Diffuse optical reconstructions of functional near infrared spectroscopy data using maximum entropy on the mean

Zhengchen Cai1, Alexis Machado2, Rasheda Arman Chowdhury2, Amanda Spilkin1, Thomas Vincent1,3,4, Ümit Aydin1,5, Giovanni Pellegrino3, Jean-Marc Lina6,7 & Christophe Grova1,2,3,7

Functional near-infrared spectroscopy (fNIRS) measures the hemoglobin concentration changes associated with neuronal activity. Diffuse optical tomography (DOT) consists of reconstructing the optical density changes measured from scalp channels to the oxy-/deoxy-hemoglobin concentration changes within the cortical regions. In the present study, we adapted a nonlinear source localization method developed and validated in the context of Electro- and Magneto-Encephalography (EEG/MEG): the Maximum Entropy on the Mean (MEM), to solve the inverse problem of DOT reconstruction. We first introduced depth weighting strategy within the MEM framework for DOT reconstruction to avoid biasing the reconstruction results of DOT towards superficial regions. We also proposed a new initialization of the MEM model improving the temporal accuracy of the original MEM framework. To evaluate MEM performance and compare with widely used depth weighted Minimum Norm Estimate (MNE) inverse solution, we applied a realistic simulation scheme which contained 4000 simulations generated by 250 different seeds at different locations and 4 spatial extents ranging from 3 to 40 cm² along the cortical surface. Our results showed that overall MEM provided more accurate DOT reconstructions than MNE. Moreover, we found that MEM was remained particularly robust in low signal-to-noise ratio (SNR) conditions. The proposed method was further illustrated by comparing to functional Magnetic Resonance Imaging (fMRI) activation maps, on real data involving finger tapping tasks with two different montages. The results showed that MEM provided more accurate HbO and HbR reconstructions in spatial agreement with the main fMRI cluster, when compared to MNE.

Functional Near-infrared spectroscopy (fNIRS) is an non-invasive functional neuroimaging modality. It detects changes in oxy-/deoxy-hemoglobin (i.e., HbO/HbR) concentration within head tissues through the measurement of near-infrared light absorption using sources and detectors placed on the surface of the head. In continuous wave fNIRS, the conventional way to transform variations in optical density to HbO/HbR concentration changes at the level of each source-detector channel, is to apply the modified Beer Lambert Law (mBLL). This model assumes homogeneous concentration changes within the detecting region, i.e., ignoring the partial volume effects which indicates the absorption of light within the illuminated regions varies locally. This assumption reduces quantitative accuracy of HbO/HbR concentration changes when dealing with focal hemodynamic changes.

In order to handle these important quantification biases associated with sensor level based analysis, diffuse optical tomography (DOT) has been proposed to reconstruct, from sensor level measures of the optical density, the fluctuations of HbO/HbR concentrations within the brain. This technique not only provides better spatial localization accuracy and resolution of the underlying hemodynamic responses, but also avoids partial volume effect in classical mBLL, hence achieves better quantitative estimation of HbO/HbR concentration changes.
DOT has been applied to reconstruct hemodynamic responses in sensory and motor cortex during median-nerve stimulation and finger tapping, to conduct visual cortex retinotopic mapping and to simultaneous image hemodynamic responses over the motor and visual cortex.

To formalize DOT reconstruction, one needs to solve two main problems. The first one is the forward problem which estimates a forward model or sensitivity matrix that maps local absorption changes within the brain to variations of optical density changes measured by each channel. The second problem is the inverse problem which aims at reconstructing the fluctuations of hemodynamic activity within the brain from scalp measurements.

The forward problem can be solved by generating a subject specific anatomical model, describing accurately propagation of light within the head. Such anatomical model is obtained by segmenting anatomical Magnetic Resonance Imaging (MRI) data, typically into five tissues (i.e., scalp, skull, cerebro-spinal fluid (CSF), white matter and gray matter), before initializing absorption and scattering coefficients values for each tissue type and for each wavelength. Solving the inverse problem relies on solving an ill-posed problem which does not provide a unique solution, unless specific additional constraints are added. The most widely used inverse method in DOT is a linear approach based on Minimum Norm Estimate (MNE) originally proposed for solving the inverse problem of MagnetoencephaloGraphy (MEG) and Electroencephalography (EEG) source localization. It minimizes the $L_2$ norm of the reconstruction error along with Tikhonov regularization. Other strategies to solve DOT inverse problem have also been considered, such as sparse regularization using the $L_1$ norm and Expectation Maximization (EM) algorithm. A non-linear method based on hierarchical Bayesian model for which inference is obtained through an iterative process has been proposed and applied on finger tapping experiments.

Maximum Entropy on the Mean (MEM) framework was first proposed by Ambia et al. and then applied and carefully evaluated by our group in the context of EEG/MEG source imaging. The MEM framework was specifically designed and evaluated for its ability to recover spatially extended generators and when applied on clinical epilepsy data. In addition to its unique ability to recover the spatial extent of the underlying generators, we also demonstrated MEM’s excellent reconstruction spatial accuracy in low SNR conditions, with the ability to limit the influence of distant spurious sources.

We believe that these important aspects should be carefully considered in the context of fNIRS reconstruction. The first one is the ability to accurately recover the spatial extent of the underlying hemodynamic activity for both focal and extended generators. The second one is to provide robust reconstruction results when data SNR decreases, especially when considering the fact that it is challenging to maintain a good intra-subject consistency using continuous-wave fNIRS due to its relatively low SNR. Therefore, our main objective was to adapt the MEM framework for fNIRS reconstruction and carefully evaluate its performance. Moreover, fNIRS reconstruction results tend to be biased towards more superficial regions, because the light sensitivity profile decreases exponentially with the depth of the generators. To overcome this bias, we implemented and evaluated a depth weighted variant of the MEM framework.

The article is organized as follows. The methodology of depth weighted MEM for DOT is first presented. Then, we described our validation framework using realistic simulations and associated validation metrics. fNIRS reconstruction using MEM was compared with widely used depth weighted Minimum Norm Estimate (MNE) inverse solution. Finally, illustrations of the methods on finger tapping fNIRS data set acquired with two different montages from 6 healthy subjects are provided and compared with functional Magnetic Resonance Imaging (fMRI) results.

**Material and methods**

**Statement.** All methods were carried out in accordance with relevant guidelines and regulations. All subjects have signed written informed consent forms for this study which was approved by the Central Committee of Research Ethics of the Minister of Health and Social Services Research Ethics Board, Québec, Canada.

**fNIRS reconstruction.** To perform fNIRS reconstructions, the relationship between measured optical density changes on the scalp and wavelength specific absorption changes within head tissue is usually expressed using the following linear model:

$$Y = AX + e$$

where $Y$ is a matrix ($p \times t$) which represents the wavelength specific measurement of optical density changes in $p$ channels at $t$ time samples. $X$ ($q \times t$) represents the unknown wavelength specific absorption changes in $q$ locations along the cortex at time $t$. $A$ ($p \times q$) is called the light sensitivity matrix which is actually the forward model relating absorption changes in the head to optical density changes measured in each channel. $e$ ($p \times t$) models the additive measurement noise. Solving the fNIRS tomographic reconstruction problem consists in solving an inverse problem which can be seen as the estimation of matrix $X$ (i.e. the amplitude for each location $q$ at time $t$). However, this problem is ill-posed and admits an infinite number of possible solutions. Therefore, solving the DOT inverse problem requires adding additional prior information or regularization constraints to identify a unique solution.

In DOT studies, anatomical constraints can be considered by defining the reconstruction solution space (i.e. where $q$ is located) within the gray matter volume or along the cortical surface. In EEG and MEG source localization studies, it is also common to constrain the reconstruction along the cortical surface. In this study, the reconstruction space was considered as the mid surface defined as the middle layer between gray matter/pial and gray/white matter interfaces.
**Minimum norm estimation (MNE).** Minimum norm estimation is one of the most widely used reconstruction methods in DOT\(^{9,11-15,22}\). Such estimation can be expressed using a Bayesian formulation which solves the inverse problem by estimating the posterior distribution 
P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}\) (i.e. the probability distribution of parameter X conditioned on data Y). A solution can be computed by imposing Gaussian distribution priors on the generators \(X (P(X = N(0, \Sigma^{-1}))\) and the noise \(e (P(e = N(0, \Sigma^{-1})))\). \(\Sigma\) is the inverse of the noise covariance which could be estimated from baseline recordings. \(\Sigma\) is the inverse of the source covariance which is assumed to be an identity matrix in conventional MNE.

The Maximum a Posteriori (MAP) estimator of the posterior distribution \(P(X|Y)\) can be obtained using maximum likelihood estimation:

\[
\hat{X}_{\text{MNE}} = \arg\min_{\lambda} (||Y - AX||^2_{\Sigma_d} + \lambda ||X||^2_{\Sigma})
\]

\[
= (A^T \Sigma_d A + \lambda \Sigma)^{-1} A^T \Sigma_d Y
\]

where \(\hat{X}_{\text{MNE}}\) is the reconstructed absorption changes along the cortical surface. \(\lambda\) is a hyperparameter to regularize the inversion using the priori minimum norm constraint \(||X||^2_{\Sigma}\). In this study, we applied the standard L-Curve method\(^{50}\) to estimate \(\lambda\).

**Depth weighted MNE.** Standard MNE solutions assumes \(\Sigma = I\), which then tends to bias the inverse solution towards the generators exhibiting large sensitivity in the forward model, therefore the most superficial ones\(^3\). When compared to EEG-MEG source localization, such bias is even more pronounced in fNIRS since within the forward model light sensitivity values decrease exponentially with the depth\(^4\). This bias can be compensated by scaling the source covariance matrix such that the variances are equalized\(^{51,52}\). In the context of DOT, depth weighted MNE has been proposed by Culver et al.\(^{53}\) as an approach to compensate this effect and applied in different studies\(^{9,11,12,14-15,22}\). In practice, depth weighting can be formulated differently, here we consider a generalization expression for the implementation of depth weighted MNE as proposed in Lin et al.\(^{14}\). It consists in initializing the source covariance matrix as \(\Sigma^{-1/2} = \Lambda\), resulting in a so called depth weighted MNE solution, described as follows:

\[
\hat{X}_{\text{D-MNE}} = \arg\min_{\lambda} (||Y - AX||^2_{\Sigma_d^\omega} + \lambda ||X||^2_{\Sigma})
\]

\[
diag(\Lambda) = \frac{1}{\text{diag} \left( (A^T \Sigma_d A)^\omega \right)^{1/2}}
\]

Depth weighted MNE solution takes into account the forward model A for each position in the brain and therefore penalizes most superficial regions exhibiting larger amplitude in A, by enhancing the contribution to deeper regions. \(\omega\) is a weighting parameter tuning the amount of depth compensation to be applied. The larger is \(\omega\), the more depth compensation is considered. \(\omega = 0\) would therefore refer to no depth compensation and an identity source covariance model. \(\omega = 0.5\) refers to standard depth weighting approach mentioned above. In the present study, we carefully evaluated the impact of this parameter on DOT accuracy with a set of \(\omega\) values (i.e. \(\omega = 0, 0.1, 0.3, 0.5, 0.7\) and 0.9).

**Maximum entropy on the mean (MEM) for fNIRS 3D reconstruction.** **MEM framework.** The main contribution of this study is the first adaptation and evaluation of MEM method\(^{11-23}\) to perform DOT reconstructions in fNIRS. Within the MEM framework, the intensity of x, i.e. amplitude of X at each location q in Eq. (1), is considered as a random variable, described by the following probability distribution \(dp(x) = p(x)dx\). The Kullback-Leibler divergence or \(\nu\)-entropy of \(dp(x)\) relative to a prior distribution \(dv(x)\) is defined as,

\[
S_v(dp(x)) = -\int_x \log \left( \frac{dp(x)}{dv(x)} \right) dp(x) = -\int_x f(x)\log(f(x))dv(x)
\]

where \(f(x)\) is the \(\nu\)-density of \(dp(x)\) defined as \(dp(x) = f(x)dv(x)\). Following a Bayesian approach to introduce the data fit, we denote \(C_m\) as the set of probability distributions on x that explains the data on average:

\[
Y = [A|_{\text{Vol.}}] E_{dp}[x]_e \quad \text{for} \quad dp \in C_m
\]

where \(Y\) represents the measured optical density changes, \(E_{dp}[x] = \int x dp(x)\) represents the statistical expectation of x under the probability distribution \(dp\), and \(I_q\) is an identity matrix of \((q \times q)\) dimension. Therefore, within the MEM framework, a unique solution of \(dp(x)\) could be obtained,

\[
dp^\star(x) = \arg\max_{dp(x)\in C_m} \left( S_v(dp(x)) \right)
\]

The solution of \(dp^\star(x)\) can be solved by maximizing the \(\nu\)-entropy which is a convex function. It is equivalent to minimize an unconstrained concave Lagrangian function i.e., \(L(dp(x), \kappa, \lambda)\), along with two Lagrangian constraint parameters, i.e., \(\kappa\) and \(\lambda\). It is finally equivalent to maximize a cost function \(D(\lambda)\) which is described as,
\[ D(\lambda) = \lambda^T Y - F(\lambda^T) - \frac{1}{2} \lambda^T \Sigma^{-1}_d (\Sigma^{-1}_d)^T \lambda \]  

(7)

where \( \Sigma^{-1}_d \) is the noise covariance matrix, \( F(\cdot) \) represents the free energy associated with reference \( dv(x) \). It is important to mention that \( D(\lambda) \) is now an optimization problem within a space of dimension equal to the number of sensors. Therefore, if we estimate \( \hat{\lambda}^* = \text{argmax}_x D(\lambda) \), the unique solution of MEM framework is then obtained from the gradient of the free energy:

\[ \hat{X}_{\text{MEM}} = \nabla_x F_\lambda^*(\lambda) | _{\lambda = \hat{\lambda}^*} \]  

(8)

For further details on MEM implementation and theory we refer the reader to\textsuperscript{31–33}.

Construction of the prior distribution for MEM estimation. To define the prior distribution \( dv(x) \) mentioned above, we assumed that brain activity can be depicted by a set of \( K \) non-overlapping and independent cortical parcels. Then the reference distribution \( dv(x) \) can be modeled as,

\[ dv(x) = \prod_{k=1}^{K} \left( 1 - \alpha_k \right) \delta(x_k) + \alpha_k N(\mu_k, \Sigma_k) dx_k, \quad 0 < \alpha_k < 1 \]  

(9)

Each cortical parcel \( k \) is characterized by an activation state, defined by the hidden variable \( S_k \), describing if the parcel is active or not. Therefore we denote \( \alpha_k \) as the probability of \( k \) th parcel to be active, i.e., \( \text{Prob}(S_k = 1) \). \( \delta_k \) is a Dirac function that allows to “switch off” the parcel when considered as inactive (i.e., \( S_k = 0 \)). \( N(\mu_k, \Sigma_k) \) is a Gaussian distribution, describing the distribution of absorptions changes within the \( k \) th parcel, when the parcel is considered as active (\( S_k = 1 \)). This prior model, which is specific to our MEM inference, offers a unique opportunity to switch off some parcels of the model, resulting in accurate spatial reconstructions of the underlying activity patterns with their spatial extent, as carefully studied and compared with other Bayesian methods in Chowdhury et al.\textsuperscript{33}.

The spatial clustering of the cortical surface into \( K \) non-overlapping parcel was obtained using a data driven parcellisation (DDP) technique\textsuperscript{35}. DDP consisted in first applying a projection method, the multivariate source prelocalization (MSP)\textsuperscript{56}, estimating a probability like coefficient (MSP score) between 0 and 1 for each vertex of the cortical surface, characterizing its contribution to the data. DDP is then obtained by using a region growing algorithm, along the tessellated cortical surface, starting from local MSP maxima. Once the parcellization is done, the prior distribution \( dv(x) \) is then a joint distribution expressed as the multiplication of individual distribution of each parcel in Eq. (9) assuming statistical independence between parcels,

\[ dv(x) = dv_1(q_1) dv_2(q_2) ... dv_k(q_k) ... dv_K(q_K) \]  

(10)

where \( dv(x) \) is the joint probability distribution of the prior, \( dv_k(q_k) \) is the individual distribution of the parcel \( k \) described as Eq. (9).

To initialize the prior in Eq. (9), \( \mu_k, \Sigma_k \) which is the mean of the Gaussian distribution, \( N(\mu_k, \Sigma_k) \), was set to zero. \( \Sigma_k \) at each time point \( t \), i.e. \( \Sigma_k(t) \), was defined by Eq. (11) according to Chowdhury et al.\textsuperscript{33},

\[ \Sigma_k(t) = \eta(t) W_k(\sigma)^T W_k(\sigma) \]  

\[ \eta(t) = 0.05 \frac{1}{\sum_{\theta_k} \hat{X}^2_{\text{MNE}}(\theta_k, t)} \]  

(11)

where \( W_k(\sigma) \) is a spatial smootheness matrix, defined by Friston et al.\textsuperscript{57}, which controls the local spatial smoothness within the parcel according to the geodesic surface neighborhood order. Same value of \( \sigma = 0.6 \) was used as in Chowdhury et al.\textsuperscript{33}, \( \eta(t) \) was defined as 5% of the averaged energy of MNE solution within each parcel \( \theta_k \) at time \( t \). Finally, we can substitute this initialization into Eq. (9) to construct the prior distribution \( dv(x) \), and then obtain the MEM solution using Eq. (8).

It is worth mentioning that we did not use MNE solution as the prior of \( \mu_k \) in Eq. (9) at all, which was actually initialized to 0 in our framework. We only used 5% of the averaged energy of MNE solution, over the parcel \( k \), to set the prior for covariance \( \Sigma_k \). The posterior estimation of parameter \( \mu_k \) was estimated from the Bayesian framework by conditioning with data. Moreover, the prior of MEM framework is a mixture of activation probability \( \alpha_k \) and a Gaussian distribution [see Eq. (9)], in which the prior for \( \alpha_k \) was informed by a spatio-temporal extension of the MSP score (see Chowdhury et al.\textsuperscript{33} for further details). These aspects completely differentiate MEM from approaches that iteratively update reconstruction results initialized by a MNE solution.

Depth weighted MEM. In addition to adapting MEM for fNIRS reconstruction, we also implemented for the first time, depth weighting within the MEM framework. Two depth weighting parameters, \( o_{01} \) and \( o_{02} \), were involved in this process. \( o_{01} \) was used to apply depth weighting on the source covariance matrix \( \Sigma_k \) of each parcel \( k \) in Eq. (11). \( o_{02} \) was applied to solve the depth weighted MNE, as described in Eq. (3), before using those prior to initialize the source covariance model within each parcel of the MEM model. Therefore, the standard MNE solution \( \hat{X}_{\text{MNE}}(\theta_k, t) \) in Eq. (11) was replaced by the depth weighted version of MNE solution \( \hat{X}_{\text{dMNE}}(\theta_k, t) \) described by Eq. (3). Consequently, the depth weighted version of \( \Sigma_k(t) \) is now defined as,
\[
\Sigma_k(t)_{dw} = \Lambda_{\omega_k} \eta(t)_{dw} \left( W_k(\sigma) \right)^T W_k(\sigma) \\
\eta(t)_{dw} = 0.05 \frac{1}{|\Omega_k|} \sum_{i \in \Omega_k} \hat{X}^2_{\text{MNE}}(i, t)
\]  

(12)

where \(\Lambda_{\omega_k}\) is the depth weighting matrix for each pacel \(k\), in which \(\omega_k\) was involved to construct this scaling matrix as described in Eq. (3). This initialization followed the logic that depth weighting is in fact achieved by scaling the source covariance matrix. The other depth weighting parameter, \(\omega_2\), was considered when solving \(\hat{X}_{\text{MNE}}(i, t)\), therefore avoiding biasing the initialization of the source covariance with a standard MNE solution.

To comprehensively compare MEM and MNE and also to investigate the behavior of depth weighting, we first evaluated the reconstruction performance of MNE with different \(\omega_2\) (i.e. step of 0.1 from 0 to 0.9). Then two of these values (i.e. \(\omega_2\) = 0.3 and 0.5) were selected for the comparison with MEM since they performed better than the others. Note that the following expressions of depth weighted MEM will be denoted as MEM(\(\omega_1, \omega_2\)) to represent the different depth weighting strategies.

**Accuracy of temporal dynamics.** The last contribution of this study was to improve the temporal accuracy of MEM solutions. In classical MEM approach\(^{33}\), \(\hat{X}_{\text{MNE}}(i, t)\) in Eq. (12) was globally normalized by \(\max_{i \in \Omega_j, t \in T} \hat{X}_{\text{MNE}}(i, t)\), where \(\Omega\) represents all the possible locations along the cortical surface and \(T\) is the whole time segment. Therefore, the constructed prior along the time actually contained the temporal scaled dynamics from MNE solution. To remove this effect, we performed local normalization for \(\hat{X}_{\text{MNE}}(i, t)\) at each time instance \(t\), i.e., by dividing by \(\max_{i \in \Omega_j, t \in T} \hat{X}_{\text{MNE}}(i, t)\). This new feature would preserve the spatial information provided by prior distribution, while allowing MEM to estimate the temporal dynamics only from the data.

**Validation of fNIRS reconstruction methods.** We evaluated the performance of the two fNIRS reconstruction methods (i.e., MEM and MNE), first within a fully controlled environment involving the use of realistic simulations of fNIRS data using montage 1, followed by evaluations on real data acquired with a well-controlled finger tapping paradigm using montage 2. One subject was involved in acquisitions using montage 1 and five subjects participated in acquisitions using montage 2.

**Montage 1** A full Double Density (DD) montage (see Fig. 1) which is a widely used fNIRS montage, was considered given that it allows sufficient dense spatial coverage of fNIRS channel to allow local DOT\(^{58}\). One healthy subject (20 years old, right handed) underwent fNIRS acquisitions with this DD montage, involving the two following sessions,

- A 10 minutes resting state session was acquired to add realistic physiology noise to corrupt our noise-free simulations, thus generating highly realistic fNIRS simulations. These resting-state fNIRS data captured spontaneous fluctuations in fNIRS signals that are related to intrinsic brain activity as well as the physiological noise of non-cerebral origin, associated with systemic blood circulations\(^4\). The subject was seating on a comfortable armchair and instructed to keep the eyes open and to remain awake. The optodes of the full DD montage (i.e. 8 sources and 10 detectors resulting in 50 fNIRS channels) are presented in Fig. 1e. The montage composed of 6 second-order distance channels (1.5 cm), 24 third-order channels(3 cm) and 12 fourth-order channels with 3.35 cm distance. In addition, we also added one proximity detector paired for each source to construct close distance channels (0.7 cm) in order to measure superficial signals within extra-cerebral tissues. To place the montage with respect to the region of interest, the center of the montage was aligned with the center of the right “hand knob” area, which controls the left hand movement\(^6\), projected on the scalp surface and then each optode was projected on the scalp surface (see Fig. 1d).

- The subject was asked to sequentially tap the left thumb against the other digits around 2Hz, therefore the main elicited hemodynamic response was indeed expected over the right hand knob area. The finger tapping paradigm consisted in 10 blocks of 30s tapping task and each of them was followed by a 30 to 35s resting period. The beginning/end of each block was informed by an auditory cue.

**Montage 2:** Five subjects underwent fNIRS acquisitions with personalized optimal montage\(^19\) during a similar aforementioned finger tapping task. Personalized optimal montage was applied to maximize the fNIRS sensitivity to the hand knob within right primary motor cortex of each participant. Please find further details in Supplementary material S4.

**MRI and fMRI Data acquisitions.** Anatomical MRI data were acquired on those 6 healthy subjects (25 ± 6 years old, right-handed) and were considered to generate realistic anatomical head models. MRI data were acquired in a GE 3T scanner at the PERFORM Center of Concordia University, Montreal, Canada. T1-weighted anatomical images were acquired using the 3D BRAVO sequence (1 × 1 × 1 mm\(^3\), 192 axial slices, 256 × 256 matrix), whereas T2-weighted anatomical images were acquired using the 3D Cube T2 sequence (1 × 1 × 1 mm\(^3\) voxels, 168 sagittal slices, 256 × 256 matrix).

Participants also underwent functional MRI acquisition while performing the same finger opposition tasks considered in fNIRS. fNIRS and fMRI data were acquired in two different sessions, one week apart from each other. fMRI acquisition consisted in a gradient echo EPI sequence (3.7 × 3.7 × 3.7 mm\(^3\) voxels, 32 axial slices,
TE = 25 ms, TR = 2000 ms). fMRI Z-maps were generated by standard first-level fMRI generalize linear model analysis using FEAT from FSL v6.0.0 software61,62.

fNIRS data acquisition. fNIRS acquisitions were conducted at the PERFORM Center of Concordia University using a BrainSight fNIRS device (Rogue Research Inc., Montréal, Canada), equipped with 16 dual wavelength sources (685 nm and 830 nm), 32 detectors and 16 proximity detectors (for short distance channels). All montages (i.e., double density and optimal montages) were installed to cover the right motor cortex. Knowing a priori the exact positions of fNIRS channels estimated on the anatomical MRI of each participant, we then used a 3D neuronavigation system (Brainsight TMS navigation system, Rogue Research Inc.) to guide the installation of the sensors on the scalp. This neuronavigation system provided real-time feedback of the optodes targeted positions, while installing them on the subject's head. Finally, every sensor was glued on the scalp using a clinical adhesive, collodion, to prevent motion and ensure good contact to the scalp19,63. For further details about this personalized installation procedure, please refer to our previous publication19.

fNIRS forward model estimation. T1 and T2 weighted anatomical images were processed using FreeSurfer V6.049 and Brain Extraction Tool2 (BET2)61 in FMRIB Software Library (FSL) to segment the head into 5 tissues (i.e. scalp, skull, Cerebrospinal fluid (CSF), gray matter and white matter see Fig. 1a).

Same optical coefficients used in19,63 for the two wavelengths considered during our fNIRS acquisition, 685 nm and 830 nm, were assigned to each tissue type mentioned above. Fluences of light for each optode (see Fig. 1b) was estimated by Monte Carlo simulations with 10^8 photons using MCXLAB developed by Fang and Boas64 and Yu et al.65 (http://mcx.space/). Sensitivity values were then computed using the adjoint formulation and were normalized by the Rytov approximation6.

For each source-detector pair of our montages, the corresponding light sensitivity map was first estimated in a volume space, and then further constrained to the 3D mask of gray matter tissue (see Fig. 1c), as suggested by
Boas and Dale. Then, these sensitivity values within the gray matter volume were projected along the cortical surface (see Figs. 1d and S4c) using the Voronoi based method proposed by32. We considered the mid-surface from FreeSurfer as the cortical surface. This surface was downsampled to 25,000 vertices. This volume to surface interpolation method has the ability to preserve sulco-gyral morphology32. After the interpolation, the sensitivity value of each vertex of the surface mesh represents the mean sensitivity of the corresponding volumetric Voronoi cell (i.e., a set of voxels that have closest distances to a certain vertex than to all other vertices).

fNIRS data preprocessing. Using the coefficient of variation of the fNIRS data, channels exhibiting a standard deviation larger than 8% of the signal mean were rejected14,66–68. Superficial physiological fluctuations were regressed out at each channel using the average of all proximity channels’ (0.7 cm) signals12. All channels were then band-pass filtered between 0.01 Hz and 0.1 Hz using a 3rd order Butterworth filter. Changes in optical density (i.e., $\Delta OD$) were calculated using the conversion to log-ratio. Finally, $\Delta OD$ of finger tapping data were block averaged around the task onsets. Note that since sensors were glued with collodion, we observed very minimal motion during the acquisitions. Real background signal considered to generate realistic simulations also underwent the same preprocessing.

Realistic simulations of fNIRS data. We first considered realistic simulations of fNIRS data to evaluate DOT methods within a fully controlled environment. To do so, theoretical task-induced HbO/HbR concentration changes were simulated within cortical surface regions with a variety of locations, areas and depths. Corresponding optical density changes in the channel space were then computed by applying the corresponding fNIRS forward model, before corrupting noise-free simulation using real resting state fNIRS baseline signal, allowing to add physiological fNIRS signal of cerebral and non-cerebral origin at different signal to noise ratio (SNR) levels. As presented in Fig. 2a, we defined three sets of evenly distributed seeds within the field of view of DOT reconstruction. The locations were selected with respect to the depth relative to the skull, namely we simulated 100 “superficial seeds”, 100 “middle seeds” and 50 “deep seeds”. The cortical regions in which we simulated an hemodynamic response were generated by region growing around those seeds, along the cortical surface. To

Figure 2. Workflow describing our proposed realistic fNIRS simulation framework. (a) 100 Superficial seeds (black dots), 100 Middle seeds (red dots), 50 Deep seeds (blue dots) with spatial extent of $Se = 3, 5, 7, 9$ neighbourhood order within the field of view. (b) Convolution of a canonical HRF model with an experimental block paradigm (60s before and 50s after the onset). (c) Simulated theoretical HbO/HbR fluctuations along the cortical surface within the corresponding generator. (d) Realistic simulations obtained by applying the fNIRS forward model and addition of the average of 10 trials of real fNIRS background measurements at 830 nm. Time course of $\Delta OD$ of all channels with SNR of 5, 3, 2 and 1 respectively are presented. Figure created by Brainstorm59 using the NIRSTORM plugin developed by our team (https://github.com/Nirstorm).
simulate generators with different spatial extents (denoted here as Se), we considered four levels of neighborhood orders, growing geodesically along the cortical surface, resulting in spatial extents ranging from Se = 3, 5, 7, 9 (corresponding areas of 3 to 40 cm²). For simplification, these cortical regions within which an hemodynamic response was simulated will be denoted as “generators” in this paper. For each vertex within a “generator”, a canonical Hemodynamic Response Function (HRF) was convolved with a simulated experimental paradigm which consisted in one block of 20s task surrounded by 60s pre-/post- baseline period (Fig. 2b). Simulated HbO/HbR fluctuations within the theoretical generator (Fig. 2c) were then converted to the corresponding absorption changes of two wavelengths (i.e., 685nm and 830nm). After applying the forward model matrix A in Eq. (1), we estimated the simulated, noise free, task induced ΔOD in all channels. Such a simulation procedure provided a fully controlled access to ground truth since the location and size of the generator along the cortical surface and the corresponding simulated hemodynamic response time courses (HbO/HbR) within each generator were fully known. Therefore, this controlled ground truth, defined in space along the cortical surface and along time, was then considered for quantitative validations of fNIRS reconstructions, when assessing localization error, spatial extent accuracy and accuracy of temporal reconstructions.

ΔOD of real resting state data were then considered as baseline data and used to add realistic fluctuations (noise) to these simulated noise-free signals. Over the 10min of recording, we randomly selected 10 baseline epochs of 120s each, free from any motion artifact by visual inspection. To mimic a standard fNIRS block average response, realistic simulations were obtained by adding the average of these 10 real baseline epochs to the theoretical noise-free simulated ΔOD, at five SNR levels (i.e. SNR = 5, 3, 2, 1). SNR was calculated through the following equation,

$$SNR = \frac{\max(\Delta OD_\lambda[0,t_1])}{mean(\Delta OD_\lambda[-t_0,0])}$$

where $\Delta OD_\lambda[0,t_1]$ is the optical density changes of a certain wavelength $\lambda$ in all channels during the period from 0s to $t_1 = 60s$. $mean(\Delta OD_\lambda[-t_0,0])$ is the standard deviation of $\Delta OD_\lambda$ during baseline period along all channels.

Simulated trials for each of four different SNR levels are illustrated in Fig. 2d. A total number of 4000 realistic simulations were considered for this evaluation study, i.e., 250 (seeds) × 4 (spatial extents) × 4 (SNR levels). Note that resting state fNIRS baseline signal was preprocessed before adding to the simulated signals.

Validation metric. Following the previously described validation metrics, we applied 4 quantitative metrics to access the spatial and temporal accuracy of fNIRS 3D reconstructions. Further details on the computation of those four validation metrics are reported in Supplementary material S1.

- **Area Under the Receiver Operating Characteristic (ROC) curve (AUC)** was used to assess general reconstruction accuracy considering both sensitivity and specificity. AUC score was estimated as the area under the ROC curve, which was obtained by plotting sensitivity as a function of (1-specificity). AUC ranges from 0 to 1, the higher it is the more accurate the reconstruction is.
- **Minimum geodesic distance (Dmin)** measuring the geodesic distance in millimeters, following the circulations of the cortical surface, from the vertex that exhibited maximum of reconstructed activity to the border of the ground truth. Low Dmin values indicate better accuracy in estimating the location of the generator.
- **Spatial Dispersion (SD)** assessed the spatial spread of the estimated generator distribution and the localization error. It is expressed in millimeters. A reconstructed map with either large spatial spread around the ground truth or large localization error would result in large SD values.
- **Shape error(SE)** evaluated the temporal accuracy of the reconstruction. It was calculated as the root mean square of the difference between the normalized reconstructed time course and the normalized ground truth time course. Low SE values indicate high temporal accuracy of the reconstruction.

Statistics. Throughout all of the quantitative evaluations among different methods involving different depth weighting factors $w$ in the results section, Wilcoxon signed rank test was applied to test the significance of the paired differences between each comparison. For each statistical test, we reported the median value of paired differences, together with its p-value (Bonferroni corrected). We are only showing results at 830nm for simulations, since the ones from 690nm under the same SNR level would have provided similar reconstructed spatiotemporal maps except for the reversed amplitudes. However, reconstruction results on real data indeed involved both wavelengths.

Results

**Evaluation of MEM versus MNE using realistic simulations.** We first investigated the effects of depth weighting factor $\omega_2$ selection for depth weighted MNE. To do so, we evaluated spatial and temporal performances of DOT reconstruction for a set of $\omega_2$ (step of 0.1 from 0 to 0.9). Based on those results reported in the Supplementary material S2 and Fig. S1, we decided to considered that most accurate fNIRS reconstructions were obtained when considering $\omega_2 = 0.3$ and 0.5 for depth weighted MNE. Therefore only those two values were further considered for comparison with MEM reconstructions.

Comparison of the performance of MEM and MNE on superficial realistic simulations are presented in Table 1 and Fig. 3, for 4 levels of spatial extent (Se = 3, 5, 7, 9), using boxplot distribution of the 4 validation metrics. We evaluated 3 depth weighted implementations of MEM, namely MEM($\omega_1 = 0.3$, $\omega_2 = 0.3$), MEM($0.3$, 0.5) and MEM(0.5, 0.5), as well as 2 depth weighted implementations of MNE, namely, MNE(0.3) and MNE(0.5).
For spatial accuracy, results evaluated using Dmin, we obtained median Dmin values of 0 mm for all methods, indicating the peak of the reconstructed map, was indeed accurately localized inside the simulated generator. It is worth mentioning that MEM(0.5, 0.5) provided few Dmin values larger than 0 mm in Se = 3 and Se = 5 cases, which consisted of superficial and focal generators. Since MEM accurately estimated the spatial extent, more depth weighting considered for MEM(0.5, 0.5) could result in focal and deeper reconstruction, hence resulting in non-zero Dmin values. On the other hand, MNE would over-estimate the size of the underlying generators, therefore resulting in 0 mm Dmin, but larger SD values in similar conditions.

When considering the general reconstruction accuracy using AUC, for focal generators such as Se = 3 and 5, we found significant larger AUC (see Table 1) for MEM(0.3, 0.3) and MEM(0.3, 0.5) when compared to the most accurate version of MNE, i.e., MNE(0.3). When considering more extended generators, i.e., Se = 7 and 9, MEM(0.3, 0.5) and MEM(0.5, 0.5) achieved significantly larger AUC than MNE(0.3). However, the AUC of
MNE(0.5) was significantly larger than MEM(0.3, 0.3) for $Se = 7$ as well as significantly larger than MEM(0.3, 0.5) and MEM(0.5, 0.5) for $Se = 9$.

In terms of spatial extent of the estimated generator distribution and the localization error, MEM provided significantly smaller SD values among all the comparisons. Finally, for temporal accuracy of the reconstruction represented by SE, MNE provided significantly lower values, but with a small difference (e.g., 0.01 or 0.02, see results on real data as a reference of this effect size), than MEM among all comparisons when $Se = 3, 5$.

Similar comparison between MEM and MNE were conducted respectively for middle seed simulated generators and deep seed simulated generators. Results were overall reporting similar trends when comparing MEM and MNE methods for middle and deep seeds, and as expected more depth weighting resulted in more accurate reconstructions (described in detail in supplementary material, Fig. S2 and Table S1 for middle seeds, Fig. S3 and Table S2 for deep seeds).

To further illustrate the performance of MEM and MNE as a function of the depth of the generator, we are presenting some reconstruction results in Fig. 4. Three generators with a spatial extent $Se = 5$ ($11 \text{ cm}^2$) were selected for this illustration. They were all located around the right “hand knob” area, and were generated from a superficial, middle and deep seed respectively. The first column presents the locations and the size of the generator along the cortical surface. (a) Superficial seed case with reconstructed maps reconstructed using all MEM and MNE implementations considered in this study. (b) Middle seed case with reconstructed maps reconstructed using all MEM and MNE implementations considered in this study. (c) Deep seed case with reconstructed maps reconstructed using all MEM and MNE implementations considered in this study. 20% inflated and zoomed maps are presented on the left corner of each figure. 100% inflated right hemisphere are presented on the right side. All the maps were normalized by their own global maximum and no threshold was applied. Figure created by Brainstorm59 using the NIRSTORM plugin developed by our team (https://github.com/Nirstorm).

Figure 4. Comparisons of the reconstruction maps using MEM and MNE in realistic simulations. Three theoretical regions with spatial extent $Se = 5$ ($11 \text{ cm}^2$) were selected near the ‘hand knob’ at different depth. The first column presents the locations and the size of the generator along the cortical surface. (a) Superficial seed case with reconstructed maps reconstructed using all MEM and MNE implementations considered in this study. (b) Middle seed case with reconstructed maps reconstructed using all MEM and MNE implementations considered in this study. (c) Deep seed case with reconstructed maps reconstructed using all MEM and MNE implementations considered in this study. 20% inflated and zoomed maps are presented on the left corner of each figure. 100% inflated right hemisphere are presented on the right side. All the maps were normalized by their own global maximum and no threshold was applied. Figure created by Brainstorm59 using the NIRSTORM plugin developed by our team (https://github.com/Nirstorm).
MNE(0.3) was not able to recover the deepest aspects of the generator, but also exhibited a large spread outside the ground truth area as suggested by a large SD = 9.69 mm. MNE(0.5) was able to retrieve the main generator, but also exhibited a large spatial spread of SD = 10.16 mm. When considering the generators obtained from the deep seed, MNE(0.3) only reconstructed part of gyrus, missing completely the main sulcus aspect of the generator, resulting in low AUC of 0.57 and a SD of 10.34 mm. MEM(0.3, 0.5) was not able to recover the deepest aspects of the sulcus, but reconstructed accurately the sulci walls, resulting in an AUC of 0.89 and a SD of 2.71 mm. MEM(0.5, 0.5) recovered the deep simulated generator very accurately, as demonstrated by the excellent scores (AUC = 0.97, SD = 2.11 mm) when compared to MNE(0.5). For those three simulations, all methods recovered the underlying time course of the activity with similar accuracy (i.e., similar SE values). In supplementary material, we added Video.1, illustrating the behavior of all the simulations and all methods, following the same layout provided in Fig. 4.

Figure 5. Effects of depth weighting on the depth and size of the simulated generators. First row demonstrates the validation matrices, AUC, SD and SE, as a function of depth of generators. We selected 250 generators created from all 250 seeds with a spatial extent of SD = 5. Depth was calculated by the average of minimum Euclidean distance from each vertex, within each generator, to the head surface. Second row demonstrates the validation matrices, AUC, SD and SE, as a function of size of generators. Involving 400 generators which constructed from 100 superficial seeds with 4 different spatial extend of Sc = 3, 5, 7, 9. Line fittings were performed via a 4 knots spline function to estimate the smoothed trend and the shade areas represent 95% confident interval. Color coded points represent the values of validation matrices of all involved generators. Figure created by MATLAB version (R2016a) https://www.mathworks.com/products/matlab.html.

Note that for this quantitative evaluation of fNIRS reconstruction methods using realistic simulation framework, we considered fNIRS data at only one wavelength (830 nm). Using single wavelength in the context simulation based evaluation is a common procedure in DOT literature9,13,23,25,29,30,69, since we may expect overall similar performances for 685 nm wavelength under the same SNR level.

Effects of depth weighting on the reconstructed generator as a function of the depth and size of the simulated generators. To summarize the effects of depth weighting in 3D fNIRS reconstructions, we further investigated the validation metrics, AUC, SD and SE, as a function of depth and size of the simulated generators. Dmin was not included due to the fact that we did not find clear differences among methods throughout all simulation parameters from previous results. In the top row of Fig. 5, 250 generators created from all 250 seeds with a spatial extent of SD = 5 were selected to demonstrate the performance of different versions of depth weighting as a function of the average depth of the generator. Whereas in the bottom row of Fig. 5, we considered 400 generators constructed from all 100 superficial seeds with 4 different spatial extents of Sc = 3, 5, 7, 9, to illustrate the performance of different versions of depth weighting as a function of the size of the generator. According to AUC, depth weighting was indeed necessary for all methods when the generator moved to deeper regions (> 2 cm) as well as when the size was larger than 20 cm². Moreover, any version of MEM always exhibited clearly less false positives, as indicated by lower SD values, than all of MNE versions, whatever was the depth
or the size of the underlying generator. We found no clear trend and difference of temporal accuracy among methods when reconstructing generators of different depths and sizes.

**Robustness of 3D reconstructions to the noise level.** All previous investigations were obtained from simulations obtained with a SNR of 5, in this section we compared the effect of the SNR level in Fig. 6, on depth weighted versions of MNE and MEM, for superficial seeds only and generators of spatial extent $Se = 5$. We only compared MEM(0.3, 0.5) and MNE(0.5) considering the observation from previous results that these two methods were overall exhibiting best performances in this condition. Regarding Dmin, paired differences were not significant but MNE exhibited more Dmin values above 0 mm than MEM at all SNR levels, suggesting that MNE often missed the main generators while MEM was more accurate in reconstructing the maximum of activity within the simulated generator. Regarding AUC, MEM(0.3, 0.5) exhibited values higher than 0.8 at all SNR levels, whereas MNE(0.5) failed to recover accurately the generator for $SNR = 1$. Besides, in Table 2, we found that difference of AUC between MEM and MNE increased when SNR level decreased, suggesting the good robustness of MEM when decreasing the SNR level. The difference of SD also increased when SNR levels decreased. Indeed, MEM exhibited stable SD values among most SNR levels (except $SNR = 1$), whereas for MNE SD values were highly influenced by the SNR level. Finally, for both methods, decreasing SNR levels resulted in less accurate time course estimation (SE increased), slightly more for MEM when compared to MNE.

**Illustration of MEM and MNE reconstructions on real fNIRS data.** For all finger tapping fNIRS data considered in our evaluations, two wavelength (i.e., 685nm and 830nm) were reconstructed first and then converted to HbO/HbR concentration changes along cortical surface using the standard absorption coefficients for each wavelength and each hemoglobin chromophore (HbO, HbR), as reported in our previous publications[19,47,70]. All the processes from fNIRS preprocessing to 3D reconstruction were completed in Brainstorm[59] using the NIRSTORM plugin developed by our team (https://github.com/Nirstorm). For full double density montage...
(montage 1), reconstructed HbR amplitudes were reversed to positive phase and normalized to their own global maximum, to facilitate comparisons. In Fig. 7a, we showed the reconstructed HbR maps at the peak of the time course (i.e., 31s) for MEM and MNE by considering the 4 depth weighted versions, previously evaluated, i.e., MEM(0.3, 0.3), MEM(0.3, 0.5), MNE(0.3) and MNE(0.5). The two depth weighted versions of MEM clearly localized well the “hand knob” region, while exhibiting very little false positives in its surrounding. On the other hand, both depth weighted version of MNE clearly overestimated the size of the hand knob region and were also exhibiting some distant possibly spurious activity. The fMRI Z-map obtained during the corresponding fMRI task is presented on Fig. 7b, after projection of the volume Z-map on the cortical surface. Fig. 7c showed the time courses within the region of interest representing the “hand knob”. Each curve represents the reconstructed time course of one vertex of the hand knob region and the amplitude were normalized by the peak value within the whole region. Further illustrations of MEM and MNE performance on finger tapping for 5 subjects with montage 2 are presented in Supplementary material S4.

**Discussion**

In the present study, we first adapted the MEM framework in the context of 3D fNIRS reconstruction and extensively validated its performance. The spatial performance of reconstructions can be considered in two aspects, 1) correctly localizing the peak of the reconstructed map close enough to the ground truth area, 2) accurately recovering the spatial extent of the generator. According to our comprehensive evaluations of the proposed depth-weighted implementations of MEM and MNE methods, accurate localization was overall not difficult to achieve as suggested by our results using Dmin metric. Almost all methods provided median value of Dmin to be 0 mm in all simulation conditions except for the lowest SNR = 1 condition where more localization error was found. On the other hand, recovering the actual spatial extent of the underlying generator is actually the most challenging task in fNIRS reconstruction. When considering the results of MNE on both realistic simulations and real finger tapping tasks, either from visual inspection (Figs. 4, 7 and S4) or quantitative evaluation by SD (Fig. 3, Table 1 and supplementary Sect. S2), we found that MNE overall reconstructed well the main generator...
but largely overestimated the size of the underlying generator. MEM was specifically developed, in the context of EEG/MEG source imaging, as a method able to recover the spatial extent of the underlying generators, which has been proved not to be the case for MNE-based approaches. A recent review in the context of EEG/MEG source imaging also suggests that the Bayesian approach with sparsity constraints is required to accurately estimate the spatial extent. These important properties of MEM were successfully demonstrated in our results on fNIRS reconstructions. These excellent performances on spatial accuracy and sensitivity to the spatial extent of the underlying generators, as quantified using Dmin, AUC and SD metrics, were reliable for different sizes and depths of simulated generators, and for real finger tapping fNIRS data as well.

In this study, we performed a detailed evaluation of depth-weighted MNE reconstruction and we also proposed for the first time a depth weighting strategy within the MEM framework, by introducing two parameters: \( \omega_1 \) acting on scaling the source covariance matrix, and \( \omega_2 \) tuning the initialization of the reference for MEM. When compared to depth weighted MNE, the MEM framework demonstrated its ability to reconstruct, different depth of focal generators as well as larger size generators, exhibiting excellent spatial accuracy to recover generators of different depths and spatial extent, as quantified using large AUC values (e.g., high AUC values) and few false positives (e.g., low SD values, see Fig. 5). When considering deeper focal generators (depth \( >2 \) cm), MEM(0.5, 0.5) clearly outperformed all other methods (see AUC and SD values in Fig. 5). In summary, for a large range of depths and spatial extents of the underlying generators, MEM methods exhibited accurate results (large AUC values) and less false positives (lower SD values) when compared to MNE methods. In practice, we would suggest to consider either \( \omega_2 = 0.3 \) or 0.5 for the initialization of MEM in all cases and only tune \( \omega_1 \). This is due to the fact that MNE(0.3 or 0.5) provided a generally good reconstruction with larger true positive rate in most scenarios, therefore providing MEM an accurate reference model, \( \delta t(x) \), to start with. Even when considering the most focal simulated generators (\( \text{Se} = 3 \) case) (see Figs. 3, 5 and Table 1), MEM(0.3, 0.3) and MEM(0.3, 0.5) were actually exhibiting very similar performances. This suggests further confirmed when considered results obtained from real data. For both montages, MEM(0.3, 0.3) results in excellent spatial agreement with fMRI Z-maps. Note that depth weighting was also considered in DOT studies using MNE and a hierarchical Bayesian DOT algorithm. A spatially-variant regularization parameter \( \beta \) was added to a diagonal regularization matrix featuring the sensitivity of every generator (forward model), and the value of \( \beta \) was tuned according to the sensitivity value of a certain depth. In practice, this strategy would result in similar depth compensation as ours, but we preferred the depth weighting parameter \( \omega \) which mapped the amount of compensation from 0 to 1 [as described in Eq. (3)] for easier interpretation and comparison. This is also a standard procedure introduced in EEG/MEG source localization studies. Finally, using the depth weighted MNE solution as the prior is a common consideration in Hierarchical Bayesian framework based fNIRS reconstructions.

Another important contribution of this study was that we improved the temporal accuracy time courses estimated within the MEM framework, resulting in similar temporal accuracy the one obtained with MNE. The largest significant SE difference between MEM and MNE was only 0.02 for \( \text{Se} = 3 \) and 0.01 for \( \text{Se} = 5 \). Corresponding time course estimations are also reported for MEM and MNE in real data (Figs. 7 and S4), suggesting again very similar performances. For instance, SE between MEM and MNE HbO time course was estimated as 0.02 for \( \text{Se} = 3 \) case (see Figs. 3, 5 and Table 1), MEM(0.3, 0.3) and MEM(0.3, 0.5) were actually exhibiting very similar performances. This suggests further confirmed when considered results obtained from real data. For both montages, MEM(0.3, 0.3) results in excellent spatial agreement with fMRI Z-maps. Note that depth weighting was also considered in DOT studies using MNE and a hierarchical Bayesian DOT algorithm. A spatially-variant regularization parameter \( \beta \) was added to a diagonal regularization matrix featuring the sensitivity of every generator (forward model), and the value of \( \beta \) was tuned according to the sensitivity value of a certain depth. In practice, this strategy would result in similar depth compensation as ours, but we preferred the depth weighting parameter \( \omega \) which mapped the amount of compensation from 0 to 1 [as described in Eq. (3)] for easier interpretation and comparison. This is also a standard procedure introduced in EEG/MEG source localization studies. Finally, using the depth weighted MNE solution as the prior is a common consideration in Hierarchical Bayesian framework based fNIRS reconstructions.

To further investigate the effects of the amount of realistic noise in our reconstructions on both reconstruction methods, we performed the comparisons along different SNR levels, i.e., \( \text{SNR} = 1, 2, 3, 5 \). As shown in Fig. 6 and Table 2, we found that MEM was overall more robust than MNE when dealing with simulated signals at lower SNR levels. This is actually a very important result since when reconstructing HbO/HbR responses, one has to consider at least two \( \Delta \text{OD} \) of different wavelengths exhibiting different SNR levels. For the simulation results, we reported reconstruction results obtained from 830nm data, whereas when considering real data (Figs. 7 and S4), we had to convert the reconstruction absorption changes at 685nm and 830nm into HbO/HbR concentration changes. Therefore, our final results were influenced by the SNR of all involved wavelengths. fNIRS is inherently sensitive to inter-subject variability, as also suggested in our application on real data presented in Fig. S4. Data from Sub05 were exhibiting a good SNR level and therefore both MEM and MNE reconstructed accurately the main cluster of the activation, while MNE presented more spatial spread and false positive activation outside the fMRI ROI. When considering subjects for whom we obtained lower SNR data, e.g., Sub02 and Sub03, MEM still recovered an activation map similar to fMRI map. In those cases, MEM not only reported suspicious activation pattern but also incorrectly reconstruct the peak amplitude outside the fMRI ROI. Our results suggesting MEM robustness in low SNR conditions for DOT are actually aligned with similar findings suggested for EEG/MEG source imaging, when considering source localization of single trial data.

To perform a detailed evaluation of our proposed fNIRS reconstructions methods, we developed a fully controlled simulation environment, similar to the one proposed by our team to validate EEG/MEG source localization methods. The fNIRS resting state data, acquired by the same montage (montage1) and underwent the same preprocessing as conducted for the real data, was added to the simulated true hemodynamic response for each channel. Indeed such environment provided us access to a ground truth, which is not possible when considering real fNIRS data set. Previous studies validated tomography results by comparing with fMRI activation map which can indeed be considered as a ground truth, but only for well controlled and reliable paradigms. Since fMRI also measures a signal of hemodynamic origin, it is reasonable to check the concordance between fMRI results and DOT reconstructions. Therefore, as preliminary illustrations, we also compared our MEM and MNE results to fMRI Z-maps obtained during finger tapping tasks on 6 healthy participants (Figs. 7 and S4), suggesting...
overall excellent performances on spatial accuracy (by qualitative visual inspections) of MEM when compared to 
MNE. Further quantitative comparison between fMRI and fNIRS 3D reconstruction, on well controlled finger 
tapping data using MEM and MNE, was conducted in our follow up study73.

Several software packages have been proposed to provide fNIRS reconstruction pipelines, as for instance 
NeuroDOT5,24, AtlasViewer25 and NIRS-SPM26. To ensure an easy access of our MEM methodology to the 
fNIRS community, we developed and released a fNIRS processing toolbox - NIRSTORM (https://github.com/
Nirstorm), as a plugin of Brainstorm software9, which is a renown software package dedicated for EEG/MEG 
analysis and source imaging. Our package NIRSTORM offers standard preprocessing, analysis and visualization 
as well as more advanced features such as personalized optimal montage design, access to forward model 
estimation using MCXlab64,65 and the MNE and MEM implementations considered in this study.

Previously, Tremblay et al.23 had comprehensively compared a variety of fNIRS reconstruction methods 
using large number of realistic simulations. Since introducing MEM was our main goal of this study, we did not 
consider such wide range of methodological comparisons. We decided to carefully compare MEM with MNE 
since MNE remains the main method considered for DOT, and is available in several software packages. As 
suggested in Tremblay et al.23, DOT reconstruction methods based on Tikhonov regularization, such as least 
square regularization in MNE, usually allow great sensitivity, but performed poorly in term of spatial extent - 
largely overestimating the size of the underlying generator. On the other hand, L1-based regularization24–27 could 
achieve more focal solutions with high specificity but much lower sensitivity. As demonstrated in our results, the 
proposed MEM framework allows reaching good sensitivity and accurate reconstruction of the spatial extent 
of the underlying generator. Bayesian model averaging (BMA) originally proposed for EEG source imaging by 
Trujillo-Barreto et al.27, also allows accurate DOT reconstructions with less false positives when compared to 
MNE. Similarly, we carefully compared MEM to Bayesian multiple priors approaches in Chowdhury et al.31 in 
the context of MEG source imaging. Comparing MEM with more advanced DOT reconstruction methods, includ-
ing also the one proposed by Yamashita et al.11, would be of great interest but was out of the scope of this study.

Considering the main contribution of this study was to introduce and adapt the MEM framework for 3D 
fNIRS reconstruction, we decided to first carefully evaluate the performance of MEM using well controlled real-
istic simulations. We also included few real data set reconstructions to illustrate the performance of the MEM 
reconstruction, whereas quantitative evaluation of MEM reconstructions on a larger database were conducted 
in our follow-up study73. In this complementary study, we conducted both individual and group-level quanti-
tative comparisons between fNIRS reconstructed maps and fMRI activation maps, both at the individual and 
group levels. In agreement with our detailed evaluation on realistic simulations, our results reported in73 also 
showed that MEM provided better spatial accuracy than MNE, while both methods offered similar temporal 
accuracy when reconstructing HbO/HbR concentration changes evoked by finger-tapping. In previous reported 
studies12–14,22–27, a high density montage was considered which was proved to be able to provide high spatial 
resolution and robustness to low SNR conditions23, evaluating the performance of MEM when considering high 
density fNIRS montage would be of great interest but was out of the scope of this present study.

Conclusion

In this study, we introduced a new fNIRS reconstruction method entitled Maximum Entropy on the Mean 
(MEM). We first implemented depth weighting into MEM framework and improved its temporal accuracy. To 
carefully validate the method, we applied a large number (n = 4000) of realistic simulations with various spatial 
extents and depths. We also evaluated the robustness of the method when dealing with low SNR signals. The 
comparison of the proposed method with the widely used depth weighted MNE was performed by applying 
four different quantification validation metrics. We found that MEM framework provided accurate and robust 
reconstruction results, relatively stable for a large range of spatial extents, depths and SNRs of the underlying 
generator. Moreover, we implemented the proposed method into a new fNIRS processing plugin - NIRSTORM in 
Brainstorm software to provide the access of the method to users for applications, validations and comparisons.

References

Received: 24 April 2021; Accepted: 24 January 2022
Published online: 10 February 2022
48. Dale, A. M. & Sereno, M. I. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. J. Cogn. Neurosci. 5, 162–176. https://doi.org/10.1162/jocn.1993.5.2.162 (1993).

49. Fischl, B. et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355. https://doi.org/10.1016/S0896-6273(02)00569-X (2002).

50. Hansen, P. C. The L-Curve and its Use in the Numerical Treatment of Inverse Problems. In Computational Inverse Problems in ElectrodCARDIOLOGY Vol. 4, (ed. Johnston, P) 119–142. http://www2.computer.dtu.dk/pubdb/pubs/2231-full.html (WIT in Press, 2000).

51. Fuchs, M., Wagner, M., Köhler, T. & Wischmann, H. A. Linear and nonlinear current density reconstructions. J. Clin. Neurophysiol. 16, 267–295. https://doi.org/10.1097/00004691-199905000-00006 (1999).

52. van der Sluis, A. Condition numbers and equilibration of matrices. Numer. Math. 14, 14–23. https://doi.org/10.1007/BF02159096 (1969).

53. Culver, J. P. et al. Diffuse optical tomography of cerebral blood flow, oxygenation, and metabolism in rat under focal ischemia. J. Cereb. Blood Flow Metab. 23, 911–924. https://doi.org/10.1097/01.WCB.0000076703.71231.BB (2003).

54. Lin, F. H. et al. Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. Neuroimage 31, 160–171. https://doi.org/10.1016/j.neuroimage.2005.11.054 (2006).

55. Lapalme, E., Lina, J. M. & Mattout, J. Data-driven parceling and entropic inference in MEG. Neuroimage 30, 160–171. https://doi.org/10.1016/j.neuroimage.2005.08.067 (2006).

56. Mattout, J., Pélégri-Nissac, M., Garnero, L. & Benali, H. Multivariate source prelocalization (MSP): Use of functionally informed basis functions for better conditioning the MEG inverse problem. Neuroimage 26, 356–373. https://doi.org/10.1016/j.neuroimage.2005.01.026 (2005).

57. Friston, K. et al. Multiple sparse priors for the M/EEG inverse problem. Neuroimage 39, 1104–1120. https://doi.org/10.1016/j.neuroimage.2007.09.048 (2008).

58. Kawaguchi, H., Koyama, T. & Okada, E. Effect of probe arrangement on reproducibility of images by near-infrared topography evaluated by a virtual head phantom. Appl. Opt. 46, 1658. https://doi.org/10.1364/AO.46.001658 (2007).

59. Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D. & Leahy, R. M. Brainstorm: a user-friendly application for MEG/EEG analysis. Comput. Intell. Neurosci. 2011, 879716. https://doi.org/10.1155/2011/879716 (2011).

60. Raffin, E., Pellegrino, G., Di Lazzaro, V., Thielker, A. & Siebner, H. R. Bringing transcranial mapping into shape: Sulcus-aligned mapping captures motor somatotopy in human primary motor hand area. Neuroimage 120, 164–175. https://doi.org/10.1016/j.neuroimage.2015.07.024 (2015).

61. Smith, S. M. et al. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23, 5208–5219. https://doi.org/10.1016/S1053-8119(04)00035-0 (2004).

62. Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W. & Smith, S. M. Review FSL. Neuroimage 62, 782–790. https://doi.org/10.1016/j.neuroimage.2012.01.013 (2012).

63. Yucel, M. A., Selb, J., Boas, D. A., Cash, S. S. & Cooper, R. J. Reducing motion artifacts for long-term clinical NIRS monitoring using collodion-fixed prism-based optical fibers. Neuroimage 85, 192–201. https://doi.org/10.1016/j.neuroimage.2013.06.054 (2014).

64. Fang, Q. & Boas, D. A. Monte Carlo Simulation of Photon Migration in 3D Turbid Media Accelerated by Graphics Processing Units. Opt. Express 17, 20103–20118. https://doi.org/10.1364/OE.17.020103 (2009).

65. Yu, L., Nina-Paravecino, F., Kaeli, D. & Fang, Q. Scalable and massively parallel Monte Carlo photon transport simulations for heterogeneous computing platforms. J. Biomed. Opt. 20, 011054 (2015).

66. Schmitz, C. H. et al. Design and implementation of dynamic near-infrared optical tomographic imaging instrumentation for simultaneous dual-breadth measurements. Appl. Opt. 44, 2140. https://doi.org/10.1364/AO.44.002140 (2005).

67. Schneider, P. et al. Fast 3D near-infrared breast imaging using indocyanine green for detection and characterization of breast lesions. RoFo Fortschritte auf dem Gebiet der Rontgenstrahlen und der Bildgebenden Verfahren 183, 956–963. https://doi.org/10.1055/s-0031-1281726 (2011).

68. Piper, S. K. et al. A wearable multi-channel NIRS system for brain imaging in freely moving subjects. Neuroimage 85, 64–71. https://doi.org/10.1016/j.neuroimage.2016.03.062 (2014).

69. Zhan, Y., Eggebrecht, A. T., Culver, J. P. & Dehghani, H. Image quality analysis of high-density diffuse optical tomography incorporating a subject-specific head model. Front. Neuroenergetics 4, 6. https://doi.org/10.3389/fnene.2012.00006 (2012).

70. Machado, A., Marcotte, O., Lina, J. M., Kobayashi, E. & Grova, C. Optimal optode montage on electroencephalograph/function near-infrared spectroscopy caps dedicated to study epileptic discharges. J. Biomed. Opt. 19, 026010. https://doi.org/10.1117/1.JBO.19.2.026010 (2014).

71. Sohrabpour, A. & He, B. Exploring the extent of source imaging: recent advances in noninvasive electromagnetic brain imaging. Curr. Opin. Biomed. Eng. https://doi.org/10.1016/j.cobe.2021.100277 (2021).

72. Novi, S. L. et al. Integration of spatial information increases reproducibility in functional near-infrared spectroscopy. Front. Neurosci. 14, 1–12. https://doi.org/10.3389/fnins.2020.00746 (2020).

73. Cai, Z. et al. Evaluation of a personalized functional near infra-red optical tomography workflow using maximum entropy on the mean. Human Brain Mapp. https://doi.org/10.1002/hbm.24566 (2021).

74. Eggebrecht, A. T., Muscigrosso, D. & Culver, J. P. NeuroDOT: an extendible Matlab toolbox for streamlined optical brain mapping (Conference Presentation). In Fantini, S., Taroni, P., Tromberg, B. J. & Sevick-Muraca, E. M. (eds.) Optical Tomography and Spectroscopy of Tissue XIII, vol. 10874, p. 51 https://doi.org/10.1117/12.2510526 (SPIE, 2019).

75. Aasted, C. M. et al. Anatomical guidance for functional near-infrared spectroscopy: AtlasViewer tutorial. Neurophotonics 2, 020801. https://doi.org/10.1002/1087-5119.20801 (2015).

76. Ye, J. C., Tak, S., Jiang, K. E., Jung, J. & Jang, J. NIRS-SPM: statistical parametric mapping for near-infrared spectroscopy. Neuroimage 44, 428–447. https://doi.org/10.1016/j.neuroimage.2008.08.036 (2009).

77. Trujillo-Barreto, N. J., Aubert-Vázquez, E. & Valdés-Sosa, P. A. Bayesian model averaging in EEG/MEG imaging. Neuroimage 21, 1300–1319. https://doi.org/10.1016/j.neuroimage.2003.11.008 (2004).

Acknowledgements

This work was supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant Program (C.G. and J.M.L.) and an operating grant from the Canadian Institutes for Health Research (CHMR MOP 135319) (C.G.). NIRS equipment was acquired using grants from NSERC Research Tools and Instrumentation Program and the Canadian Foundation for Innovation (C.G.). Z.C. is funded by the Fonds de recherche du Québec - Sante (FRQS) Doctoral Training Scholarship and the PERFORM Graduate Scholarship in Preventive Health Research. G.P. is funded by Stratus Canada Foundation.

Author contributions

Z.C. and C.G. conceived the original idea. Z.C., A.M., R.A.C., J.M.L. and C.G. developed the theoretical formalism. T.V., U.A. and G.P. designed the task experiment protocol. Z.C. and A.S. conducted the data acquisitions.
Z.C. performed the data analysis. All authors contributed to review the manuscript. Z.C. wrote the manuscript with input from all authors. C.G. supervised the project.

**Competing interests**
The authors declare no competing interests.

**Additional information**

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-06082-1.

**Correspondence** and requests for materials should be addressed to Z.C.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022