalized solution $\mathbf{X}_{[i]} = \lim_{k \rightarrow \infty} \mathbf{X}_{[i,k]}$ (see Appendix III for the technical details). At iteration i , the $\mathbf{X}_{[i]}$ normalized MAP estimate is calculated using the following linked vector series:

$$\begin{cases} \text{vec}(\mathbf{X}_{[i,k+1]}) = \text{vec}(\mathbf{X}_{[i,k]}^*) \\ \quad - \mathbf{B}_{[i,k]} \mathbf{N}_{[i,k]}^\top (\mathbf{A}_{[i,k+1]} - \mathbf{A}_{[i,k]}) \\ \mathbf{A}_{[i,k+1]} - \mathbf{A}_{[i,k]} = \left(\mathbf{N}_{[i,k]} \mathbf{B}_{[i,k]} \mathbf{N}_{[i,k]}^\top \right)^{-1} \\ \quad \times \left[\mathbf{N}_{[i,k]} \left(\text{vec}(\mathbf{X}_{[i,k]}^*) - \frac{1}{2} \text{vec}(\mathbf{X}_{[i,k]}) \right) - \frac{1}{2} \mathbf{1}_q \right] \end{cases} \quad (24)$$

where $\mathbf{A}_{[i,k]}$ is the estimated Lagrange parameter vector (cf Appendix B), $\mathbf{N}_{[i,k]}$ is the operator defined at (19) and evaluated for $\mathbf{X}_{[i,k]}$ and

$$\begin{cases} \text{vec}(\mathbf{X}_{[i,k]}^*) = \mathbf{B}_{[i,k]} \left(\mathbf{I}_t \otimes \mathbf{A}_{[i-1,k]}^\top \right) \text{vec}(\mathbf{M}) \\ \mathbf{B}_{[i,k]} = \left(\mathbf{I}_t \otimes \mathbf{A}_{[i-1,k]}^\top \mathbf{A}_{[i-1,k]} \right. \\ \quad \left. + \frac{\beta_{[i-1]}}{\alpha_{[i-1]}} \mathbf{T}^\top \mathbf{T} + \mathbf{I}_t \otimes \text{diag}(\mathbf{A}_{[i,k]}) \right)^{-1} \end{cases} \quad (25)$$

Note that \mathbf{X}^* indicates the unconstrained estimate of \mathbf{X} . Indeed, in the absence of the normalization constraint, \mathbf{B} would correspond to the covariance matrix of the posterior pdf of \mathbf{X} and whose eigenvalues would have been translated (regularized) by \mathbf{A} . Then, $\mathbf{X}_{[i]}$ is simply built as the nonconstrained estimate, perturbed by the application of the constraint correction term $\mathbf{B}\mathbf{N}^\top \mathbf{A}$.

B. Spatial Prior

Spatial constraints now concern the time invariant variable \mathbf{w} . We here consider a first-order Markov random field to introduce specific spatial priors on the source spatial extent. More precisely, the prior assumption is such that coefficient w_j for dipole j can be approximated by the mean of the coefficients $w_{j'}$ in the first-order neighborhood V_j of dipole j

$$w_j - \frac{1}{\text{card}[V_j]} \sum_{j' \in V_j} w_{j'} \sim \mathcal{N}\left(0, \frac{1}{\gamma}\right) \quad (26)$$

where $\text{card}[V_j]$ is the number of elements of the set V_j , γ is such that $1/\gamma(\mathbf{S}^\top \mathbf{S})^{-1}$ is the prior covariance matrix of \mathbf{w} and the $n \times n$ matrix \mathbf{S} is given by

$$\begin{cases} \mathbf{S}_{jj} = 1 \\ \mathbf{S}_{jj'} = -\frac{1}{\text{card}[V_j]}, & \text{if } j' \in V_j \\ \mathbf{S}_{jj'} = 0, & \text{otherwise} \end{cases} \quad (27)$$

It is then straightforward to derive the following Gaussian spatial prior law on \mathbf{w} :

$$p(\mathbf{w}|\gamma) \propto \gamma^{\frac{n}{2}} \exp\left(-\frac{\gamma}{2} \|\mathbf{S}\mathbf{w}\|^2\right). \quad (28)$$

Then using (16) and (17) yields the conditional posterior pdf of \mathbf{w} at the i th iteration

$$\begin{aligned} & p(\mathbf{w}|\mathbf{M}, \mathbf{X}_{[i]}, \alpha_{[i-1]}, \gamma_{[i-1]}) \\ & \propto p(\mathbf{M}|\mathbf{X}_{[i]}, \mathbf{W}, \alpha_{[i-1]}) \cdot p(\mathbf{w}|\gamma_{[i-1]}) \propto \alpha_{[i-1]}^{\frac{qt}{2}} \gamma_{[i-1]}^{\frac{n}{2}} \\ & \quad \times \exp\left(-\frac{\alpha_{[i-1]}}{2} \|\text{vec}(\mathbf{M}) - \mathbf{D}_{[i]}\mathbf{w}\|^2 - \frac{\gamma_{[i-1]}}{2} \|\mathbf{S}\mathbf{w}\|^2\right) \end{aligned} \quad (29)$$

where \mathbf{D} is defined by (13). Finally, the analytic expression of the corresponding conditional MAP estimate is given by

$$\mathbf{w}_{[i]} = \left(\mathbf{D}_{[i]}^T \mathbf{D}_{[i]} + \frac{\gamma_{[i-1]} \mathbf{S}^T \mathbf{S}}{\alpha_{[i-1]}} \right)^{-1} \mathbf{D}_{[i]}^T \text{vec}(\mathbf{M}). \quad (30)$$

C. Estimating the Hyperparameters

The hyperparameters are also updated at each iteration by maximizing their conditional posterior pdf. We here consider an empirical Bayesian estimation based on both the likelihood of the model and specific prior laws.

Since all the hyperparameters should be positive, we consider Gamma laws $G(a, b)$ as pdf priors, where the shape parameter a remains constant over hyperparameters.

We first need to specify the prior pdf $G(a, b)$ of the inverse noise variance α . Let \mathbf{M}_0 be a $p \times t_0$ data matrix that only contains background noise.¹ Using the previous definition of the likelihood and an initial Jeffrey's prior law on σ^2 (principle of scale parameter equivalence [38]), it can be shown that the initial posterior pdf of σ^2 is an inverse Gamma law whose parameters are

$$\begin{cases} a = \frac{pt_0}{2} \\ b_1 = \frac{\text{tr}(\mathbf{M}_0^T \mathbf{M}_0)}{2} \end{cases} \quad (31)$$

By definition, these are also the parameters of the corresponding Gamma law of α . In the BASTA estimation framework, we, therefore, define the prior pdf of α using those pre-estimated parameters.²

Regarding the two other hyperparameters, we *a priori* enforce a balance weighting of the likelihood and the regularization terms in the first conditional estimate of \mathbf{X} and \mathbf{w} . This sets the prior expectation of these hyperparameters and hence yields the following estimate of parameter b_2 (respectively, b_3) of the Gamma prior pdf associated with β (respectively, γ)

$$\begin{cases} b_2 = b_1 \frac{\text{tr}(\mathbf{T}^T \mathbf{T})}{t \cdot \text{tr}(\mathbf{A}_{[0]}^T \mathbf{A}_{[0]})} \\ b_3 = b_1 \frac{\text{tr}(\mathbf{S}^T \mathbf{S})}{\text{tr}(\mathbf{D}_{[0]}^T \mathbf{D}_{[0]})} \end{cases} \quad (32)$$

Then, using (16), (17), (22), (28), and assuming that $\mathbf{X}_{[i]}$ does satisfy the normalization constraint, the conditional posterior pdf of the three hyperparameters are given by

$$\begin{cases} p(\alpha | \mathbf{M}, \mathbf{X}_{[i]}, \mathbf{w}_{[i]}) \propto \alpha^{\frac{pt}{2} + a - 1} \\ \quad \times \exp\left(-\alpha \left(\frac{\text{tr}((\mathbf{M} - \mathbf{G}\mathbf{W}_{[i]}\mathbf{C}\mathbf{X}_{[i]})^2)}{2} + b_1 \right)\right) \\ p(\beta | \mathbf{X}_{[i]}) \propto \beta^{\frac{qt}{2} + a - 1} \exp\left(-\beta \left(\frac{\|\mathbf{T}\text{vec}(\mathbf{X}_{[i]})\|^2}{2} + b_2 \right)\right) \\ p(\gamma | \mathbf{w}_{[i]}) \propto \gamma^{\frac{nt}{2} + a - 1} \exp\left(-\gamma \left(\frac{\|\mathbf{S}\mathbf{w}_{[i]}\|^2}{2} + b_3 \right)\right) \end{cases} \quad (33)$$

¹Note that \mathbf{M}_0 could even be approximated by the prestimulus data of the considered data time window.

²The further refinement of α can be seen as a learning algorithm that updates our knowledge of the system.

Finally, the corresponding conditional MAP estimates are given by

$$\begin{cases} \alpha_{[i]} = \frac{pt+2a-2}{\text{tr}((\mathbf{M} - \mathbf{G}\mathbf{W}_{[i]}\mathbf{C}\mathbf{X}_{[i]})^2) + 2b_1} \\ \beta_{[i]} = \frac{qt+2a-2}{\|\mathbf{T}\text{vec}(\mathbf{X}_{[i]})\|^2 + 2b_2} \\ \gamma_{[i]} = \frac{nt+2a-2}{\|\mathbf{S}\mathbf{w}_{[i]}\|^2 + 2b_3} \end{cases} \quad (34)$$

D. Initialization

We initialized \mathbf{X} at iteration 0 in the least informative way, i.e., each ECD time course is flat: $\mathbf{X}_{[0]} = (1/\sqrt{(t)})\mathbf{1}_{q \times t}$.

$\mathbf{X}_{[0]}$ being defined (and normalized), the next step consists of estimating $\mathbf{w}_{[0]}$ knowing $\mathbf{X}_{[0]}$. Finally, we set the initial value of each hyperparameter to its prior expectation.

E. Assessing the Uncertainty of the MAP Estimates

Once the mode of the posterior pdf has been derived, one may be concerned in quantifying the uncertainty associated to this point estimate of the model parameters. For jointly distributed Gaussian variables, the posterior covariance matrix contains all information required to build some inference. Therefore, we chose to use the Laplace approximation of the joint pdf, since it provides us with a fair estimation of the joint covariance matrix of (\mathbf{w}, \mathbf{X}) . The derivation of the approximated joint covariance matrix using Laplace's method is given in Appendix II.

IV. SIMULATION AND EVALUATION PROCEDURES

Numerical Monte-Carlo simulations were performed in order to evaluate the proposed approach and to compare it with the following three well-known inverse estimators.

- LORETA [15]'s solution minimizes a smoothness criterion (minimum Laplacian field) under the constraint $\mathbf{M} = \mathbf{G}\mathbf{J}$. Its analytical expression is

$$\hat{\mathbf{J}}_{\text{LORETA}} = (\mathbf{L}^T \mathbf{L})^{-1} \mathbf{G}^T (\mathbf{G}(\mathbf{L}^T \mathbf{L})^{-1} \mathbf{G}^T)^+ \mathbf{M} \quad (35)$$

where $+$ is the pseudo-inverse operator and \mathbf{L} is a discrete Laplacian estimator (constrained onto the three-dimensional (3-D) cortical surface, then the method is also referred to as cLORETA [25])

$$\begin{cases} \mathbf{L}(j, j) = \sum_{j' \in V_j} \frac{1}{d(j, j')} \\ \mathbf{L}(j, k) = -\frac{1}{d(j, k)}, & \text{if } k \in V_j \\ \mathbf{L}(j, k) = 0, & \text{otherwise} \end{cases} \quad (36)$$

where $d(j, k)$ is the Euclidian distance between dipole j and dipole k .

- The MN method's solution is the one that fits the data best under the constraint of a minimized L_2 -norm [14]. Its analytic expression is

$$\hat{\mathbf{J}}_{\text{MN}} = \mathbf{G}^T (\mathbf{G}\mathbf{G}^T + \alpha \mathbf{I}_p)^{-1} \mathbf{M} \quad (37)$$

where α is evaluated using the L-curve heuristic [16], [39].

- The WMN, which is a variant of the MN method, compensates the bias of deep sources MN estimates using the following weighting procedure [14]:

$$\hat{\mathbf{J}}_{\text{WMN}} = \mathbf{F}^{-1} \mathbf{G}^{\top} (\mathbf{G} \mathbf{F}^{-1} \mathbf{G}^{\top} + \alpha \mathbf{I}_p)^{-1} \mathbf{M} \quad (38)$$

where \mathbf{F} is a $n \times n$ diagonal matrix such that $F_{ii} = \mathbf{G}_i^{\top} \mathbf{G}_i$, and α is evaluated using the L-curve heuristic.

A. Monte-Carlo Simulations

Since the EEG sources are widely believed to be restricted to the pyramidal neuron cells of the cortical strip [1], a common approach within the distributed model framework consists of constraining the dipoles to be distributed onto the cortical surface extracted from a structural magnetic resonance imaging (MRI) volume [13]. Following the segmentation of the MRI volume, dipoles are typically located at each node of a triangular mesh of the white/grey matter interface [40]. Furthermore, since the apical dendrites of these cortical neurons are organized perpendicularly to the surface, the corresponding dipoles are often constrained to have this particular orientation.

To simulate EEG data on 59 sensors uniformly distributed over the scalp, we considered a 3-D cortical mesh made of 3050 vertices uniformly spread over the two hemispheres. The corresponding forward operator \mathbf{G} was computed based on a three-sphere analytic model provided by the BrainStorm software [41]. The corresponding simulated data were generated using the forward operator \mathbf{G} . Then, a zero-mean, i.i.d. and Gaussian noise \mathbf{E} was added in order to produce the final simulated measures.

We ran two series of simulations, dedicated to evaluate the influence of the noise level (SNR-series), and the number of active sources (NS-series), respectively.

- SNR-series: four levels of SNRs were considered ($SNR = 1, 3, 5, 10$). The SNR was defined as the variance ratio $SNR = (\sigma_M / \sigma_E)$, where σ_M and σ_E are the variance of the induced signal without noise, and of the noise, respectively. For each SNR level, 40 situations were simulated. For each simulation, two extended sources (around 2.5 cm^2 each) were randomly chosen from the whole cortical surface. Each source had a half-sine time course, with a temporal offset of $\Delta t_i = i \cdot (T/4)$, where i is the index of the source ($i = 1, 2$) and T is the period of its sine time course.
- NS-series: five numbers of active sources were considered ($NS = 1, 2, 3, 4, 5$). In these series, the SNR was set to $SNR = 1.3$. For each number of active sources NS , 40 situations were simulated. For each simulation, the NS extended sources (around 2.5 cm^2 each) were randomly chosen from the whole cortical surface. Each source had a half-sine time course, with a temporal offset of $\Delta t_i = i \cdot (T/4)$, where i is the index of the source ($i = 1, \dots, NS$) and T is the period of its sine time course.

In all simulations, the active sources were chosen without considering the cortical patches, i.e., they might (or not) coincide with the cortical parcelling underlying the extended source mixing model. The Fig. 1 depicts an illustrative example of a simulation configuration with $NS = 2$ and $SNR = 5$.

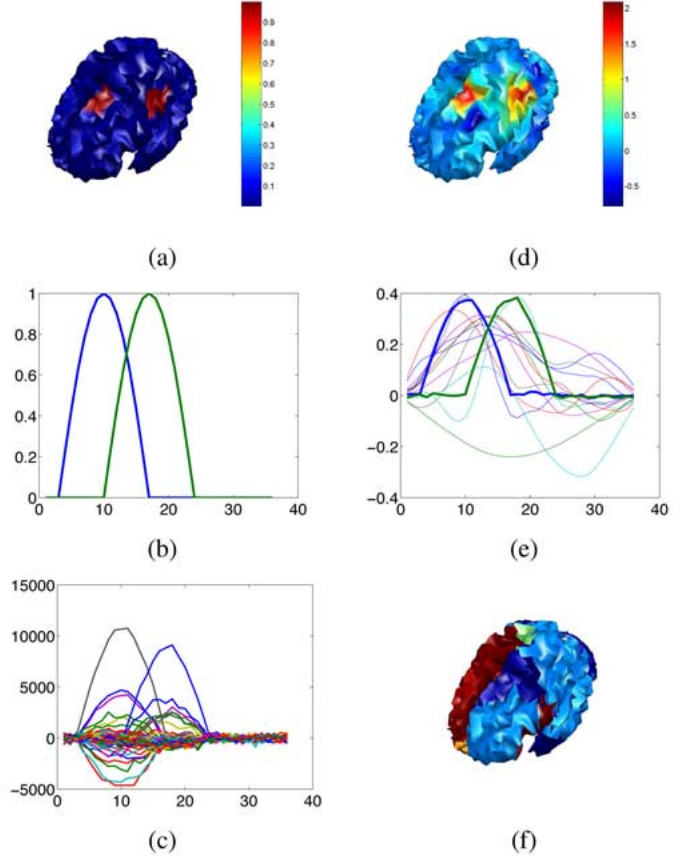


Fig. 1. Illustrative example of the simulation framework. Two activity sources (a) were simulated, with sine-like time courses (b), shifted by a quarter period. The corresponding simulated scalp data (c) had a signal-to-noise ratio (SNR) of 5. The right side of the figure show the temporal invariant map w (d), the reconstructed time courses of the extended sources [(e) (in bold) the two activated clusters] and (f) the parcelling C , which is graphically depicted by associating the same random color to all dipoles belonging to each of the cortical clusters. Note that w is an efficient estimate of the spatial profile of the simulated sources, in terms of both localization and spatial extent estimation. Moreover, the estimates of the time courses of the clusters corresponding to the activated extended areas identified using the w map perfectly match the simulated temporal dynamics. Last but not least, the spatio-temporal decomposition of BASTA provide an intuitive and accurate way of interpreting and visualizing the sources.

B. Quantitative Evaluation Metrics

Let \mathbf{J} denote the true simulated time courses of the elementary dipoles of the DL model and $\hat{\mathbf{J}}$ a given estimate of \mathbf{J} . We considered the following evaluation metrics in order to evaluate and compare the performance of the four inverse estimators.

- 1) The topological distance (TD) between the center of mass of a simulated source and the corresponding local maximum of the reconstructed source distribution. For each simulated area, TD was estimated by considering the time sample where the simulated source amplitude was maximum. In the following, we denote by TD the mean topological distance over the two simulated sources for each synthetic data set. It quantifies the localization error. In the results section, TD is given in meters.
- 2) Area Under the receiver operating characteristic (ROC) Curve (AUC): for each simulation, a ROC analysis was

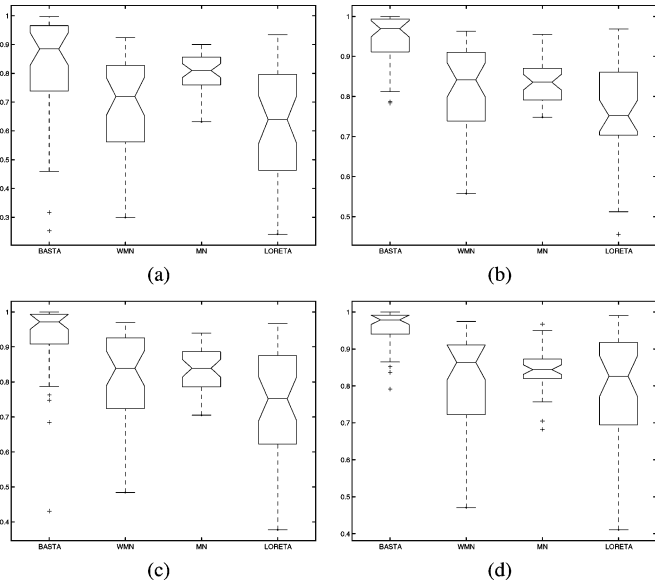


Fig. 2. Boxplot representations of the observed distributions of AUC for the SNR-series of Monte-Carlo simulations. The noise level were: (a) SNR = 1.3, (b) SNR = 3, (c) SNR = 5, and (d) SNR = 10. In each subplot, the reconstructions algorithms are, from left to right: BASTA, WMN, MN, LORETA. Boxplot representations show the median, first and third quantiles and minimum/maximum values of the empirical distribution (apart from possible outliers depicted as single points).

performed, comparing the cortical binary mask corresponding to the simulated areas and the map $\hat{\mathbf{J}}_{\max}$ defined as

$$\hat{\mathbf{J}}_{\max}(i) = \max \left[\left\{ \left\| \hat{\mathbf{J}}_{ij} \right\| \right\}_{j=1}^t \right], \quad i = 1, \dots, n. \quad (39)$$

AUC: is the probability to correctly classify an elementary dipole (either as activated or as inactivated). We refer the interested reader to [42] and [43] for details. Note that this index provides an accurate evaluation of the spatial behavior of an estimator (localization and extent estimation).

For each Monte-Carlo simulation, the two accuracy metrics were calculated for the four reconstruction methods. An analysis of variance (ANOVA) was then computed, comprising an F-test assessing the presence of a global effect of the reconstruction method, and subsequent T-tests for pairwise comparisons³ (significance level: $p < 0.05$).

V. RESULTS

For each simulations series, results of the ANOVA (boxplots⁴ of the distribution of the accuracy metric for each method, value of the global F-test p-value, T-statistic, and p-values of pairwise comparisons) are given.

- *SNR-Series*: Figs. 2 and 3 show boxplot diagrams (40 simulations for each) for every signal-to-noise level (SNR = 1.3, 3, 5, 10) of the area under the ROC curve (AUC) and of the TD, respectively. Table I (respectively, Table II) shows the results of the ANOVA comparing the reconstructions

³The F-test showed in all series of simulations a significant effect of the reconstruction method.

⁴Boxplot representations show the median, first and third quantiles and minimum/maximum values of the empirical distribution (apart from possible outliers depicted as single points).

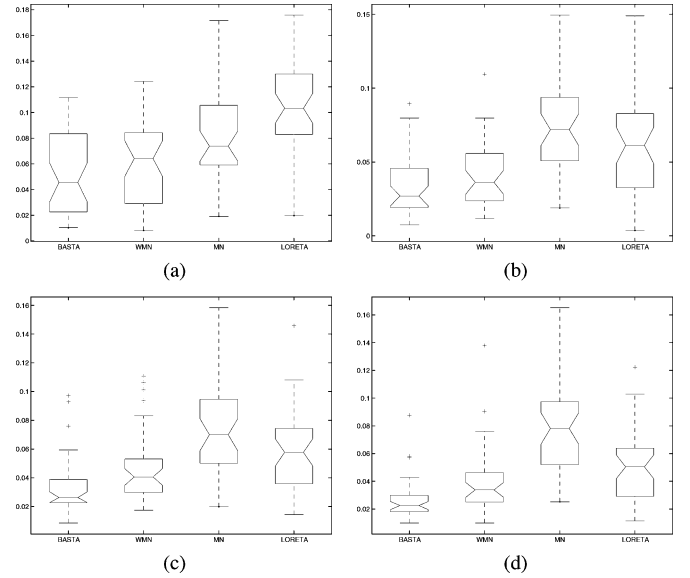


Fig. 3. Boxplot representations of the observed distributions of TD for the SNR-series of Monte-Carlo simulations. The noise level were: (a) SNR = 1.3, (b) SNR = 3, (c) SNR = 5, and (d) SNR = 10. In each subplot, the reconstructions algorithms are, from left to right: BASTA, WMN, MN, LORETA. TD is given in meters. For details on boxplot representation, see Fig. 2.

approaches through the AUC index (respectively, the TD index).

Regarding the AUC index, the only case where BASTA did not perform significantly better than the three other methods was at the lowest SNR level, with the MN method [see Table I(a)]. However, the boxplot distributions show a clear behavioral gap between the two methods [see Fig. 3(a)]. Otherwise, MN and WMN outperformed LORETA, but were generally not significantly contrasted. As for the TD index, BASTA did not behave significantly better than WMN for SNR levels SNR = 1.3 and SNR = 3. According to that metric, LORETA behaved generally significantly better than MN, and significantly worse than WMN. Since the latter was not the second more efficient method regarding the AUC index, this may reflect an accuracy in localization associated to a poor capability of WMN to estimate spatial extents. In any case, no method was identified as at least not significantly different from BASTA with regards to both accuracy metrics.

- *NS-Series*: Figs. 4 and 5 show boxplot diagrams (40 simulations for each) for all numbers of active sources (NS = 1, 2, 3, 4, 5) of the area under the ROC curve (AUC) and of the TD, respectively. Table III (respectively, Table IV) shows the results of the ANOVA comparing the reconstructions approaches through the AUC index (respectively, the TD index).

The AUC value obtained for BASTA was significantly higher than that obtained for all other methods (see Table III and Fig. 4). For all four NS values, the AUC value was significantly higher for MN than for WMN. Moreover, LORETA behaved generally worse than all three other methods. Both remarks are coherent with the results obtained for the SNR-series. As for the TD

TABLE I

ANOVA PAIRWISE COMPARISONS OF AUC FOR THE SNR-SERIES OF MONTE-CARLO SIMULATIONS. THE NOISE LEVELS WERE: (a) SNR = 1.3, (b) SNR = 3, (c) SNR = 5, AND (d) SNR = 10. THE CORRESPONDING p-VALUES FOR THE GLOBAL F-TEST OF THE ANOVA WERE: $p = 2.1e^{-7}$, $p = 1.47e^{-13}$, $p = 1.5e^{-8}$, AND $p = 2.5e^{-10}$, RESPECTIVELY, I.E., ALL GLOBAL F-TEST SHOWED A SIGNIFICANT EFFECT OF THE RECONSTRUCTION APPROACH. IN EACH SUBTABLE, EACH (i,j) ENTRY SHOWS THE T-STATISTIC (p-VALUE) RESULTING FROM THE PAIRWISE COMPARISON OF METHOD j MINUS METHOD i

	BASTA	WMN	MN
WMN	5.55 ($< 1e^{-5}$)	—	—
MN	<u>0.82</u> (0.41)	-3.81 ($5e^{-4}$)	—
LORETA	5.80 ($< 1e^{-5}$)	1.68 (0.10)	4.88 ($< 1e^{-5}$)

(a)

	BASTA	WMN	MN
WMN	5.27 ($< 1e^{-5}$)	—	—
MN	4.27 ($1e^{-4}$)	-1.17 (0.24)	—
LORETA	7.84 ($< 1e^{-5}$)	3.55 ($1.0e^{-3}$)	3.65 ($8.0e^{-4}$)

(c)

	BASTA	WMN	MN
WMN	7.32 ($< 1e^{-5}$)	—	—
MN	7.13 ($< 1e^{-5}$)	-0.94 (0.35)	—
LORETA	9.88 ($< 1e^{-5}$)	3.24 ($2.5e^{-3}$)	3.22 ($2.6e^{-3}$)

(b)

	BASTA	WMN	MN
WMN	8.06 ($< 1e^{-5}$)	—	—
MN	8.85 ($< 1e^{-5}$)	-1.13 (0.26)	—
LORETA	7.49 ($< 1e^{-5}$)	1.46 (0.15)	1.96 (0.056)

(d)

TABLE II

ANOVA PAIRWISE COMPARISONS OF TD FOR THE SNR-SERIES OF MONTE-CARLO SIMULATIONS. THE NOISE LEVELS WERE: (a) SNR = 1.3, (b) SNR = 3, (c) SNR = 5, AND (d) SNR = 10. THE CORRESPONDING p-VALUES FOR THE GLOBAL F-TEST OF THE ANOVA WERE: $p = 1.8e^{-9}$, $p = 5.3e^{-9}$, $p = 3.3e^{-10}$, AND $p = 1.1e^{-15}$, RESPECTIVELY, I.E., ALL GLOBAL F-TEST SHOWED A SIGNIFICANT EFFECT OF THE RECONSTRUCTION APPROACH. IN EACH SUBTABLE, EACH (i,j) ENTRY SHOWS THE T-STATISTIC (p-VALUE) RESULTING FROM THE PAIRWISE COMPARISON OF METHOD j MINUS METHOD i

	BASTA	WMN	MN
WMN	-1.42 (0.16)	—	—
MN	-3.74 ($6e^{-4}$)	-3.10 ($3.5e^{-3}$)	—
LORETA	-6.36 ($< 1e^{-5}$)	-5.24 ($< 1e^{-5}$)	-2.40 (0.019)

(a)

	BASTA	WMN	MN
WMN	-2.83 ($7.2e^{-3}$)	—	—
MN	-7.17 ($< 1e^{-5}$)	-5.39 ($< 1e^{-5}$)	—
LORETA	-5.18 ($< 1e^{-5}$)	-2.06 (0.045)	2.93 (0.055)

(c)

	BASTA	WMN	MN
WMN	-1.70 (0.097)	—	—
MN	-6.68 ($< 1e^{-5}$)	-5.02 ($< 1e^{-5}$)	—
LORETA	-3.84 ($4e^{-4}$)	-3.17 ($2.9e^{-3}$)	1.46 (0.15)

(b)

	BASTA	WMN	MN
WMN	-2.93 ($5.6e^{-3}$)	—	—
MN	-8.63 ($< 1e^{-5}$)	-7.05 ($< 1e^{-5}$)	—
LORETA	-4.90 ($< 1e^{-5}$)	-1.80 (0.08)	4.33 ($1e^{-4}$)

(d)

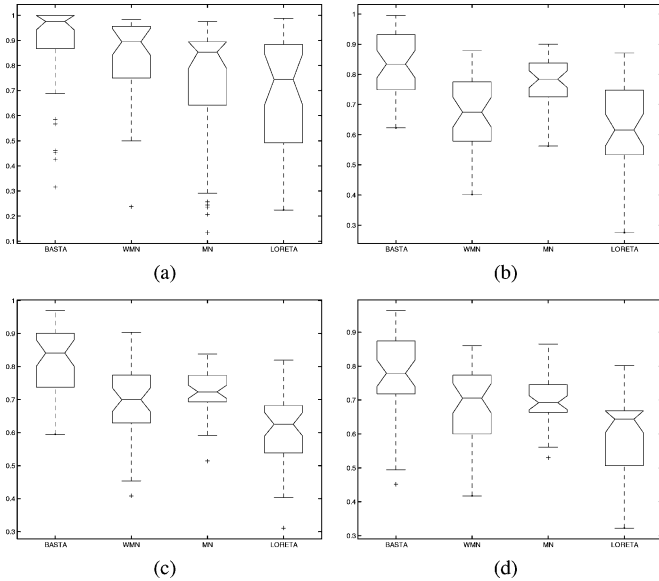


Fig. 4. Boxplot representations of the observed distributions of AUC for the NS-series of Monte-Carlo simulations. The number of simulated sources were: (a) NS = 1, (b) NS = 3, (c) NS = 4, and (d) NS = 5. The case NS = 2 is shown on the SNR-series figure, in the subplot showing the series corresponding to SNR = 1.3 [Fig. 2(a)]. In each subplot, the reconstructions algorithms are, from left to right: BASTA, WMN, MN, LORETA. For details on boxplot representation, see Fig. 2.

index, BASTA outperformed significantly all methods in all cases (see Table IV and Fig. 3), except WMN for the NS = 1 case, despite the clear gap shown in Fig. 5(a). Moreover, WMN had a significantly better TD than MN, which was in turn generally significantly better than that

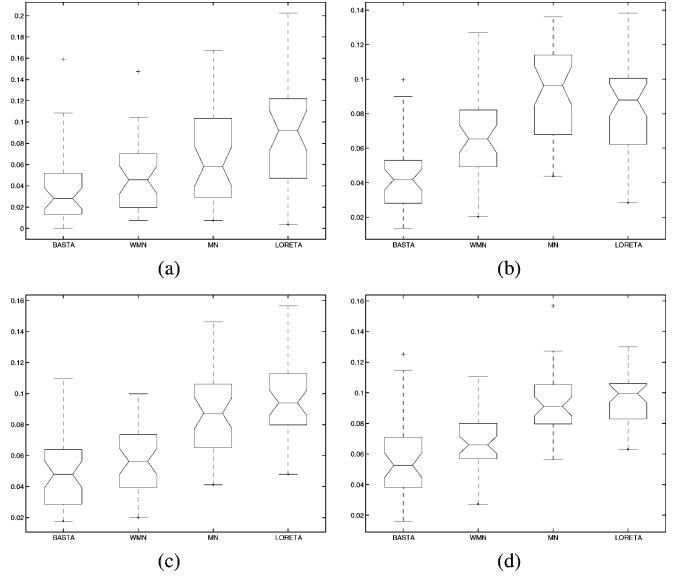


Fig. 5. Boxplot representations of the observed distributions of TD for the NS-series of Monte-Carlo simulations. The number of simulated sources were: (a) NS = 1, (b) NS = 3, (c) NS = 4, and (d) NS = 5. The case NS = 2 is shown on the SNR-series figure, in the subplot showing the series corresponding to SNR = 1.3 [Fig. 3(a)]. In each subplot, the reconstructions algorithms are, from left to right: BASTA, WMN, MN, LORETA. TD is given in m. For details on boxplot representation, see Fig. 2.

of LORETA. Again, in those series, no method was identified as at least not significantly different from BASTA with regards to both accuracy metrics.

Briefly, for all these series of simulations, BASTA gave the best results, i.e., the means and medians of both accuracy metrics were

TABLE III

ANOVA PAIRWISE COMPARISONS OF AUC FOR THE NS-SERIES OF MONTE-CARLO SIMULATIONS. THE NUMBER OF SIMULATED SOURCES WERE: (a) NS = 1, (b) NS = 3, (c) NS = 4, AND (d) NS = 5. THE CORRESPONDING p-VALUES FOR THE GLOBAL F-TEST OF THE ANOVA WERE: $p = 5.0e^{-5}$, $p = 3.0e^{-14}$, $p = 9.2e^{-14}$, AND $p = 1.2e^{-10}$, RESPECTIVELY, I.E., ALL GLOBAL F-TEST SHOWED A SIGNIFICANT EFFECT OF THE RECONSTRUCTION APPROACH. IN EACH SUBTABLE, EACH (i,j) ENTRY SHOWS THE T-STATISTIC (p-VALUE) RESULTING FROM THE PAIRWISE COMPARISON OF METHOD j MINUS METHOD i

	BASTA		WMN		MN	
WMN	2.49	(0.017)	—	—	—	—
MN	3.38	(< 1e ⁻⁵)	4.68	(< 1e ⁻⁵)	—	—
LORETA	6.15	(< 1e ⁻⁵)	4.23	(1e ⁻⁴)	0.79	(0.42)

(a)

	BASTA		WMN		MN	
WMN	9.76	(< 1e ⁻⁵)	—	—	—	—
MN	3.15	(3e ⁻³)	-3.87	(4e ⁻⁴)	—	—
LORETA	10.24	(< 1e ⁻⁵)	2.12	(3.9e ⁻²)	6.18	(< 1e ⁻⁵)

(b)

	BASTA		WMN		MN	
WMN	5.45	(< 1e ⁻⁵)	—	—	—	—
MN	4.48	(1e ⁻⁴)	-1.86	(0.12)	—	—
LORETA	9.04	(< 1e ⁻⁵)	3.45	(3e ⁻⁴)	6.41	(< 1e ⁻⁵)

(c)

	BASTA		WMN		MN	
WMN	4.34	(1e ⁻⁴)	—	—	—	—
MN	4.85	(1e ⁻⁴)	-0.20	(0.84)	—	—
LORETA	8.47	(< 1e ⁻⁵)	4.81	(< 1e ⁻⁵)	4.52	(1e ⁻⁴)

(d)

TABLE IV

ANOVA PAIRWISE COMPARISONS OF TD FOR THE NS-SERIES OF MONTE-CARLO SIMULATIONS. THE NUMBER OF SIMULATED SOURCES WERE: (a) NS = 1, (b) NS = 3, (c) NS = 4, AND (d) NS = 5. THE CORRESPONDING p-VALUES FOR THE GLOBAL F-TEST OF THE ANOVA WERE: $p = 1.7e^{-5}$, $p = 1.2e^{-13}$, $p = 0$, AND $p = 2.2e^{-16}$, RESPECTIVELY, I.E., ALL GLOBAL F-TEST SHOWED A SIGNIFICANT EFFECT OF THE RECONSTRUCTION APPROACH. IN EACH SUBTABLE, EACH (i,j) ENTRY SHOWS THE T-STATISTIC (p-VALUE) RESULTING FROM THE PAIRWISE COMPARISON OF METHOD j MINUS METHOD i

	BASTA		WMN		MN	
WMN	-1.73	(0.09)	—	—	—	—
MN	-3.67	(8e ⁻⁴)	-3.36	(1.8e ⁻³)	—	—
LORETA	-5.33	(< 1e ⁻⁵)	-3.93	(4e ⁻⁴)	-1.80	(0.08)

(a)

	BASTA		WMN		MN	
WMN	-4.86	(< 1e ⁻⁵)	—	—	—	—
MN	-8.70	(< 1e ⁻⁵)	-4.04	(2e ⁻⁴)	—	—
LORETA	-8.64	(< 1e ⁻⁵)	-2.91	(5.9e ⁻³)	1.89	(0.065)

(b)

	BASTA		WMN		MN	
WMN	-2.51	(0.016)	—	—	—	—
MN	-6.36	(< 1e ⁻⁵)	-5.83	(< 1e ⁻⁵)	—	—
LORETA	-9.39	(< 1e ⁻⁵)	-6.75	(< 1e ⁻⁵)	-1.02	(0.31)

(c)

	BASTA		WMN		MN	
WMN	-2.13	(0.038)	—	—	—	—
MN	-8.89	(< 1e ⁻⁵)	-6.73	(< 1e ⁻⁵)	—	—
LORETA	-7.60	(< 1e ⁻⁵)	-6.39	(< 1e ⁻⁵)	-0.74	(0.45)

(d)

in any case better than those of the other approaches. Moreover, except for four pairwise comparisons (see underlined p-values in the results tables), BASTA behaved significantly better than the other approaches. However, part of the observed distributions showed a significant tail: some simulated source configurations were too hard to recover correctly, whatever the method used.

VI. DISCUSSION

The localization of EEG/MEG activity source has been intensively studied during the past two decades along two directions: ECD and DL modeling. Those two approaches have their own drawbacks, but they can be combined using a cortical parcelling, leading to a new and accurate hybrid approach; namely, an extended source mixing model. Under both specific and realistic underlying assumptions (temporal coherence for ECD; prior cortical restriction for DL), we propose a new decomposition of the activity sources by separating the information to be restored into temporally and spatially specific quantities. Consequently, the source estimation benefits from both advantages of the two techniques: few estimated parameters for ECD (by taking a sufficiently wide data window, the problem could even—as for ECD models—become over-determined), and fine anatomical description for distributed modeling (allowing an accurate estimation of the spatial extent of the sources). Moreover, distinct temporal and spatial prior information about the sources can be flexibly introduced through the Bayesian framework. Indeed, associated to a dedicated hierarchical data generative model, Bayesian inference mechanics allows us to estimate from the data both the parameters and the hyperparameters of

the model. Moreover, we derived the Laplace approximation of the posterior covariance matrix of the model parameters. This provides us with a reliable tool to assess the significance of the MAP estimates using any statistical hypothesis testing strategy.

The main contribution of this work is the spatio-temporal decomposition of the source parameters, as a corollary of the extended source mixing model. This model relies upon a strong hypothesis: the true active areas are understood as clusters of elementary dipoles that are highly temporally correlated. In fact, we assume that their relative amplitude strictly does not change with time. This assumption leads us to define two quantities to be estimated: the temporal invariant map \mathbf{w} (which represents the spatial profile of the active areas) and the temporal dynamics embodied in matrix \mathbf{X} (which describes the extended source time courses). Such a model has the following implications.

- 1) The spatial profile of the sources is estimated using the entire data window. By selecting a wide enough data window, the problem may even become over-determined as for ECD models.
- 2) When estimating the spatial profile \mathbf{w} of the active sources, the linear dependence between the columns of \mathbf{G} is partially lifted by the introduction of a time course contrast between dipoles belonging to different patches in the matrix \mathbf{D} . Naturally, within a patch one cannot overcome the ill-conditioning of the gain matrix, and the inverse problem remains ill-posed. However, enforcing the spatial structure to be time invariant provides a natural and efficient regularization process.
- 3) The two variables \mathbf{w} and \mathbf{X} provide an easy and accurate way of interpreting and visualizing the sources. Moreover,

specific spatial and temporal constraints can be introduced in their estimation, leading to realistic estimates of the sources. Many other priors could be here considered. For instance, the use of statistical interdependence constraints on the extended sources should be further investigated.

- 4) The spatio-temporal decomposition of the sources could provide a particularly suitable framework for data fusion. For instance, the statistical parametric map (SPM) provided by classical fMRI data analysis is time invariant and reflects the spatial profile of activation. Under the assumption that the same cortical network explains both the EEG/MEG and the fMRI data, one could, therefore, use fMRI SPMs as soft and controlled constraints on the spatial variable \mathbf{w} [44]. Moreover, the time invariant map of activation \mathbf{w} could even be estimated using both the EEG/MEG and the fMRI data sets, without overweighting one of them, which is a well-balanced data fusion procedure.
- 5) As for the dimensionality of the problem, let us consider the range of admissible numbers of extended sources and time samples. If $q = 1$, a unique time course is relevant to explain the temporal variability of the data. In the absence of noise, the different time samples are, thus, colinear. Therefore, if $q = 1$, the (only) gain of the extended source mixing model is about noise robustness (we are looking for the data structure that is time invariant). At the opposite, if $q = n$ (the number of the DL elementary dipoles), the model is nothing but the original distributed model. Hence, a kind of optimal q should be estimated or picked up using heuristic considerations. In order to ensure a well-conditioned estimation of the matrix \mathbf{X} , q has to be smaller than the number of sensors p . For instance, q could be estimated using preprocessing data analysis, like ICA [45], [46] or MUSIC [47], [48] algorithms.

We also assume noise stationarity in the definition of the data likelihood. This assumption becomes questionable for a “long” period of time: t cannot be as large as we would like to. On the other hand, we need sufficient statistics in order to estimate the temporal invariant map \mathbf{w} (accurate in case of equality between the number of parameters to be estimated and the available data). This leads to a lower bound for the size of the data window: $t \geq (n/p)$.

For instance, we used the entire simulated data (50 time samples) and about thirty patches for the cortical parcelling.

Last but not least, in contrast to classical regularization techniques (see [39], [49]), we propose a simple and flexible Bayesian estimation of the model hyperparameters (α, β, γ) . Note however that we keep constant the number of extended sources, omitting it in the posterior pdf. Despite the fact that a reasonable tuning of this hyperparameter did not seem to have a significant influence on the estimation (as long as different active areas do not belong to the same patch), this higher inference level may be incorporated in the chain rule, in order to take this uncertainty into account. Indeed, since decreasing the patches number q is equivalent to increasing their mean size, the coupling of our method with a Bayesian integration procedure (see [35] and specially [25] for a close application) over the different spatial q -scales could lead to an interesting multiresolution framework. Of course, this would considerably increase the computation time of the algorithm and

the benefit of such a refinement of our approach has to be further investigated.

APPENDIX I

CONVERGENCE OF THE ESTIMATION ALGORITHM

Searching the MAP estimates of \mathbf{w} , \mathbf{X} , α , β and γ means maximizing their joint posterior pdf. Now, the extended source mixing model is a bilinear equation of the parameters. It is, therefore, not straightforward to maximize the joint posterior pdf (because of the data likelihood term). Hence, we will use an additional consideration. Let us write: $U(\mathbf{X}, \mathbf{w}, \alpha, \beta, \gamma) = \log(p(\mathbf{X}, \mathbf{w}, \alpha, \beta, \gamma | \mathbf{M}))$. Then, for any two sets of parameters $(\mathbf{X}^*, \mathbf{w}^*, \alpha^*, \beta^*, \gamma^*)$ and $(\hat{\mathbf{X}}, \hat{\mathbf{w}}^*, \hat{\alpha}^*, \hat{\beta}^*, \hat{\gamma}^*)$ we have

$$\begin{aligned} & U(\hat{\mathbf{X}}, \hat{\mathbf{w}}^*, \hat{\alpha}^*, \hat{\beta}^*, \hat{\gamma}^*) - U(\mathbf{X}^*, \mathbf{w}^*, \alpha^*, \beta^*, \gamma^*) \\ &= \log \left(\frac{p(\hat{\mathbf{X}}, \hat{\mathbf{w}}^*, \hat{\alpha}^*, \hat{\beta}^*, \hat{\gamma}^* | \mathbf{M})}{p(\mathbf{X}^*, \mathbf{w}^*, \alpha^*, \beta^*, \gamma^* | \mathbf{M})} \right) \\ &= \log \left(\frac{p(\hat{\mathbf{X}} | \mathbf{M}, \hat{\mathbf{w}}^*, \hat{\alpha}^*, \hat{\beta}^*, \hat{\gamma}^*) \cdot p(\hat{\mathbf{w}}^*, \hat{\alpha}^*, \hat{\beta}^*, \hat{\gamma}^* | \mathbf{M})}{p(\mathbf{X}^* | \mathbf{M}, \mathbf{w}^*, \alpha^*, \beta^*, \gamma^*) \cdot p(\mathbf{w}^*, \alpha^*, \beta^*, \gamma^* | \mathbf{M})} \right) \\ &= \log \left(p(\hat{\mathbf{X}} | \mathbf{M}, \hat{\mathbf{w}}^*, \hat{\alpha}^*, \hat{\beta}^*, \hat{\gamma}^*) p(\mathbf{X}^* | \mathbf{M}, \mathbf{w}^*, \alpha^*, \beta^*, \gamma^*) \right). \end{aligned}$$

Hence, starting from an initial guess $(\mathbf{X}^*, \mathbf{w}^*, \alpha^*, \beta^*, \gamma^*)$, we are guaranteed to increase the joint posterior pdf $p(\mathbf{X}, \mathbf{w}, \alpha, \beta, \gamma | \mathbf{M})$ evaluated at $(\mathbf{X}^*, \mathbf{w}^*, \alpha^*, \beta^*, \gamma^*)$ if we can find any value $\hat{\mathbf{X}}$ of the parameter \mathbf{X} such that $p(\hat{\mathbf{X}} | \mathbf{M}, \hat{\mathbf{w}}^*, \hat{\alpha}^*, \hat{\beta}^*, \hat{\gamma}^*) \geq p(\mathbf{X}^* | \mathbf{M}, \mathbf{w}^*, \alpha^*, \beta^*, \gamma^*)$. In other words, increasing the posterior pdf of a given variable conditional on the others increases their joint posterior pdf. This allows to define our iterative maximization algorithm of the joint posterior pdf as a split-step conditional MAP estimation. This defines a fix point algorithm, meaning that the joint posterior pdf is monotonically increasing during the successive iterations. However, strict global optimization is ensured only if the sampled joint posterior pdf is monomodal.

In our simulations, the convergence criterion of the iterative conditional pdf maximization was set as: $((U_{[t]} - U_{[t-1]}) / U_{[t-1]}) \leq 10^{-5}$, where $U_{[t]} = \log(p(\mathbf{w}_{[t]}, \mathbf{X}_{[t]}, \alpha_{[t]}, \beta_{[t]}, \gamma_{[t]} | \mathbf{M}))$.

APPENDIX II

LAPLACE APPROXIMATION OF THE JOINT COVARIANCE MATRIX

In this section, we derive a method dedicated to approximate the joint posterior distribution of \mathbf{w} and \mathbf{X} in order to derive their joint covariance matrix. Let us write the second order Taylor series of the joint posterior distribution of (\mathbf{w}, \mathbf{X}) in the neighborhood of the mode $(\hat{\mathbf{w}}, \hat{\mathbf{X}}, \hat{\alpha}, \hat{\beta}, \hat{\gamma})$

$$\begin{aligned} -\log p(\mathbf{w}, \mathbf{X} | \mathbf{M}) &\approx -\log p(\hat{\mathbf{w}}, \hat{\mathbf{X}}, \hat{\alpha}, \hat{\beta}, \hat{\gamma} | \mathbf{M}) \\ &\quad + \frac{1}{2} \begin{pmatrix} \text{vec}(\mathbf{X}) - \text{vec}(\hat{\mathbf{X}}) \\ \mathbf{w} - \hat{\mathbf{w}} \end{pmatrix}^T \mathbf{H} \begin{pmatrix} \text{vec}(\mathbf{X}) - \text{vec}(\hat{\mathbf{X}}) \\ \mathbf{w} - \hat{\mathbf{w}} \end{pmatrix} \end{aligned}$$

knowing that the Hessian matrix \mathbf{H} has the following expression:

$$\mathbf{H} = - \begin{pmatrix} \hat{\alpha} \mathbf{I}_t \otimes \hat{\mathbf{A}}^T \hat{\mathbf{A}} + \hat{\beta} \mathbf{T}^T \mathbf{T} & \hat{\alpha} \left((\mathbf{I}_t \otimes \hat{\mathbf{A}}^T) \hat{\mathbf{D}} - \mathbf{Q} \right) \\ \hat{\alpha} \left((\mathbf{I}_t \otimes \hat{\mathbf{A}}^T) \hat{\mathbf{D}} - \mathbf{Q} \right)^T & \hat{\alpha} \hat{\mathbf{D}}^T \hat{\mathbf{D}} + \hat{\gamma} \mathbf{S}^T \mathbf{S} \end{pmatrix} \quad (40)$$

where \mathbf{Q} is a $qt \times n$ matrix such that

$$\mathbf{Q} = \begin{pmatrix} \mathbf{C}^\top \text{diag}(\mathbf{G}^\top \mathbf{M}_1) & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \mathbf{C}^\top \text{diag}(\mathbf{G}^\top \mathbf{M}_t) \end{pmatrix}. \quad (41)$$

The bloc diagonal of the Hessian \mathbf{H} correspond to the conditional (marginal) posterior inverse covariance matrices of $\text{vec}(\mathbf{X})$ and \mathbf{w} . The (symmetrical) off-diagonal terms relate to the structure of their (mutual) covariance, which has not been taken into account during the alternate conditional maximization algorithm.

The subsequent so-called Laplace approximation [50] states that the joint posterior pdf may be replaced by a Gaussian density with covariance matrix equal to minus the inverse of the Hessian matrix

$$p(\mathbf{w}, \mathbf{X} | \mathbf{M}) \approx \mathcal{N} \left(\begin{pmatrix} \text{vec}(\hat{\mathbf{X}}) \\ \hat{\mathbf{w}} \end{pmatrix}, -\mathbf{H}^{-1} \right). \quad (42)$$

This approximation can be used in order to build some inference about \mathbf{w} and/or \mathbf{X} . For example, one may easily rely on this Gaussian form of the joint posterior pdf to assess the significance of the MAP parameter estimates using any statistical hypothesis testing strategy.

APPENDIX III

MAP estimate of time variable \mathbf{X}

This appendix describes the technical details relative to the derivation of the conditional MAP estimate of \mathbf{X} , using a Lagrange multiplier approach in order to enforce the normalization constraint. For the sake of clarity, we here omit the index of BASTA iterations $[j]$ in the description of the following iterative optimization method. Given the mixing matrix \mathbf{A} , the conditional MAP estimation of \mathbf{X} can be expressed as a constrained minimization

$$\begin{cases} \hat{\mathbf{X}} = \underset{\mathbf{X}}{\text{argmin}} f(\mathbf{X}) \\ \mathbf{X} : h_j(\mathbf{X}) = 0, \quad j = 1, \dots, q \end{cases} \quad (43)$$

where

$$\begin{cases} f(\mathbf{X}) = \|\text{vec}(\mathbf{M}) - (\mathbf{I}_t \otimes \mathbf{A})\text{vec}(\mathbf{X})\|^2 + \frac{\beta}{\alpha} \|\mathbf{T}\text{vec}(\mathbf{X})\|^2 \\ \mathbf{h}(\mathbf{X}) = \mathbf{N}\text{vec}(\mathbf{X}) - \mathbf{1}_q \end{cases}. \quad (44)$$

This problem is usefully rewritten thanks to the Lagrange equations

$$\begin{cases} \nabla L(\mathbf{X}, \boldsymbol{\Lambda}) = \mathbf{0}_{qt} \\ \mathbf{h}(\mathbf{X}) = \mathbf{0}_q \end{cases} \quad (45)$$

where the Lagrangian function $L(\mathbf{X}, \boldsymbol{\Lambda})$ is defined as

$$L(\mathbf{X}, \boldsymbol{\Lambda}) = f(\mathbf{X}) + \boldsymbol{\Lambda}^\top \mathbf{h}(\mathbf{X}) \quad (46)$$

$\boldsymbol{\Lambda}$ being the $q \times 1$ vector of Lagrange parameters that is to be quantified in order to make the MAP estimate exactly satisfy the normalization constraint. Newton's method [37] consists of linearizing (45) in the neighborhood of a point $(\mathbf{X}_{[k]}, \boldsymbol{\Lambda}_{[k]})$, and defining $(\mathbf{X}_{[k+1]}, \boldsymbol{\Lambda}_{[k+1]})$ as the solution of the following resulting system of linear equations:

$$\begin{cases} \nabla L(\mathbf{X}_{[k]}, \boldsymbol{\Lambda}_{[k]}) + \nabla^2 L(\mathbf{X}_{[k]}, \boldsymbol{\Lambda}_{[k]}) (\text{vec}(\mathbf{X}_{[k+1]}) \\ \quad - \text{vec}(\mathbf{X}_{[k]})) + \nabla \mathbf{h}(\mathbf{X}_{[k]}) (\boldsymbol{\Lambda}_{[k+1]} - \boldsymbol{\Lambda}_{[k]}) = \mathbf{0}_{qt} \\ \mathbf{h}(\mathbf{X}_{[k]}) + \nabla \mathbf{h}(\mathbf{X}_{[k]}) (\text{vec}(\mathbf{X}_{[k+1]}) - \text{vec}(\mathbf{X}_{[k]})) = \mathbf{0}_q \end{cases} \quad (47)$$

where $\nabla L(\mathbf{X}_{[k]}, \boldsymbol{\Lambda}_{[k]}) = \nabla f(\mathbf{X}_{[k]}) + \nabla \mathbf{h}(\mathbf{X}_{[k]}) \boldsymbol{\Lambda}_{[k]}$. This system is rewritten in a matrix form as

$$\begin{pmatrix} \nabla^2 L(\mathbf{X}_{[k]}, \boldsymbol{\Lambda}_{[k]}) & \nabla \mathbf{h}(\mathbf{X}_{[k]}) \\ \nabla \mathbf{h}(\mathbf{X}_{[k]})^\top & \mathbf{0}_{q \times q} \end{pmatrix} \times \begin{pmatrix} \text{vec}(\mathbf{X}_{[k+1]}) - \text{vec}(\mathbf{X}_{[k]}) \\ \boldsymbol{\Lambda}_{[k+1]} \end{pmatrix} = \begin{pmatrix} -\nabla f(\mathbf{X}_{[k]}) \\ -\mathbf{h}(\mathbf{X}_{[k]}) \end{pmatrix}. \quad (48)$$

In order to solve it, we need the expressions of the gradients and Hessians in our particular case

$$\begin{cases} \nabla f(\mathbf{X}_{[k]}) = -2(\mathbf{I}_t \otimes \mathbf{A}^\top) \text{vec}(\mathbf{M}) + 2(\mathbf{I}_t \otimes \mathbf{A}^\top \mathbf{A}) \\ \quad \times \text{vec}(\mathbf{X}_{[k]}) + 2\frac{\beta}{\alpha} \mathbf{T}^\top \mathbf{T} \text{vec}(\mathbf{X}_{[k]}) \\ \nabla^2 f(\mathbf{X}_{[k]}) = 2\mathbf{I}_t \otimes \mathbf{A}^\top \mathbf{A} + 2\frac{\beta}{\alpha} \mathbf{T}^\top \mathbf{T} \\ \nabla \mathbf{h}(\mathbf{X}_{[k]}) = 2\mathbf{N}_{[k]}^\top \\ \sum_{i=1}^q \lambda_{i[k]} \nabla^2 h_i(\mathbf{X}_{[k]}) = 2\mathbf{I}_t \otimes \text{diag}(\boldsymbol{\Lambda}_{[k]}) \end{cases} \quad (49)$$

where $\mathbf{N}_{[k]}$ stands for \mathbf{N} evaluated at $\mathbf{X}_{[k]}$.

Hence, the solution of the system (48) is unique and expressed as

$$\begin{pmatrix} \text{vec}(\mathbf{X}_{[k+1]}) - \text{vec}(\mathbf{X}_{[k]}) \\ \boldsymbol{\Lambda}_{[k+1]} \end{pmatrix} = \mathbf{A}_{h[k]}^{-1} \mathbf{M}_{h[k]} \quad (50)$$

with $\mathbf{A}_{h[k]}$ a $q(t+1) \times q(t+1)$ partitioned full rank matrix and $\mathbf{M}_{h[k]}$ a $q(t+1) \times 1$ vector such that (51), shown at the bottom of the page, holds where $\mathbf{B}_{[k]}^{-1} = \mathbf{I}_t \otimes \mathbf{A}^\top \mathbf{A} + (\beta/\alpha) \mathbf{T}^\top \mathbf{T} + \mathbf{I}_t \otimes \text{diag}(\boldsymbol{\Lambda}_{[k]})$ is a $qt \times qt$ full rank matrix. Now, note that we can rewrite the inverse of the matrix $\mathbf{A}_{h[k]}$ as

$$\mathbf{A}_{h[k]}^{-1} = \begin{pmatrix} \mathbf{B}_{[k]} + \mathbf{B}_{[k]} \mathbf{N}_{[k]}^\top \mathbf{V}^{-1} \mathbf{N}_{[k]} \mathbf{B}_{[k]} & -\mathbf{B}_{[k]} \mathbf{N}_{[k]}^\top \mathbf{V}^{-1} \\ -\mathbf{V}^{-1} \mathbf{N}_{[k]} \mathbf{B}_{[k]} & \mathbf{V}^{-1} \end{pmatrix} \quad (52)$$

where \mathbf{V} is such that $\mathbf{V} = \mathbf{0}_{q \times q} - \mathbf{N}_{[k]} \mathbf{B}_{[k]} \mathbf{N}_{[k]}^\top = -\mathbf{N}_{[k]} \mathbf{B}_{[k]} \mathbf{N}_{[k]}^\top$. Then, reordering the different terms

$$\begin{cases} \mathbf{A}_{h[k]} = \begin{pmatrix} \mathbf{B}_{[k]}^{-1} & \mathbf{N}_{[k]}^\top \\ \mathbf{N}_{[k]} & \mathbf{0}_{q \times q} \end{pmatrix} \\ \mathbf{M}_{h[k]} = \begin{pmatrix} (\mathbf{I}_t \otimes \mathbf{A}^\top) \text{vec}(\mathbf{M}) - \left(\mathbf{I}_t \otimes \mathbf{A}^\top \mathbf{A} + \frac{\beta}{\alpha} \mathbf{T}^\top \mathbf{T} \right) \text{vec}(\mathbf{X}_{[k]}) \\ \frac{1}{2} (\mathbf{1}_q - \mathbf{N}_{[k]} \text{vec}(\hat{\mathbf{X}}_{[k]})) \end{pmatrix} \end{cases} \quad (51)$$

of the equation system (50), and noting that $\mathbf{N}^T \mathbf{A} = (\mathbf{I}_t \otimes \text{diag}(\mathbf{A})) \text{vec}(\mathbf{X})$, finally yields

$$\begin{cases} \text{vec}(\mathbf{X}_{[k+1]}) = \text{vec}(\mathbf{X}_{[k]}^*) - \mathbf{B}_{[k]} \mathbf{N}_{[k]}^T (\mathbf{A}_{[k+1]} - \mathbf{A}_{[k]}) \\ \mathbf{A}_{[k+1]} = \left(\mathbf{N}_{[k]} \mathbf{B}_{[k]} \mathbf{N}_{[k]}^T \right)^{-1} \\ \quad \times \left[\mathbf{N}_{[k]} \left(\text{vec}(\mathbf{X}_{[k]}^*) - \frac{1}{2} \text{vec}(\hat{\mathbf{X}}_{[k]}) \right) - \frac{1}{2} \mathbf{1}_q \right] + \mathbf{A}_{[k]} \end{cases} \quad (53)$$

where we denoted: $\text{vec}(\mathbf{X}_{[k]}^*) = \mathbf{B}_{[k]} (\mathbf{I}_t \otimes \mathbf{A}^T) \text{vec}(\mathbf{M})$. These two linked vector series are computed iteratively until a stable point is reached, hence providing the conditional estimate of \mathbf{X} under the normalization constraint. Note that the seed estimate of \mathbf{X} (say $\mathbf{X}_{[1]}$) is derived by normalizing afterwards the non constrained estimate of \mathbf{X} . For the Lagrange multipliers vector initialization, we simply used the null vector: $\mathbf{A}_{[1]} = \mathbf{0}_q$.

ACKNOWLEDGMENT

The authors would like to thank S. Jbabdi, P. Bellec, and G. Marrelec for helpful discussions. They would also like to especially thank M. Péligrini-Issac for her thorough review.

REFERENCES

- [1] P. L. Nunez and R. B. Silberstein, "On the relationship of synaptic activity to macroscopic measurements: does co-registration of EEG with fMRI make sense?," *Brain Topogr.*, vol. 13, pp. 79–96, 2000.
- [2] M. Hämäläinen, R. Hari, R. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa, "Magnetoencephalography-theory, instrumentation, and application to noninvasive studies of the working human brain," *Rev. Modern Physics*, vol. 65, pp. 413–497, 1993.
- [3] J. Malmivuo and R. Plonsey, *Bioelectromagnetism*. New York: Oxford Univ. Press, 1981.
- [4] S. Baillet, J. C. Mosher, and R. M. Leahy, "Electromagnetic brain mapping," *IEEE Signal. Process. Mag.*, vol. 18, no. 6, pp. 14–30, Nov. 2001.
- [5] K. J. Koles, "Trends in EEG source localization," *Electroenceph. Clin. Neurophysiol.*, vol. 106, pp. 219–230, 1998.
- [6] R. D. Pascual-Marqui. (1999) Review of methods for solving the EEG inverse problem. *Int. J. Bioelectromagn.* [Online], pp. 75–86. Available: <http://www.tut.fi/ijbem>
- [7] P. L. Nunez, *Electric Fields of the Brain*. New York: Oxford Univ. Press, 1981.
- [8] J. Sarvas, "Basic mathematical and electromagnetic concepts of the bi-magnetic inverse problem," *Phys. Med. Biol.*, vol. 32, pp. 11–22, 1987.
- [9] M. Scherg and J. S. Ebersole, "Models of brain electromagnetic source analysis," *Brain Topogr.*, vol. 5, pp. 419–423, 1993.
- [10] M. Scherg and D. von Cramon, "Evoked dipole source potentials of the human auditory cortex," *Electroenceph. Clin. Neurophysiol.*, vol. 65, pp. 344–360, 1986.
- [11] J. C. de Munck, "The estimation of time-varying dipoles on the basis of evoked potentials," *Electroencephalogr. Clin. Neurophysiol.*, vol. 77, pp. 156–160, 1990.
- [12] J. C. Mosher and R. M. Leahy, "Multiple dipole modeling and localization from spatio-temporal MEG data," *IEEE Trans. Biomed. Eng.*, vol. 39, no. 6, pp. 541–557, Jun. 1992.
- [13] A. M. Dale and M. Sereno, "Improved localization of cortical activity by combining EEG and MEG with MRI surface reconstruction: a linear approach," *J. Cognit. Neurosci.*, vol. 5, pp. 162–176, 1993.
- [14] M. S. Hämäläinen and R. J. Ilmoniemi, "Interpreting magnetic fields of the brain—minimum norm estimates," *Med. Biol. Eng. Comput.*, vol. 32, pp. 35–42, 1994.
- [15] R. D. Pascual-Marqui, C. M. Michel, and D. Lehmann, "Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain," *Int. J. Psychophysiol.*, vol. 18, pp. 49–65, 1994.
- [16] I. E. Gorodnitsky, J. S. George, and B. D. Rao, "Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm," *Electroenceph. Clin. Neurophysiol.*, vol. 95, pp. 231–251, 1995.
- [17] S. Baillet and L. Garnero, "A Bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 5, pp. 374–385, May 1997.
- [18] J. J. Riera, M. E. Fuentes, P. A. Valdes, and Y. Oharriz, "EEG-distributed inverse solutions for a spherical head model," *Inverse Prob.*, vol. 14, pp. 1009–1019, 1998.
- [19] S. Baillet, L. Garnero, G. Marin, and J. P. Hugonin, "Combined MEG and EEG source imaging by minimization of mutual information," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 5, pp. 522–534, May 1999.
- [20] A. M. Dale, A. K. Liu, B. R. Fischl, R. L. Buckner, J. W. Belliveau, J. D. Lewine, and E. Halgren, "Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity," *Neuron*, vol. 26, pp. 55–67, 2000.
- [21] K. Sekihara, S. S. Nagarajan, D. Poeppel, A. Marantz, and Y. Miyashita, "Reconstructing spatio-temporal activities of neural sources using an MEG vector beamformer technique," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 7, pp. 760–771, Jul. 2001.
- [22] R. D. Pascual-Marqui, "Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details," *Meth. Find. Exp. Clin. Pharmacol.*, vol. 24, pp. 5–12, 2002.
- [23] J. Hori, M. Aiba, and B. He, "Spatio-temporal cortical source imaging of brain electrical activity by means of time-varying parametric projection filter," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 5, pp. 768–777, May 2004.
- [24] C. Amblard, E. Lalpalmé, and J. M. Lina, "Biomagnetic source detection by maximum entropy and graphical models," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 3, pp. 427–442, Mar. 2004.
- [25] N. J. Trujillo-Barreto, E. Aubert-Vasquez, and P. A. Valdes-Sosa, "Bayesian model averaging in EEG/MEG imaging," *Neuroimage*, vol. 21, pp. 1300–1319, 2004.
- [26] L. Gavit, S. Baillet, J. F. Mangin, J. P. Pechet, and L. Garnero, "A multiresolution framework to MEG/EEG source imaging," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 10, pp. 1080–1087, Oct. 2001.
- [27] O. Yamashita, A. Galka, T. Ozaki, R. Biscay, and P. Valdes-Sosa, "Recursive penalized least squares solution for dynamical inverse problems of EEG generation," *Hum. Brain Mapp.*, vol. 21, no. , pp. 221–235, 2004.
- [28] J. C. Mosher, R. M. Leahy, D. W. Shattuck, and S. Baillet, "MEG source imaging using multipolar expansions," in *Proc. 16th Conf. Information Processing in Medical Imaging (IPMI 2003)*, C. J. Taylor and J. A. Noble, Eds., 1999, pp. 15–28.
- [29] S. Baillet, J. C. Mosher, K. Jerbi, and R. M. Leahy, Hybrid MEG source characterization by cortical remapping and imaging of parametric source models. presented at Proc. 12th International Conf. on Biomagnetism (BIOMAG'2000). [Online]. Available: citeseer.ist.psu.edu/baillet00hybrid.html
- [30] J. C. de Munck, "The potential distribution in a layered spheroidal volume conductor," *J. Appl. Phys.*, vol. 64, pp. 464–470, 1988.
- [31] M. S. Hämäläinen and J. Sarvas, "Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data," *IEEE Trans. Biomed. Eng.*, vol. 36, no. 2, pp. 165–171, Feb. 1989.
- [32] J. Daunizeau, J. Mattout, B. Goulard, J. M. Lina, and H. Benali, "Data-driven cortex parcelling: a regularization tool for the EEG/MEG inverse problem," in *Proc. Int. Symp. Biomedical Imaging*, 2003, pp. 1343–1346.
- [33] J. Mattout, M. Péligrini-Issac, A. Bellio, J. Daunizeau, and H. Benali, "Localization Estimation Algorithm (LEA): a supervised prior-based approach for solving the EEG/MEG inverse problem," in *Proc. Information Processing in Medical Imaging Conf. (IPMI'2003)*, C. J. Taylor and J. A. Noble, Eds., 2003, pp. 536–547.
- [34] H. H. Harman, *Modern Factor Analysis, Third Edition Revised*. Chicago, IL: The Univ. Chicago Press, 1989.
- [35] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*. New York: Springer-Verlag, 1985.
- [36] K. J. Friston, W. Penny, S. Kiebel, G. Hinton, and J. Ashburner, "Classical and Bayesian inference in neuroimaging: theory," *Neuroimage*, vol. 16, pp. 465–483, 2002.
- [37] J. Ortega and W. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*. New York: Academic, 1970.
- [38] G. L. Bretthorst, "Bayesian spectrum analysis and parameter estimation," in *Lecture Notes in Statistics*. New York: Springer-Verlag, 1988.
- [39] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," *SIAM Rev.*, vol. 34, pp. 561–580, 1992.
- [40] J. F. Mangin, V. Frouin, I. Bloch, J. Régis, and J. Lopez-Krahe, "From 3D magnetic resonance images to structural representations of the cortex topography using topology preserving deformations," *J. Math. Imag. Vis.*, vol. 5, pp. 297–318, 1995.
- [41] BrainStorm Software. [Online]. Available: <http://neuroimage.usc.edu/brainstorm/>
- [42] J. P. Egan, *Signal Detection Theory and ROC Analysis*. New York: Academic, 1975.

- [43] J. A. Swets, "Measuring the accuracy of diagnostic systems," *Science*, vol. 240, p. 1285, 1988.
- [44] J. Daunizeau, C. Grova, J. Mattout, G. Marrelec, D. Clonda, B. Goulard, J. M. Lina, and H. Benali, "Assessing the relevance of fMRI-based prior in the EEG inverse problem: a Bayesian model comparison approach," *IEEE Trans. Signal. Process.*, no. 9, pp. 3461–3472, Sep. 2005.
- [45] A. Hyvarinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Netw.*, vol. 13, pp. 411–430, 2000.
- [46] T. P. Jung, S. Makeig, M. J. McKeown, A. J. Bell, T. W. Lee, and T. J. Sejnowski, "Imaging brain dynamics using independent component analysis," *Proc. IEEE*, vol. 89, no. 7, pp. 1107–1121, Jul. 2001.
- [47] J. C. Mosher and R. M. Leahy, "Recursive MUSIC: a framework for EEG and MEG source localization," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 11, pp. 1342–1354, Nov. 1998.
- [48] ———, "Source localization using recursively applied and projected (RAP) MUSIC," *IEEE Trans. Signal. Process.*, vol. 37, no. 2, pp. 332–340, Feb. 1999.
- [49] A. N. Tikhonov, *Solutions of Ill-Posed Problems*, A. N. , Ed. Washington, D.C.: H. W. Winston, 1977.
- [50] J. K. O. Ruanaidh and W. J. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing*. New York: Springer-Verlag, 1996.

Jean Daunizeau was born in Boulogne, France, in 1977. In 2002, he received the M.Sc. degree in physics from Université Paris-Sud, France, with a specialization in medical imaging. Since 2002, he has been working towards the Ph.D. degree in the Unité 678 "Laboratoire d'imagerie fonctionnelle," INSERM, Paris, France, and in the Centre de Recherches Mathématiques, Université de Montréal, Montréal, QC, Canada.

His research interests include statistical modeling, inverse problems and information fusion with applications mainly in EEG/MEG/fMRI activation/connectivity data analysis.

Jérémie Mattout was born in France in 1974. From 1995 to 1998, he was a student in applied physics at the Institut Supérieur de la Matière et du Rayonnement, Caen, France. He completed the Ph.D. degree in 2002, at the Quantitative Medical Imaging INSERM Unit and the Cognitive Neuroscience and Brain Imaging CNRS Laboratory, Pitié-Salpêtrière Hospital, Paris, France.

He is now a Research Fellow at the Wellcome Department of Imaging Neuroscience, London, U.K.. His current research focuses on the use of electromagnetic neuroimaging data to study brain functional integration.

Diego Clonda was born in Montréal, QC, Canada, in 1974. He received the B.Sc. degree in physics from Université de Montréal in 1995 and the M.Sc. degree in medical physics from McGill University, Montreal, in 1998. He is currently working towards the Ph.D. degree in physics at the Centre de Recherches Mathématiques (Montréal).

From 1996 to 1998 he worked in the Image-Guided Neurosurgery Lab of McGill's biomedical engineering department, where he did his Master's research project. His main research interests are wavelet-based signal processing, statistical modeling in imagery and Markovian models.

Bernard Goulard was born in Paris, France. He received the undergraduate degree from the University of Nancy and University of Grenoble. He received the Ph.D. degree from the University of Pennsylvania, Philadelphia, PA, on "Weak Interactions in Nuclei" in 1964.

As a Professor at Université de Montréal, he extended this field to weak and electromagnetic interactions in nuclei, till 1990. Then, he applied his knowledge of theoretical methods to treatment of signals and more recently to medical imaging problems.

Habib Benali received the Ph.D. degree in signal processing and statistics from Rennes University, France, in 1984, and the "Habilitation à Diriger des Recherches" (diploma to lead research project) in 1995 from Paris Dauphine University, Paris, France.

From 1984–1987 he has been an Assistant Professor in the department of Computer Science and Statistics, IUT Vannes. He joined the INSERM (French National of Health and Medical Research) in 1987. He is currently a Researcher Director in the Unit 678 "Laboratory of Functional Imaging" of the INSERM, where he leads the group "Parametric Imaging of the Cerebral Activity and Metabolism." Since 1999, he is affiliated as a regular Researcher at the Mathematical Research Center of the University of Montreal, Montreal, QC, Canada. He has published more than 70 papers including original publications in peer-reviewed journals, didactic publications, and books. His current research interests are in human brain mapping and functional connectivity analysis using fMRI/EEG/MEG data, statistical modeling of the neural networks involved in motor skill learning in the human.

Dr. Benali is a member of the OHBM, ASU, SFdS, SFGMB societies.

Jean-Marc Lina received the Electrical Eng. Diploma from the Institut National Polytechnique de Grenoble, Grenoble, France, in 1982, and the M.Sc. and Ph.D degree in mathematical physics from the University of Montréal, Montréal, QC, Canada, in 1984 and 1990 respectively.

He is currently Professor in the Electrical Engineering Department of the Ecole Supérieure de Technologie in Montreal. His research interest include wavelets, learning algorithms, statistical modeling, and entropic techniques in signal processing with applications in various domain of image processing and brain imaging.