

# Hierarchical Bayesian regularization of reconstructions for diffuse optical tomography using multiple priors

Farras Abdelnour,<sup>1</sup> Christopher Genovese,<sup>2</sup>  
and Theodore Huppert<sup>1,3\*</sup>

<sup>1</sup>Department of Radiology, University of Pittsburgh, 200 Lothrop St. Pittsburgh PA 15213, USA

<sup>2</sup>Department of Statistics, Carnegie Mellon University, 5000 Forbes Ave. Pittsburgh PA 15213, USA

<sup>3</sup>Department of Bioengineering University of Pittsburgh, 300 Technology Dr. Pittsburgh PA 15219, USA  
\*huppert@upmc.edu

**Abstract:** Diffuse optical tomography (DOT) is a non-invasive brain imaging technique that uses low-levels of near-infrared light to measure optical absorption changes due to regional blood flow and blood oxygen saturation in the brain. By arranging light sources and detectors in a grid over the surface of the scalp, DOT studies attempt to spatially localize changes in oxy- and deoxy-hemoglobin in the brain that result from evoked brain activity during functional experiments. However, the reconstruction of accurate spatial images of hemoglobin changes from DOT data is an ill-posed linearized inverse problem, which requires model regularization to yield appropriate solutions. In this work, we describe and demonstrate the application of a parametric restricted maximum likelihood method (ReML) to incorporate multiple statistical priors into the recovery of optical images. This work is based on similar methods that have been applied to the inverse problem for magnetoencephalography (MEG). Herein, we discuss the adaptation of this model to DOT and demonstrate that this approach provides a means to objectively incorporate reconstruction constraints and demonstrate this approach through a series of simulated numerical examples.

©2010 Optical Society of America

**OCIS codes:** (170.3010) Image reconstruction techniques; (170.2655) Functional monitoring and imaging.

---

## References and links

1. D. A. Boas, A. M. Dale, and M. A. Franceschini, "Diffuse optical imaging of brain activation: approaches to optimizing image sensitivity, resolution, and accuracy," *Neuroimage* **23**(Suppl 1), S275–S288 (2004).
2. T. J. Huppert, M. S. Allen, S. G. Diamond, and D. A. Boas, "Estimating cerebral oxygen metabolism from fMRI with a dynamic multicompartment Windkessel model," *Hum. Brain Mapp.* **30**(5), 1548–1567 (2009) (PMCID: 2670946).
3. T. Wilcox, H. Bortfeld, R. Woods, E. Wruck, and D. A. Boas, "Using near-infrared spectroscopy to assess neural activation during object processing in infants," *J. Biomed. Opt.* **10**(1), 011010 (2005).
4. S. Perrey, "Non-invasive NIR spectroscopy of human brain function during exercise," *Methods* **45**(4), 289–299 (2008).
5. I. Miyai, H. C. Tanabe, I. Sase, H. Eda, I. Oda, I. Konishi, Y. Tsunazawa, T. Suzuki, T. Yanagida, and K. Kubota, "Cortical mapping of gait in humans: a near-infrared spectroscopic topography study," *Neuroimage* **14**(5), 1186–1192 (2001).
6. U. Sunar, S. Makonnen, C. Zhou, T. Durduran, G. Yu, H. W. Wang, W. M. Lee, and A. G. Yodh, "Hemodynamic responses to antivasular therapy and ionizing radiation assessed by diffuse optical spectroscopies," *Opt. Express* **15**(23), 15507–15516 (2007).
7. U. Sunar, H. Quon, T. Durduran, J. Zhang, J. Du, C. Zhou, G. Yu, R. Choe, A. Kilger, R. Lustig, L. Loevner, S. Nioka, B. Chance, and A. G. Yodh, "Noninvasive diffuse optical measurement of blood flow and blood oxygenation for monitoring radiation therapy in patients with head and neck tumors: a pilot study," *J. Biomed. Opt.* **11**(6), 064021 (2006).
8. S. R. Arridge, "Optical tomography in medical imaging," *Inverse Probl.* **15**(2), 14–93 (1999).
9. A. P. Gibson, J. C. Hebden, and S. R. Arridge, "Recent advances in diffuse optical imaging," *Phys. Med. Biol.* **50**(4), R1–R43 (2005).

10. J. Mattout, C. Phillips, W. D. Penny, M. D. Rugg, and K. J. Friston, "MEG source localization under multiple constraints: an extended Bayesian framework," *Neuroimage* **30**(3), 753–767 (2006).
11. K. J. Friston, *Statistical parametric mapping: the analysis of functional brain images*. 2007, London: Academic. vii, 647.
12. R. W. Cox, "AFNI: software for analysis and visualization of functional magnetic resonance neuroimages," *Comput. Biomed. Res.* **29**(3), 162–173 (1996).
13. A. Li, Q. Zhang, J. P. Culver, E. L. Miller, and D. A. Boas, "Reconstructing chromosphere concentration images directly by continuous-wave diffuse optical tomography," *Opt. Lett.* **29**(3), 256–258 (2004).
14. D. Harville, "Maximum likelihood approaches to variance component estimation and related problems," *J. Am. Stat. Assoc.* **72**(358), 320–338 (1977).
15. K. J. Friston, W. Penny, C. Phillips, S. Kiebel, G. Hinton, and J. Ashburner, "Classical and Bayesian inference in neuroimaging: theory," *Neuroimage* **16**(2), 465–483 (2002).
16. K. J. Friston, D. E. Glaser, R. N. Henson, S. Kiebel, C. Phillips, and J. Ashburner, "Classical and Bayesian inference in neuroimaging: applications," *Neuroimage* **16**(2), 484–512 (2002).
17. A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc., B* **39**(1), 1–38 (1977).
18. D. K. Joseph, T. J. Huppert, M. A. Franceschini, and D. A. Boas, "Diffuse optical tomography system to image brain activation with improved spatial resolution and validation with functional magnetic resonance imaging," *Appl. Opt.* **45**(31), 8142–8151 (2006).
19. F. Abdelnour, B. Schmidt, and T. J. Huppert, "Topographic localization of brain activation in diffuse optical imaging using spherical wavelets," *Phys. Med. Biol.* **54**(20), 6383–6413 (2009) (PMCID: 2806654.).
20. I. Daubechies, *Ten Lectures On Wavelets*. SIAM, 1992.
21. T. J. Huppert, S. G. Diamond, and D. A. Boas, "Direct estimation of evoked hemoglobin changes by multimodality fusion imaging," *J. Biomed. Opt.* **13**(5), 054031 (2008).
22. D. A. Boas, and A. M. Dale, "Simulation study of magnetic resonance imaging-guided cortically constrained diffuse optical tomography of human brain function," *Appl. Opt.* **44**(10), 1957–1968 (2005).
23. B. W. Pogue, S. C. Davis, X. Song, B. A. Brooksby, H. Dehghani, and K. D. Paulsen, "Image analysis methods for diffuse optical tomography," *J. Biomed. Opt.* **11**(3), 033001 (2006).

---

## 1. Introduction

Diffuse optical tomography (DOT) is a non-invasive technology which uses low-levels of non-ionizing light in the range of 650-900nm to record changes in the optical absorption and scattering of tissue. Over the past thirty years, the use of DOT for non-invasively imaging the human brain has been steadily growing as reviewed in [1]. As compared with functional MRI (fMRI), DOT is less costly, more portable, and allows for a wider range of experimental scenarios because it does not require a dedicated scanner nor require the subject to lay supine. Moreover, optical imaging has the ability to resolve changes in both oxy- and deoxy-hemoglobin (denoted HbO<sub>2</sub> and Hb respectively) within the brain using multiple wavelengths of light, which can potentially lead to the ability to discriminate blood flow and oxygen metabolism changes [2]. Examples of niche applications for optical brain research have included studies on infant and child brain activation [3], studies of activation during exercise and mobility [4,5], bedside monitoring of clinical patients [6,7], as well as more traditional cognitive testing and psychology studies. Because of its low cost of operation and portability, DOT has been growing in popularity in these fields over the last several years.

Although a strength of optical imaging is its temporal resolution (several hertz) and ability to detect both oxy- and deoxy-hemoglobin, optical imaging can also provide some degree of specificity to spatially localize regions of brain activity through the reconstruction of images from data collected via multiple optical light emitter and detector pairs. In a typical optical brain imaging experiment, a grid of light emitters and detectors is placed on the surface of the scalp as shown in Fig. 1. The optical absorption changes recorded from light diffusely traveling between emitter and detector pairs can be used to recover low-resolution spatial images of the underlying blood flow changes. However, such images are often difficult to accurately recover due to optical scattering in the tissue, the limited number of measurements typically available, and the inverse problem of estimating changes within the underlying volume of tissue (brain) from measurements made only on the surface of the head. The estimation of optical images is generally both under-determined (more unknowns than measurements) and ill-posed (no single unique solution). The optical inverse problem has been reviewed in [8].

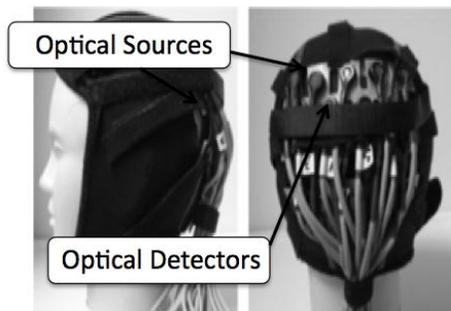


Fig. 1. Diffuse optical imaging uses fiber optic based light sources and detectors to record changes in the optical absorption of underlying tissue. A grid of sensors is placed non-invasively on the head of a participant and used to measure changes in oxy- and deoxy-hemoglobin in the brain during task-evoked activation. The source-detector arrangement in the probe above is shown in Fig. 2.

Active research on the optical inverse problem has led to continued improvements in recent years and for an overview of image reconstruction techniques see [9]. To date, much of the work on the optical inverse problem has involved the use of regularization priors, such as the weighted minimum norm (WMN) or Tikhonov regularization. In general, regularization methods require an estimate of weight (trust) given to the regularization penalty (prior) that is usually predefined or manually tuned to yield acceptable images. One of the difficulties of these techniques, therefore, is the need for an *a priori* choice of this weight or weights in the case of multiple priors.

In this paper, we will describe how the restricted maximum likelihood (ReML) method and a Bayesian formulation can be used to stabilize the optical inverse problem and introduce multiple statistical priors. ReML has been introduced previously in the field of magnetoencephalography (MEG) for a similar inverse model [10]. This method is also the basis of a deconvolution model currently implemented for the time-series analysis of fMRI data within the software packages SPM (statistical parametric mapping [11]) and AFNI (analysis of functional neuroimages [12]). The purpose of this paper is not to exhaustively cover the theory behind the ReML method, which can be found in previous literature (see [11] for a review). Instead, we will briefly outline the concept, describe how to implement this method for optical spatial inverse problems, and provide several demonstrations of how this approach can be used to optimally introduce multiple priors into optical reconstructions. Throughout this paper, we will present numerical examples of this model with increasing complexity in order to gradually introduce components of this method. First, in initial demonstrations, we will first show that for the case of a single minimum norm prior model (similar to the widely used Tikhonov regularization to the optical inverse problem), the ReML method produces identical results to those obtained through optimized regularization via the L-curve method. In later sections of this work, we generalize this model to show that the ReML method allows the incorporation of additional priors including assumptions about the physiology of the brain's response and region-of-interest information. We will finally show how the ReML method can be used to independently tune the Bayesian noise model in the brain and superficial layers to create depth discrimination and to separate superficial noise and brain activity signals.

## 2. Theory

### *The optical forward model*

Diffuse optical brain imaging has been described in several recent reviews [1]. In this section, we will only briefly describe the setup and recording of optical data as it pertains to the optical spatial inverse problem. In a typical DOT experiment, a grid of light emitters and detectors is

positioned on the surface of the scalp of the subject (see Fig. 1). At each of these emitter positions, light is sent into the tissue at two or more wavelengths. Due to the highly scattering nature of biological tissue, this light spreads as it enters the tissue. For samples thicker than a few millimeters, the propagation of light through tissue is often approximated by a diffusion-based random walk of the photons of light and can be modeled through Monte Carlo, finite difference, finite element, or boundary element methods. The propagation of the light depends on the structure of the underlying layers of the tissue (in our case; the scalp, skull, cerebral spinal fluid, and gray/white matter layers of the head), which defines the volume sampled by each optical emitter and detector combination. The optical measurement model is approximated by the expression

$$OD_{i,j}^{\lambda} = L_{i,j}^{\lambda} \left( \mu_A^{\lambda}, \mu_S^{\lambda} \right) + G_{i,j}^{\lambda} + v_{i,j}^{\lambda} \quad (1)$$

where OD is the measured optical density and is defined as

$$OD_{i,j}^{\lambda} = -\text{Log} \left( \frac{I_{i,j}^{\lambda}}{I_0^{\lambda}} \right) \quad (2)$$

In Eq. (2),  $I$  is the intensity of light exiting the tissue and  $I_0$  is the light entering the tissue.  $G$  is a geometry dependent factor. In Eq. (1),  $v$  is an additive noise in the measurement space (e.g. instrument noise), which will be emphasized further in the context of the ReML model.  $L_{ij}^{\lambda}$  is the optical measurement model obtained from estimation of the ensemble path of photons through the tissue and describes the summation of absorption values  $\mu_A$  along the diffuse path traveled by the light going from a particular light emitter to a detector pair (i,j). Both  $\mu_A$  and  $\mu_s$  are vectors of the absorption and scattering values at each position in the volume and can be reshaped as an image of these changes.

During brain activity, regional changes in blood flow alter the concentration of hemoglobin in the brain and in turn change the absorption of the tissue. In this work, we will ignore scattering changes associated with brain activity although the model we will describe can be easily extended to deal with scattering as well. For small changes in absorption, as typically observed for brain activity studies, the change in optical density  $\Delta OD$  is approximated by linearization of Eq. (1) around the baseline values of  $\mu_A$  and  $\mu_s$  and subtraction of the baseline absorption.

$$\Delta OD_{i,j}^{\lambda} = A_{i,j}^{\lambda} \left( \mu_A^{\lambda}, \mu_S^{\lambda} \right) \cdot \Delta \mu_A^{\lambda} + v_{i,j}^{\lambda} \quad (3)$$

$\Delta \mu_A^{\lambda}$  is a vector of the changes in absorption at each position (voxel) in the underlying tissue.  $A_{ij}^{\lambda}$  is the Jacobian of the optical measurement model. Equation (3), describes the optical forward model describing the change in optical signal caused by changes in the absorption in the underlying tissue for one particular wavelength of light and set of baseline optical properties. Typically in brain imaging studies, two or more wavelengths of light are used to provide an ability to distinguish changes in both oxy-hemoglobin ( $\text{HbO}_2$ ) and deoxy-hemoglobin (Hb). The overall absorption at each wavelength is a linear combination of the contributions from each chromophore and is given by the Beer-Lambert expression

$$\Delta \mu_A^{\lambda} = \epsilon_{\text{HbO}_2}^{\lambda} \cdot \left( \Delta [\text{HbO}_2] + \omega_{\text{HbO}_2} \right) + \epsilon_{\text{Hb}}^{\lambda} \cdot \left( \Delta [\text{Hb}] + \omega_{\text{Hb}} \right), \quad (4)$$

where  $\epsilon_{\text{HbX}}^{\lambda}$  is the molar extinction coefficient for oxy- or deoxy-hemoglobin at the particular wavelength and describes the wavelength specific absorption properties of these chromophores per molar unit of concentration.  $\omega_{\text{HbX}}$  is a second type of additive noise (or uncertainty error) term acting in the image (brain) space and is distinct from the measurement space noise ( $v$ ). We will clarify this distinction later in the context of the ReML model. Again,  $\Delta \mu_A$ ,  $\Delta [\text{HbX}]$ , and  $\omega$  are vectors representing these changes at each position in the tissue. Equation (4) can be substituted into the optical forward model to produce the optical measurement model with spectral priors (e.g. Li et al [13])

$$\Delta OD_{i,j}^{\lambda} = A_{i,j}^{\lambda} \cdot \left( \varepsilon_{HbO_2}^{\lambda} \cdot (\Delta[HbO_2] + \omega_{HbO_2}) + \varepsilon_{Hb}^{\lambda} \cdot (\Delta[Hb] + \omega_{Hb}) \right) + \nu_{i,j}^{\lambda} \quad (5)$$

From here on, the dependence of the forward model ( $A_{i,j}$ ) on baseline absorption and scattering coefficients will be no longer explicitly written (e.g.  $A_{i,j}^{\lambda} = A_{i,j}^{\lambda}(\mu_A^{\lambda}, \mu_s^{\lambda})$ ). Changes in oxy- and deoxy-hemoglobin can be inferred from optical measurements at multiple ( $N$ ) wavelengths by means of solving a set of linear equations given by

$$\begin{bmatrix} \Delta OD_{i,j}^{\lambda_1} \\ \Delta OD_{i,j}^{\lambda_2} \\ \Delta OD_{i,j}^{\lambda_N} \end{bmatrix} = \begin{bmatrix} A_{i,j}^{\lambda_1} \cdot \varepsilon_{HbO_2}^{\lambda_1} & A_{i,j}^{\lambda_1} \cdot \varepsilon_{Hb}^{\lambda_1} \\ A_{i,j}^{\lambda_2} \cdot \varepsilon_{HbO_2}^{\lambda_2} & A_{i,j}^{\lambda_2} \cdot \varepsilon_{Hb}^{\lambda_2} \\ A_{i,j}^{\lambda_N} \cdot \varepsilon_{HbO_2}^{\lambda_N} & A_{i,j}^{\lambda_N} \cdot \varepsilon_{Hb}^{\lambda_N} \end{bmatrix} \cdot \left( \begin{bmatrix} \Delta[HbO_2] \\ \Delta[Hb] \end{bmatrix} + \begin{bmatrix} \omega_{HbO_2} \\ \omega_{Hb} \end{bmatrix} \right) + \begin{bmatrix} \nu_{i,j}^{\lambda_1} \\ \nu_{i,j}^{\lambda_2} \\ \nu_{i,j}^{\lambda_N} \end{bmatrix} \quad (6)$$

and  $\lambda_i$  denotes the  $i^{\text{th}}$ -wavelength. Hereafter, the optical forward model will be written in the more compact form

$$Y = H \cdot (\beta + \omega) + \nu \quad (7)$$

where the new variable  $\beta$  has been introduced to describe the unknown values of the combination of oxy- and deoxy-hemoglobin changes in the tissue, given by

$$\beta = \begin{bmatrix} \Delta[HbO_2] \\ \Delta[Hb] \end{bmatrix} \quad (8)$$

In summary, Eqs. (7) and (8) describe the linear relationship between changes in concentrations of HbO<sub>2</sub> and Hb in the tissue, and the changes in optical density as recorded on the surface between optical sources and detectors. It is this equation that must be inverted in order to reconstruct an image (volume) of the hemodynamic changes in the brain.

#### *The optical inverse problem*

The estimation of optical images by the inversion of Eq. (7) entails an underdetermined problem where there generally are significantly less available measurements ( $Y$ ) than unknown parameters ( $\beta$ ) in the image to-be-estimated. This means that, in general, there is not enough information in the measurements alone to yield accurate and unique estimates of images of brain activity. There are two general approaches to solving this problem—regularization and Bayesian theories. In general, regularization theory (including Tikhonov regularization) has been most widely used and is more familiar to practitioners of the optical inverse problem. On the other hand, ReML and our current work are based on the Bayesian interpretation. For this reason, we will briefly attempt to reconcile these two theories noting that for a subset of regularization models in the class of linear-quadratic regularization (which includes many of the current optical inverse models), there is an equivalent Bayesian interpretation of the model.

Regularization methods attempt to stabilize the solution of the inverse problem by extending the objective function used to minimize the problem by adding additional penalty terms. For example, in conventional least squares inverse models, the least-squares cost function minimizes the mean squared error of the model to fit the data. In the case of the linear model  $Y = H \cdot \beta$ , the least squares solution is given by

$$\arg \min_{\beta} \|Y - H \cdot \beta\|_N^2 \quad (9)$$

In other words, Eq. (9) aims to find the value of the parameter set ( $\beta$ ) that minimizes the error to the measured data. The notation  $\|X\|_N^2$  denotes the weighted norm

( $\|X\|_N^2 \equiv X^T \cdot N \cdot X$ ). Generalized Tikhonov regularization extends this least-squares cost function by adding a penalty for deviations of  $\beta$  from some prior expected value of the parameter ( $\beta_0$ ) and is given by the minimization expression

$$\arg \min_{\beta} \|Y - H \cdot \beta\|_N^2 + \lambda \cdot \|\beta - \beta_0\|_P^2 \quad (10)$$

The term  $\lambda$  is a hyper-parameter, which is used to tune the model. In the case that  $\lambda$  is small, the solution favors minimizing the residual error with the measured data and likewise when  $\lambda$  is large, the solution is biased towards matching the prior ( $\beta_0$ ). A typical assumption in the optical inverse model is that  $\beta_0$  is zero, which results in what is called the minimum norm solution. The regularization model can be extended to add additional penalty terms. For example, Li *et al* [13] extended this model with a second penalty term which applied only to parameters outside of a predefined region-of-interest (for example regions selected from MRI segmentation). The cost function proposed by Li et al was:

$$\arg \min_{\beta} \|Y - H \cdot \beta\|_r^2 + \lambda_1 \cdot \|M \cdot \beta\|_r^2 + \lambda_2 \cdot \|(1 - M) \cdot \beta\|_r^2 \quad (11)$$

where  $M$  specifies a binary mask of a predefined region-of-interest such that

$$M = \begin{cases} 1 & \text{if in region-of-interest} \\ 0 & \text{else} \end{cases} \quad (12)$$

The Li et al model thus had two regularization hyperparameters ( $\lambda_1$  and  $\lambda_2$ ), which applied penalties to the parameters inside and outside of the region-of-interest respectively. Alternative regularization models have been proposed to add low-pass or high-pass operators to impose smoothness on the solution. In regularization models, the L-curve technique and generalized cross-validation can be used to optimally select the hyperparameters of the model. However to date, many optical reconstruction methods have used  $\lambda$  as a manual tuning parameter allowing images to be adjusted in a subjective optimization. In general terms, the regularization hyperparameter ( $\lambda$ ) is a weight that is assigned to that penalty term in the cost function.

Linear quadratic regularization models are a subset of regularization cost functions that only contain L2 norm penalty terms such as those described in Eqs. (9)–(11). The cost function can be viewed as a weighted penalization for the distance of the solution from the priors (either the measurement itself or the prior on the solution; e.g.  $\beta_0$ ). In the regularization model, these distance penalties (e.g.  $N$  and  $\lambda \cdot P$  in Eq. (10)) can be somewhat arbitrary provided that they are symmetric matrices. In contrast, the Bayesian model offers an alternative interpretation by suggesting that the optimal distance weight should be the inverse of a covariance matrix. For example, in Eq. (10), the weighted norm penalty  $N$  should be the inverse of the measurement noise covariance and from the second term, the value of  $\lambda \cdot P$  should be the inverse of the parameter covariance. In terms of the optical inverse model, these two terms are the covariance of  $v$  and  $\omega$  respectively from Eq. (7).

Whether one interprets Eq. (10) from a (linear-quadratic) regularization or Bayesian perspective, the solution to the linear model ( $Y = H \cdot \beta$ ) is the same and is given by the Gauss-Markov equation:

$$\beta = \beta_0 + (H^T \cdot N \cdot H + \lambda \cdot P)^{-1} H^T \cdot N \cdot (Y - H \cdot \beta_0) \quad (13)$$

In the case of the Bayesian model,  $N$  and  $\lambda \cdot P$  would be given by the inverse of the measurement and parameter covariance models (which we will later denote  $C_N$  and  $C_P$  following the convention of the SPM software).

### Restricted Maximum Likelihood (ReML)

While regularization models are not restricted to linear quadratic expressions and are thus more general than Bayesian models, the Bayesian point-of-view offers additional optimization methods to select the hyperparameters of the model (e.g.  $\lambda$ ) as alternatives to L-curve or cross-validation methods. In particular, under the Bayesian point-of-view and the interpretation of  $N$  and  $P$  as the inverse of covariance matrices, additional objective functions such as the maximum likelihood of the model can be used to optimize the hyperparameters. Restricted maximum likelihood was introduced by Harville [14] as a method to produce unbiased estimates of the covariance parameters of a linear mixed model under Bayesian assumptions. This approach is implemented in several commercial statistical packages such as SPSS in the *MIXED* function. In the context of neuroimaging ReML was introduced by Friston *et al* for the stabilization of the temporal deconvolution model used for analysis of brain activity images in functional MRI [15,16]. This is implemented in the software SPM (Statistical Parametric Mapping [11]; ). This method was later implemented in the context of the ill-posed image reconstruction inverse model for magnetoencephalography (MEG) and electroencephalography (EEG) also within the SPM software [10].

Details of the derivation and theory behind the ReML model are described in several papers by Friston *et al* [15,16]. In brief, ReML is based on the maximization of the log-likelihood of the data conditional on the set of hyperparameters (e.g.  $p(Y|\lambda)$ ). It can be shown that (see appendix 3 of [11]) maximization of the log-likelihood function is equivalent to maximizing the free-energy of the model and is given by the expression:

$$\arg \max_{\{\beta, C_N, C_P\}} -\frac{1}{2} \|Y - H \cdot \beta\|_{C_N}^2 - \frac{1}{2} \|\beta - \beta_0\|_{C_P}^2 - \frac{1}{2} \text{Log} |C_N| - \frac{1}{2} \text{Log} |C_P| \quad (14)$$

Note that this is similar to the previous weighted least-squares cost function expression with the addition of log-determinant penalties for the covariance terms and a change in sign of all the terms.

Rather than solving for the full covariance matrices ( $C_N$  and  $C_P$ ) from Eq. (14), the covariance models can be parameterized as a linear combination of covariance components. For example:

$$\begin{aligned} C_N &= \sum_i \Lambda_i \cdot Q_{N,i} \\ C_P &= \sum_j \Lambda_j \cdot Q_{P,j} \end{aligned} \quad (15)$$

where  $Q_N$  and  $Q_P$  are symmetric matrices that can be used to build up the covariance model. In the example of the optical model,  $C_N$  represents the covariance of the measurement noise and thus, two (or more) diagonal covariance components ( $Q_N$ ) might be used with each representing the variance on one of the two (or more) measured optical wavelengths. In the methods section, we will further detail the selection of these components for the optical model. The hyper-parameters ( $\Lambda$ ; upper-case lambda) in Eq. (15) adjust the weighting of these covariance components. Again, in the context of the two wavelength optical model, there would be two hyper-parameters allowing adjustment of the noise at the two wavelengths. Note in reference to the work by Friston *et al* [15] for the derivation of the ReML model, a lower-case lambda ( $\lambda$ ) is used and has been switched here for distinction from the term as used in the context of the regularization model.

In order to solve the ReML model, the expectation-maximization (EM) algorithm [17] can be used. In order to minimize the parameterized form of Eq. (14) for both the parameters ( $\beta$ ) and the hyperparameters ( $\Lambda$ ), the EM model alternates between estimation of both types. First given an estimate of the hyperparameters, the Gauss-Markov expression is solved to estimate the parameters:

$$\beta = \beta_0 + \left( H^T \cdot C_N^{-1} \cdot H + C_P^{-1} \right)^{-1} H^T \cdot C_N^{-1} \cdot (Y - H \cdot \beta_0) \quad (16)$$

Note, Eqs. (13) and (16) are identical, where we have just substituted to the specific Bayesian form of the model.

Once the parameters of the model are estimated from Eq. (16) (expectation step), the Eq. (14) is maximized for the hyperparameters ( $\Lambda$ ) in the maximization step. The matrix derivative of Eq. (14) with respect to the vector of hyperparameters is set to zero and solved to yield an updated estimate of these hyperparameters. These are substituted back into Eq. (16) to re-estimate the model parameters. This is repeated until convergence is met. In our model, we defined convergence by the change in the free energy of the model [Eq. (14)], which we describe further in the methods section.

### 3. Methods

In this paper, we will demonstrate the application of a ReML model to the optical inverse problem using several numerical examples. In this section, we will detail the procedure to generate these simulations and the practical implementation of the ReML model as specifically related to our optical model.

#### *Calculation of optical forward model*

In this study, we used a semi-infinite homogeneous slab model for the calculation of the optical forward model. We have chosen to use this model for computational reasons, because this geometry has a known analytic solution, however, this regularization model can be extended for any such forward model. For the simulations, we used an overlapping (tomography) imaging probe as described in Joseph et al [18]. A total of 8 sources and 15 detectors were used with a nearest neighbor distance of 2.5mm and a second nearest neighbor distance of 4.2mm. The probe geometry used in this study is shown in Fig. 2. Measurements were simulated at 830nm and 690nm matching the optical system in our lab.

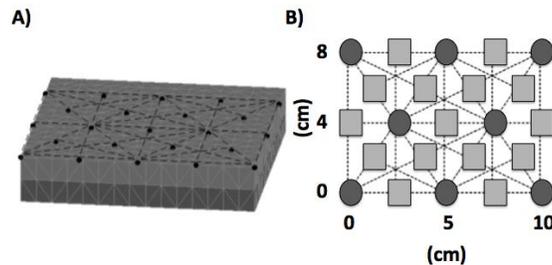


Fig. 2. Simulation (A) and optical probe geometry (B) used in the construction of sample problems in this work. This probe was based on a tomography (over-lapping measurement) design described in [18] consisting of eight source positions (circles) and fifteen detector positions (squares).

We generated four types of simulated activations. In the first examples of this model, we used a single layered model of size 16x16x1 voxels [6.7mm x 6.7mm x 10mm]. An activation spot (13.4mm x 13.4mm) was added in oxy-hemoglobin (+ 1.0 $\mu$ M) and deoxy-hemoglobin (-0.25 $\mu$ M). Measurement noise was added to the measurement vector to reach the desired signal-to-noise ratio for each section. In the later set of examples, a more realistic two-layer model was used to mimic the superficial skin and brain layers. Low-frequency additive image-space noise was placed in the superficial layer to mimic systemic noise in some simulations. Finally, in the last examples the reconstruction of multiple loci of brain activity was examined.

#### *Wavelet reparameterization of DOT inverse model*

As noted above, the optical forward model is the product of a non-square sensitivity matrix  $H$  and the vector of unknowns  $\beta$ . The matrix  $H$  projects the hemoglobin concentration changes from points in the volume of tissue to the expected optical density changes measured at the surface by a particular grid of optical source-detector pairs. In a recent paper [19], we

described the use of wavelets as basis set on which to reconstruct optical images. The advantage of the wavelet-based reparameterization of the model is that the specific wavelet filter banks (low-pass, band-pass(es), high-pass) can be biased in the reconstruction as a means of imposing spatial smoothing on the reconstructed image. The wavelet transform can be thought of as a reparameterization of a signal or image in a sense similar to a Fourier transform. However, unlike the Fourier transform, wavelets allow localization in both spatial location and spatial frequency (or time and temporal frequency). By representing the model in the wavelet domain, the different levels of spatial frequencies can be separately regularized through the ReML method. Thus, in the same way that the choice of two  $\Lambda$ 's can be used to adjust the variance (regularization) in the skin and deeper layers, the wavelet model allows us to independently adjust the bias towards or away from a specific spatial frequency band. In the context of diffuse optical image reconstruction, we will show that the wavelet representation offers an ability to distinguish superficial physiology and evoked brain signals on the basis of a priori knowledge that these two signals compose different spatial characteristics. In our work in [19] we described a detailed wavelet model based on the extracted curvature of the cortex. In this current work, we will use a much simpler (conventional) set of wavelets as a means to better demonstrate the current model with less confusion. In our model, we will find  $\beta$  by first reparametrizing the model using orthogonal wavelet transform and then estimating the coefficients  $\beta_w$  of the wavelet transform of  $\beta$ . The orthogonal wavelet transformation is a reversible rotation which can be expressed in a matrix notation such that

$$\begin{aligned}\beta_w &= W \cdot \beta \\ \beta &= W^T \cdot \beta_w\end{aligned}\tag{17}$$

where  $W$  is the wavelet analysis model (transformation from image to wavelet space). Here,  $W^T \cdot W = I$ . In this work we use the Daubechies wavelet [20] generated by FIR orthogonal filters of length 2 coefficients and separable in the x-y (in-plane) dimension. Equation (7) is now restated for each layer in terms of wavelets as

$$Y = H \cdot W^T \cdot (\beta_w + \omega_w) + \nu\tag{18}$$

Equation (18) is the optical forward model in the wavelet domain. The parameter noise term ( $\omega_w$ ) is now also in the wavelet domain. Since we intend to use a structure to the covariance components, which allows a non-white spatial frequency distribution (particularly to model systemic physiological noise), we will define the covariance components of the ReML model directly in the wavelet domain. The structure of the wavelet matrix  $W$  is shown in Fig. 3. The low-pass, band-pass, and high-pass components map to regions of the matrix. Figure 3 shows the structure of this model for the one-dimensional with only 2 stages for simplicity. The actual model used a two-dimensional structure (in the x/y plane) with three stages.

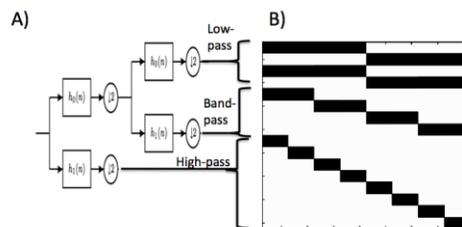


Fig. 3. The optical inverse model was reparameterized in terms of wavelet coefficients. In the wavelet representation, the original image is described as a linear combination of low-pass, band-pass, and high-pass filter banks (left; for 1-dimensional case). The wavelet transform can be implemented in matrix form, which has the same filter structure (right) and will be used to apply a frequency bias to the superficial and deeper layers of the reconstruction.

### Example Covariance Components

In the parametric Bayesian model, the covariance of the extended measurement vector is described as a weighted linear combination of covariance components, which capture specific a priori features of the model. In this section, we will detail the covariance components that will be discussed in the remainder of this paper.

*Minimum norm prior.* In order to compare against the Tikhonov regularization methods, the minimum norm (covariance) prior should take the form  $C_N = \Lambda_1 \cdot I$  and  $C_P = \Lambda_2 \cdot I$  where both oxy- and oxy-hemoglobin are modeled by a single covariance component and hyperparameter.  $C_N$  and  $C_P$  define the total noise model via Eq. (15). This produces the effect of a single hyperparameter to tune the model and is equivalent to the Tikhonov regularization model

$$\beta = (H^T \cdot H + \lambda \cdot I)^{-1} \cdot H^T \cdot Y \quad (19)$$

where the ratio of  $\Lambda_1$  and  $\Lambda_2$  from the Bayesian model are replaced by the single regularization term  $\lambda$ .

*Measurement noise prior.* In general, optical recordings will have different noise depending on the wavelength. While this is less of a concern for systems with only two measured wavelengths, combining more than two wavelengths into estimates of oxy- and deoxy-hemoglobin via the modified Beer-Lambert law requires an estimate of the noise at each wavelength leading to the weighted least-squares model. In order to model this,  $C_N$  is modeled by a separate component for each of the measurement types with unity values allow the corresponding diagonal elements for each wavelength type. Thus, the two-wavelength optical model, which we will use in this work, will have two hyper-parameters to define the measurement covariance ( $C_N$ ). While this paper is concerned with optical-only reconstructions, we note that this approach is amendable to multimodal data as well, for example, the joint image reconstruction of brain activity from concurrent optical and functional MRI data as shown in Huppert et al [21].

*Separation of HbO<sub>2</sub> and Hb.* One of the limitations of the Tikhonov approach is that the same level of variance is assumed for the entire parameter space, namely both oxy- and deoxy-hemoglobin. To model different noise levels for both oxy- and deoxy-hemoglobin, the  $C_P$  component of the covariance matrix can be modeled as a linear combination of two unity-diagonal covariance components corresponding to the two chromophores of the model. We can also impose additional statistical relationships between oxy- and deoxy-hemoglobin, such as the observation that these changes are often anti-correlated (e.g. the typical hemodynamic response involves oxy-hemoglobin increasing and deoxy-hemoglobin decreasing). This can be modeled by negative-signed off-diagonal elements to the covariance components, e.g.

$$Q_{HbO_2/Hb} = \begin{bmatrix} 0 & -I_{HbO_2/Hb} \\ -I_{HbO_2/Hb} & 0 \end{bmatrix} \quad (20)$$

Note that this term will be multiplied by a to-be-estimated hyperparameter, which will rescale the covariance.

*Depth specific spatial frequency priors.* As previously discussed, the reparameterization of the optical forward model via the wavelet transform allows statistical priors to impose relationships between the levels of spatial frequency. Namely, the variance of the corresponding low-pass, band-pass, and high-pass wavelet coefficients can be reweighted according to a priori assumptions, such as the expectation that superficial (systemic) signals will be low frequency. By weighting the variance between each frequency band, a covariance component acting as a low-pass filter can be constructed.

$$Q_{Low-pass} = \begin{bmatrix} I_{Low-pass} & & \\ & \frac{1}{\sigma^2} I_{Band-pass} & \\ & & \frac{1}{\sigma^4} I_{High-pass} \end{bmatrix} \quad (21)$$

The parameter  $\sigma$  is a smoothing factor. In principal, this term could be included in the list of hyper-parameters and solved, however, this would create a non-linear model and is beyond the current scope of this work. Here, we have used a fixed value of  $\sigma = 15\text{mm}$  ( $= 2.2$  voxel dimensions) in the model.

*Incorporating prior knowledge of location of ROI.* Finally, the covariance components of the ReML model can be used to impose a priori knowledge of regions-of-interest for the location of activation. Such prior information can be obtained for example from experience or from alternate modality such as functional MRI or atlas based priors. The resulting Q's can then be given by the diagonal matrices for HbO<sub>2</sub> and Hb:

$$Q_{ROI(i,j)} = \begin{cases} 1 & \text{if } \{i,j\} \text{ in region-of-interest} \\ 0 & \text{else} \end{cases} \quad (22)$$

#### 4. Results

In this section, we will demonstrate the utility of the ReML method through several examples from simplistic to complex. We will first show that the proposed EM approach produces nearly identical results to the Tikhonov model [Eq. (19)] optimized by the L-curve method in the trivial case of a single covariance component. From here, we will then show how additional covariance components can be used to add information about wavelength and hemoglobin specific noise. In the first few examples, we will demonstrate this model with a single-layered image. Later, we will switch to a two-layered model simulating the skin and brain layers. We will show that the ReML model is able to provide depth-dependent regularization in the case of either no superficial noise or the difficult case of spatially structured superficial noise. Through this discussion, we will gradually introduce components of the model, building to the final most complete model for the most difficult case. While we do this in order to emphasize each feature of the method, we do note that the final model we will describe performs equally well and in some cases better than the simple model used to demonstrate the earlier examples.

##### *Comparison of ReML and L-curve*

As an initial demonstration of the ReML method, we implemented a covariance model consisting of solely the minimum norm prior ( $C_N = \Lambda_1 \cdot I$  and  $C_P = \Lambda_2 \cdot I$ ). This model allows direct comparison to the L-curve approach to defining  $\lambda$  in Eq. (19) ( $\lambda = \Lambda_1/\Lambda_2$ ). A single layered image with a depth of 1cm was generated (16 x 16 x 1 voxels [6.7mm6.7xmmx10mm]) and a colocalized oxy-hemoglobin [ $1\mu\text{M}$ ; Fig. 4(A1)] and deoxy-hemoglobin [ $-0.25\mu\text{M}$ ; Fig. 4(B1)] perturbation was added. In Fig. 4, data was generated contrast-to-noise ratio of 100:1 by adding random zero-mean measurement noise and reconstructed using the ReML procedure [Fig. 4(A2) and 4(B2)]. A L-curve was generated and used to select the optimal regularization [Fig. 4(A3) and 4(B3)]. In Fig. 5, the same model is shown but for a lower signal-to-noise level of 5:1. In the higher noise simulations, background noise is more clearly pronounced in the reconstructed images. As expected for this trivial case of minimum norm (covariance) prior, the L-curve and ReML estimation routines produced quantitatively similar reconstructions of both oxy- and deoxy-hemoglobin at both 50:1 and 5:1 signal-to-noise levels. In Fig. 6, we further compare the performance of the L-curve and ReML models through a range of contrast-to-noise levels from 100,000:1 (little noise) to 1:10 (more noise than signal). Over the majority of this range, the two methods

agree closely with each other and the theoretical optimal parameter. At very low single-to-noise levels, the L-curve tended to overestimate the regularization, which was the result of numerical instabilities in finding the corner of the L-curve. Nevertheless, we concluded that the two approaches were comparable over a large range of noise. This result was actually expected since discussion in Mattout et al [10] details the theoretical relationship between these two methods.

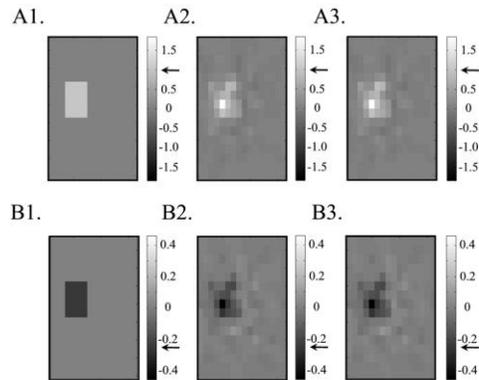


Fig. 4. This figure shows a comparison the ReML and L-curve tuned Tikhonov regularized reconstructions for simulations at low noise (signal-to-noise ratio of 100:1). In the top row (row-A), the original target (A1), the EM-reconstructed image (A2) and the L-curve reconstructed image (A3) of oxy-hemoglobin (+ 1 $\mu$ M simulated) is shown. In the bottom row (row-B) the original and reconstructed images of deoxy-hemoglobin (-0.25 $\mu$ M simulated). Notably, the ReML and L-curve techniques are nearly identical for this trivial case of only a single regularization hyper-parameter ( $\lambda = \Lambda_1/\Lambda_2$ ).

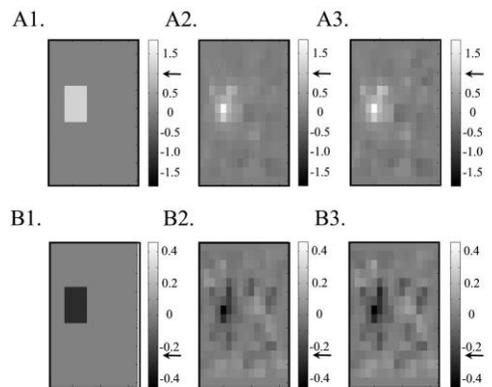


Fig. 5. This figure shows a comparison the EM and L-curve tuned Tikhonov regularized reconstructions for simulations at high noise (signal-to-noise ratio of 5:1). The definitions of the subplots are identical to Fig. 4.

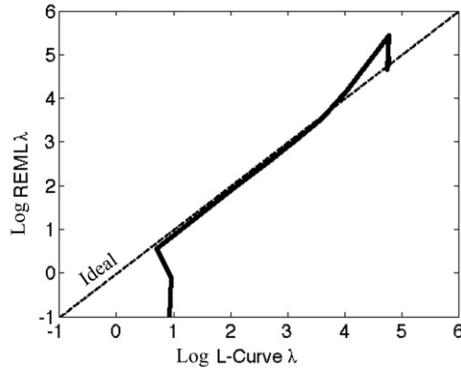


Fig. 6. In this figure, we compare the value of the hyperparameter ( $\lambda$ ) determined by the L-curve and ReML technique ( $\text{REML } \lambda = \Lambda_1/\Lambda_2$ ) for simulations with a contrast-to-noise ranging from 1:10 to 100,000:1 (half decade intervals). The L-curve and ReML techniques agree closely over this range implying that the ReML technique performs as well as the L-curve method for the trivial example of a single covariance component.

### *Incorporation of physiological priors*

As previously noted, one of the limitations of the single minimum norm regularization prior [Tikhonov prior; Eq. (19)] is that the same variance is assumed across all parameters. For example, oxy- and deoxy-hemoglobin maps will have the same level of regularization (minimum norm) penalty. In the context of DOT, measurements at different wavelengths are expected to have different levels of noise. However, the level of noise in each may not be known *a priori*. In addition, oxy- and deoxy-hemoglobin changes are also subject to different noise contributions from superficial and systemic physiology; e.g. cardiac pulsation which preferentially contributes to noise in oxy-hemoglobin. In order to account for this, the covariance of oxy- and deoxy-hemoglobin parameters can be independently estimated through the inclusion of separate covariance components for each. In the context of our current simulations, this introduces a total of four covariance components (one per each of the two wavelengths measured and one per oxy- and deoxy-hemoglobin across the image).

To demonstrate separate regularization for oxy- and deoxy-hemoglobin, a perturbation in oxy-hemoglobin (only) was added to the simulation as before and reconstructed using the ReML and L-curve models. Additive measurement noise was again simulated at a signal-to-noise ratio of 5:1. The reconstructions based on the Tikhonov prior [Eq. (19)] demonstrate this limitation of the approach [Fig. 7(A3) and 7(B3)]. In this reconstruction, the value of  $\lambda$  is selected via the L-curve method to provide reasonable reconstruction of the oxy-hemoglobin component. However, because this  $\lambda$  is also applied to the deoxy-hemoglobin component, the reconstructed deoxy-hemoglobin image [Fig. 7(B3)] shows significant noise and artifacts of similar magnitude to  $\lambda$ . In contrast, in the ReML method, because the regularization of oxy- and deoxy-hemoglobin is individually determined, a lower variance in the deoxy-hemoglobin model is adapted and the resulting artifacts are considerably lower. In the case of the EM model, the cross-talk in the deoxy-hemoglobin is close to negligible (<0.1%).

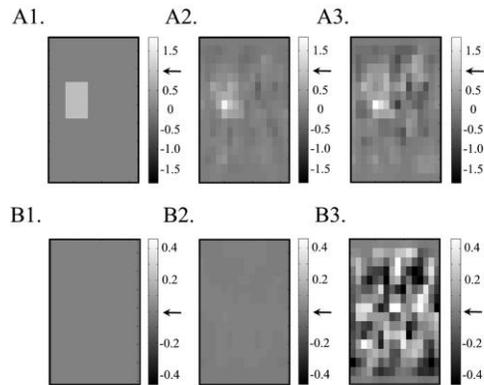


Fig. 7. In this figure, a perturbation in oxy-hemoglobin only (row A) was simulated. No deoxy-hemoglobin changes were simulated (row B). In the Tikhonov regularized inverse [Eq. (19)], which applies the same regularization factor to both the oxy- and deoxy-hemoglobin parameters, the L-curve technique (A3 and B3) gave a reliable estimate for oxy-hemoglobin, but this same level of regularization resulted in a very noisy deoxy-hemoglobin image. The ReML approach (A2 and B2) used separate hyperparameters to regularize the two hemoglobin species and resulted in close estimation of both images. Row B shows the deoxy-hemoglobin results.

As a further example of the incorporation of physiological priors into the optical inverse model, one of the common observations of evoked brain activity signals is that oxy- and deoxy-hemoglobin change in opposite directions. Typical of a change in blood flow, oxy-hemoglobin increases are accompanied by decreases in deoxy-hemoglobin. This can be statistically imposed upon the model by the inclusion of a further covariance component of the model consisting of negative covariance between the oxy- and deoxy-hemoglobin entries of the model. The advantage of this component is particularly realized in the case of data with a very low signal-to-noise ratio. In Fig. 8, we show the reconstructions for data simulated with a signal-to-noise ratio of 2:1 for the 830nm simulated measurements and 1:2 (more noise than signal) for the 690nm simulated measurements. This model used separate oxy- and deoxy-hemoglobin components [Fig. 8(A2) and 8(B2)] plus a component to explicitly model the covariance between oxy- and deoxy-hemoglobin [Fig. 8(A3) and 8(B3)]. This example is experimentally realistic since the body absorbs about twice as much light at the 690nm wavelength resulting in a generally lower signal-to-noise ratio at this wavelength. For this reason, the deoxy-hemoglobin estimate is generally higher in noise. By adding the additional covariance term, the oxy- and deoxy-hemoglobin changes are statistically related and thus the quality of both images improves by virtue of reducing much of the disinformation between the two images.

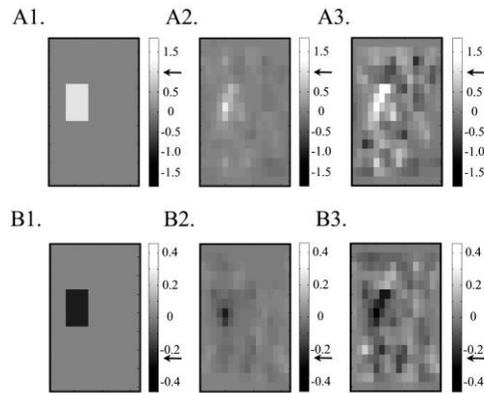


Fig. 8. In this figure, measurements were simulated to have a signal-to-noise ratio of 2:1 at the 830nm wavelength and only 1:2 at the 690nm wavelength. The resulting image reconstructions obtained via the ReML regularization using separate covariance components for oxy- and deoxy-hemoglobin (A3 and B3) was very noisy (as expected at this very low SNR). The noise in the images was reduced when a third covariance component modeling the negative-covariance between oxy- and deoxy-hemoglobin was also included (A2 and B2). Row A and B show the oxy- and deoxy-hemoglobin images respectively. Subplot A1 and B1 are the simulated (truth) images.

#### *Example of depth-specific regularization*

In all the previous examples, a single layered model was used in order to compare the ReML and L-curve approaches to optimal regularization. In this section, we used a multiple layered forward model (16 x 16 x 2 voxels [6.7 x 6.7 x 10 mm]) to mimic the layered structure of the head. Changes in oxy- and deoxy-hemoglobin were simulated in the deeper layer (10-20mm) and in this initial example, no additional image-space noise was added to the upper layer. Random measurement noise was added to generate a signal-to-noise ratio of 10:1. In order to account for depth-dependent (or region-dependent) regularization in the model, covariance components for each region are used. In our simulated example, a total of six covariance components are included (one per each of the two measured wavelengths; and 2 x 2 for oxy- and deoxy-hemoglobin in the top and bottom layers of the model).

In Fig. 9, we show the reconstructed images from the superficial (row A) and deep layers (row B). In Fig. 9(A2) and 9(B2), the reconstruction based on the L-curve tuned Tikhonov prior [Eq. (19)] is shown which applies the same regularization level to both layers. As anticipated, the minimum norm solution is heavily biased to the upper layer and results in an underestimation of the magnitude of the activation pattern. The image is over an order of magnitude diminished from the target. In Fig. 9(A3) and 9(B3), we show the reconstructions from the ReML approach using the four-covariance component matrices. Here, the reconstruction is correctly placed in the deeper layer, while the upper layer remains close to (but not entirely equal to) zero. One of the key features of this model is that, while in this case the reconstructed values are correctly near-zero in the upper layer, this is not inherent to the model. As we will demonstrate in the next section, the same models can also estimate non-zero superficial noise signals if supported by the data. In contrast, depth-dependent regularization has been demonstrated using cortically constrained solutions (e.g [22]). This approach assumes that the signal from the upper layer is zero and can produce accurate images provided that this assumption is correct. Namely, the cortically constrained model does not allow for non-zero superficial noise as any such noise would artificially projected into the brain layer. Our ReML reconstruction of the two-level model is quantitatively similar to a cortically constrained solution [Fig. 9(B4)] in which the superficial layer was masked and a L-curve tuned regularization was applied.

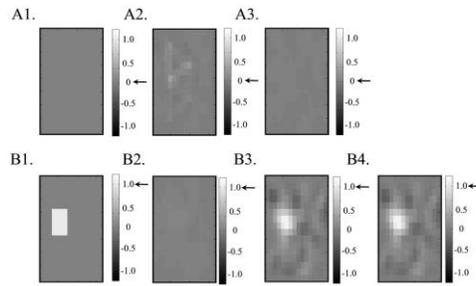


Fig. 9. In this figure, we compare reconstructions of the two-layered model. In rows A and B the superficial and deeper layers are shown. Only the oxy-hemoglobin results are shown. A perturbation was simulated only in deeper layer (B1). In A2 and B2, we show the reconstruction using a covariance component that spans both layers (akin to conventional Tikhonov regularization). Here, the same regularization is applied to both layers and the reconstructed image is heavily biased to the upper layer and underestimated. In A3 and B3, we show reconstructions using separate covariance components for the upper and lower level, which allows a total of four (2 layers x oxy- and deoxy-hemoglobin) hyperparameters to be estimated via the ReML method. This allows an empirically determined spatially distributed regularization of the model that results in correct placement of the reconstructed object in the bottom layer. This result is nearly identical to a cortically constrained reconstruction (B4) where the top layer is masked and only the bottom layer reconstructed.

#### *Contamination from superficial noise*

In this example, a low spatial-frequency noise image was added to the upper layer of our model in order to simulate superficial systemic physiology. Here, we will assume that the superficial image represents largely low spatial-frequency information and use this prior information to design the covariance components of the wavelet levels to impose a bias towards superficial low-frequency solutions. In this case, we model the covariance components such that they attenuate the spatial frequency-bands of the wavelet components in the two layers at different rates. Since it is known *a priori* that the upper layer has lower spatial frequency activities than in the lower layer, we impose covariance components that act as low-pass filters in this layer as described in Eq. (21). The wavelet coefficients for a given layer are increasingly attenuated as the spatial frequency increases. The two layers are assigned different attenuation rates, with the upper layer having  $\sigma = 2.2$  voxels (15mm), and the lower layer with  $\sigma = 1$  voxels (no low-pass filtering). This leads to a total of four covariance components, taking into account HbO<sub>2</sub> and Hb with the general form as given in Eq. (21).

Figure 10 depicts the reconstructed images with activities in the upper layer. Figures 10(A1) and 10(B1) show the images to be estimated. In Figs. 10(A2) and 10(B2) the image reconstruction based on the L-curve tuned Tikhonov prior is shown where, as in the above example, the same regularization level is applied to both layers. As to be expected, the minimum norm solution is heavily biased to the upper layer and results in underestimation of the magnitude of the activation pattern in the lower layer. In Figs. 10(A3) and 10(B3), we show the reconstructions from the EM approach using the four-parameter covariance component matrices described above. Here, the reconstruction correctly identifies and separates the upper from the lower layer where we have exploited the wavelet components' properties at different frequencies. As in the above example we next restrict our reconstruction to the cortically constrained solution [Fig. 10(B4)]. Clearly, this model does not allow for non-zero superficial noise, as any such noise would artificially be projected into the brain layer. This example makes it clear that in order to properly reconstruct the functional activities the superficial layer must be taken into account.

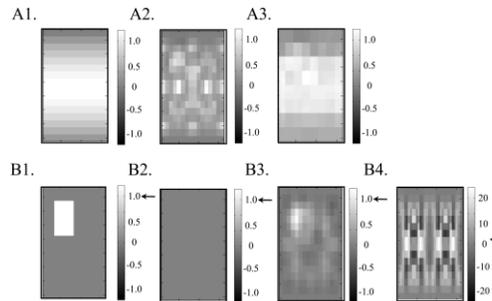


Fig. 10. In this figure, we compare image reconstructions in the case of a two-layered model with non-zero noise structure ( $+ 1\mu\text{M}$ ) in the superficial layer (A1). Row A shows the top layer and row B shows the deeper ("brain") layer. Only oxy-hemoglobin results are shown. In A2 and B2, we show the results using the reconstruction using the ReML approach with covariance components for the two layers but without any frequency bias (e.g.  $\sigma_1 = \sigma_2 = 1$  voxel). In A3 and B3, the reconstruction using a low-frequency bias in the top layer is shown ( $\sigma_1 = 2.2$  voxels and  $\sigma_2 = 1$  voxel). In B4, the reconstruction with a cortical-constraint is shown, which artificially pulls the superficial noise into the bottom layer and results in a grossly overestimated signal.

### *Incorporation of a priori region-of-interest information*

Incorporating a prior knowledge about a region of interest, when available, can be greatly helpful for accurate reconstruction of the image. If the region-of-interest was known with certainty, then this can be imposed with a hard-constraint allowing only these regions and nowhere else to possess non-zero values. However, the hard-constraint is rather trivial and unrealistic since most of the time, the region-of-interest is not known with absolute certainty. Using the ReML approach, regions-of-interest can be statistically imposed and the weight of this imposition can be empirically determined. In this first example, we will demonstrate the case where the a priori regions-of-interest are correct. In the second example, we will show that the model can also handle the more difficult case where the regions-of-interest are incorrect.

Using the covariance components given in Eq. (22), where the Q matrices describe the wavelet transform of the functional activities, a region-of-interest derived structure can be imposed on the covariance of the model. This leads to a total of two measurement noise ( $C_N$ ) and six parameter noise ( $C_p$ ) related hyperparameters. Four covariance components are used to model  $\text{HbO}_2$  and  $\text{Hb}$  in the top and bottom layers (same as previous example). The final two covariance components describe the region of interest for both  $\text{HbO}_2$  and  $\text{Hb}$  as given in Eq. (22).

We first consider an example where we have two active regions and with no activities in the upper layer as shown in Fig. 11. We consider the case where the regions-of-interest are near (2 voxels) and far (7 voxels) from each other. For comparison, we also reconstructed images using ReML with only covariance matrices for the  $\text{HbO}_2$  and  $\text{Hb}$  for each layer (as shown previously). Figure 11(A1) shows the case of the image to be estimated with two distinct regions of activities, while Fig. 11(B1) shows the case of two regions of interest located near each other. Figures 11(A2) and 11(B2) depict the reconstructed images for the case where only covariance matrices separating the layers and  $\text{HbO}_2$  and  $\text{Hb}$  for each layer. In Fig. 11(A2) the recovered image shows accurate recovery of the functional activities. In Fig. 11(B2) the spatially close regions of activities lead to a degree of ambiguity in the reconstructed image, where the boundaries of the two regions of activities tend to overlap. By using in addition covariance components describing the regions of interest for both oxy and deoxy-hemoglobin, visibly improved reconstruction is possible, as shown in Figs. 11(A3) and 11-B3 where in the latter it is possible to discriminate the two reconstructed activities.

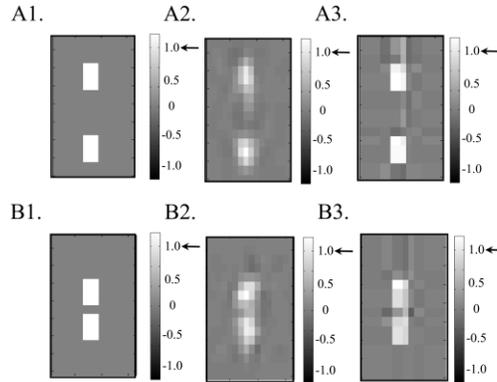


Fig. 11. In this figure, a two layer model with two perturbations placed either 6 voxels (40mm; row A) or 2 voxels (1.3 mm; row B) apart. Only the deeper layer and only the oxy-hemoglobin results are shown. In A2 and B2, we show the reconstructions using the ReML model without any specific region-of-interest priors. In A3 and B3, we show the reconstructions using a statistical region-of-interest prior. The arrows indicate the magnitude of the simulated values.

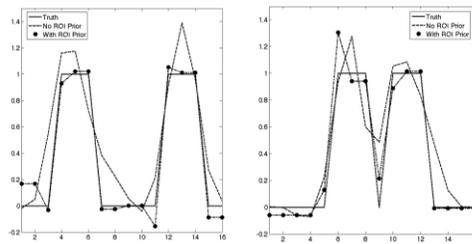


Fig. 12. Here, we show cross-sections of the reconstructions shown in Fig. 11.

Figure 12 shows cross sections of the reconstructed images in this example. Figure 12(A1) and 12(A2) depict the cross sections of the original image, the reconstruction without region of interest prior, and the reconstruction with ROI prior. It can be seen in Fig. 12(A2) that the addition of the ROI prior improves the separation of the two regions of interest.

As a final example, we consider the robustness of the ReML reconstruction scheme to incorrect regions-of-interest. A single perturbation was added to the target image [Fig. 13(A1)]. This data was then reconstructed using no region-of-interest information [Fig. 13(A2)], the correct region-of-interest [Fig. 13(B1)] and an incorrect region-of-interest [Fig. 13(B2)]. The correct and incorrect regions are also overlain on the target image [Fig. 13(A1)]. While clearly, the ReML scheme benefits from the prior region-of-interest information leading to an improved reconstruction, the method is not harmed by the incorrect region support. The case of false ROI depicted in Fig. 13(B2) still leads to a reconstruction comparable with that of Fig. 13(A2) where no region-of-interest prior was supplied. The ReML reconstruction scheme empirically learned not to use the false region-of-interest prior.

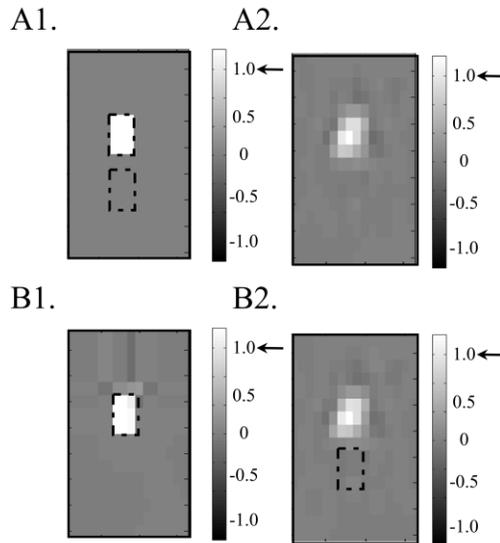


Fig. 13. In this figure, we demonstrate the effects of using an incorrect region-of-interest prior. The simulated true image (A1; SNR = 10:1) had a single perturbation in the second layer. The top layer and deoxy-hemoglobin results are not shown. In A2, we show the reconstructed image without any region-of-interest priors. In B1, we show the reconstructed image using the correct region-of-interest as a prior (prior is outlined in black). Finally in B2, we show the reconstruction using an incorrectly placed region-of-interest prior (outlined in black). Using the incorrect prior produced nearly identical results to the case in which no prior was used demonstrating that the ReML method correctly assigned a near-zero weight to the incorrect prior.

## 5. Discussion

In this study, we have described the application of restricted maximum likelihood (ReML) methods to determine the optimal weighting of multiple statistical priors into the reconstruction of diffuse optical imaging data. While this approach offers similar results to an optimally chosen regularization via traditional weighted minimum norm or multiple priors (e.g [13].) penalties, this new approach derives the optimal weights by an empirical algorithm that maximizes the log-likelihood of the data. The advantage of this approach is that it is fully objective in comparison to the manual tuning applied to many DOT reconstructions; particularly using multiple priors. This provides a means to introduce multiple priors or even select between multiple competing models. Even in the case where improper assumptions were made, such as the incorrect region-of-interest or presence of non-zero noise in the superficial layer, the EM model obtained accurate solutions.

It is important to recognize that we have not invented this technique anew for the application to DOT, but rather have merely adapted this method from a broad range of uses for the EM and ReML algorithm, including its application to functional MRI analysis in the program SPM (Statistical Parametric Mapping [11];) and for image reconstruction of magnetoencephalography (MEG) as demonstrated by Mattout et al [10]. Although the application of this method to DOT required a few modifications to account for mixed noise statistics inherent in optical measurements at multiple wavelengths and the estimation of multiple chromophores, this approach does not fundamentally differ from these previous papers.

This approach offers the potential for several extensions as a method to incorporate a range of prior information into the ill-posed inverse problem. One of the applications for this method is the field of diffuse optical mammography where the blood oxygenation and optical scattering properties of breast tissue are estimated by a similar inverse model to the one demonstrated here for the case of brain imaging (see [23] for review). The ReML algorithm

can be extended to the non-linear case (an iteratively linearized inverse model is usually used in optical mammography) and used to statistical priors about structure from MRI or x-ray images. While Li et al [13] demonstrated the utility of multiple such priors, they used a manual adjustment of the weights which could now be replaced with the proposed ReML method outlined here for optimal selection. Using this approach, physiological priors can also be introduced such as a priori expectations to the blood oxygen saturation levels or relative magnitudes of changes (similar to the imposition of a negative correlation between oxy- and deoxy-hemoglobin offered here).

A second application for this method is the analysis of multimodal data. In Huppert et al [21] we described pseudo-Bayesian joint-reconstruction for concurrent optical and functional MRI data, showing that the high temporal resolution of the optical data could be combined with the spatial information of the MRI to generate movies of brain activity matching the benefits of both modalities. In that early work, however, the measurement and parameter covariance were modeled using a priori estimates of these terms. Now, using ReML, the relative noise of the two modalities can be empirically estimated and an optimal combination of the two modalities can be fused to yield the same sorts of movies.

In conclusion, we have shown that the EM and ReML methods offer an empirical method for optimizing the inclusion of multiple prior information in the reconstruction of images from diffuse optical tomography.

### **Acknowledgements**

This work was funded by the University of Pittsburgh Department of Radiology and the National Institutes of Health (NIH) (R21NS064457).