

Dynamic Filtering of Static Dipoles in MagnetoEncephaloGraphy

Alberto Sorrentino^{*}, Adam M. Johansen[†],
John A. D. Aston[‡], Thomas E. Nichols[§],
and Wilfrid S. Kendall

*Department of Statistics,
University of Warwick
Coventry
CV4 7AL
UK*

e-mail: A.Sorrentino@warwick.ac.uk

e-mail: A.M.Johansen@warwick.ac.uk

e-mail: J.A.D.Aston@warwick.ac.uk

e-mail: T.E.Nichols@warwick.ac.uk

e-mail: W.S.Kendall@warwick.ac.uk

Abstract:

We consider the problem of estimating neural activity from measurements of the magnetic fields recorded by magnetoencephalography. We exploit the temporal structure of the problem and model the neural current as a collection of evolving current dipoles, which appear and disappear, but whose locations are constant throughout their lifetime. This fully reflects the physiological interpretation of the model.

In order to conduct inference under this proposed model, it was necessary to develop an algorithm based around state-of-the-art sequential Monte Carlo methods employing carefully designed importance distributions. Previous work employed a bootstrap filter and an artificial dynamic structure where dipoles performed a random walk in space, yielding non-physical artefacts in the reconstructions; such artefacts are not observed when using the proposed model. The algorithm is validated with simulated data, in which it provided an average localisation error which is approximately half that of the bootstrap filter. An application to complex real data derived from a somatosensory experiment is presented. Assessment of model fit via marginal likelihood showed a clear preference for the proposed model and the associated reconstructions show better localisation.

AMS 2000 subject classifications: Primary 60G35; secondary 62M05;92C55.

Keywords and phrases: Magnetoencephalography, Multi-object tracking, Particle filtering, Resample-Move.

^{*}A.S. was supported by a Marie Curie Intra European Fellowship within the 7th European Community Framework Programme.

[†]A.M.J. was partially supported by Engineering and Physical Sciences Research Council grant EP/I017984/1.

[‡]J.A.D.A. was partially supported by Engineering and Physical Sciences Research Council grant EP/H016856/1 as well as the EPSRC/HEFCE CRiSM grant.

[§]T.E.N. was partially supported by Medical Research Council grant G0900908.

1. Introduction

Magnetoencephalography (MEG) [16] is an imaging technique which uses a helmet-shaped array of superconducting sensors to measure, non-invasively, magnetic fields produced by underlying neural currents in a human brain. The sampling rate of MEG recordings is typically around 1,000 Hz, which allows observation of neural dynamics at the millisecond scale. Among other non-invasive neuroimaging tools, only electroencephalography (EEG) features a comparable temporal resolution. EEG can be considered complementary to MEG, due to its different sensitivity to source orientation and depth [6]. Note that estimation of the neural currents from the measured electric or magnetic fields is an ill-posed inverse problem [26]: specifically, there are infinitely many possible solutions, because there exist source configurations that do not produce any detectable field outside the head.

There are two well-established approaches to source modeling of MEG data. In the *distributed source* approach, the neural current is modeled as a continuous vector field inside the head, discretized on a large set of voxels; in this case, the inverse problem is linear, and standard regularization algorithms can be applied: commonly used methods include Minimum Norm Estimation [17, 18], a Tikhonov-regularized solution corresponding to the Bayesian maximum a posteriori (MAP) solution under a Gaussian prior; Minimum Current Estimation (MCE) [35], an L^1 minimization that corresponds to the MAP associated with an exponential prior in the Bayesian framework; and beamforming [36]. In this work we use the other approach, a *current dipole* model, where neural current is modeled as a small set of point sources, or current dipoles; each dipole represents the activity of a small patch of the brain cortex as an electric current concentrated at a single point. A current dipole is a six-dimensional object: three coordinates define the dipole location within the brain, a further three coordinates define the dipole orientation and strength (the *dipole moment*). The dipole model is a useful low-dimensional representation of brain activity: in typical MEG experiments aimed at studying the brain response to external stimuli [23, 27], the neural activity is modeled with a very small number of dipoles, whose locations are fixed but which have activity that evolves over time. However, estimation of dipole parameters is mathematically more challenging than estimation of the whole vector field, for at least two reasons: first, the number of dipoles is generally unknown, and must be estimated from the data; second, the measured signals depend non-linearly on the dipole location. For these two reasons, dipole estimation is still largely performed with simple non-linear optimization algorithms that have to be initialized and supervised by expert users.

Distributed source methods are more prevalent, and most of them estimate the source distribution independently at each time point. However, since the time interval between two subsequent recordings is so small — about one millisecond — the underlying brain activity does not much change between consecutive measurements. Spatio-temporal modeling can incorporate this prior knowledge by requiring that the solution satisfy some form of temporal continu-

ity. The availability of increasing computational resources has made it possible to explicitly account for the dynamic nature of the problem; [24, 34, 33] employ spatio-temporal regularisation.

Recently, MEG source estimation has been cast as a filtering problem within a state-space modeling framework. This approach has the further appeal that, in principle, it can be used on-line, producing sequential updating of the posterior distribution that incorporates the new data as they become available at a computational cost (per measurement update) which does not increase unboundedly over time. In [21, 22], a distributed source model was used and inference obtained with a high-dimensional Kalman filter. In [29, 2, 31, 25] a dipole model was used, and the posterior distribution of the multi-dipole configuration was approximated either with a bootstrap or with an approximation to a Rao-Blackwellized bootstrap particle filter; however the nature of the approximation to the Rao-Blackwellized filter was such that it yields underestimated uncertainty. However, in the interests of computational expediency, all of these studies employed an artificial dynamic model in which dipole locations were modeled as performing a random walk in the brain.

In this study we present a novel state space model for MEG data, based on current dipoles. Unlike most other work in this area, the proposed approach explicitly models the number of dipoles as a random variable, allowing new dipoles to become active and existing dipoles to stop producing a signal. In contrast to previous work on Bayesian filtering of multi-dipole models, we treat the location of a dipole source as fixed over the lifetime of the source. This is in accordance with the general neurophysiological interpretation of a dipole as arising from the coherent activity of neurons in a small patch of cortex. This is not a minor modification: it significantly influences the results, their interpretation and the computational techniques which are required in order to perform inference. The fact that dipole locations do not change over time raises a computational challenge: while it would seem natural to adopt a sequential Monte Carlo algorithm to approximate the posterior distribution for our state-space model, it is well known that these methods are not well suited to the direct estimation of static parameters although a number of algorithms have been developed to address this particular problem in recent years [19]. Standard particle filters are well-suited to the estimation of time-varying parameters in ergodic state space models, as they can exploit knowledge of the dynamics of the system itself to provide a good exploration of the parameter space. In order to perform inference for the proposed model effectively, we adopt a strategy based upon the Resample-Move algorithm [14]: we introduce a Markov Chain Monte Carlo move, formally targeting the whole posterior distribution in the path space, as a means to explore the state space while working with near-static parameters. We note that the appearance and disappearance of dipoles provides some level of ergodicity and ensures that there are no truly static parameters within the state vector; this also implies that algorithms appropriate for the estimation of true static parameters are not applicable in the current context. The proposed dynamic structure is exploited to allow us to implement MCMC moves on this space which mix adequately for the estimation task at hand without having

computational cost which increases unboundedly as more observations become available. In addition, we improve the basic importance sampling step with the introduction of a carefully designed importance distribution.

The remainder of this paper has the following structure: Section 2 provides a very brief summary of filtering in general and particle filtering in particular, Section 3 introduces the proposed models and associated algorithms and Sections 4 and 5 provides validation of these algorithms via a simulation study and an illustration of performance on real data. A brief discussion is provided in the final section.

2. Bayesian and Particle Filtering

Bayesian filtering is a general approach to Bayesian inference for Hidden Markov Models: one is interested in the sequence of posterior distributions $\{p(j_{0:t}|b_{1:t})\}_{t=1,\dots,T}$, and particularly the associated marginal distributions $\{p(j_t|b_{1:t})\}_{t=1,\dots,T}$, for the unobserved process $\{J_1, \dots, J_t, \dots\}$ given realizations of the measurements $\{B_1, \dots, B_t, \dots\}$, where j_t and b_t are instances of the corresponding random variables. If one assumes that

- the stochastic process $\{J_t\}$ is a first order Markov process with initial distribution $p(j_0)$ and homogeneous transition probabilities $p_t(j_{t+1}|j_t) = p(j_{t+1}|j_t)$; such that $p(j_{0:t}) = p(j_0) \prod_k p(j_{k+1}|j_k)$; in MEG, this corresponds to the model for evolution of current dipoles
- each observation B_t is statistically independent of the past observations given the current state j_t , and has conditional distribution $p_t(b_t|j_t)$, which it is convenient to treat as time homogeneous, $p_t(b_t|j_t) = p(b_t|j_t)$; in MEG, the observations are thus assumed to only depend on the current neural configuration.

then the posterior distribution at time t is given by

$$p(j_{0:t}|b_{1:t}) = \frac{p(j_{0:t}, b_{1:t})}{p(b_{1:t})} = \frac{p(j_0) \prod_{k=1}^t p(j_k|j_{k-1})p(b_k|j_k)}{p(b_{1:t})}, \quad (2.1)$$

and satisfies the recursion

$$p(j_{0:t}|b_{1:t}) = p(j_{0:t-1}|b_{1:t-1}) \frac{p(b_t|j_t)p(j_t|j_{t-1})}{p(b_t|b_{1:t-1})}. \quad (2.2)$$

Unfortunately, this recursion is only a formal solution as it is not possible to evaluate the denominator except in a few special cases, notably linear Gaussian and finite state space models. It is, therefore, necessary to resort to numerical approximations to perform inference in these systems. Particle filters (see [15, 4, 14] for original work of particular relevance to the present paper and [12] for a recent review) are a class of methods that combine importance sampling and resampling steps in a sequential framework, in order to obtain samples approximately distributed according to each of the filtering densities in turn.

These algorithms are especially well-suited to applications in which a time-series of measurements is available and interest is focussed on obtaining on-line updates to the information about the current state of the unobservable system.

Importance sampling is one basic element of particle filtering; it is a standard technique for approximating the expectation $\int f(x)p(x)dx$ of a reasonably well-behaved function f under a density $p(x)$ when i.i.d. samples from $p(x)$ are unavailable; the strategy consists in sampling $\{x^i\}_{i=1,\dots,N}$ from an *importance* density $q(x)$ such that $q(x)/p(x) > 0$, and then approximating

$$\int f(x)p(x)dx = \int f(x)\frac{p(x)}{q(x)}q(x)dx \simeq \sum_i f(x^i)w^i; \quad (2.3)$$

where the weights $w^i \propto p(x^i)/q(x^i)$ correct for the use of the importance density, and are normalized such that $\sum_i w^i = 1$. Conditions of boundedness of the weight function $p(x)/q(x)$, and finiteness of the variance of $f(X)$ for $X \sim p(\cdot)$, are together sufficient to ensure that this estimator obeys a central limit theorem with finite variance [13].

To apply importance sampling to the posterior density, one could sample $N_{\text{particles}}$ points (or *particles*) $\{j_{0:t}^1, \dots, j_{0:t}^{N_{\text{particles}}}\}$ from a proposal density $q(j_{0:t}|b_{1:t})$ and associate a weight $w_t^i \propto \frac{p(j_{0:t}^i|b_{1:t})}{q(j_{0:t}^i|b_{1:t})}$ to each particle. In the sequential framework, importance sampling can be simplified by a proper choice of the importance density: if the importance density is such that $q(j_{0:t}|b_{1:t}) = q(j_0) \prod_k q(j_k|j_{1:k-1}, b_k)$, then given the sample set at time $t-1$, $\{j_{0:t-1}^1, \dots, j_{0:t-1}^{N_{\text{particles}}}\}$, which is appropriately weighted to approximate $p(j_{0:t-1}|b_{1:t-1})$, one can draw $\{j_t^i\}$ from $q(j_t|j_{0:t-1}^i, b_t)$ and set $j_{0:t}^i = (j_{0:t-1}^i, j_t^i)$. Furthermore, thanks to the recursion (2.2) one can update the particle weight recursively

$$\begin{aligned} w_t^i &\propto \frac{p(j_{0:t}^i|b_{1:t})}{q(j_{0:t}^i|b_{1:t})} = \frac{p(j_{0:t-1}^i|b_{1:t-1})p(j_t^i|j_{0:t-1}^i)p(b_t|j_t^i)/p(b_t|b_{1:t-1})}{q(j_{0:t-1}^i)q(j_t^i|j_{0:t-1}^i, b_t)} \\ &\propto w_{t-1}^i \frac{p(b_t|j_t^i)p(j_t^i|j_{0:t-1}^i)}{q(j_t^i|j_{0:t-1}^i, b_t)}. \end{aligned} \quad (2.4)$$

Resampling is a technique, first suggested in the particle filtering literature by [15], which attempts to address the inevitable increase in the variance of the importance sampling estimator as the length of the time series being analysed increases. Resampling is a stochastic procedure by which particles with high weights are replicated and those with small weights eliminated, in such a way that the expected number of ‘‘offspring’’ of each particle is precisely the product of $N_{\text{particles}}$ and its weight before resampling. The unweighted population produced by resampling is then propagated forward by the recursive mechanism described above. The simplest approach to resampling, termed *multinomial* resampling as it corresponds to drawing a vector of offspring numbers from a multinomial distribution, is to draw N times from the empirical measure associated with the weighted sample. Although transparent, such an approach unnecessarily increases Monte Carlo variance and a number of lower-variance

alternatives have been developed — see [10] for a comparison between some of the most common approaches. In the experiments below the *systematic* resampling scheme of [4] was employed, that provides a good compromise between low variance and cheap computational cost.

The resampling step has two main drawbacks: first, it causes an immediate increase in Monte Carlo variance; second, resampling can only maintain or decrease the number of distinct sampled values for the past states, so that eventually, after a sufficiently large number of steps, all the particles will be descended from a single particle at time 1. However, the benefits of resampling outweigh its drawbacks. Intuitively, resampling concentrates the sample points in the high probability region of the state space, avoiding wasteful use of computational resources in updating of states with negligible likelihood. Also it can be proved that resampling helps to control the variance of the estimates over time [5, 9].

One issue in the application of particle filtering is the choice of the importance distribution. The simplest choice, leading to the so called *bootstrap* filter, is to use the prior distribution as importance distribution, setting $q(j_t|j_{0:t-1}, b_t) = p(j_t|j_{t-1})$. However, when the likelihood is informative, this choice will lead to an extremely high variance estimator. A good importance density should produce reasonably uniform importance weights, i.e. the variance of the importance weights should be small. The optimal importance distribution that minimizes the conditional variance of the importance weights whilst factorising appropriately, is given [11] by

$$q^*(j_t|b_t, j_{t-1}) = p(j_t|b_t, j_{t-1}) = \frac{p(j_t|j_{t-1})p(b_t|j_t)}{p(b_t|j_{t-1})}; \quad (2.5)$$

in practice, one should always try to approximate this distribution as well as is computationally feasible. Furthermore, a convenient way to monitor the variance of the importance weights is by looking at the effective sample size [4, ESS], defined as

$$\text{ESS} = \left(\sum_{i=1}^N (w_t^i)^2 \right)^{-1}. \quad (2.6)$$

The ESS ranges between 1 and $N_{\text{particles}}$ and can be thought of as an estimate (good only when the particle set is able to provide a good approximation of the posterior density of interest) of the number of independent samples from the posterior which would be expected to produce an estimator with the same variance as the importance sampling estimator which was actually used (resampling somewhat complicates the picture in the SMC setting). While a low ESS provides evidence that a weighted sample is dominated by a small number of particles with high weights, a high ESS provides no guarantee of a sample being good.

It is possible to obtain an estimate of the marginal likelihood (which, remarkably, is unbiased [9]) from the particle filter output,

$$p(b_{1:t}) = p(b_1) \prod_{n=1}^{t-1} p(b_{n+1}|b_{1:n}), \quad (2.7)$$

using the direct approximation of the *conditional likelihood*,

$$p(b_{n+1}|b_{1:n}) \approx \sum_i \tilde{w}_{n+1}^i \Rightarrow p(b_{1:t}) \approx \prod_{n=1}^t \sum_i \tilde{w}_n^i, \quad (2.8)$$

where \tilde{w}_{n+1}^i are the unnormalized weights at time $n + 1$.

3. Filtering of Static Dipoles

3.1. Statistical Model

In MEG, we are given a sequence of recordings of the magnetic field $\{b_t\}_{t=1,\dots,T}$, and wish to perform inference on the underlying neural current $\{j_t\}_{t=1,\dots,T}$ that has produced the measured fields.

In this study we model the neural current as a set of current dipoles $j_t = \{d_t^{(1)}, \dots, d_t^{(N_t)}\}$; here and in the rest of the paper, superscripted parenthetical indices label individual dipoles within a dipole set. Each current dipole $d_t^{(i)} = (r_t^{(i)}, q_t^{(i)})$ is characterized by a location $r_t^{(i)}$ within the brain volume, and a dipole moment $q_t^{(i)}$, representing direction and strength of the neural current at that location.

In order to perform Bayesian inference, we need to specify two distributions: the prior distribution for the neural current in time, and the likelihood function.

Prior Distributions

We specify the prior distribution for the spatio-temporal evolution of the neural current by providing the prior distribution at $t = 0$, and a homogeneous transition kernel $p(j_t|j_{t-1})$. We devise our prior model for the neural current path following two basic principles:

- at any time point t , the number of active dipoles N_t is expected to be small and the average dipole lifetime is around 30 milliseconds;
- dipole moments change (continuously) in time, to model increasing/diminishing activity of a given neural population, but dipole locations are fixed; for this reason, we term this the Static model.

In addition, for computational reasons we impose an upper bound on the number of simultaneously active dipoles N_t : in the experiments below we set this upper bound to $N_{\max} = 7$, as our informal prior expectation on the number of

dipoles is markedly less than 7, and this is born out by the data. Finally, for both computational and modeling reasons, dipole locations are required to belong to a finite set of pre-defined values $r^{(i)} \in R_{\text{grid}}$ with $R_{\text{grid}} = \{r_{\text{grid}}^k\}_{k=1}^{N_{\text{grid}}}$. It is customary in MEG research to use this kind of grid, where points are distributed along the cortical surface, the part of the brain where the neural currents flow. At the computational level, the use of these grids allows pre-calculation of the forward problem, i.e. of the magnetic field produced by unit dipoles, as described later. Automated routines for segmentation and reconstruction of the cortical surface from Magnetic Resonance images have been available for over ten years; in the experiments below we used Freesurfer [8] to obtain the tessellation of the cortical surface from the Magnetic Resonance data; we then used the MNE software package (<http://www.martinos.org/mne/>) to get a sub-sample of this tessellation with a spacing of 5 mm.

At time $t = 0$ the initial number of dipoles N_0 is assumed to follow a truncated Poisson distribution with rate parameter 1 and maximum N_{max} ; we then specify a uniform distribution over the grid points for the dipole locations, and a Gaussian distribution for the dipole moments, leading to the joint prior distribution:

$$p(j_0) = \sum_k P(N_0 = k) \prod_{n=1}^k U_{R_{\text{grid}}}(r_0^{(n)}) \mathcal{N}(q_0^{(n)}; 0, \sigma_q \mathbf{I}) \quad (3.1)$$

where $U_{R_{\text{grid}}}(\cdot)$ is the uniform distribution over the set R_{grid} and $\mathcal{N}(\cdot; \mu, \Sigma)$ is the Gaussian density of mean μ and covariance Σ .

The transition density accounts for the possibility of dipole birth and dipole death, as well as for the evolution of individual dipoles. It is assumed that only one birth or one death can happen at any time point. The transition density is composed of three summands as follows:

$$\begin{aligned} p(j_t | j_{t-1}) = & \\ & P_{\text{birth}} \times U_{R_{\text{grid}}}(r_t^{(N_t)}) \mathcal{N}(q_t^{(N_t)}; 0, \Delta) \times \prod_{n=1}^{N_{t-1}} \delta_{r_t^{(n)}, r_{t-1}^{(n)}} \mathcal{N}(q_t^{(n)}; q_{t-1}^{(n)}, \Delta) + \\ & + P_{\text{death}} \times \frac{1}{N_{t-1}} \sum_{j=1}^{N_{t-1}} \prod_{n=1}^{N_{t-1}-1} \delta_{r_t^{(n)}, r_{t-1}^{(a_{j,n})}} \mathcal{N}(q_t^{(n)}; q_{t-1}^{(a_{j,n})}, \Delta) + \\ & + (1 - P_{\text{birth}} - P_{\text{death}}) \times \prod_{n=1}^{N_{t-1}} \delta_{r_t^{(n)}, r_{t-1}^{(n)}} \mathcal{N}(q_t^{(n)}; q_{t-1}^{(n)}, \Delta). \quad (3.2) \end{aligned}$$

where $\delta_{\cdot, \cdot}$ is the Kronecker delta function. The first term in equation (3.2) accounts for the possibility that a new dipole appears, with probability P_{birth} ; the location of the new dipole, for convenience the N_t th dipole of the set, is uniformly distributed in the brain, while the dipole moment has a Gaussian distribution. All other dipoles evolve independently: dipole locations remain the

same as in the previous time point, while dipole moments perform a Gaussian random walk. The second summand in equation (3.2) accounts for the possibility that one of the existing dipoles disappears: one of the dipoles in the set at time $t - 1$ is excluded from the set at time t ; all existing dipoles have equal probability of disappearing; surviving dipoles evolve according to the same rules described earlier. The disappearance of a dipole entails a re-arrangement of the dipole labels; namely, the label of a dipole changes if a dipole with a smaller label disappears from the set; here $a_{j,n}$ is the label of the ancestor of the n th dipole after the death of the j th dipole, and is given by

$$a_{j,n} = \begin{cases} n & \text{if } n < j, \\ n + 1 & \text{if } n \geq j. \end{cases} \quad (3.3)$$

Finally, in the last term the number of dipoles in the set remains the same.

The parameters of these prior distributions were set to reflect our informal (and neurophysiologically-motivated) prior expectations for the number of dipoles and their temporal evolution. Birth and death probabilities were set, respectively, to $P_{\text{birth}} = 1/100$ and $P_{\text{death}} = (1 - (1 - 1/30)^{N_t})$, as the expected lifetime of a single dipole is about 30 time points, since simultaneous deaths are neglected; in addition, due to the presence of an upper bound to the number of simultaneous dipoles, the birth probability is zero when N_t is equal to this upper bound. In simulation experiments we found that, as a consequence of depending upon a finite sample approximation of the posterior, better estimation of the precise time of dipole disappearance could be obtained by increasing the death probability to a substantially larger value. However, such large death probabilities would render the average dipole lifetime unrealistically short, and thus we preferred to use a value that makes our prior as close as possible to the underlying physiological process. The transition density for the dipole moment is Gaussian, but the covariance matrix is not isotropic: the variance is ten times larger in the direction of the dipole moment itself, thus giving preference to changes in strength relative to changes in the orientation.

Likelihood

The magnetic field distribution is measured by an array of SQUID-based (Superconducting QUantum Interference Device) sensors, arranged around the subject's head producing, at time t , a column vector b_t containing one entry for each of the N_{sensors} sensors.

The relationship between the neural current parameters and the experimental data is contained in the *leadfield* or *gain* matrix G , whose size is $N_{\text{sensors}} \times 3N_{\text{grid}}$. Each column of the leadfield matrix contains the magnetic field produced by a unit dipole placed in one of the N_{grid} grid points and oriented along one of the three orthogonal directions; we denote by $G(r^k)$ the matrix of size $N_{\text{sensors}} \times 3$

that contains the fields produced by the three orthogonal dipoles at r^k . The measurement model is

$$b_t = \sum_{i=1}^{N_t} G(r_t^{(i)})q_t^{(i)} + \epsilon_t \quad (3.4)$$

where ϵ_t is an additive noise vector. Assuming that the ϵ_t are independent and Gaussian with covariance Σ_{noise} leads to the likelihood:

$$p(b_t|j_t) = \mathcal{N}\left(b_t; \sum_{i=1}^{N_t} G(r_t^{(i)})q_t^{(i)}, \Sigma_{\text{noise}}\right). \quad (3.5)$$

3.2. Computational Algorithm

The principal difficulty with the development of effective particle filtering algorithms for the static model described in the previous section is as follows: the dipole locations, except at the times of appearance and disappearance, behave as static parameters. Standard sequential Monte Carlo algorithms operating as filters/smoothers are not well suited to inference in the presence of static parameters and a variety of techniques have been devised to address that particular problem [19]. If one has a Hidden Markov Model with unknown static parameters, if one simply augments the state vector with the unknown static parameters and introduces an additional degenerate element in the transition kernel then one quickly suffers from degeneracy — in the case of the sequential importance resampling algorithm, for example, the algorithm is dependent upon sampling good values for the static parameters at the beginning of the sampling procedure as there is no mechanism for subsequently introducing any additional diversity. The problem is exacerbated by the fact that this initial sampling stage is extremely difficult in the MEG context, as the state space is large and the relationship between the likelihood and the underlying states is complex and nonlinear. Below, we develop strategies which exploit the fact that the dipole locations are not really static parameters, as they persist for only a random subset of the time sequence being analysed, together with more sophisticated sequential Monte Carlo techniques.

Here we propose a computational algorithm characterized by two main features. First, a mechanism that exploits the Resample-Move idea [14] in order to mitigate considerably the degeneracy effect produced by the static parameters; the idea is to introduce a Markov Chain Monte Carlo move at each iteration, targeting the whole posterior distribution. Second, a well designed importance distribution, in which birth locations and deaths are drawn from approximations to the optimal importance density (2.5).

Resample-Move

The Resample-Move algorithm is an approach for addressing degeneracy in sequential Monte Carlo algorithms. The idea is to use a Markov kernel $K(j_{0:t}^y|j_{0:t})$

of invariant distribution $p(j_{0:t}|b_{1:t})$ to provide diversity in the sample set. If $J_{0:t}$ is distributed according to $p(j_{0:t}|b_{1:t})$, and $J'_{0:t}|J_{0:t}$ is distributed according to $K(j'_{0:t}|j_{0:t})$, then $J'_{0:t}$ is still marginally distributed according to $p(j_{0:t}|b_{1:t})$ and more generally the distribution of $J'_{0:t}$ cannot differ more from $p(j_{0:t}|b_{1:t})$ in total variation than does the distribution of $J_{0:t}$.

In this study, the Markov kernel is constructed following the standard Metropolis-Hastings algorithm: proposed samples $j'_{0:t}$ are drawn from a proposal distribution $L(j'_{0:t}|j_{0:t})$ and then accepted with probability α , with

$$\alpha = \min \left(1, \frac{p(j'_{0:t}|b_{1:t})L(j_{0:t}|j'_{0:t})}{p(j_{0:t}|b_{1:t})L(j'_{0:t}|j_{0:t})} \right). \quad (3.6)$$

Since the purpose of this move is to introduce diversity for the dipole locations, we devised a simple proposal distribution that involves only a modification of the dipole locations, modifying one dipole at a time. Specifically, at every time point and for each particle we propose sequentially N_t^i moves, where N_t^i is the number of dipoles in the particle: for each dipole we choose one of its neighbours at random (one of the grid points within a fixed radius of 1 cm); the proposed particle $j'_{0:t}$ differs from the original particle $j_{0:t}$ only in the location of the proposed dipole; the acceptance probability α is dominated by the ratio of the likelihood of the original and the displaced particles, that can only differ for time points after the appearance of the dipole at (say) time $t = t_0$,

$$\alpha = \min \left(1, \frac{|S| \prod_{n=t_0}^t p(b_n|j'_n)}{|S'| \prod_{n=t_0}^t p(b_n|j_n)} \right) \quad (3.7)$$

where $|S|$ is the number of neighbours of the dipole in $j_{0:t}$ and $|S'|$ is the number of neighbours of the dipole in $j'_{0:t}$. Note that the $|S|/|S'|$ factor arises from the asymmetric proposal — although it may, initially, appear symmetric the restriction to an irregular discretisation grid induces asymmetry.

Importance Distribution

As mentioned in Section 2, having a good importance distribution is important in order to make a particle filter work in practice. At the same time, in our case the optimal importance distribution (2.5) is intractable — as it generally is for realistic models. Here we propose an importance density that features an acceptable computational cost but substantially improves the statistical efficiency in two crucial points. When birth is proposed, instead of drawing uniformly from the brain, the new dipole location is sampled according to a heuristic distribution based on the data; the exact nature of this distribution is unimportant as it is corrected for by importance sampling. Conditional on not proposing a birth, a death is proposed with approximately optimal probability. More precisely, we

propose to use the following importance distribution:

$$\begin{aligned}
q(j_t | j_{t-1}, b_t) = & \\
& Q_{\text{birth}} \times q(r_t^{(N_t)}, q_t^{(N_t)} | b_t, j_{t-1}) \times \prod_{n=1}^{N_{t-1}} \delta_{r_t^{(n)}, r_{t-1}^{(n)}} \mathcal{N}(q_t^{(n)}; q_{t-1}^{(n)}, \Delta) + \\
& Q_{\text{death}}(j_{t-1}, b_t) \times \sum_{j=1}^{N_{t-1}} P_{\text{dying}}(d^{(j)} | j_{t-1}, b_t) \times \prod_{n=1}^{N_{t-1}-1} \delta_{r_t^{(n)}, r_{t-1}^{(a_{j,n})}} \mathcal{N}(q_t^{(n)}; q_{t-1}^{(a_{j,n})}, \Delta) + \\
& (1 - Q_{\text{birth}} - Q_{\text{death}}(j_{t-1}, b_t)) \times \prod_{n=1}^{N_{t-1}} \delta_{r_t^{(n)}, r_{t-1}^{(n)}} \mathcal{N}(q_t^{(n)}; q_{t-1}^{(n)}, \Delta). \quad (3.8)
\end{aligned}$$

Birth is proposed at a fixed rate Q_{birth} , because evaluating the optimal birth probability would require the evaluation of intractable integrals and even obtaining a reasonable approximation would be computationally prohibitive; we use $Q_{\text{birth}} = 1/3$ in our algorithm. In the absence of a (near) optimal proposal, detecting new dipoles is the most challenging task faced by the algorithm, it is therefore appropriate to dedicate a substantial proportion of the computing resources to this task and so we use a value rather larger than $P_{\text{birth}} = 1/100$. When birth is proposed, the new dipole location is proposed from a heuristic proposal distribution $q(r_t^{(N_t)}, q_t^{(N_t)} | b_t, j_{t-1})$ computed from the data and obtained as follows: consider the linear inverse problem

$$b_t = GJ_t + \epsilon_t \quad (3.9)$$

where G is the whole leadfield matrix and J_t is a vector whose entries $J_t^k = J_t(r_k)$ represent the current strength at each point r_k of the grid; this inverse problem can be solved with a Tichonov regularization method,

$$\hat{J}_t = RG^T(GRG^T + \lambda_{\text{reg}}I)b_t \quad (3.10)$$

where R is a weighting matrix which mitigates the bias toward superficial sources [20] and λ_{reg} is the regularization parameter. The Tichonov solution provides a widespread estimate of neural activity within the brain; by normalizing the Tichonov solution, we obtain a spatial distribution which should be largest in the regions in which there is highest probability that a source is present:

$$q(r | b_t) = \hat{J}_t(r) / \sum_{r'} \hat{J}_t(r') \quad (3.11)$$

This density does not depend on the actual particle state, which is a significant computational advantage: it can be calculated just once per iteration rather than once per particle per iteration. However, there is a drawback in that its performance is expected to worsen as the number of dipoles increases (and much of the mass associated with q is located close to existing dipoles). We approximate the optimal death probability by assuming that the dipole parameters do

not change from $t - 1$ to t : death is proposed with probability

$$Q_{\text{death}}(j_{t-1}, b_t) = \frac{(1 - Q_{\text{birth}}) \times 1/N_{t-1} \sum_{k=1}^{N_{t-1}} p(b_t | j_{t-1}^{(-k)}) P_{\text{death}}}{1/N_{t-1} \sum_{k=1}^{N_{t-1}} p(b_t | j_{t-1}^{(-k)}) P_{\text{death}} + p(b_t | j_{t-1}) (1 - P_{\text{birth}} - P_{\text{death}})}, \quad (3.12)$$

where $j_{t-1}^{(-k)} = \{d_{t-1}^{(1)}, \dots, d_{t-1}^{(k-1)}, d_{t-1}^{(k+1)}, \dots, d_{t-1}^{(N_{t-1})}\}$ is the dipole set at time $t - 1$ without the k th dipole; if death is proposed the dipole to be killed is drawn according to:

$$P_{\text{dying}}(d^{(k)} | j_{t-1}, b_t) \propto p(b_t | j_{t-1}^{(-k)}). \quad (3.13)$$

Otherwise, with probability $1 - Q_{\text{birth}} - Q_{\text{death}}(j_{t-1}, b_t)$ the number of dipoles remains the same. The overall approach is outlined in Algorithm 1.

Algorithm 1 Outline of the Resample-Move algorithm

```

for  $i = 1, \dots, N$  do
  draw  $j_0^i$  from  $p(j_0)$ ;
end for
for  $t = 1, \dots, T$  do
  for  $i = 1, \dots, N$  do (importance sampling)
    draw  $j_t^i$  from  $q(j_t | \tilde{j}_{0:t-1}^i, b_t)$ ,
    set  $\tilde{j}_{0:t}^i = (j_{0,t-1}^i, j_t^i)$ 
    compute the unnormalized weights  $\tilde{w}_t^i = \frac{p(b_t | j_t^i) p(j_t^i | \tilde{j}_{0:t-1}^i)}{q(j_t^i | \tilde{j}_{0:t-1}^i, b_t)}$ 
  end for
  for  $i = 1, \dots, N$  do (normalize the weights)
     $w_t^i = \tilde{w}_t^i / W_t$ , with  $W_t = \sum_i \tilde{w}_t^i$ 
  end for
  for  $i = 1, \dots, N$  do (resample)
    draw  $\tilde{j}_{0:t}^i$  from  $\{\tilde{j}_{0:t}^i\}$ , with  $P(\tilde{j}_t = j_t^k) = w_t^k \forall k$ 
  end for
  for  $i = 1, \dots, N$  do (move)
    for  $k = 1, \dots, N_t^i$  do
      draw  $r^*$  from the neighbours of  $r_t^{(k),i}$ 
      accept the jump, replacing  $r_t^{(k),i}$  with  $r^*$  with probability given by Eq. (3.7)
    end for
  end for
end for

```

3.3. Connections with Previous Work

Application of particle filtering for estimation of current dipole parameters from MEG data has been described in [2, 3, 25, 31, 30]. A fundamental difference between our work and previous studies is that they all used a Random Walk model, that is, dipole locations were allowed to change in time, according to a random walk. In addition, in previous studies birth and death probabilities were

set to $P_{\text{birth}} = P_{\text{death}} = 1/3$. The computation was performed with a standard bootstrap particle filter, but a heuristic factor was used to penalize models with large number of dipoles: the particle weight, rather than being proportional to the likelihood alone, was in fact proportional to $\frac{1}{N_t^i} p(b_t | j_t^i)$, where N_t^i is the number of dipoles.

Our proposed strategy has a number of benefits: it is fully Bayesian and hence admits a clear interpretation and, most importantly, the statistical model is consistent with the biophysical understanding of the system being modeled. Experimentally, we found that models which incorporated artificial dynamics (random-walk type models) lead to significant artefacts in the reconstruction in which dipoles moved gradually from one side of the brain to the other in opposition to the interpretation of those dipoles as arising from fixed neural populations. Although the Resample-Move mechanism and Random Walk models may appear superficially similar they have very different interpretations and consequences: using the Random-Walk model is equivalent to performing inference under the assumption that the dipole location changes from one iteration to the next; using the Resample-Move algorithm leads to inference consistent with a model in which the dipoles do not move.

Below the Static model is compared with the Random Walk model described in previous studies; in our implementation of the Random Walk model, dipoles can jump between neighbouring points, with a transition probability proportional to $\exp(-d^2/2\sigma_d^2)$, where d is the distance between the two points and $\sigma_d = 0.5$ cm in the simulations below.

The use of improved importance distributions is also possible in the context of the Random Walk model and we have employed the importance distributions described in the following section, which improved its performance in comparison with the bootstrap approach employed previously.

Importance Distributions for the Random Walk Model

In the Random Walk Model, the transition probability distribution allows current dipole locations to jump within the set of neighbouring points; the use of bootstrap proposals to implement this, in conjunction with random change of dipole moment, will certainly result in a loss of sample points in the high probability region, even in the course of a single step. In our implementation of the Random Walk model we use the following approach in order to improve importance sampling efficiency: for each dipole contained in the particle — starting from the one most recently born — we first sample the dipole moment according to the dynamic model, and then calculate the probability that a dipole with the sampled dipole moment is at any of the neighbouring locations, given the data and the other dipoles. At each step the most recent values of the remaining parameters are always used, hence the k th dipole is sampled conditional on the current values of the dipoles with larger label, and on the previous values of the dipoles with smaller label.

The improved birth and death moves developed in the previous section can

also be employed without modification in the Random Walk model.

3.4. Computational Considerations

We end this section with a very brief examination of the computational cost of the proposed algorithms. In the MEG application, likelihood evaluation is responsible for the vast majority of computational effort. The only additional computation in the proposed method apart from these evaluations is the Tichonov inverse solution, which is quite fast, and is carried out once per iteration rather than once per particle per iteration. Because of this, the relative cost of the Tichonov inverse computation can be treated as negligible. Consequently we use the number of likelihood evaluations required by the proposed algorithms as a proxy for computational effort. We itemize this effort as follows:

- The total number of likelihood computation required by the bootstrap filter is TN , where T is the total number of time points and N the number of particles.
- The Resample-Move algorithm requires calculation of the likelihood for the whole past history of each dipole, hence requiring an additional $TN\bar{N}_{\text{dip}}T_{\text{life}}/2$, where T_{life} is the average lifetime of a dipole.
- The death proposal requires a number of additional likelihood evaluations of $TN\bar{N}_{\text{dip}}$, where \bar{N}_{dip} is the average number of dipoles.
- Finally, for the Random Walk model, the proposed conditional importance sampling requires the calculation of a number of likelihoods equal to the average number of neighbours; this is done at every time step, for each active dipole, hence brings an additional cost of $TN\bar{N}_{\text{dip}}N_{\text{neighbours}}$.

Relative computational costs depend on the data set, particularly in the case of the Resample-Move algorithm. Assuming an average number of dipoles of 1, an average number of neighbours of 25 and an average lifetime of 30 time points, the Resample-Move algorithm has a computational cost that is approximately 16 times higher than the bootstrap, while the conditional importance sampling is approximately 25 times more costly than the bootstrap, when run with the same number of particles.

4. Simulation Experiment

In this section, simulated data is used to validate and assess the performance of the proposed method.

We first describe the generation of the synthetic data. Then we propose a set of estimators for extracting relevant information based on the approximation to the posterior density provided by the particle filter. Finally, we present a number of measures for evaluating discrepancies between the estimated and the target dipole configuration.

4.1. Generation of Synthetic Data

100 different data sets were produced, according to the following protocol:

1. The synthetic magnetic field is generated from static dipolar sources through the assumed forward matrix (which is taken to be the same as is used in the model); dipoles used to produce the synthetic data set belong to the grid mentioned in Section 2, and will be referred to as *target* dipoles.
2. Each data set comprises 70 time points, and contains the activity of 5 sources overall; sources appear one at a time, at regular intervals of 5 time points.
3. Source locations are random, with uniform distribution in the brain, with the constraint that no two sources in the same data set can lie within 3 centimetres of one another. The strength of the signal produced by a source depends on the distance of the source from the sensors, so that randomness of source location implies that the signal strength — and eventually the detectability of a source — is itself random.
4. Source orientations are first drawn at random, with uniform distribution in the unit sphere, and then projected along the plane orthogonal to the radial direction at the dipole location; by “radial direction” we mean the direction of the segment connecting the dipole location to the center of the sphere best approximating the brain surface. Radial dipoles in a spherical conductor do not produce magnetic field outside of the conductor [26], so this projection avoids the creation of undetectable sources among the target dipoles.
5. The intensity of the dipole moment is kept fixed throughout the lifetime of each source, as shown in Figure 1 (although fixed intensity clearly does not mimick the actual behaviour of neural sources, we adopt this simple condition as it helps to provide a definite measure of the number of active sources at any time).
6. Noise is additive, zero-mean and Gaussian.

4.2. Point Estimates for the Multi-Dipole Configuration

The posterior distribution of a point process is a multi-dimensional object that is not easy to investigate, and is hard to represent faithfully by point estimates, a problem which is well known in the multi-object tracking literature (see, for example, [37] for another setting in which a very similar problem arises). At the same time, often in practical applications one is actually interested in point estimates; here, we are particularly interested in evaluating whether the particle filters provide good estimates of the dipole configuration that produced the synthetic data. We therefore propose a set of quantities that can be used for this purpose, bearing in mind that they are only low-dimensional projections of the actual output of the particle filter:

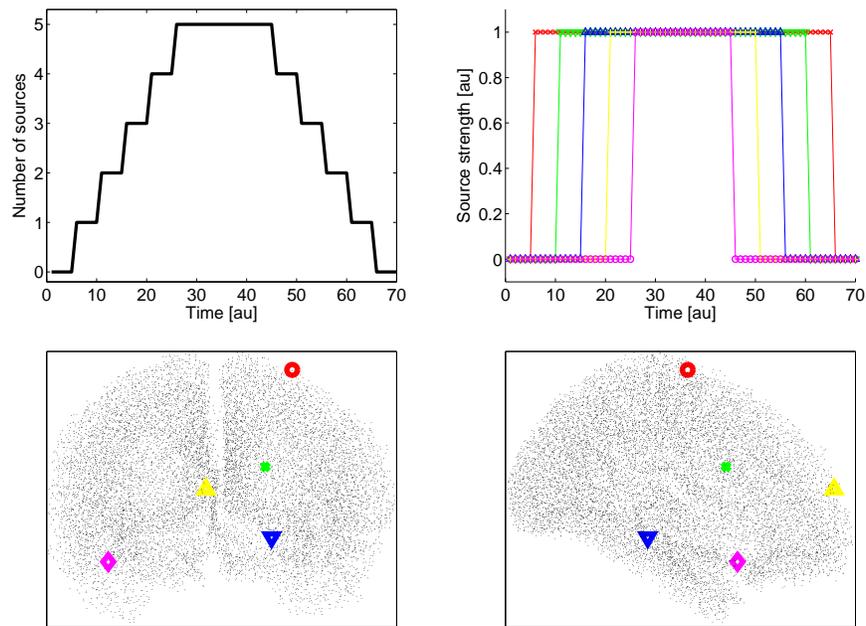


FIG 1. A sample dipole configuration generating one of the 100 data sets: in top left panel, the number of active sources as a function of time; in the top right panel, individual source waveforms; in the lower panels, the source locations and the grid points, randomly drawn from the uniform distribution in the brain.

- The number of active sources can be represented *via* the marginal distribution for the number of dipoles, which can be computed from the approximation to the posterior density as

$$P(N_t = k|b_{1:t}) = \sum_i w_t^i \delta_{k, N_t^i}. \quad (4.1)$$

- The location of the active sources can be represented *via* the intensity measure of the corresponding point process. This provides information about the dipole location which is highly suited to visual inspection; from the particle filter we get the following approximation to the intensity measure

$$p(r_t|b_{1:t}) \simeq \sum_i w_t^i \sum_{k=1}^{N_t^i} \delta_{r_t, r_t^{i,(k)}}. \quad (4.2)$$

In our implementation this approximation is defined for all the locations on the grid, but not continuously over the entire volume.

- The direction and the intensity of the estimated dipoles, i.e. the vectorial mark of the point process: one way to provide such information is to calculate the average dipole moment at each location, conditional on having a dipole at that location:

$$E[q_t|r] = \sum_i w_t^i \sum_{k=1}^{N_t^i} q_t^{i,(k)} \delta_{r, r_t^{i,(k)}}. \quad (4.3)$$

We use the following procedure to obtain a “representative set” of dipoles from the approximated posterior distribution:

1. estimate the number of dipoles in the set by taking the mode of the posterior distribution (4.1);
2. find the N highest *peaks* of the intensity measure (4.2): a *peak* is a grid point where the intensity measure is higher than that of its neighbours; we take these peaks as point estimates of the dipole locations;
3. for each estimated dipole location, the estimated dipole moment will be the average dipole moment at that location, as in (4.3).

As an alternative to the optimization in step (2), we also tried a clustering approach based on a Gaussian mixture model augmented with a uniform component, devised to model possible outliers; in the simulations below, the two approaches produced essentially the same results (not shown). While these measures are only low dimensional projections and cannot replace the rich information contained in the posterior distribution, we feel they capture the most important features relevant to the neuroscientist.

4.3. Discrepancy Measures

Once this typical set has been estimated, the discrepancy between the estimated dipole set $\hat{j}_t = \{\hat{d}_t^{(1)}, \dots, \hat{d}_t^{(\hat{N}_t)}\}$ and the target dipole set $j_t = \{d_t^{(1)}, \dots, d_t^{(N_t)}\}$

can be computed. However, measuring the distance between two point sets is a non trivial task even in the simple case when the two sets contain the same number of points, and becomes even more complicated when the two sets contain different number of points. Furthermore, in the application under study the points also have marks, or dipole moments, which should be taken into account. In the following, we list several useful measures of discrepancy between the target and the estimated dipole configurations.

- Average distance from closest target (ADCT): at first we may be interested in answering this question: how far, on average, is the estimated dipole from any of the target dipoles? To answer this question, we can calculate the ADCT, defined as

$$ADCT(t) = \frac{1}{\widehat{N}_t} \sum_{k=1}^{\widehat{N}_t} \min_j |\widehat{r}_t^{(k)} - r_t^{(j)}|, \quad (4.4)$$

where $|\cdot|$ denotes the Euclidean norm.

- Symmetrized distance (SD): we may also want to incorporate in the distance measure the presence of undetected sources; to do this, we calculate a symmetrized version of the ADCT

$$SD(t) = \frac{1}{\widehat{N}_t} \sum_{k=1}^{\widehat{N}_t} \min_j |\widehat{r}_t^{(k)} - r_t^{(j)}| + \frac{1}{N_t} \sum_{j=1}^{N_t} \min_k |\widehat{r}_t^{(k)} - r_t^{(j)}|. \quad (4.5)$$

- Optimal SubPattern Assignment metric (OSPA): if two estimated dipoles are both close to the same target dipole, neither ADCT nor SD will notice it; the OSPA metric [28] overcomes this limitation by forcing a one-to-one correspondence between the estimated and the target dipoles; the OSPA metric is defined as

$$OSPA(t) = \begin{cases} \min_{\pi \in \Pi_{\widehat{N}_t, N_t}} \frac{1}{\widehat{N}_t} \sum_{k=1}^{\widehat{N}_t} |\widehat{r}_t^{(k)} - r_t^{(\pi(k))}| & \text{if } \widehat{N}_t \leq N_t, \\ \min_{\pi \in \Pi_{N_t, \widehat{N}_t}} \frac{1}{N_t} \sum_{k=1}^{N_t} |\widehat{r}_t^{(\pi(k))} - r_t^{(k)}| & \text{if } \widehat{N}_t > N_t, \end{cases} \quad (4.6)$$

where $\Pi_{k,l}$ is the set of all permutations of k elements drawn from l elements; note that, this metric only calculates the discrepancy between the dipoles in the smaller set and the subset of dipoles in the larger set that has the smaller discrepancy with those in the smaller set.

- Widespread measure (WM): finally, we want to combine discrepancies in the source location with discrepancies in the dipole moment; the following measure does this by replacing each dipole (both in the target dipole set and in the estimated dipole set) with a Gaussian-like function in the brain, centered at the dipole location, with fixed variance, and height proportional to the dipole moment; the difference between the two spatial distributions is then integrated in the whole brain:

$$WM(t) = \int \left| \left[\sum_{k=1}^{\widehat{N}_t} |\widehat{q}_t^{(k)}| \mathcal{N}(r; \widehat{r}_t^{(k)}, \sigma) - \sum_{k=1}^{N_t} |q_t^{(k)}| \mathcal{N}(r; r_t^{(k)}, \sigma) \right] \right| dr, \quad (4.7)$$

where the integral must in practice be approximated by numerical methods.

4.4. Results

We ran the Resample-Move particle filter on the 100 synthetic data sets described at the beginning of this section. We also ran a bootstrap filter on the same data to evaluate its ability to sample the Static model's posterior. In addition, we ran both a standard bootstrap and an improved filter, as described in the previous section, implementing the Random Walk model.

All filters were run with 10,000 particles; in addition, in order to compare the performances for approximately equal computational cost, we ran both the Resample-Move filter and the improved filter for the Random Walk model with 500 particles. All filters were run with the same parameter values: the standard deviation of the Gaussian prior for the dipole moment was set to $\sigma_q = 1$ nAm; the noise covariance matrix was diagonal, with the same value σ_{noise}^2 for each channel and estimated from the first 5 time points: this was done in analogy with typical MEG experiments with external stimuli, where a pre-stimulus interval is typically used to estimate the noise variance.

We computed the discrepancy measures proposed in Section 4. The results are shown in Figure 2; the widespread measure provided results that are very similar as those of the OSPA metric, hence for brevity it is not shown here. In Figure 3 we show the estimated number of sources, the effective sample size, as given by Equation (2.6), and the conditional likelihood as a function of time.

All the discrepancy measures indicate that the Resample-Move particle filter provides a substantial improvement over the bootstrap filter. The use of three different measures, in conjunction with the observation of the estimated number of sources in Figure 3, gives more insights about the qualitative nature of the improvements. First of all, the ADCT indicates that the dipoles estimated with the Resample-Move are on average much closer to the target sources; in addition, there is a rather small difference between the results obtained running the Resample-Move with 10,000 particles and with 500 particles. The average localization error is about 7 mm with the new model in contrast to the bootstrap particle filter which achieves an average localization error of 1.4 cm. The SD provides a slightly different result: there is more difference here between 500 and 10,000 particles; this is due to the fact that using a higher number of particles allows the algorithm to explore the state space better; in addition, the relatively small difference with the bootstrap filter here is due to the fact that the Resample-Move algorithm tends to slightly underestimate the number of sources, which is penalized by the SD. Finally, in terms of the OSPA measure, the Resample-Move provides a rather large improvement over the bootstrap: that the difference is so large is due to the fact that the bootstrap filter tends to overestimate the number of dipoles, in the proximity of a target dipole (as it is unable to update dipole locations and the likelihood strongly supports the presence of some additional signal source). This does not have a big impact

on the ADCT, but is highly penalized by the OSPA. Observation of the ESS and of the conditional likelihood in Figure 3 strengthen the previous results. The Resample-Move algorithm maintains a higher ESS throughout the whole temporal window, in which the number of dipoles increases from zero to five and then returns to zero. The conditional likelihood further adds to the general evidence that the Resample Move algorithm has better explored the state space whilst the bootstrap algorithm has missed a substantial part of the mass of the posterior.

We finally compare the performance of the Static model and the Random Walk model. The discrepancy measures indicate that the two models perform rather similarly in terms of average localization accuracy; this has to be regarded as a positive fact, since the localization accuracy of the Random Walk model was already recognized as being satisfactory [31], and the Static model is in fact a model for which inference is harder; however, in most experimental conditions, the Random Walk model is not believed to be as physiologically plausible as the Static model. Notably, in this synthetic experiment in which we know that the dipoles are actually static, we observe that the Static model leads to higher conditional likelihood than the random walk model. As in the context of Bayesian model selection, this implies that the data supports the Static model in preference to the Random Walk model. However, some caution should be exercised in interpreting these results: in both cases the true static parameters (noise variance, scale of random walk) have been fixed and so the time integral of these conditional likelihoods can only be interpreted as a marginal *profile* likelihood (it is not currently feasible to consider the penalisation of more complex models afforded by a fully Bayesian method in which the unknown parameters were marginalized out).

5. Example of Application to Real Data

We applied the Resample-Move algorithm to real MEG recordings which were obtained during stimulation of a large nerve in the arm. This choice is motivated by the fact that the neural response to this type of somatosensory stimulation is relatively simple and rather well understood [23], and therefore allows a meaningful assessment of performance.

5.1. Experimental Data

We used data from a Somatosensory Evoked Fields (SEFs) mapping experiment. The recordings were performed after informed consent was obtained, and had prior approval by the local ethics committee. Data were acquired with a 306-channel MEG device (Elekta Neuromag Oy, Helsinki, Finland) comprising 204 planar gradiometers and 102 magnetometers in a helmet-shaped array. The left median nerve at wrist was electrically stimulated at the motor threshold with an interstimulus interval randomly varying between 7.0 and 9.0 s. The MEG signals were filtered to 0.1-200 Hz and sampled at 1000 Hz. Electrooculogram (EOG)

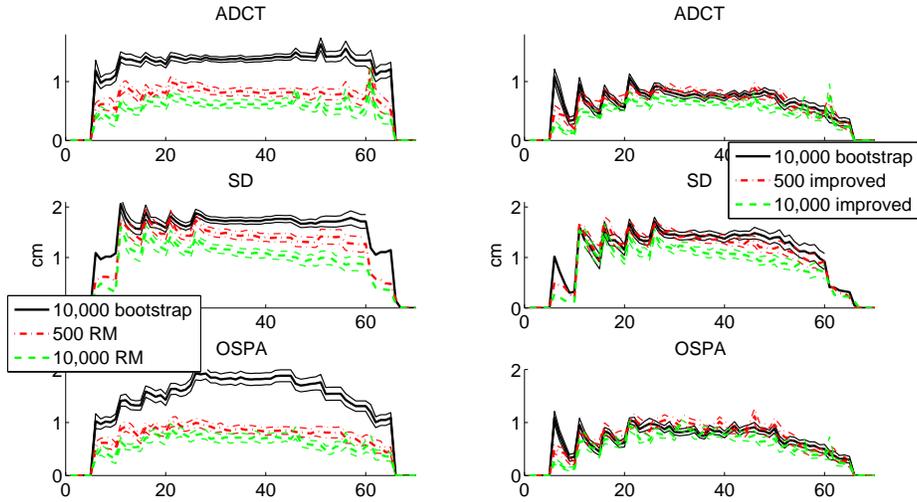


FIG 2. Discrepancy measures for the Static model (left column) and the Random Walk model (right column).

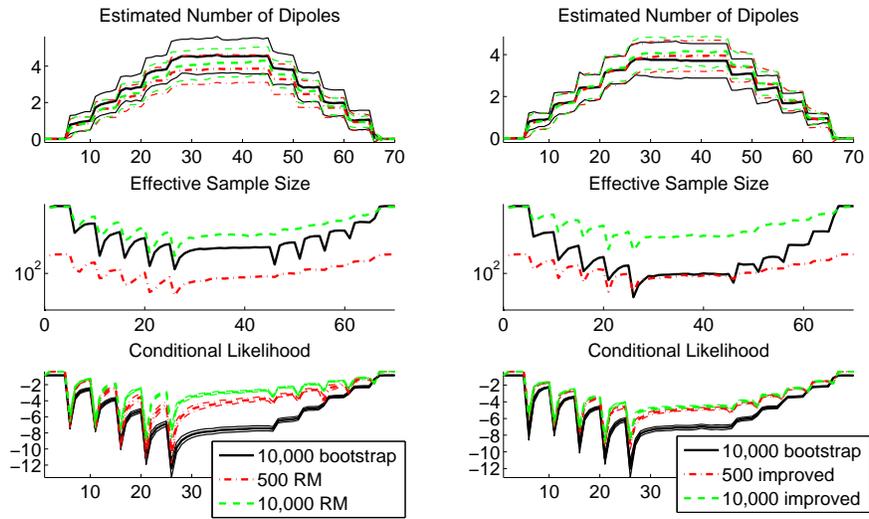


FIG 3. Estimated number of sources, effective sample size and conditional likelihood for the Static model (left column) and the Random Walk model (right column).

was used to monitor eye movements that might produce artifacts in the MEG recordings; trials with EOG or MEG exceeding 150 mV or 3 pT/cm, respectively, were excluded and 84 clean trials were averaged. To reduce external interference, signal space separation method [32] was applied to the average. A 3D digitizer and four head position indicator coils were employed to determine the position of the subjects head within the MEG helmet with respect to anatomical MRIs obtained with a 3-Tesla MRI device (General Electric, Milwaukee, USA).

5.2. Results

The Resample Move particle filter implementing the Static model was applied to the MEG recordings; for the sake of comparison, we also applied the particle filter based on the Random Walk model and the conditional sampling; finally, we compared these results to the source estimates computed with the dSPM algorithm [7] contained in the MNE software package, consisting in a noise-normalized linear inverse estimate. Both particle filters were run with the same parameter values: the standard deviation of the Gaussian prior for the dipole moment was set to $\sigma_q = 50$ nAm, which is a physiologically plausible value; the noise covariance matrix was diagonal, with values σ_{noise}^2 estimated from the prestimulus interval.

Figure 4 illustrates the localization provided by the three methods: the colour-coded map represents the marginal probability of there being a source at each location (see Equation (4.2)) for the two particle filters, and the noise-normalized source strength in the case of dSPM. We show snapshots at three time points. The very first response in the primary somatosensory area SI, at 25 ms, is localized by both the Static and the Random Walk particle filters in a very similar way; interestingly, the dSPM estimate at this time point is rather widespread and indicates activity in slightly more frontal areas. At time $t = 50$ ms after stimulus, however, we begin to see the difference between the Static model and the Random Walk model: as the SI area has been active for the past 25 ms, the posterior map of the Static model is much more concentrated now, having accumulated information on the source location; the Random Walk model indicates activity in the same area but provides a more blurred image; the estimate of dSPM is here closer to the probability maps provided by the two filters. At time $t = 85$ ms, finally, we observe more activation in SI, and the additional activity in the ipsilateral and contralateral SII: observing the posterior maps provided by the Static model we observe, as in [23], that the contralateral SII activation is more frontal than the ipsilateral SII activation; the Random Walk model provides, again, more blurred images; finally, dSPM indicates activity in a significantly larger area.

In Figure 6 we show the cumulative marginal likelihood (2.7) and the effective sample size. While the effective sample size produces rather similar results for the two models, the marginal likelihood indicates that after approximately $t = 60$ ms the Static model provides higher likelihood than the Random Walk model. Importantly, the cumulative likelihood at time t is the likelihood of the

whole time series *up to* time t . The fact that the difference between the two models tends to increase with time indicates that, as more data are gathered, the Static model is increasingly preferentially supported by the measurements. The ratio of the two likelihoods at the terminal time point indicates that the whole time series is several orders of magnitude more likely under the Static model than under the Random Walk model, thus providing a rather strong confirmation that the Static model is a much better representation of the underlying neurophysiological processes than the Random Walk model.

An additional point that deserves to be highlighted here is that not only are the probability maps provided by the Static model sparser than those provided by the Random Walk, but also (as one might reasonably expect) they show less temporal variation. To illustrate this point, in Figure 5 we show two maps that have been obtained by integrating over time the dynamic probability maps provided by the Static and the Random Walk filters: while the Static model has high probability in few small areas and negligible probability elsewhere, the Random Walk model provides a flatter image, with rather homogeneous probability values in a larger area: a consequence of the fact that the Random Walk model attaches a large part of its posterior probability mass to dipoles which move around the brain.

As we run several independent realizations of the filters with the same parameter values and 100,000 particles, we observed that for $t > 75$ ms not all the runs provide exactly the same estimates. While in the majority of the runs the mode of the posterior distribution consistently presents the source configuration depicted in Figure 4, in approximately 10% of the runs the ipsi-lateral and contra-lateral SII sources are replaced with a pair of sources in between the two hemispheres, one at the top in the motor area and one rather deep and central; the two SII areas are still represented in the posterior distribution, but with slightly lower mass, and may not appear in low-dimensional summaries of the posterior. As noted previously, accurately summarising high-dimensional multi-object posteriors is known to be a rather difficult problem. Finally, we note that if too small a sample size was employed then we found that the quality of the approximation of the posterior could deteriorate to the point that the posterior did not contain mass in a neighbourhood of the configuration shown in Figure 4. Naturally, sample-based approximations to high-dimensional distributions fail to accurately capture the important features of those distributions if the sample is too small. In the case of a sample of size 10,000 we observed this failure mode in less than 10% of the runs. In practice, such a phenomenon should be easily detected by replicating the estimation procedure a number of times.

6. Discussion

In this study we have presented a new state space model for dynamic inference on current dipole parameters from MEG data with particle filtering. The model has been devised to reflect the common neurophysiological interpretation of a current dipole as the activity of a small patch of cortex: the number of dipoles

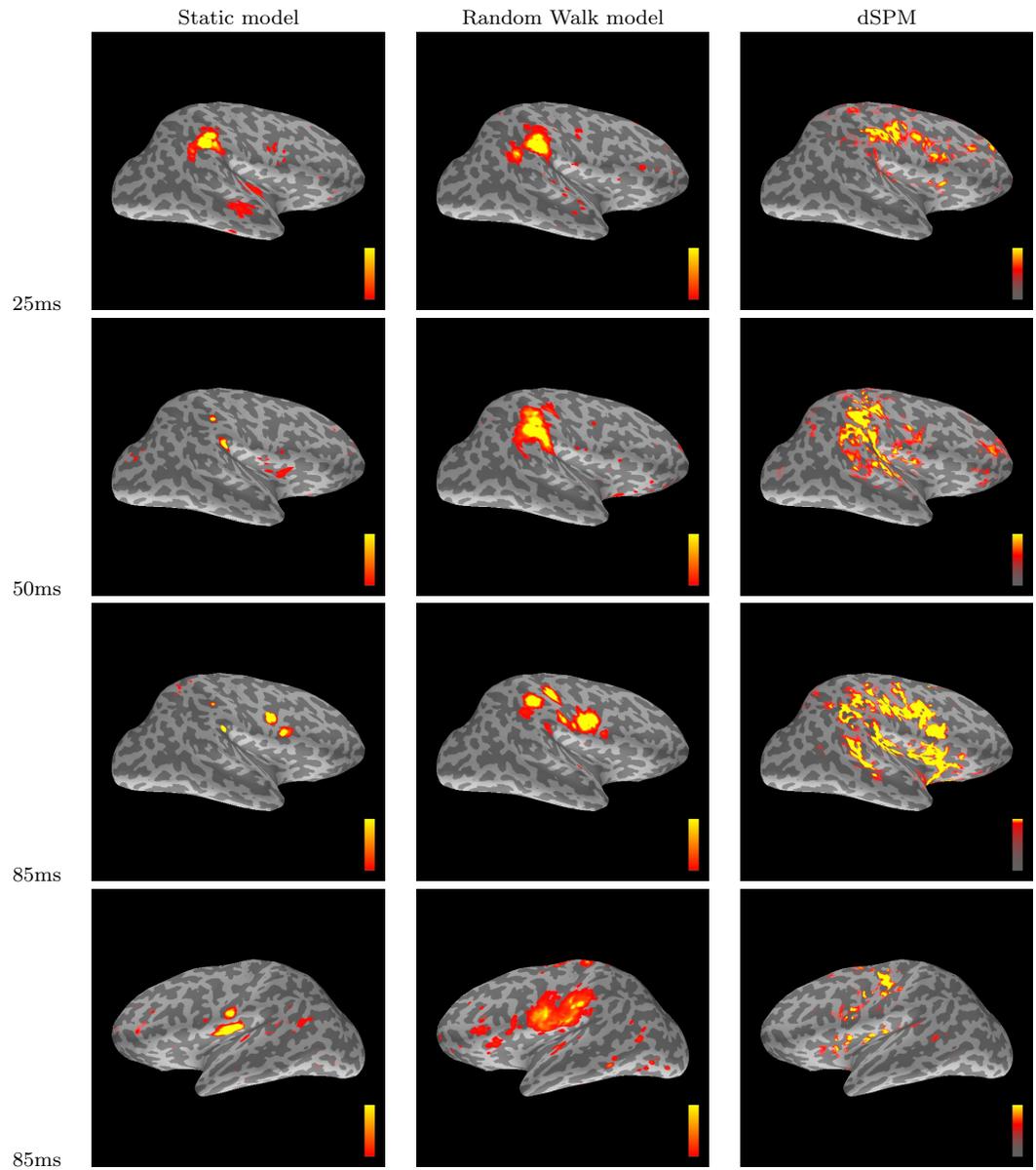


FIG 4. SEF data. Static Model (left), Random Walk Model (centre) and dSPM estimate (right).

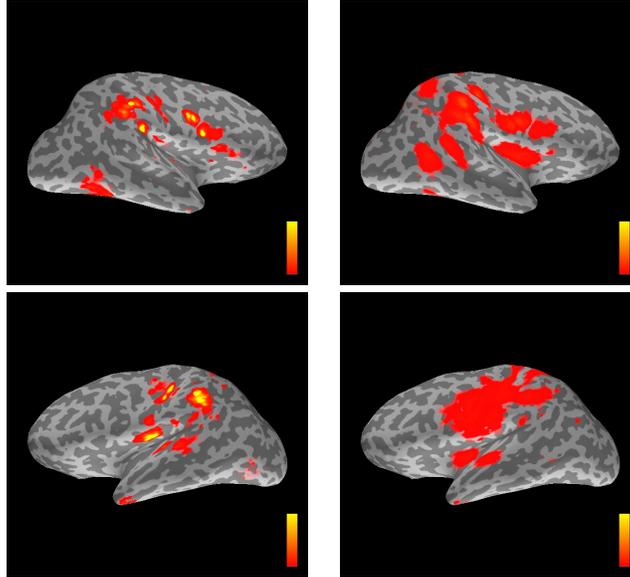


FIG 5. Time-integrated probability maps: the Static Model (left column) exhibits less temporal variation than the Random Walk Model (right column).

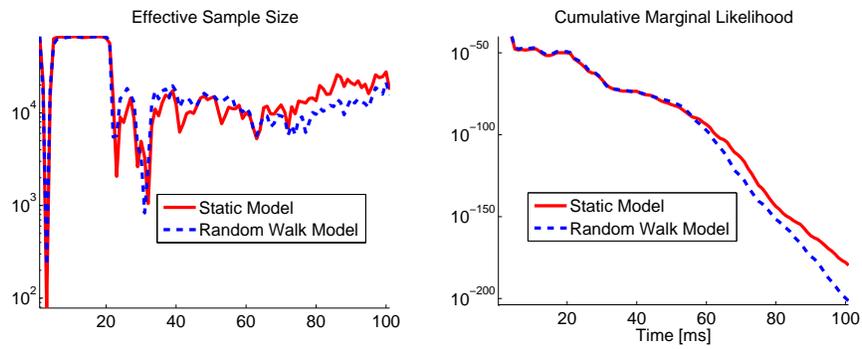


FIG 6. The Effective Sample Size and the marginal likelihood, as obtained with the Static and the Random Walk model with the SEF data.

is time-varying, as dipoles can appear and disappear, and dipole locations are fixed during the dipole life time. Standard sequential Monte Carlo algorithms are not well suited to “filtering” of static parameters; for the same reasons simple sequential importance resampling is not able to efficiently approximate the posterior distributions associated with these near-static objects. We have developed a Resample-Move type algorithm with carefully-designed proposal distributions which are appropriate for inference in this type of model.

We have used synthetic data to show that the average localization error provided by the Resample Move algorithm is close to 5 mm, i.e. the average grid spacing, even when the data are produced by five simultaneous dipoles. In addition, the effective sample size remains high even when the filter explores the high-dimensional parameter space of five dipoles, consistent with a good approximation of the posterior distribution. Although the quality of the approximation naturally depends on the sample size, we demonstrated good results can be obtained with a realistic computational cost. Finally, comparison of the conditional likelihood of our Static dipole model with that of a Random Walk model indicates that the proposed method is actually capable of providing a better explanation of the data.

Application of the proposed method to an experimental data set has produced similar results: the effective sample size and the conditional likelihood remain high throughout the whole time series; the posterior probability maps are well in accordance with what is understood to be the usual brain response to median nerve stimulation. The Static model leads to physiologically-interpretable output which is consistent with the biomedical understanding of the dipole model. We did not observe the type of artefacts found with the Random Walk model in which dipoles slowly moved across the brain surface when using this model.

Future research will concentrate on increasing the number of samples and decreasing the computational time. Implementation on GPUs should provide a viable way to reduce the computational time exploiting massive parallelization. Improving the efficiency of the MCMC step is also of interest. Other possible interesting research directions include the use of smoothing [1] techniques and estimation of the static parameters [19], here fixed a priori using approaches prevalent in the literature, as well as generalization of the source model.

Acknowledgments

We gratefully acknowledge the help of Dr. Lauri Parkkonen and Dr. Annalisa Pascarella, who together with AS undertook collection, post-processing and analysis of the experimental data.

References

- [1] M. Briers, A. Doucet, and S. Maskell, *Smoothing algorithms for state space models*, Annals of the Institute of Statistical Mathematics **62** (2010), no. 1, 61–89.

- [2] C. Campi, A. Pascarella, A. Sorrentino, and M. Piana, *A Rao-Blackwellized particle filter for magnetoencephalography*, *Inverse Problems* **24** (2008), 025023.
- [3] ———, *Highly automated dipole estimation*, *Computational Intelligence and Neuroscience* **2011** (2011), 982185.
- [4] J. Carpenter, P. Clifford, and P. Fearnhead, *An improved particle filter for non-linear problems*, *IEE Proceedings Radar, Sonar & Navigation* **146** (1999), 2–7.
- [5] N. Chopin, *Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference*, *The Annals of Statistics* **32** (2004), 2385–2411.
- [6] D. Cohen and B.N. Cuffin, *Demonstration of useful differences between magnetoencephalogram and electroencephalogram*, *Electroencephalography and Clinical Neurophysiology* **56** (1983), 38–51.
- [7] A. Dale, A.K. Liu, B.R. Fischl, R.L. Buckner, J.W. Beldaveau, J.D. Lewine, and E. Halgren, *Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity*, *Neuron* **26** (2000), 55–67.
- [8] A.M. Dale, B. Fischl, and M.I. Sereno, *Cortical Surface Based Analysis i. segmentation and Surface Reconstruction*, *NeuroImage* **9** (1999), 179–194.
- [9] P. Del Moral, *Feynman-Kac formulæ*, Springer, 2004.
- [10] R. Douc and O. Cappé, *Comparison of resampling schemes for particle filters*, *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis*, vol. I, IEEE, 2005, pp. 64–69.
- [11] A. Doucet, S. Godsill, and C. Andrieu, *On sequential Monte Carlo sampling methods for Bayesian filtering*, *Statistics and Computing* **10** (2000), 197–208.
- [12] A. Doucet and A.M. Johansen, *A tutorial on particle filtering and smoothing: Fifteen years later*, *The Oxford Handbook of Nonlinear Filtering*, Oxford University Press, 2011.
- [13] J. Geweke, *Bayesian inference in econometrics models using Monte Carlo integration*, *Econometric* **57** (1989), 1317–1339.
- [14] W.R. Gilks and C. Berzuini, *Following a moving target - Monte Carlo inference for dynamic Bayesian models*, *Journal of the Royal Statistical Society. Series B* **63** (2001), 127–146.
- [15] N.J. Gordon, D.J. Salmond, and A.F.M. Smith, *Novel approach to nonlinear/non-Gaussian Bayesian estimation*, *IEE Proceedings F Radar and Signal Processing* **140** (1993), 107–113.
- [16] M. Hämmäläinen, R. Hari, J. Knuutila, and O.V. Lounasmaa, *Magnetoencephalography: theory, instrumentation and applications to non-invasive studies of the working human brain*, *Reviews of Modern Physics* **65** (1993), 413–498.
- [17] M. Hämmäläinen and R.J. Ilmoniemi, *Interpreting measured magnetic fields of the brain: estimates of current distributions*, Tech. report, Helsinki University of Technology, 1984.
- [18] ———, *Interpreting magnetic fields of the brain: minimum norm estimates*,

- Medical & Biological Engineering & Computing **32** (1994), 35–42.
- [19] N. Kantas, A. Doucet, S.S. Singh, and J.M. Maciejowski, *An overview of sequential Monte Carlo methods for parameter estimation in general state-space models*, Proceedings IFAC System Identification (SysId) Meeting, 2009.
 - [20] F.H. Lin, T. Witzel, S.P. Ahlfors, S.M. Stufflebeam, J.V. Belliveau, and M.S. Hamalainen, *Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates*, NeuroImage **31** (2006), 160–171.
 - [21] C.J. Long, P.L. Purdon, S. Temeranca, N.U. Desai, M. Hämäläinen, and E.N. Brown, *Large scale Kalman filtering solutions to the electrophysiological source localization problem — a MEG case study*, Proceedings of the 28th IEEE EMBS Annual International Conference, vol. 5, 2006, pp. 4532–4535.
 - [22] ———, *State-space solutions to the dynamic magnetoencephalography inverse problem using high performance computing*, The Annals of Applied Statistics **5** (2011), 1207–1228.
 - [23] F. Mauguiere, I. Merlet, N. Forss, S. Vanni, V. Jousmaki, P. Adeleine, and R. Hari, *Activation of a distributed somatosensory cortical network in the human brain. A dipole modelling study of magnetic fields evoked by median nerve stimulation. part I: location and activation timing of SEF sources*, Electroencephalography and Clinical Neurophysiology **104** (1997), 281–289.
 - [24] W. Ou, M.S. Hämäläinen, and P. Golland, *A distributed spatio-temporal EEG/MEG inverse solver*, NeuroImage **44** (2009), 932–946.
 - [25] A. Pascarella, A. Sorrentino, C. Campi, and M. Piana, *Particle filtering, beamforming and rap-music in the analysis of magnetoencephalography time series: a comparison of algorithms*, Inverse Problems and Imaging **4** (2010), 169–190.
 - [26] J. Sarvas, *Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem*, Physics in Medicine and Biology **32** (1987), 11–22.
 - [27] M. Scherg and D. von Cramon, *Evoked dipole source potentials of the human auditory cortex*, Electroencephalography and Clinical Neurophysiology **65** (1986), 344–360.
 - [28] D. Schuhmacher, B.T. Vo, and B.N. Vo, *A consistent metric for performance evaluation of multi-object filters*, IEEE Transactions on Signal Processing **56** (2008), 3447–3457.
 - [29] E. Somersalo, A. Voutilainen, and J.P. Kaipio, *Non-stationary magnetoencephalography by Bayesian filtering of dipole models*, Inverse Problems **19** (2003), 1047–1063.
 - [30] A. Sorrentino, *Particle Filters for Magnetoencephalography*, Archives of Computational Methods in Engineering **17** (2010), 213–251.
 - [31] A. Sorrentino, L. Parkkonen, A. Pascarella, C. Campi, and M. Piana, *Dynamical MEG source modeling with multi-target Bayesian filtering*, Human Brain Mapping **30** (2009), 1911–1921.
 - [32] S. Taulu, M. Kajola, and J. Simola, *Suppression of interference and arti-*

- facts by the signal space separation method*, Brain Topography **4** (2004), 269–275.
- [33] T.S. Tian, J.Z. Huang, H. Shen, and Z. Li, *A two-way regularization method for MEG source reconstruction*, The Annals of Applied Statistics **X** (to appear), pp–pp.
- [34] T.S. Tian and Z. Li, *A spatio-temporal solution for the EEG/MEG inverse problem using group penalization methods*, Statistics and its Interface **4** (2011), 521–533.
- [35] K. Uutela, M. Hämäläinen, and E. Somersalo, *Visualization of magnetoencephalographic data using minimum current estimates*, NeuroImage **10** (1999), 173–180.
- [36] B.D. Van Veen, W. van Drongelen, M. Yuchtman, and A. Suzuki, *Localization of brain electrical activity via linearly constrained minimum variance spatial filtering*, IEEE Transactions on Biomedical Engineering **44** (1997), 867–880.
- [37] B.N. Vo, S. Singh, and A. Doucet, *Sequential Monte Carlo methods for multi-target filtering with random finite sets*, IEEE Transactions on Aerospace and Electronic Systems **41** (2005), 1224–1245.