Assessing the Relevance of fMRI-Based Prior in the EEG Inverse Problem: A Bayesian Model Comparison Approach

Jean Daunizeau, Christophe Grova, Jérémie Mattout, Guillaume Marrelec, Diego Clonda, Bernard Goulard, Mélanie Pélégrini-Issac, Jean-Marc Lina, and Habib Benali, *Member, IEEE*

***Abstract*—Characterizing the cortical activity from electro- and magneto-encephalography (EEG/MEG) data requires solving an ill-posed inverse problem that does not admit a unique solution. As a consequence, the use of functional neuroimaging, for instance, functional Magnetic Resonance Imaging (fMRI), constitutes an appealing way of constraining the solution. However, the match between bioelectric and metabolic activities is desirable but not assured. Therefore, the introduction of spatial priors derived from other functional modalities in the EEG/MEG inverse problem should be considered with caution. In this paper, we propose a Bayesian characterization of the relevance of fMRI-derived prior information regarding the EEG/MEG data. This is done by quantifying the adequacy of this prior to the data, compared with that obtained using an noninformative prior instead. This quantitative comparison, using the so-called Bayes factor, allows us to decide whether the informative prior should (or not) be included in the inverse solution. We validate our approach using extensive simulations, where fMRI-derived priors are built as perturbed versions of the simulated EEG sources. Moreover, we show how this inference framework can be generalized to optimize the way we should incorporate the informative prior.**

***Index Terms*—Bayes factor, EEG, fMRI, fusion, MEG, prior, relevance.**

I. Introduction

BECAUSE neuroimaging gives access to both the localization and the dynamics of the sources of cerebral activity, it is nowadays a major tool for the investigation of any cognitive process. In contrast to positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), which measure cerebral vascular and metabolic variations resulting from changes in neuronal activity, electroencephalography (EEG) and magnetoencephalography (MEG) are direct physical measurements of neuronal currents. Moreover, they are the only modalities capable of resolving temporal patterns of neuronal activity in the millisecond range [1]. However, knowledge of the scalp electric (resp. magnetic) field does not allow an estimation of, in an unequivocal manner, the current generators with which this field is associated. The so-called EEG/MEG inverse problem is said to be mathematically ill-posed; it has no unique solution in the most general unconstrained case.

Relying on the physical concept of current dipole, which is a plausible bioelectric source model [2], distributed methods consist of describing a predefined dense set of dipoles, typically spread all over the cortical sheet [3]. Each of these elementary dipoles models a neuronal macrocolumn whose activation could potentially explain the measurements. Given this anatomical constraint, estimating the amplitude of each elementary dipole is a linear problem and should allow not only the localizaton of the activated areas but the quantification of their spatial extent as well. However, the major drawback of this approach is the huge number of parameters to be estimated (approximately 10^3) compared to the available data (10^2 sensors). Thus, the problem remains underdetermined, and regularization is all the more needed to constrain the solution space by specifying some *a priori* information [4] (see also [5] and [6] for typical Bayesian approaches). It is extremely appealing to introduce spatial priors obtained from different neuroimaging techniques (like fMRI) in order to improve the spatial accuracy of the EEG/MEG inverse problem methods. The basic assumption underlying such an approach is the existence of a coupling between the bioelectric activity and the Blood Oxygenation Level Dependent (BOLD) response, which is the main biophysical effect exploited by fMRI to infer part of the brain metabolic activity. Previous EEG/MEG-fMRI fusion approaches have relied on this implicit link to build their methodology [7]–[9]. Yet divergences between the anatomical localization obtained by functional techniques and those obtained from electrocortical stimulations are not unfrequent (cf. [10] or [11] for review and insights). This has insidious consequences in the constrained estimation of the EEG/MEG sources. For instance, in [7], the authors have recognized that when fMRI was considered as the "truth" for spatial information, serious bias might occur when

Manuscript received September 30, 2004; revised March 2, 2005. J. Daunizeau was supported by the Association pour la Recherche contre le Cancer. C. Grova was supported by the J. Timmins fellowship from the Montreal Neurological Institute. G. Marrelec was supported by the Fondation Fyssen. The associate editor coordinating the review of this paper and approving it for publication was Guest Editor Dr. Guido Nolte.

J. Daunizeau, and H. Benali are with INSERM UMR-S U678/UPMC, Paris, France, and also with the Centre de Recherches Mathématiques, Montréal, QC, H3C 1K3 Canada (e-mail: jean.daunizeau@imed.jussieu.fr; habib.benali@imed.jussieu.fr).

C. Grova is with the Montréal Neurological Institute, McGill University, Montréal, QC, H3A 2T5 Canada. (e-mail: christophe.grova@mail.mcgill.ca)

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

G. Marrelec is with INSERM U494, Paris, France, and also with the Functional Neuroimaging Unit, Université de Montréal, Montréal, QC, H3W 1W5 Canada. (e-mail: guillaume.marrelec@umontreal.ca)

D. Clonda and B. Goulard are with the Centre de Recherches Mathématiques, Montréal, QC, H3C 3J7 Canada (e-mail: clonda@crm.umontreal.ca, goulard@crm.umontreal.ca).

M. Pélégrini-Issac is with the INSERM UMR-S U678/UPMC, Paris, France. (e-mail: melanie.pelegrini@imed.jussieu.fr)

J.-M. Lina is with the Ecole de Technologie Supérieure, Montréal, QC, H3C 1K3 Canada, INSERM UMR-S U678/UPMC, Paris, France, and also with the Centre de Recherches Mathématiques, Montréal, QC, H3A 2T5 Canada (e-mail: jmlina@ele.etsmtl.ca).

Digital Object Identifier 10.1109/TSP.2005.853220

an actual EEG/MEG source did not induce significant variations of the BOLD signal. Therefore, as it has been claimed in [11]: "integration of functional modalities into the solution of the neuroelectromagnetic inverse problem should be cautiously considered until a more tight coupling between BOLD effects and electrophysiological measurements could be established".

Since the estimates of the EEG/MEG sources will be different depending on the chosen prior, the reliability of the inverse reconstruction should depend on the credibility of the prior information, but given solutions constrained or unconstrained by fMRI, which one should be chosen? One should evaluate how much one can trust the constrained estimation, since it is known that fMRI and EEG/MEG may not be fully concordant. Any regularized estimation is the result of a tradeoff between a data adequacy term (likelihood) and a regularization term (prior). Intuitively, if the informative prior derived from fMRI is more or less concordant with the underlying EEG/MEG activity sources (which is desirable, but not necessarily so), both the EEG/MEG data and the prior should "pull in the same direction." Conversely, the higher the mismatch between the fMRI-derived prior and the EEG/MEG underlying sources, the worse the conflict between the EEG/MEG data adequacy and the fMRI-derived prior, compared with the result obtained using a noninformative prior instead. This conflict, if it could be quantified, should assess the relevance of the informative prior assumption. In this paper, we propose to use the only EEG/MEG data to characterize the relevance of any fRMI-derived prior that could be incorporated in the EEG/MEG inverse problem. This is done in a hierarchical Bayesian inference framework [12], which allows the taking into account of the uncertainty arising from any parametrization of the probabilistic model (e.g., through the use of unknown variance hyperparameters). The so-called Bayes factor then compares the posterior probability of the informative prior to that of a noninformative prior. This index of the relevance of the informative prior allows us to decide whether a given fMRI-derived prior should be included (or not) in the inverse problem. Furthermore, we show how this inference can be used to optimize the way we include the informative prior.

In Section II, we detail the Bayesian framework built in order to assess the relevance of the fMRI-derived prior model. In Section III, we describe the strategy used to evaluate the proposed methodology. The results are presented in Section IV and finally discussed, together with the method in Section V.

II. Theory: Assessing the Relevance of fMRI-Derived Priors

Let us assume that we have a linear forward model at our disposal, which links the available data (EEG/MEG measurements) to the parameters of interest (the intensities of the dipoles):

$$\mathbf{M} = \mathbf{GJ} + \mathbf{E} \tag{1}$$

where $\mathbf{M}$ is the $p \times t$ data matrix (p is the number of sensors and t the number of time samples), $\mathbf{E}$ models an additive noise, $\mathbf{J}$ is the $n \times t$ matrix of unknown parameters (n is the number of dipoles), and $\mathbf{G}$ is the $p \times n$ linear transform operator (forward operator). In the distributed source framework, $\mathbf{G}$ is obtained by solving the so-called forward problem for a given set of dipoles with fixed position and orientation (spread perpendicularly to the cortical surface) [13]. Each column of $\mathbf{G}$ gives the expected measurements on the scalp associated with the corresponding unit dipole (its so-called forward field).

The ill-posed estimation of $\mathbf{J}(n \gg p)$ is done under additional constraints derived from a given prior assumption. The tradeoff between the data likelihood and the prior terms is driven through additional parameters denoted as "hyperparameters" that have to be estimated. We wish to quantify the posterior probability of any prior assumption included in the inverse problem. Hence, we have to account for the uncertainty arising from any varying parameter of the model. This is done by considering all unknowns as nuisance parameters and integrating the joint posterior probability density function (pdf) upon them.

In the following, tr[$\mathbf{M}$] and $\mathbf{M}^T$ will denote the trace and the transpose of the matrix $\mathbf{M}$, respectively. Given any $n \times 1$ column vector $\mathbf{V}$, diag($\mathbf{V}$) is the $n \times n$ diagonal matrix whose diagonal elements are those of the vector $\mathbf{V}$. $\Gamma(x)$ denotes the gamma function evaluated at x. For two variables x and y, $x \mid y$ stands for "x given y" and $p(x)$ for the probability of x.

A. First Level of Inference: Estimating the Parameters

Assuming a specific prior model for the unknown parameters of interest ($\mathbf{J}$), the Bayesian inference framework allows us to derive their estimation. Let us write Bayes' rule:

$$p(\mathbf{J} \mid \sigma^2, \epsilon^2, \mathbf{M}, H) = \frac{p(\mathbf{M} \mid \mathbf{J}, \sigma^2) \cdot p(\mathbf{J} \mid \sigma^2, \epsilon^2, H)}{p(\mathbf{M} \mid \sigma^2, \epsilon^2, H)} \tag{2}$$

where we have the following.

- $p(\mathbf{J} \mid \sigma^2, \epsilon^2, \mathbf{M}, H)$ is the pdf of $\mathbf{J}$, given the data, a set of mutually independent hyperparameters (σ^2, ϵ^2), and a certain hypothesis H to be defined. The Maximum *A Posteriori* (MAP) of $\mathbf{J}$ maximizes this quantity.
- $p(\mathbf{M} \mid \mathbf{J}, \sigma^2)$ is the data likelihood, knowing the model parameters ($\mathbf{J}$) and the hyperparameter σ^2. Here, we build it, assuming both independence between (ϵ^2, H) and $\mathbf{M}$ and an i.i.d., zero-mean, and Gaussian noise $\mathbf{E}$

$$p(\mathbf{M} \mid \mathbf{J}, \sigma^2) = (2\pi)^{-(pt/2)} \cdot (\sigma^2)^{-(pt/2)} \cdot \exp\left(-\frac{1}{2\sigma^2}\text{tr}\left[(\mathbf{M} - \mathbf{GJ})^T(\mathbf{M} - \mathbf{GJ})\right]\right) \tag{3}$$

  where the hyperparameter σ^2 is the noise variance.
- $p(\mathbf{J} \mid \sigma^2, \epsilon^2, H)$ is the prior pdf of $\mathbf{J}$, dependent on the hyperpameters (σ^2, ϵ^2), and on the hypothesis H. We chose to express the hyperparameter of the Gaussian prior pdf of $\mathbf{J}$ as a formal rescaling of the noise variance (σ^2/ϵ^2) [14]. This allows us to infer directly the relative weight of the prior term in the posterior covariance pdf (see Section II-B). Then, the prior pdf of $\mathbf{J}$ is built as a zero-mean Gaussian pdf, with the covariance matrix dependent on H, $(\sigma^2/\epsilon^2)(\mathbf{L}^{(H)T}\mathbf{L}^{(H)})^{-1}$, as follows:

$$p(\mathbf{J} \mid \epsilon^2, \sigma^2, H) = (2\pi)^{-(nt/2)} \cdot \left(\frac{\epsilon^2}{\sigma^2}\right)^{(nt/2)} \cdot \mid \mathbf{L}^{(H)} \mid^t \cdot \exp\left(-\frac{\epsilon^2}{2\sigma^2} tr[\mathbf{J}^T\mathbf{L}^{(H)T}\mathbf{L}^{(H)}\mathbf{J}]\right). \tag{4}$$

Note that we further assume the independence of the different time samples (as in the likelihood definition).

- $p(\mathbf{M} \mid \sigma^2, \epsilon^2, H)$ is the data "evidence" that is conditional on the hyperparameters and hypothesis H. This quantity does not play any role at this level of inference.

Given a set of hyperparameters (σ^2, ϵ^2), it is possible to define $\hat{\mathbf{J}}$, which is the MAP estimate of $\mathbf{J}$, as the $\mathbf{J}$ value that maximizes $p(\mathbf{J} \mid \sigma^2, \epsilon^2, \mathbf{M}, H)$. This is equivalent to minimizing the following functional, which is the sum of the log-likelihood and a regularization term weighted by ϵ^2:

$$\hat{\mathbf{J}} = \arg\min_{\mathbf{J}}[\mathrm{tr}\left[(\mathbf{M}-\mathbf{GJ})^T(\mathbf{M}-\mathbf{GJ})\right] + \epsilon^2 \mathrm{tr}[\mathbf{J}^T\mathbf{L}^{(H)T}\mathbf{L}^{(H)}\mathbf{J}]]. \quad (5)$$

Its analytical expression is straightforward:

$$\hat{\mathbf{J}} = \mathbf{\Sigma}^{-1}\mathbf{G}^T\mathbf{M} \quad (6)$$

where $\mathbf{\Sigma}$ is defined as

$$\mathbf{\Sigma} = \mathbf{G}^T\mathbf{G} + \epsilon^2 \cdot \mathbf{L}^{(H)T}\mathbf{L}^{(H)}. \quad (7)$$

In that parametrization of the problem, $\hat{\mathbf{J}}$ is formally independent from the noise variance σ^2, in contrast to the posterior covariance matrix of J, $\sigma^2\mathbf{\Sigma}^{-1}$, which states that the efficiency of the estimation decreases with the power of the noise.

B. Second Level of Inference: Estimating the Hyperparameters

Equation (6) shows that the estimation of $\mathbf{J}$ remains conditional on ϵ^2 and hypothesis H. Therefore, in order to optimize the previous level of inference, we have to find an estimate for the hyperparameter ϵ^2. Let us write Bayes' rule for this second level of inference:

$$p(\sigma^2, \epsilon^2 \mid \mathbf{M}, H) = \frac{p(\mathbf{M} \mid \sigma^2, \epsilon^2, H) \cdot p(\sigma^2, \epsilon^2)}{p(\mathbf{M} \mid H)} \quad (8)$$

where we have the following.

- $p(\sigma^2, \epsilon^2 \mid \mathbf{M}, H)$ is the posterior pdf of the hyperparameters conditional on the data and hypothesis H. We will maximize this law to define an estimate of the hyperparameters.
- $p(\mathbf{M} \mid \sigma^2, \epsilon^2, H)$ is the data likelihood of the hyperparameters and hypothesis H. Since this quantity is the normalization factor of the first level of inference (cf. (2)), it is derived by integration of the numerator of (2) over $\mathbf{J}$ (cf. Appendix A):

$$p(\mathbf{M} \mid \sigma^2, \epsilon^2, H) = (2\pi)^{-(pt/2)} \cdot (\sigma^2)^{-(pt/2)} \cdot \epsilon^{nt} \cdot \mid \mathbf{\Sigma} \mid^{-(t/2)} \cdot \mid \mathbf{L}^{(H)} \mid^t \times \exp\left(-\frac{1}{2\sigma^2} \cdot \left(\mathrm{tr}[\mathbf{M}^T\mathbf{M}] - \mathrm{tr}\left[\hat{\mathbf{J}}^T\mathbf{\Sigma}\hat{\mathbf{J}}\right]\right)\right). \quad (9)$$

- $p(\sigma^2, \epsilon^2)$ is the prior pdf of the hyperparameters. This law is taken as Jeffreys' law, which is a noninformative law (uniform pdf over the log of the hyperparameter) invariant to one-to-one reparametrization [15]):

$$p(\sigma^2, \epsilon^2) = p(\sigma^2) \cdot p(\epsilon^2) = R_\sigma \cdot R_\epsilon \cdot (\sigma^2)^{-1} \cdot (\epsilon^2)^{-1}$$
$$\sigma^2 \in \left[\sigma^2_{\min}, \sigma^2_{\max}\right] \text{ and } \epsilon^2 \in \left[\epsilon^2_{\min}, \epsilon^2_{\max}\right]$$

where σ^2 and ϵ^2 are assumed to be *a priori* statistically independent from each other and from the hypothesis H. Furthermore, R_σ and R_ϵ are defined as

$$R_\sigma = \left(\ln \frac{\sigma^2_{\max}}{\sigma^2_{\min}}\right)^{-1}, \quad R_\epsilon = \left(\ln \frac{\epsilon^2_{\max}}{\epsilon^2_{\min}}\right)^{-1}.$$

- $p(\mathbf{M} \mid H)$ is the data "evidence" given the hypothesis H, which does not have any influence on the inference at this level.

Applying Bayes' rule [see (8)] and integrating over σ^2 yields the posterior joint pdf of the hyperparameter ϵ^2 conditional on the hypothesis H (cf. Appendix B):

$$p(\epsilon^2 \mid \mathbf{M}, H) = C(H) \cdot \epsilon^{nt-2} \cdot \mid \mathbf{\Sigma} \mid^{-(t/2)} \cdot \left(\mathrm{tr}[\mathbf{M}^T\mathbf{M}] - \mathrm{tr}\left[\hat{\mathbf{J}}^T\mathbf{\Sigma}\hat{\mathbf{J}}\right]\right)^{-(pt/2)} \quad (10)$$

where $C(H)$ is a normalization coefficient (that does not depend on ϵ^2) such that

$$C(H) = \frac{1}{p(\mathbf{M} \mid H)} \cdot \pi^{-(pt/2)} \cdot R_\sigma \cdot R_\epsilon \cdot \Gamma\left(\frac{pt}{2}\right) \cdot \mid \mathbf{L}^{(H)} \mid^t.$$

We finally define $\hat{\epsilon}^2$, which is the MAP estimate of ϵ^2, as the value that maximizes the posterior pdf of ϵ^2 given the data (and hypothesis H):

$$\hat{\epsilon}^2 = \arg\max_{\epsilon^2}[p(\epsilon^2 \mid \mathbf{M}, H)]. \quad (11)$$

No analytic expression exists for this estimate, but since this pdf is unidimensional, we can evaluate it numerically by scanning the admissible values of ϵ^2. Due to the high underdetermination of the model, numerical difficulties complicate this computation (see Appendix C). However, this so-called empirical Bayesian strategy allows us to define conditional MAP estimates of $\mathbf{J}$ and σ^2. The former comes from (6), where $\mathbf{\Sigma}$ is evaluated at $\epsilon^2 = \hat{\epsilon}^2$ using (7). The latter is derived by noting that $(\sigma^2 \mid \mathbf{M}, H, \epsilon^2 = \hat{\epsilon}^2)$ is inverse-Gamma distributed, with scale and shape parameters $b(\epsilon^2) = \left(\mathrm{tr}[\mathbf{M}^T\mathbf{M}] - \mathrm{tr}\left[\hat{\mathbf{J}}^T\mathbf{\Sigma}\hat{\mathbf{J}}\right]\right)/2$ and $a = pt/2$, respectively (cf. Appendix B). This yields the following estimator for σ^2:

$$\hat{\sigma}^2 = \arg\max_{\sigma^2}\left[p(\sigma^2 \mid \mathbf{M}, H, \epsilon^2 = \hat{\epsilon}^2)\right] = \frac{b(\hat{\epsilon}^2)}{a+1}. \quad (12)$$

C. Third Level of Inference: Assessing the Relevance of the fMRI-Derived Prior Model

All that we have previously described was conditional on a certain hypothesis H, which was not explicitly stated, but was introduced in the prior pdf of $\mathbf{J}$. In order to assess the relevance of an fMRI-derived prior, we have to compare its data likelihood

to that of a noninformative prior. Therefore, we define two hypotheses H_0 and H_1:

1) H_0: "*The source intensities at time sample j are independent and have the same power.*" Under the Gaussian assumption, this is modeled as

$$\mathbf{J}_j \sim \mathcal{N}\left(\mathbf{0}_n, \frac{\sigma^2}{\epsilon^2}\mathbf{I}_n\right), \quad j = 1, \ldots, t. \tag{13}$$

In other words, the prior covariance matrix relative to the hypothesis H_0 [cf. Section II-A, (4)] is $\mathbf{L}^{(H_0)} = \mathbf{I}_n$. Since this hypothesis does not favor any particular source, H_0 will be denoted as the "noninformative" hypothesis (null hypothesis).

2) H_1: "*The source intensities at time sample j are independent, with their power linked to the activation map $\mathbf{Z}$ derived from fMRI data.*" Under the Gaussian assumption, this is modeled as

$$\mathbf{J}_j \sim \mathcal{N}\left(\mathbf{0}_n, \frac{\sigma^2}{\epsilon^2}\mathbf{f}(\mathbf{Z})\right), \quad j = 1, \ldots, t \tag{14}$$

where $\mathbf{f}$ is the linking function. Again, we can write $\mathbf{L}^{(H_1)} = \mathbf{f}(\mathbf{Z})^{-(1/2)}$. This will be denoted as the "informative" prior hypothesis. Since the prior variance of each source is the deviation from its null prior mean, the informative hypothesis H_1 may indeed be interpreted as a parametrization of the prior power of the sources, whatever the proper statistical nature of the fRMI activation map $\mathbf{Z}$.

According to our empirical Bayesian strategy, we wish to quantify the posterior probability of H_0 and H_1, given the only EEG/MEG data (without any dependence on other model parameters). Thus, by comparing those two posterior probabilities, we can define a measure of the relevance of the informative prior.

We then write Bayes' rule for this third level of inference:

$$p(H_i \mid \mathbf{M}) = \frac{p(\mathbf{M} \mid H_i) \cdot p(H_i)}{p(\mathbf{M})}, \quad i = 0, 1 \tag{15}$$

where we have the following.

- $p(H_i \mid \mathbf{M})$ is the posterior probability of the ith hypothesis given the data.
- $p(\mathbf{M} \mid H_i)$ is the data likelihood of H_i. Since it is the normalization factor of the second level of inference [cf. (8)], it is derived by integrating upon ϵ^2 the joint pdf of the hyperparameter and hypothesis H_i conditional on the data (cf. Appendix D):

$$p(\mathbf{M} \mid H_i) = K \cdot \mid \mathbf{L}^{(H_i)} \mid^t \cdot I_i \tag{16}$$

where the integral I_i and the constant K are defined as follows:

$$\begin{aligned} I_i &= \int \epsilon^{nt-2} \cdot \mid \boldsymbol{\Sigma} \mid^{-(t/2)} \\ &\quad \cdot \left(tr[\mathbf{M}^T\mathbf{M}] - \operatorname{tr}\left[\hat{\mathbf{J}}^T\boldsymbol{\Sigma}\hat{\mathbf{J}}\right]\right)^{-(pt/2)} d\epsilon^2 \\ K &= \pi^{-(pt/2)} \cdot \Gamma\left(\frac{pt}{2}\right) \cdot R_\sigma \cdot R_\epsilon. \end{aligned} \tag{17}$$

The integral I_i has no analytical solution, but we can estimate it by scanning on the admissible values of ϵ^2.

- $p(H_i)$ is the prior law upon the considered assumptions. In our case, we prefer not to privilege any of those and, therefore, set a uniform law $p(H_0) = p(H_1) = 1/2$.
- $p(\mathbf{M})$ is the data evidence. In order to normalize the posterior law on our hypothesis space, we write

$$p(\mathbf{M}) = \frac{1}{2}\left(p(\mathbf{M} \mid H_0) + p(\mathbf{M} \mid H_1)\right). \tag{18}$$

The posterior probability of each H_i is, hence, defined as

$$p(H_i \mid \mathbf{M}) = \frac{p(\mathbf{M} \mid H_i)}{p(\mathbf{M} \mid H_0) + p(\mathbf{M} \mid H_1)}.$$

Then, we define the prior relevance index α as the logarithm of the so-called Bayes factor [16], which, given that H_0 and H_1 are *a priori* equiprobable, is written as

$$\alpha = \ln\left(\frac{p(H_1 \mid \mathbf{M})}{p(H_0 \mid \mathbf{M})}\right). \tag{19}$$

A positive α value would mean that, given the only EEG/MEG data, the hypothesis H_1 is more probable (or relevant) than the hypothesis H_0. In other words, one can "trust" the active areas of the fMRI map and, consequently, favor the fMRI-constrained estimate. Ideally, this should, for instance, correspond to situations where fMRI and EEG/MEG "see concordant sources" and where the introduction of the fMRI-derived prior allows the finding of (true) sources that are not "seen" when using the noninformative prior. Conversely, a negative α value would mean that the sources identified by fMRI may not strictly correspond to the sources underlying the EEG/MEG measurements, yielding a biased (thus less probable) fMRI-constrained estimate.

III. Evaluation Using Simulated Data

1) Objective: The question the methodology is meant to answer concerns the choice of the estimate when both noninformative and informative priors are available.

We have therefore focused on studying the adequacy between the relevance index α, which is measured with respect to the data only and some "gold-standard"-driven relevance metrics that require the "ground truth" (the simulated sources).

2) Data Sets: In order to investigate different levels of EEG/fMRI concordance/discordance, several fMRI maps $\mathbf{Z}$ were built as perturbed versions of the actual distribution of EEG sources $\mathbf{J}_{\mathrm{EEG}}$. This allowed us to study the behavior of the prior relevance index α as a function of the "truth" of the prior we defined. Three perturbations of the fMRI map were considered:

- Random perturbation. The fMRI map was built as a random perturbation of the simulated sources $\mathbf{Z} = \left\{(\mathbf{J}_{\mathrm{EEG}}(i) + \mathbf{F}(i))^2; i = 1, \ldots, n\right\}$, where $\mathbf{F}$ was a $n \times 1$ random vector that followed a zero-mean Gaussian law $\mathbf{F} \sim \mathcal{N}\left(\mathbf{0}_n, \sigma_F^2\mathbf{I}_n\right)$. The higher the perturbation level σ_F^2, the less "relevant" the resulting $\mathbf{Z}$ map. For each EEG simulated data, ten prior covariance

matrices $\mathbf{L}^{(H)}$ were then built, including the noninformative one (identity), and nine informative ones based on nine $\mathbf{Z}$ maps perturbed at different levels ($\sigma_F^2 = 0$, 0.5, 1, 2, 4, 8, 16, 32, and 64). For each noise level, we performed 50 simulations, and in each of them, two extended (approximately 2.5 cm^2) sources were randomly chosen onto the cortical sheet.

- Distance between the underlying EEG source and the fMRI activation focus. For each simulated EEG source $\mathbf{J}_{\mathrm{EEG}}$, 13 fMRI foci $\mathbf{J}_{\mathrm{fMRI}}$ were randomly drawn within a corresponding increasing spatial neighborhood of the EEG source ($d_o = 1.3$, 1.9, 2.5, 3.2, 3.8, 4.4, 5.1, 5.5, 6.2, 6.9, 7.5, 8.1, and 8.8 cm, where d_o denotes the distance between the center of the simulated active area and the center of the fMRI activity focus). Both $\mathbf{J}_{\mathrm{EEG}}$ and $\mathbf{J}_{\mathrm{fMRI}}$ had the same spatial extent. Fifty pairs $(\mathbf{J}_{\mathrm{EEG}}, \mathbf{J}_{\mathrm{fMRI}})$ per spatial neighborhood were simulated. In this case, the fMRI prior was defined as $\mathbf{Z} = \{\mathbf{J}_{\mathrm{fMRI}}(i)^2; i = 1, \ldots, n\}$.
- Occurrence of spurious fMRI activations not seen in EEG. For each simulated EEG source $\mathbf{J}_{\mathrm{EEG}}$, we built two fMRI activation foci: one concordant (corresponding to the simulated map $\mathbf{J}_{\mathrm{EEG}}$) and one discordant (embodied in the so-called $\mathbf{J}_{\mathrm{fMRI}}$ map). The discordant fMRI focus, which is randomly drawn on the cortical sheet, did not correspond to any EEG activation and will be referred to as a spurious fMRI activation. All sources had the same spatial intensity profile, but different intensity ratios λ between the spurious and the concordant fMRI activations were investigated. The informative prior was then defined using the fMRI activation map $\mathbf{Z} = \left\{(\mathbf{J}_{\mathrm{EEG}}(i) + \lambda \mathbf{J}_{\mathrm{fMRI}}(i))^2; i = 1, \ldots, n\right\}$. For each of the seven intensity ratios ($\lambda = 0.01$, 0.1, 0.5, 1, 5, 10, and 100), 50 simulations were generated.

The head model used to simulate EEG signals consisted of a smoothed three-dimensional (3-D) cortical surface composed of 1148 sources. As it is commonly done in distributed modeling, the orientation of each source was set to the normal to the surface. An acquisition system of 117 EEG sensors was simulated. The forward matrix $\mathbf{G}$ was calculated according to a three-layer spherical model [13] using the BrainStorm software [17]. In all our simulations, the maximum value of the simulated EEG sources was $\max[\mathbf{J}_{\mathrm{EEG}}] = 1$. A white, zero-mean, i.i.d. Gaussian noise ($\sigma^2 = \max[\mathbf{M}]/10$) was added to the induced signal.

For each simulated EEG data and fMRI $\mathbf{Z}$ map, both MAP estimators $\hat{\mathbf{J}}_0$ and $\hat{\mathbf{J}}_1$ (derived from H_0 and H_1 hypotheses, respectively) were then computed, and the relevance of the informative prior was assessed using the index α. The EEG/fMRI coupling function $\mathbf{f}$ was defined as the heuristic:

$$\mathbf{f}(\mathbf{Z}) = \mathbf{I}_n + \frac{\delta - 1}{\max(\mathbf{Z})} \mathrm{diag}\,(\mathbf{Z}). \tag{20}$$

As in [8] and [18], the parameter δ, which tunes the weight of the fMRI constraint in the prior covariance matrix, was empirically set to $\delta = 10$.

3) Criteria: The sum of square errors $\left(\mathrm{SSE}(\hat{\mathbf{J}}) = \|\hat{\mathbf{J}} - \mathbf{J}_{\mathrm{EEG}}\|^2\right)$ assesses the accuracy of estimation of the current amplitudes. For the random perturbation, we also made use of the area under the receiver operating characteristic curve (AUC) [19] in order to quantify the detection accuracy. The received operating characteristic (ROC) curve was constructed by calculating the pair of sensitivity/specificity values (linked to the true/false positive/negative detection rates) for the overall range of possible thresholds of the reconstructed map $\hat{\mathbf{J}}$. In a frequentist point of view, the area under the ROC curve is, hence, an estimate of the probability to classify correctly an active dipole of the model.

From those evaluation metrics, we introduced two scores β and γ to compare the "fMRI-constrained" estimator $\hat{\mathbf{J}}_1$ and the corresponding unconstrained estimator $\hat{\mathbf{J}}_0$:

$$\beta = \ln\left(\frac{\mathrm{AUC}(\hat{\mathbf{J}}_1)}{\mathrm{AUC}(\hat{\mathbf{J}}_0)}\right), \quad \gamma = \ln\left(\frac{\mathrm{SSE}(\hat{\mathbf{J}}_1)}{\mathrm{SSE}(\hat{\mathbf{J}}_0)}\right).$$

Either $\beta > 0$ or $\gamma < 0$ means that the "fMRI-constrained" solution is the most accurate.

Finally, we introduced a third "gold standard"-based criterion d defined as $d = \ln(d_o/2)$. $d > 0$ means that the distance between the prior spot and the simulated area was greater than 2 cm.

Given the simulated sources, β, γ, and d allowed us to quantify the relevance of the informative prior. Therefore, the "data-driven" index of relevance α given by (19) was compared with those three "gold standard"-based validation metrics. In order to quantify the accuracy of the proposed methodology, the agreement between the "gold standard"-based criteria and the "data-driven" index of relevance was assessed by calculating a χ^2 score (using the decision threshold 0).

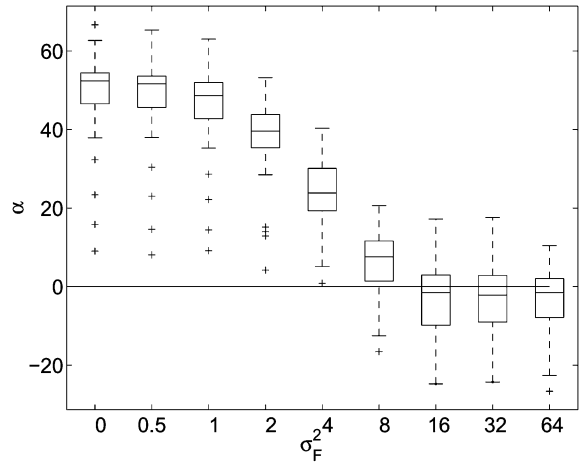
IV. Results

A. Random Perturbation

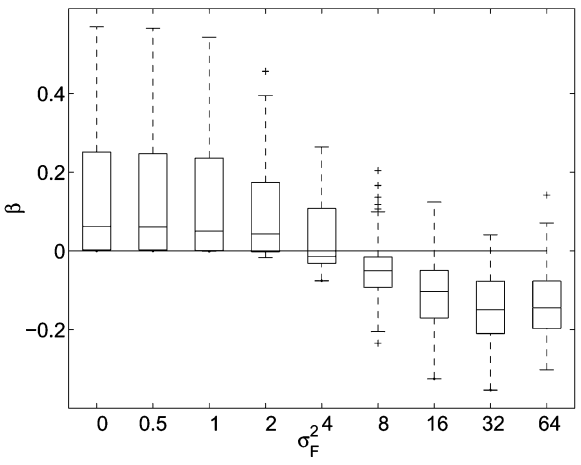
Boxplot representations of the distributions of α, β, and γ are presented in Fig. 1, as functions of the perturbation level σ_F^2. In all of our figures, boxplots show the median, first, and third quantiles and minimum/maximum values of the empirical distribution (apart from possible outliers depicted as single points). Up to a noise level of variance $\sigma_F^2 = 8$, the data-driven relevance index α favored the fMRI-constrained estimate. This decision was in agreement with both "gold standard"-based relevance criteria AUC and SSE. To study the adequacy between those criteria, we plotted each relevance metric as a function of each other one for all simulations (see Fig. 2). The three criteria α, β, and γ exhibited a significant statistical dependence ($\chi^2_{\alpha,\beta} = 66.5$, $\chi^2_{\alpha,\gamma} = 94.6$, and $\chi^2_{\beta,\gamma} = 239.1$; $P < 0.001$).

B. Effect of the Distance Between the Actual EEG Source and the fMRI Focus

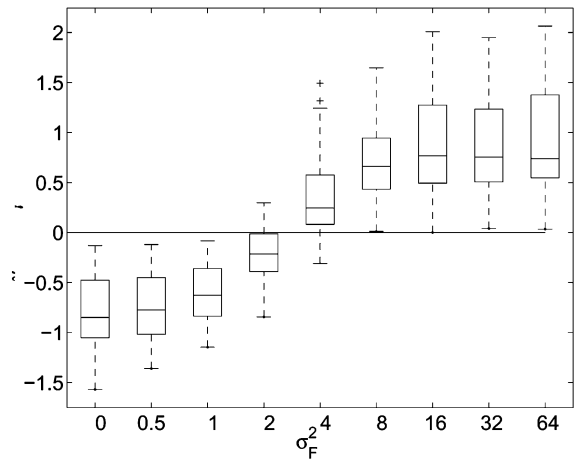
Boxplot representations of the distributions of α and γ criteria, for each distance d_o, are shown on Fig. 3. When the fMRI source was located close to the EEG source, even though the match was not perfect, using the fMRI-derived prior still



(a)

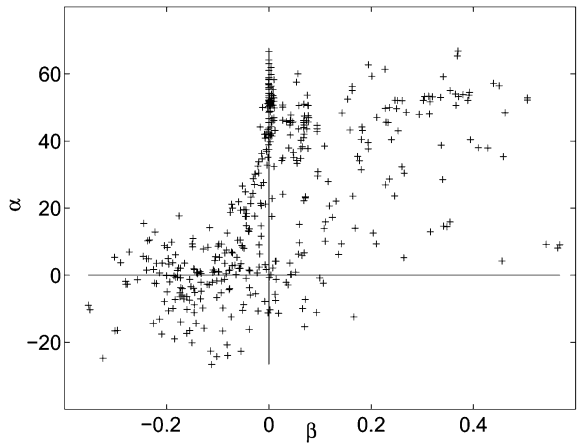


(b)

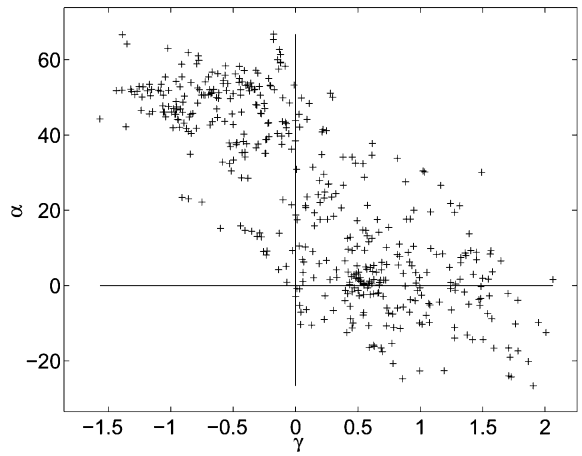


(c)

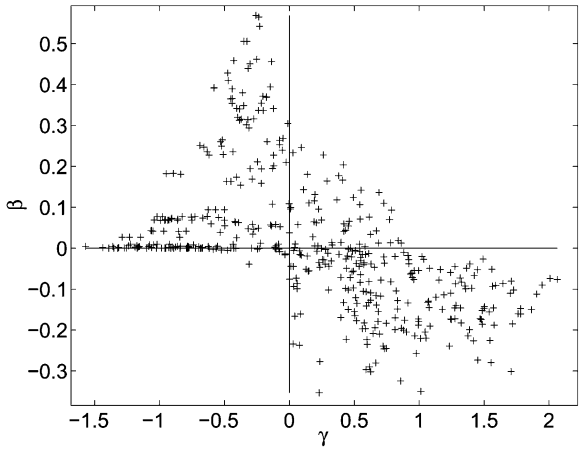
Fig. 1. Effect of the noise of the fMRI activation for $\sigma_F^2 = 0$, 0.5, 1, 2, 4, 8, 16, 32, and 64, respectively. Distribution of α, β, and γ for each noise level. Boxplot representations show the median, first, and third quantiles and minimum/maximum values of the empirical distribution (apart from possible outliers depicted as single points). (a) Distribution of α. (b) Distribution of β. (c) Distribution of γ.



(a)



(b)



(c)

Fig. 2. Random perturbation: adequacy between the relevance indices. Each point represents one of the three calculated pairs of relevance indices [(a) α versus β; (b) α versus γ; (c) β versus γ] for each source reconstruction. The zero value on each axis corresponds to the decision threshold for accepting the fMRI prior or not, based on the corresponding relevance index. This defines two pairs of quadrants on the plane, which are associated with concordance [top-right and bottom-left quadrants for (a), top-left and bottom-right quadrants for (b) and (c)) or divergence (top-left and bottom-right quadrants for (a), top-right and bottom-left quadrants for (b) and (c)] for the decision of whether the fMRI prior should be accepted or not. (a) Adequacy between α and β ($\chi^2 = 66.5, P < 0.001$). (b) Adequacy between α and γ ($\chi^2 = 94.6, P < 0.001$). (c) Adequacy between β and γ ($\chi^2 = 239.1, P < 0.001$).

enhanced the performance. A significant statistical dependence (see Fig. 4) between the three criteria α, γ, and d was exhibited ($\chi^2_{\alpha,\gamma} = 122.9$, $\chi^2_{\alpha,d} = 177.4$, and $\chi^2_{d,\gamma} = 301.8$; $P < 0.001$). This point is particularly interesting, since we have shown that the "data-driven" relevance index α was able to distinguish whether the actual EEG source and the fMRI focus were intersecting or not (the latter case corresponding here to $d_o > 2$ cm, where the noninformative prior should be favored).

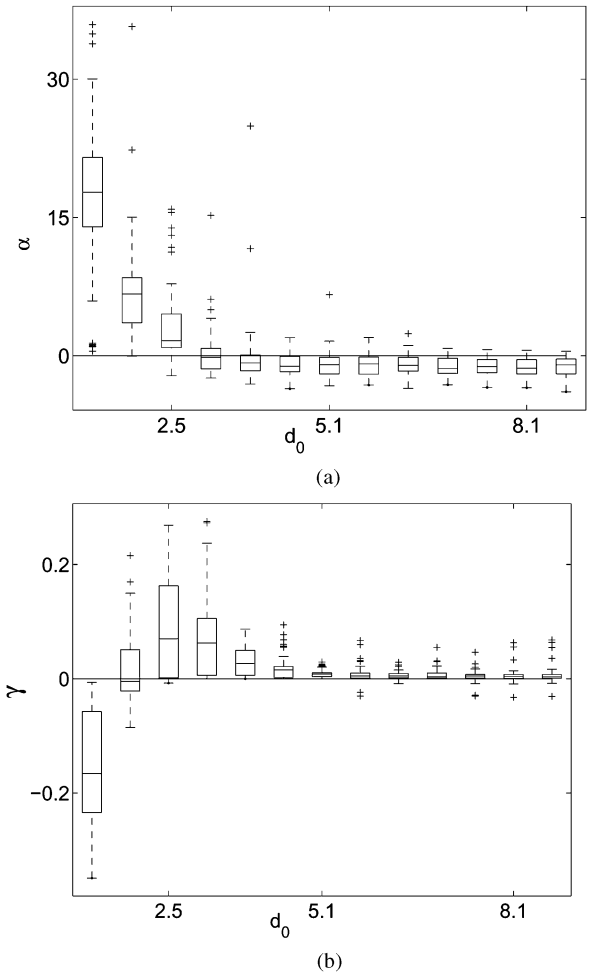


Fig. 3. Effect of the distance between the EEG source and the fMRI source. Distributions of α and γ as a function of the distance d_o. Boxplot representations show the median, first, and third quantiles and minimum/maximum values of the empirical distribution (apart from possible outliers depicted as single points). (a) Distribution of α. (b) Distribution of γ.

C. *Effect of the Occurrence of Spurious fMRI Activation Foci*

Fig. 5 shows boxplot representations of the distributions of α and γ as a function of the intensity ratio λ, as well as the adequacy between α and γ when pooling results for all intensity ratios. The χ^2 test suggests a very good agreement between α and γ ($\chi^2_{\alpha,\gamma} = 77.6$; $P < 0.001$). Moreover, both metrics showed a clear gap when $\lambda = 1$, from which the relevance of the fMRI-derived prior significantly drops. This corresponds to situations where the spurious source began to be more powerful than the truly activated one in the fMRI map $\mathbf{Z}$. Like the "gold-standard"-based metric γ, the "data-driven" relevance index α was able to distinguish between those cases. However, in almost all configurations, the solution constrained by the informative prior was favored by the "data-driven" relevance index α. This observation is in good agreement with the distribution of the "gold standard"-based metric γ.

V. Discussion

In this paper, we proposed to quantify the relevance (regarding the EEG/MEG data itself) of fMRI-derived informative priors that could be incorporated in the EEG/MEG inverse problem. Both informative (fMRI-derived) and noninformative prior hypotheses were compared through the use of Bayes factors in order to decide whether the informative prior should be included in the inverse problem. The proposed metric α

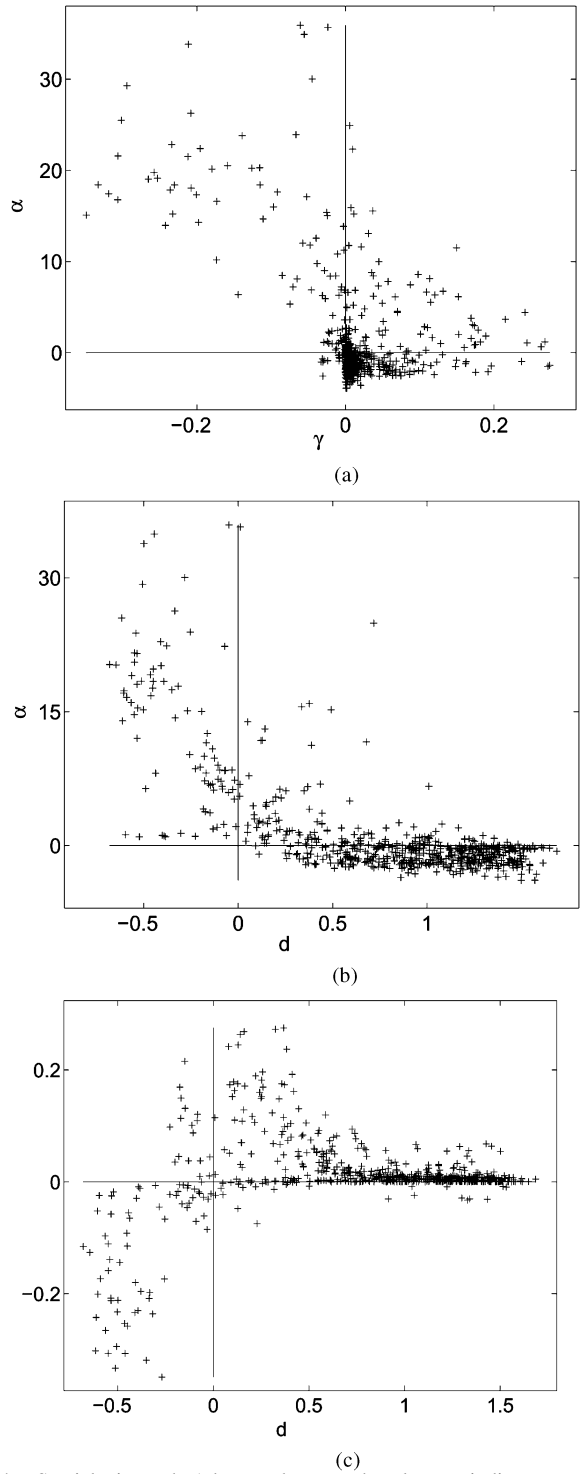


Fig. 4. Spatial mismatch: Adequacy between the relevance indices α, γ, and d. Each point represents one of the three calculated pairs of relevance indices [(a) α versus γ; (b) α versus d; (c) γ versus d] for each source reconstruction. The zero value on each axis corresponds to the decision threshold for accepting the fMRI prior or not, based on the corresponding relevance index. This defines two pairs of quadrants on the plane, which are associated to concordance [top-left and bottom-right quadrants for (a) and (b), top-right and bottom-left quadrants for (c)] or divergence [top-right and bottom-left quadrants for (a) and (b), top-left and bottom-right quadrants for (c)] for the decision of whether the fMRI prior should be accepted or not. (a) Adequacy between α and γ ($\chi^2 = 122.9$, $P < 0.001$). (b) Adequacy between α and d ($\chi^2 = 177.4$, $P < 0.001$). (c) Adequacy between γ and d ($\chi^2 = 301.8$, $P < 0.001$).

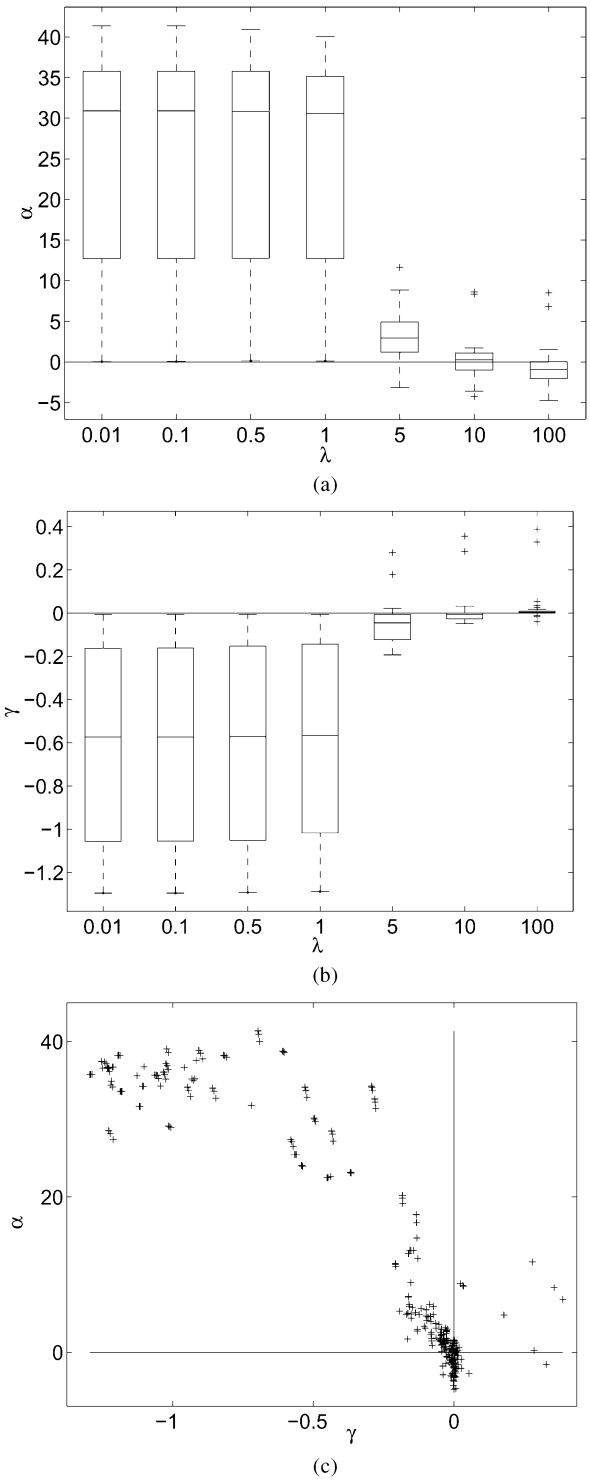


Fig. 5. Effect of the occurrence of an fMRI spurious source not seen in EEG. (a) and (b) Boxplot distributions of α and γ, respectively, for each intensity ratio λ between the spurious fMRI source and the concordant fMRI source. Boxplot representations show the median, first, and third quantiles and minimum/maximum values of the empirical distribution (apart from possible outliers depicted as single points). (c) Adequacy between α and γ. Each point represents the calculated pair of relevance indices (α versus γ) for each source reconstruction. The zero value on each axis corresponds to the decision threshold for accepting the fMRI prior or not, based on the corresponding relevance index. This defines two pairs of quadrants on the plane, which are associated with concordance (top-left and bottom-right quadrants) or divergence (top-right and bottom-left quadrants) for the decision of whether the fMRI prior should be accepted or not. (a) Distribution of α. (b) Distribution of γ. (c) Adequacy between α and γ ($\chi^2 = 77.6, P < 0.001$).

proved able to predict "gold standard"-based accuracy metrics that compare the "truth" (the simulated sources) with both constrained and unconstrained estimates. This suggests a very good agreement about the decision we take by introducing (or not) the informative prior between both the "gold standard"-based and the "data-driven" relevance indices.

As in any regularized estimation, the inference of $\mathbf{J}$ is conditional on the hyperparameter controlling the tradeoff between the data-likelihood and the prior term. Whereas existing classical techniques use an heuristic criterion (e.g., the L-curve method [20]), we provide an empirical Bayesian strategy to choose the optimal hyperparameter ϵ^2. Besides, note that we chose point MAP estimators for all unknown quantities of interest, but our approach could, without difficulty, be extended to a fully Bayesian inference scheme, e.g., based on purely marginal estimates. As for the third level of inference, our approach is linked to the Bayesian Model Averaging [16], which was already applied to the EEG/MEG inverse problem in order to reduce the support of the underlying sources in [21]. However, those authors made necessary approximations in the posterior pdf of the prior submodels (that could be here associated to different hypotheses H_i), due to the lack of unidimensional analytical form for the pdf of the hyperparameter ϵ^2. Hence, the benefits of our approach in the estimation of both the hyperparameter ϵ^2 and the posterior probability of hypotheses H_i have to be further investigated.

As we have already stated, numerical difficulties complicate the computation of the marginal pdf of the hyperparameter ϵ^2. This is essentially due to the estimation of the determinant of the source posterior covariance matrix $\mathbf{\Sigma}$. Preliminary results showed that even in a highly underdeterminated situation ($(p/n) \sim 100$), the algebraic solution we proposed (cf. Appendix C) allows us to obtain a stable calculus of the determinant. This seems promising as it should guarantee the applicability of the method, whatever the dimensionality of the problem.

In the evaluation strategy, we had to choose the threshold above which the cortical distance between the center of the underlying active area and the fMRI focus was associated with an irrelevant spatial prior. The mismatch of 2 cm has been frequently reported in the literature [10], [11], but the influence of this particular threshold was not quantitatively assessed.

Furthermore, the linking function $\mathbf{f}$ was defined using heuristic considerations as in [8]. In order to enhance the realism of the informative prior, we introduce in the EEG/MEG inverse problem, future works on real data should build this function $\mathbf{f}$ according to neurobiologically plausible considerations. In that respect, recent works proposed a parametric coupling of bioelectric and metabolic activities [22]. The third level of inference could then be used in order to infer the additional parameters of the coupling model. Note that H_0 and H_1 indeed already correspond to two limiting cases of the heuristic parametrization of the linking function $\mathbf{f}$ [cf. (20)]: $\delta = 1$ (no link) and $\delta = 10$ (high link), respectively. Applied on real datasets, the third level of inference may be used to optimize the heuristic parameter δ that weighted the fMRI contrast in the linking function, by scanning δ on its admissible values, and by defining $\hat{\delta}$ as its MAP estimate. Reject of the (fMRI-based)

informative prior hypothesis can then be associated with low δ MAP estimates.

Finally, note that while some of the fMRI activated areas may correspond to spurious EEG sources, some may be truly concordant. Therefore, the method might benefit from being adapted to evaluate not only fMRI priors globally but by considering the different areas separately. Indeed, at a higher computational expense, all combinations of activated areas may be included through a proper parametrization of the linking function $\mathbf{f}$ (which may act as a mask). Those "disturbed maps" can then be compared with each other in order to select the most probable subset of truly activated areas.

APPENDIX A
DERIVATION OF THE LIKELIHOOD OF THE HYPERPARAMETERS

According to Bayes' rule associated with the first level of inference [see (2)], the likelihood of the hyperparameters is derived as in (21), shown at the bottom of the page. Rearranging the exponential term and introducing $\hat{\mathbf{J}}$ [the MAP estimate of $\mathbf{J}$ cf. (6)], we obtain

$$\begin{aligned} p(\mathbf{M} \mid \sigma^2, \epsilon^2, H) &= A(\epsilon^2, \sigma^2, H) \\ &\quad \cdot \prod_{j=1}^{t} \underbrace{\int \exp\left(-\frac{1}{2\sigma^2} \cdot \left(\mathbf{J}_j - \hat{\mathbf{J}}_j\right)^T \boldsymbol{\Sigma} \left(\mathbf{J}_j - \hat{\mathbf{J}}_j\right)\right) d\mathbf{J}_j}_{} \\ &= A(\epsilon^2, \sigma^2, H) \cdot \prod_{j=1}^{t} a_j \end{aligned} \tag{22}$$

where $A(\epsilon^2, \sigma^2, H)$ is a normalization coefficient such that

$$\begin{aligned} A(\epsilon^2, \sigma^2, H) = (2\pi)^{-((n+p)t/2)} \cdot \sigma^{-(n+p)t} \cdot \epsilon^{nt} \cdot \mid \mathbf{L}^{(H)} \mid^t \\ \times \exp\left(-\frac{1}{2\sigma^2}\left(tr\left[\mathbf{M}^T\mathbf{M}\right] - tr\left[\hat{\mathbf{J}}^T\boldsymbol{\Sigma}\hat{\mathbf{J}}\right]\right)\right) \end{aligned} \tag{23}$$

and $\boldsymbol{\Sigma} = \mathbf{G}^T\mathbf{G} + \epsilon^2 \cdot \mathbf{L}^{(H)T}\mathbf{L}^{(H)}$. Indeed, the posterior pdf of $\mathbf{J}_j$, given the hyperparameters, is a Gaussian law of mean $\hat{\mathbf{J}}_j$ and variance-covariance matrix $\sigma^2\boldsymbol{\Sigma}^{-1}$. Then, the integral a_j is straightforward:

$$a_j = (2\pi)^{(n/2)} \cdot \sigma^n \cdot \mid \boldsymbol{\Sigma} \mid^{-(1/2)}. \tag{24}$$

This yields the data likelihood of the hyperparameters:

$$\begin{aligned} p(\mathbf{M} \mid \sigma^2, \epsilon^2, H) &= (2\pi)^{-(pt/2)} \cdot (\sigma^2)^{-(pt/2)} \\ &\quad \cdot \epsilon^{nt} \cdot \mid \boldsymbol{\Sigma} \mid^{-(t/2)} \cdot \mid \mathbf{L}^{(H)} \mid^t \\ &\quad \times \exp\left(-\frac{1}{2\sigma^2} \cdot \left(\text{tr}\left[\mathbf{M}^T\mathbf{M}\right] - \text{tr}\left[\hat{\mathbf{J}}^T\boldsymbol{\Sigma}\hat{\mathbf{J}}\right]\right)\right). \end{aligned} \tag{25}$$

APPENDIX B
DERIVATION OF THE MARGINAL PDF OF ϵ^2

Given the data likelihood knowing the set of hyperparameters, we apply Bayes' rule associated with the second level of inference [see (8)] to derive the joint posterior pdf of σ^2 and ϵ^2:

$$\begin{aligned} p(\sigma^2, \epsilon^2 \mid \mathbf{M}, H) = B(\epsilon^2, H) \cdot (\sigma^2)^{-(pt/2)-1} \\ \cdot \exp\left(-\frac{1}{2\sigma^2} \cdot \left(\text{tr}\left[\mathbf{M}^T\mathbf{M}\right] - \text{tr}\left[\hat{\mathbf{J}}^T\boldsymbol{\Sigma}\hat{\mathbf{J}}\right]\right)\right) \end{aligned} \tag{26}$$

where $B(\epsilon^2, H)$ is a normalization coefficient such that

$$\begin{aligned} B(\epsilon^2, H) = \frac{1}{p(\mathbf{M} \mid H)} \cdot (2\pi)^{-(pt/2)} \cdot R_\sigma \cdot R_\epsilon \\ \cdot \epsilon^{nt-2} \cdot \mid \boldsymbol{\Sigma} \mid^{-(t/2)} \cdot \mid \mathbf{L}^{(H)} \mid^t. \end{aligned} \tag{27}$$

We recognize for σ^2 an inverse-Gamma law $IG(a, b)$. Note that if $x \sim IG(a, b)$, then its pdf is written as $f_{IG}(x \mid a, b) = (b^a/\Gamma(a))x^{-a-1}\exp(-b/x)$. Identifying the different parameters yields

$$p(\sigma^2, \epsilon^2 \mid \mathbf{M}, H) = B(\epsilon^2, H) \cdot \frac{\Gamma(a)}{b^a} \cdot f_{IG}(\sigma^2 \mid a, b) \tag{28}$$

where the parameters of the inverse-Gamma law are $a = (pt/2)$ and $b = (1/2)\left(\text{tr}\left[\mathbf{M}^T\mathbf{M}\right] - \text{tr}\left[\hat{\mathbf{J}}^T\boldsymbol{\Sigma}\hat{\mathbf{J}}\right]\right)$. This allows us to obtain the exact law of the hyperparameter ϵ by integrating upon σ^2 (which is considered to be a nuisance parameter)

$$\begin{aligned} p(\epsilon^2 \mid \mathbf{M}, H) &= \int p(\sigma^2, \epsilon^2 \mid \mathbf{M}, H) d\sigma^2 \\ &= C(H) \cdot \epsilon^{nt-2} \cdot \mid \boldsymbol{\Sigma} \mid^{-(t/2)} \\ &\quad \cdot \left(\text{tr}\left[\mathbf{M}^T\mathbf{M}\right] - \text{tr}\left[\hat{\mathbf{J}}^T\boldsymbol{\Sigma}\hat{\mathbf{J}}\right]\right)^{-(pt/2)} \end{aligned} \tag{29}$$

$$\begin{aligned} p(\mathbf{M} \mid \sigma^2, \epsilon^2, H) &= \int p(\mathbf{M} \mid \mathbf{J}, \sigma^2, H) \cdot p(\mathbf{J} \mid \sigma^2, \epsilon^2, H) d\mathbf{J} \\ &= (2\pi)^{-((n+p)t/2)} \cdot \sigma^{-(n+p)t} \cdot \epsilon^{nt} \cdot \mid \mathbf{L}^{(H)} \mid^t \\ &\quad \times \int \exp\left(-\frac{1}{2\sigma^2} \cdot \sum_{j=1}^{t}\left(\left\|\mathbf{M}_j - \mathbf{G}\mathbf{J}_j\right\|^2 + \epsilon^2 \cdot \left\|\mathbf{L}^{(H)}\mathbf{J}_j\right\|^2\right)\right) d\mathbf{J} \\ &= (2\pi)^{-((n+p)t/2)} \cdot \sigma^{-(n+p)t} \cdot \epsilon^{nt} \cdot \mid \mathbf{L}^{(H)} \mid^t \\ &\quad \times \prod_{j=1}^{t} \int \exp\left(-\frac{1}{2\sigma^2} \cdot \left(\left\|\mathbf{M}_j - \mathbf{G}\mathbf{J}_j\right\|^2 + \epsilon^2 \cdot \left\|\mathbf{L}^{(H)}\mathbf{J}_j\right\|^2\right)\right) d\mathbf{J}_j. \end{aligned} \tag{21}$$

$$\begin{cases} p(\epsilon \mid \mathbf{M}, H_i) = \frac{K}{p(\mathbf{M}|H)} \epsilon^{-2} \mid \mathbf{I}_n - \mathbf{DG} \mid^{(t/2)} \left(\mathrm{tr}\left[\mathbf{M}^T\mathbf{M}\right] - \mathrm{tr}\left[\hat{\mathbf{J}}^T\mathbf{\Sigma}\hat{\mathbf{J}}\right]\right)^{-(pt/2)} \\ p(H_i \mid \mathbf{M}) = \frac{K}{p(H_0|\mathbf{M})+p(H_1|\mathbf{M})} \int \epsilon^{-2} \mid \mathbf{I}_n - \mathbf{DG} \mid^{(t/2)} \left(\mathrm{tr}\left[\mathbf{M}^T\mathbf{M}\right] - \mathrm{tr}\left[\hat{\mathbf{J}}^T\mathbf{\Sigma}\hat{\mathbf{J}}\right]\right)^{-(pt/2)} d\epsilon. \end{cases} \tag{36}$$

where $C(H)$ is a normalization coefficient such that

$$C(H) = \frac{1}{p(\mathbf{M} \mid H)} \cdot \pi^{-(pt/2)} \cdot R_\sigma \cdot R_\epsilon \cdot \Gamma\left(\frac{pt}{2}\right) \cdot \mid \mathbf{L}^{(H)} \mid^t . \tag{30}$$

Appendix C
Numerical Recipes

Typically, the (forward) distributed model underlying the EEG/MEG measurements is underdetermined: There are more parameters to be estimated than available data ($n \gg p$). This has a consequence in terms of the inversion of the posterior variance-covariance matrix of the sources, which is numerically instable (not to mention its determinant…). Yet, this step of numerical calculus is critical for our second and third levels of inference. We therefore have to rewrite the usual algebraic forms in order to end up with numerically stable calculus. Let us recall the MAP estimate of $\mathbf{J}$

$$\begin{aligned} \hat{\mathbf{J}} &= \left(\mathbf{G}^T\mathbf{G} + \epsilon^2\mathbf{L}^T\mathbf{L}\right)^{-1}\mathbf{G}^T\mathbf{M} \\ &= \mathbf{\Sigma}^{-1}\mathbf{G}^T\mathbf{M}. \end{aligned} \tag{31}$$

Now, the inverse of any matrix $\mathbf{B}$ of the form $\mathbf{B} = \mathbf{A} + \mathbf{U}\mathbf{V}^T$ (where $\mathbf{A}$ is an invertible matrix and for any $\mathbf{U}$ and $\mathbf{V}$ matrices) is given by the Shermann–Morrison–Woodbury formula [23]:

$$\mathbf{B}^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\left(\mathbf{U}\left(\mathbf{I} + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U}\right)^{-1}\mathbf{V}^T\right)\mathbf{A}^{-1}. \tag{32}$$

Using $\mathbf{B} = \mathbf{\Sigma}$, $\mathbf{U} = \mathbf{G}^T$, $\mathbf{V}^T = \mathbf{G}$, and $\mathbf{A} = \epsilon^2\mathbf{L}^T\mathbf{L}$, we then can write

$$\begin{aligned} \mathbf{\Sigma}^{-1} &= \frac{1}{\epsilon^2}(\mathbf{L}^T\mathbf{L})^{-1} - \frac{1}{\epsilon^4}(\mathbf{L}^T\mathbf{L})^{-1} \\ &\quad \times \left(\mathbf{G}^T\left(\mathbf{I} + \frac{1}{\epsilon^2}\mathbf{G}(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{G}^T\right)^{-1}\mathbf{G}\right)(\mathbf{L}^T\mathbf{L})^{-1} \\ &= \frac{1}{\epsilon^2}\left(\mathbf{I} - \mathbf{DG}\right)(\mathbf{L}^T\mathbf{L})^{-1} \end{aligned} \tag{33}$$

where $\mathbf{D}$ is the $n \times p$ matrix defined as

$$\mathbf{D} = (\mathbf{L}^T\mathbf{L})^{-1}\mathbf{G}^T\left(\epsilon^2\mathbf{I} + \mathbf{G}(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{G}^T\right)^{-1}. \tag{34}$$

This matrix is crucial, since it is strictly equivalent to the pseudo-inverse of $\mathbf{G}$ used in the "simple" form of the MAP estimate of $\mathbf{J}$. It is indeed straightforward to show that if $\mathbf{G}$ is of full rank, $\mathbf{D} = \mathbf{\Sigma}^{-1}\mathbf{G}^T$. Hence, we can rewrite the MAP estimate of $\mathbf{J}$ in the numerical stable form $\hat{\mathbf{J}} = \mathbf{DM}$.

Furthermore, (33) allows us also to rewrite the determinant of $\mathbf{\Sigma}$ in the following way:

$$\begin{aligned} \mid \mathbf{\Sigma} \mid &= \mid \mathbf{\Sigma}^{-1} \mid^{-1} \\ &= \epsilon^{2n} \mid \mathbf{L} \mid^2 \mid \mathbf{I} - \mathbf{DG} \mid^{-1} . \end{aligned} \tag{35}$$

This result can then be included in the posterior probabilities calculus of the second and third levels of inference as in (36), shown at the top of the page.

Appendix D
Derivation of the Likelihood of the Prior Assumption

The normalization constraint on the posterior pdf of the hyperparameter ϵ given to the data and the ith hypothesis allow us to write

$$\begin{aligned} 1 &= \int p\left(\epsilon^2 \mid \mathbf{M}, H_i\right) d\epsilon^2 \\ &= K \cdot \frac{1}{p(\mathbf{M} \mid H_i)} \cdot \mid \mathbf{L}^{(H_i)} \mid^t \\ &\quad \cdot \int \epsilon^{nt-2} \cdot \mid \mathbf{\Sigma} \mid^{-(t/2)} \\ &\quad \cdot \left(\mathrm{tr}\left[\mathbf{M}^T\mathbf{M}\right] - \mathrm{tr}\left[\hat{\mathbf{J}}^T\mathbf{\Sigma}\hat{\mathbf{J}}\right]\right)^{-(pt/2)} d\epsilon^2 \end{aligned}$$

where K is a normalization coefficient such that

$$K = \pi^{-(pt/2)} \cdot \Gamma\left(\frac{pt}{2}\right) \cdot R_\sigma \cdot R_\epsilon. \tag{37}$$

It is now possible to evaluate the data likelihood under the hypothesis H_i in the following way:

$$\begin{aligned} p(\mathbf{M} \mid H_i) &= K \cdot \mid \mathbf{L}^{(H_i)} \mid^t \cdot \int \epsilon^{nt-2} \cdot \mid \mathbf{\Sigma} \mid^{-(t/2)} \\ &\quad \cdot \left(\mathrm{tr}\left[\mathbf{M}^T\mathbf{M}\right] - \mathrm{tr}\left[\hat{\mathbf{J}}^T\mathbf{\Sigma}\hat{\mathbf{J}}\right]\right)^{-(pt/2)} d\epsilon^2 \\ &= K \cdot \mid \mathbf{L}^{(H_i)} \mid^t \cdot I_i \end{aligned} \tag{38}$$

where I_i represents the numerical evaluation of the above integral.

References

[1] 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. Mod. Phys.*, vol. 65, pp. 413–497, 1993.

[2] P. L. Nunez, *Electric Fields of the Brain*. Oxford, U.K.: New York, 1981.

[3] 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.

[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] 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.

[6] C. Bertrand, M. Ohmi, R. Suzuki, and H. Kado, "A probabilistic solution to the MEG inverse problem via MCMC methods: The reversible jump and parallel tempering algorithms," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 5, pp. 533–542, May 2001.

[7] 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.

[8] F. Babiloni, C. Babiloni, F. Carduci, L. Angelone, C. D. Gratta, G. L. Romani, P. M. Rossini, and F. Cincotti, "Linear inverse estimation of cortical sources by using high resolution EEG and fMRI priors," *Int. J. Bioelectromag.*, vol. 3, pp. 62–74, 2001.
[9] S. P. Ahlfors and G. U. Simpson, "Geometrical interpretation of fMRI-guided MEG/EEG," *Neuroimag.*, vol. 22, pp. 323–332, 2004.
[10] 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.
[11] S. L. Gonzales-Andino, O. Blanke, G. L. G. Thut, and R. G. de Peralta Merendez, "The use of functional constraints for the neuromagnetic inverse problem: Alternatives and caveats," *Int. J. Bioelectromag.*, vol. 3, pp. 103–114, 2001.
[12] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, "Bayesian Data Analysis," in *Tests in Statistical Science*. London, U.K.: Chapman and Hall, 1995.
[13] J. C. de Munck, "The potential distribution in a layered spheroidal volume conductor," *J. Appl. Phys.*, vol. 64, pp. 464–470, 1988.
[14] G. Marrelec, H. Benali, P. Ciuciu, M. Pélégrini-Issac, and J. B. Poline, "Robust bayesian estimation of the hemodynamic response function in event-related BOLD fMRI using basic physiological information," *Hum. Brain Map.*, vol. 19, pp. 1–17, 2003.
[15] G. L. Bretthorst, *Bayesian Spectrum Analysis and Parameter Estimation*. New York: Springer-Verlag, 1988, Lecture Notes in Statistics.
[16] J. A. Hoetling, D. Madigan, A. E. Raftery, and C. T. Volinsky, "Bayesian model averaging: A tutorial," *Statist. Sci.*, vol. 14, pp. 382–417, 1999.
[17] BrainStorm Software. [Online]. Available: http://neuroimage.usc.edu/brainstorm/
[18] A. K. Liu, J. W. Belliveau, and A. M. Dale, "Spatiotemporal imaging of human brain activity using functional MRI constrained magnetoencephalography data: Monte-Carlo simulations," *Proc. USA Nat. Acad. Sci.*, vol. 95, pp. 8945–8950, 1998.
[19] J. P. Egan, *Signal Detection Theory and ROC Analysis*. New York: Academic, 1975.
[20] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," *SIAM Rev.*, vol. 34, pp. 561–580, 1992.
[21] N. J. Trujillo-Barreto, E. Aubert-Vasquez, and P. A. Valdes-Sosa, "Bayesian model averaging in EEG/MEG imaging," *Neuroimag.*, vol. 21, pp. 1300–1319, 2004.
[22] M. Jones, N. Hewston-Stoate, J. Martindale, P. Redgrave, and J. Mayhew, "Nonlinear coupling of neural activity and CBF in rodent barrel cortex," *Neuroimag.*, vol. 22, pp. 956–965, 2004.
[23] W. W. Hager, "Updating the inverse of a matrix," *SIAM Rev.*, vol. 31, pp. 221–239, 1989.

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

Christophe Grova was born in Créhange, France, in 1974. In 1998, he received both the Engineering and Master degrees in biomedical engineering from the University of Technology of Compiègne, Compiègne, France, with a specialization in image and signal processing. From 1998 to 2002, he was with the IDM Laboratory, UPRES 3192, University of Rennes, Rennes, France, from which he received the Ph.D. degree in "Validation of SPECT/MRI registration methods in the context of epilepsy."

Since January 2003, he has been working with the team led by Prof. J. Gotman at the Montréal Neurological Institute, Montréal, QC, Canada, as a postdoctoral fellow. His research interests are statistical signal processing, localization of epileptic spikes using distributed source modeling, and validation of signal and image processing methods in the clinical context of epilepsy.

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 received the Ph.D. degree in 2002 from 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 with the Wellcome Department, Imaging Neuroscience, London, U.K. His current research focuses on the use of electromagnetic neuroimaging data to study brain functional integration.

Guillaume Marrelec received the M.S. in engineering and applied mathematics jointly from the École Centrale Paris, Paris, France, and the Universität Stuttgart, Stuttgart, Germany, in 1999. He then joined brain functional imaging research within the U494 INSERM, Paris, France, where he specialized in Bayesian analysis, from which he received the Ph.D. degree in 2003.

He is currently involved in post-doctoral research at the Université de Montréal, Montréal, QC, Canada. His research interests include investigation of functional brain interactivity through neuroimaging and brain plasticity induced by learning or surgery.

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, Montréal, in 1998. He is currently pursuing the Ph.D. degree in physics from the Centre de Recherches Mathématiques, Montréal.

From 1996 to 1998, he was with the Image-Guided Neursurgery Lab, Biomedical Engineering Departement, McGill University, 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. He did undergraduate studies at the University of Nancy, Nancy, France, and the University of Grenoble, Grenoble, France. He received the Ph.D. degree from the University of Pennsylvania, Philadelphia, with a thesis on on "Weak Interactions in Nuclei" in 1964.

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

Mélanie Pélégrini-Issac received the Ph.D. degree in physics from the Universif Paris 11, Orsay, France, in 1997.

In 2000, she became a Research Engineer with the French National Institute of Health and Medical Research (INSERM), Paris, France. From 2000 to 2004, she was with the Unité 483 InSERM, Paris, where she was mainly concerned with the processing of fMRI data in the context of multimodality (fMRI/EEG/MEG) experiments. Since 2005, she has been with the Functional Imaging Laboratory, Unité 678, INSERM. Her research interests include statistical image and signal processing for the detection of cerebral activation in fMRI and MEG and the analysis of brain functional connectivity in fMRI.

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 a Professor with the Electrical Engineering Department, Ecole Supérieure de Technologie, Montréal. 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.

Habib Benali (M'03) received the Ph.D. degree in signal processing and statistics in 1984 from Rennes University, Rennes, France, and the "Habilitation à Diriger des Recherches" (diploma to lead research project) degree in 1995 from Paris Dauphine University Paris, France.

From 1984 to 1987, he was an Assistant Professor with the Department of Computer Science and Statistics, IUT Vannes, Vannes, France. He joined the French National of Health and Medical Research (INSERM), Paris, in 1987. He is currently a researcher director with the Unit 678 "Laboratory of Functional Imaging," INSERM, where he leads the group "Parametric Imaging of the Cerebral Activity and Metabolism." Since 1999, he has been affiliated with the Mathematical Research Center, University of Montréal, Montréal, QC, Canada, as a regular researcher. 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 and statistical modeling of the neural networks involved in motor skill learning in humans.

Dr. Benali is a member of the OHBM, ASU, SFdS, and SFGBM.