Robust inference of baseline optical properties of the human head with 3D segmentation from magnetic resonance imaging

Alex H. Barnett[†,*], Joseph P. Culver[†], A. Gregory Sorensen[†], Anders Dale[†], and David A. Boas[†]
[†] *NMR Center, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA 02129*
[*] *Courant Institute, New York University, 251 Mercer Street, New York, NY 10012*

We model the capability of a small (6-optode) time-resolved diffuse optical tomography (DOT) system to infer baseline absorption and reduced scattering coefficients of the tissues of the human head (scalp, skull and brain). Our heterogeneous three-dimensional diffusion forward model uses tissue geometry from segmented MR data. Handling the inverse problem via Bayesian inference, and introducing a realistic noise model, we predict coefficient errorbars in terms of detected photon number and assumed model error. We demonstrate the large improvement that an MR-segmented model can provide: 2–10% error in brain coefficients (for 2×10^6 photons, 5% model error). We sample from the exact posterior, and show robustness to numerical model error. This opens up the possibility of simultaneous DOT/MR for quantitative cortically-constrained functional neuroimaging.

I. INTRODUCTION

Diffuse optical tomography (DOT) [1] is arousing growing interest as a noninvasive tool for neuroimaging, both clinical [2] and functional [3, 4], as well as other clinical applications such as breast tumor detection [5–8], tracking muscular oxygenation [9] and arthritic joint imaging [10]. By making use of spectroscopy, this modality can couple directly to hemodynamic quantities of interest, measuring both total hemoglobin concentration and oxygenation. Optical contrast can also arise from cell-based mechanisms such as the redox state of cytochrome-*c*-oxidase, and *in vivo* contrast agents such as fluorescent and voltage-sensitive dyes [4]. By relying on photon transport through tissue, DOT also accesses spatial information to a depth of several centimeters. DOT shows great promise for functional neuroimaging: the advantages of functional DOT (fDOT) include fast temporal resolution of order 10 ms, and moderate spatial resolution of order 1–2 cm, placing it in these two categories somewhere between both magneto-encephalography (MEG) and electroencephalography (EEG), and functional magnetic resonance imaging (fMRI). In addition to sensitivity to absorbing chromophores, DOT can be sensitive to cellular scattering changes during neuronal activation [4, 11, 12]. DOT is noninvasive and non-ionising. The apparatus is relatively inexpensive, is compatible with (and complementary to) other modalities allowing simultaneous imaging [13, 14], is portable, is capable of continuous monitoring, and does not require an immobile subject.

The key challenge of DOT is that of extracting spatial maps of the optical properties (absorption coefficient μ_a, and reduced scattering coefficient μ_s') within a highly-scattering tissue volume, by coupling multiple light sources and multiple detectors to the surface of the skin. The available signals comprise light intensities at each detector due to each source. For each of these source-detector pairs, it is possible to measure a DC intensity (continous-wave systems), or intensity amplitude and phase (RF-modulated systems), or obtain the intensity distribution as a function of time-of-flight (time-domain systems). Because tissue dimensions of interest are much larger than the photon mean free path, the path taken by photons from source to detector is diffusive rather than straight; this is what limits the spatial resolution of DOT. Recovery of $\mu_a(\mathbf{r})$ and $\mu_s'(\mathbf{r})$ from measured signals requires solving an inverse problem, non-linear in the optical parameters, and known to be ill-posed [15, 16]. Usually $\mu_a(\mathbf{r})$ is measured at several wavelengths within the near-infrared low-tissue-absorption window of 650–900 nm. From this, the known spectral absorption curves [17] of the chromophores HbR and HbO_2 allows spatial maps of [HbR] and [HbO_2] (and therefore of total hemoglobin concentration) to be constructed.

Our goal in this paper is to explore the limits to accuracy in measuring *baseline* optical properties of the human head using time-domain apparatus, when high-resolution anatomical information from MR is used to constrain the optical model. Absolute cerebral oximetry and blood volume measurement, which require baseline (absolute) quantification of cortical μ_a are important for study of stroke [18], head trauma, migraine, neonatal ischemia [19] and brain development [20]. Baseline properties are important for another reason. For functional DOT it is common to use a perturbation model where signal changes are taken to be a linear function of optical parameter changes. This allows rapid imaging via matrix inverse methods [15, 21–23]. However, one cannot [24] compute the required sensitivity functions without knowing both baseline (absolute) μ_a and μ_s'. We envisage using simultaneous MR and DOT to perform accurate baseline measurements, followed by cortically-constrained fDOT imaging and simultaneous BOLD (blood oxygenation level dependent) fMRI [3, 14].

In our model the scalp, skull, brain, and cerebro-spinal fluid (CSF) are assigned separate optical properties, assumed uniform within each tissue type, and their three-dimensional (3D) geometry is taken from automatically-

segmented MR images. This greatly reduces N the number of unknowns (we will use $N = 6$, compared to $N \sim 10^2$–10^4 common in pixel- or voxel-based representations), making the inverse problem much more tractable. This idea of MR-constrained reconstruction is not new to DOT [25], certainly in two dimensions [26, 27], nor to other medical inverse problems [28, 29]. Note that our approach differs from the class of segmentation methods where tissue boundaries are themselves parameterized by unknowns [30, 31].

Time-resolved intensity measurement is popular for extracting baseline properties in simple homogeneous [32–34] and layered [35–38] slab tissue models, in both transmission and reflection geometries. Time-of-flight information resolves otherwise indistinguishable effects of absorption and scattering changes [15]. For example, Kienle [39] uses a 2-layer analytic diffusion model to fit for optical coefficients of *in vivo* arm muscle, with the superficial fat layer thickness constrained by another imaging modality (in this case ultrasound). For the adult head *in vivo* measurements have usually also been analysed with simple analytical diffusion models, even though the internal geometry may be poorly represented as a slab or layered slab.

To overcome these limitations, we use a numerical forward model which can handle arbitrary 3D tissue geometries and optode locations, in the diffusion approximation. We use this to perform 'model-based' fitting via nonlinear optimization [15, 40, 41]. We use the Bayesian paradigm for inference, giving us the full posterior *probability density function* (PDF) of the parameters (which in turn could be used to get the full PDF of baseline [HbR] and [HbO_2]). Bayesian inference has been used with success in related medical inverse problems [29, 42–44]. However, fully-Bayesian [41, 45, 46] or Bayes-influenced [47] applications in DOT are less common. In contrast to most previous work we use the posterior PDF to study expected errors (*i.e.* errorbars) and the detailed distribution in the parameters, rather than presenting a single 'best' solution. We also introduce a novel realistic noise model which captures our belief in signal reliability at both small and large numbers of detected photons. This accounts for photon detection noise and forward modeling errors; the latter have been ignored in the Bayesian DOT literature thus far.

The paper is organized as follows. In Section II we outline the application of Bayesian inference to our problem, and introduce the general form of the likelihood function. In Section III we present the head system, and the forward and noise models used for inference and simulation of noisy experimental signals. In Section IV we give results showing the accuracy of inference achievable using a homogeneous head model and a segmented head model. We present achievable errorbars as a function of total collected photons and assumed model errors. We investigate in detail the posterior distribution in the optical parameters, using a PDF sampling method, and validate our choice of posterior approximation. We also demonstrate robustness to forward model errors. We conclude and suggest future directions in Section V. Two appendices contain methodological details which would otherwise impede the flow of the main text.

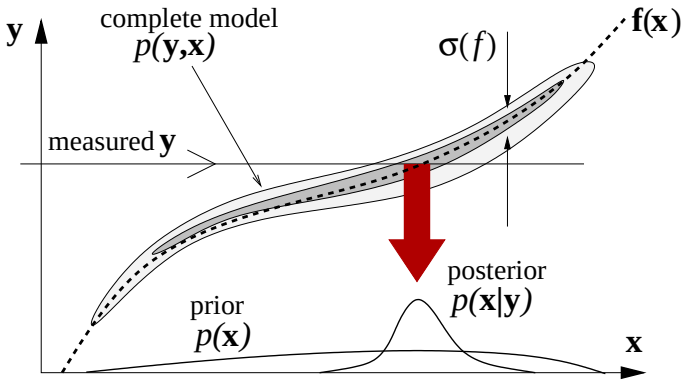

FIG. 1: Illustration of Bayesian inference of unknown parameters $\mathbf{x}$ given the measured signal vector $\mathbf{y}$. The complete model comprises a forward model $\mathbf{f}(\mathbf{x})$ (dotted line) and an inference noise model, giving the joint PDF $p(\mathbf{y}, \mathbf{x})$ which can written as a prior $p(\mathbf{x})$ multiplied by a likelihood $p(\mathbf{y}|\mathbf{x})$. The inference noise model describes all measurement and model errors; here we use independent Gaussian noise with signal-dependent width $\sigma(f)$.

II. INFERENCE FRAMEWORK

The Bayesian approach treats an inverse problem as an inference problem: our lack of knowledge about model parameters is represented by a PDF over those parameters. A model represented by $\mathcal{H}$ (for "hypothesis") contains unknown parameters given by the vector $\mathbf{x} \equiv \{x_1 \cdots x_N\}$, and includes a prior PDF on those parameters $p(\mathbf{x}|\mathcal{H})$. We consider how this PDF is modified by the arrival of the experimental data (measured signal) vector $\mathbf{y} \equiv \{y_1 \cdots y_M\}$, to give a posterior PDF

$$p(\mathbf{x}|\mathbf{y},\mathcal{H}) \;\propto\; p(\mathbf{y}|\mathbf{x},\mathcal{H})\, p(\mathbf{x}|\mathcal{H}), \tag{1}$$

where $p(\mathbf{y}|\mathbf{x},\mathcal{H})$ is the *likelihood* defined by model $\mathcal{H}$, and the constant of proportionality depends only on $\mathbf{y}$. The posterior encodes everything we now know about $\mathbf{x}$. When we are not distinguishing between different models, we will drop the conditionality on $\mathcal{H}$ in our notation.

The relevant advantages of this framework include: 1) Explicit formulation of all assumptions. Assumptions are present in *all* approaches to inverse problems but are not always expressed. 2) The ill-posed nature of the inverse problem, implying uncertainty in some parameter directions, is embraced and handled probabilistically. This contrasts with traditional approaches where this is viewed as 'unstability', and therefore removed by *ad hoc* regularization methods. 3) The posterior PDF contains all confidence intervals and error correlations of parameters. In the case of hemoglobin concentration param-

eters, their posterior could be used to infer neural activation (via fitting models of neural-hemodynamic coupling [48]), with more reliability than 'best-fit' parameter values alone. 4) Probabilistic prior information, for instance from another imaging method, can be consistently made use of, allowing optimal use of multimodal imaging information. 5) The Bayesian method is optimal in the sense of giving unbiased minimum-variance estimators when the likelihood and prior are correct. 6) Complex models (with many additional, or hidden, parameters, such as amplitude calibrations [49]) can be handled without the danger of overfitting that can occur using simple cost-function minimization. 7) Competing models $\mathcal{H}_i$ can be ranked by their ability to explain the data, even if they have very different structure and numbers of parameters. Offsetting these advantages is the main obstacle to use of Bayesian methods: they can be very computationally intensive compared to more *ad hoc* approaches. For an introduction see the informal tutorial by MacKay [50], or other more detailed reviews [51–55].

We take a forward model $\mathbf{f}(\mathbf{x})$ which encapsulates our physical model of the DOT system: given an optical parameter vector $\mathbf{x}$ it returns the expected signal vector $\mathbf{f} \equiv \{f_1 \cdots f_M\}$. We also need a noise model, giving the probability that a signal $\mathbf{y}$ could be generated by adding noise to the noise-free signal $\mathbf{f}$. We use an independent Gaussian noise model, giving the likelihood function

$$p(\mathbf{y}|\mathbf{x}) \propto e^{-L(\mathbf{y};\mathbf{x})}, \tag{2}$$

with the negative log likelihood

$$L(\mathbf{y};\mathbf{x}) = \frac{1}{2}\sum_{m=1}^{M} \ln \sigma_m^2 + \frac{1}{2}\sum_{m=1}^{M} \frac{(f_m(\mathbf{x}) - y_m)^2}{\sigma_m^2}. \tag{3}$$

Each noise standard deviation $\sigma_m \equiv \sigma(f_m(\mathbf{x}))$ is given by the same function applied to the corresponding signal level $f_m(\mathbf{x})$. The form of $\sigma(f)$ will be discussed in Section III D. The second term is analogous to a χ^2 cost function. The constant of proportionality and the log term arise from the normalization requirement $\int d^M\mathbf{y}\, p(\mathbf{y}|\mathbf{x}) = 1$. Our inference procedure is shown schematically in Fig. 1.

For this work we have used a constant prior $p(\mathbf{x})$ within certain biologically-motivated bounds on each parameter $x_n \in [x_{n,\min}, x_{n,\max}]$, and zero outside (in other words, a standard min-max prior). This reflects a lack of preference over $\mathbf{x}$ within these bounds, and zero belief in values outside the bounds. Therefore the posterior is

$$p(\mathbf{x}|\mathbf{y}) \propto \begin{cases} e^{-L(\mathbf{y};\mathbf{x})} & x_n \in [x_{n,\min}, x_{n,\max}] \ \ \forall n, \\ 0 & \text{otherwise}, \end{cases} \tag{4}$$

where we will not need to know the constant of proportionality. It would be simple to encode more specific prior beliefs than this simple constraint.

For N more that 2 or 3 the high dimensionality, and the fact that each evaluation of $\mathbf{f}(\mathbf{x})$ requires a time-consuming solution of a PDE, make it impossible in general to characterize the posterior over all $\mathbf{x}$. However, for differentiable $\mathbf{f}(\mathbf{x})$, there will always be a sufficiently small choice of $\sigma(f)$ that the bulk of the posterior probability mass will be close to the *maximum a-posteriori* (MAP) value $\mathbf{x}_{\text{MAP}}$, and will be well approximated by a multivariate Gaussian with covariance matrix Σ_{MAP} and mean $\mathbf{x}_{\text{MAP}}$ [52]. The inference task is then to locate $\mathbf{x}_{\text{MAP}}$ and measure Σ_{MAP}, from which confidence intervals on each x_n can be computed. In all but Section IV D we use this approach; details are presented in Appendix A.

We note that this search for $\mathbf{x}_{\text{MAP}}$ is analogous to model-based approaches where an objective function (cost function) $L(\mathbf{y};\mathbf{x})$, in our case differing from the weighted least squares form only by an additional log term, is minimized by an iterative approach [15, 56]. Indeed much of the machinery is in common. However the interpretation is different: those who use model-based approaches are generally interested in the single 'best' solution [41, 57, 58], rather than properties of the full posterior PDF. (Often this restriction is a practical one due to large N).

When the posterior is significant over a region where $\mathbf{f}(\mathbf{x})$ is no longer linearizable, the Gaussian approximation becomes bad. In this case we use Markov chain Monte Carlo (MCMC) sampling [55] to generate a set of uncorrelated samples from the exact posterior, providing a (noisy) cloud-like impression of the full PDF. We demonstrate this more general but computationally-intensive method in Section IV D.

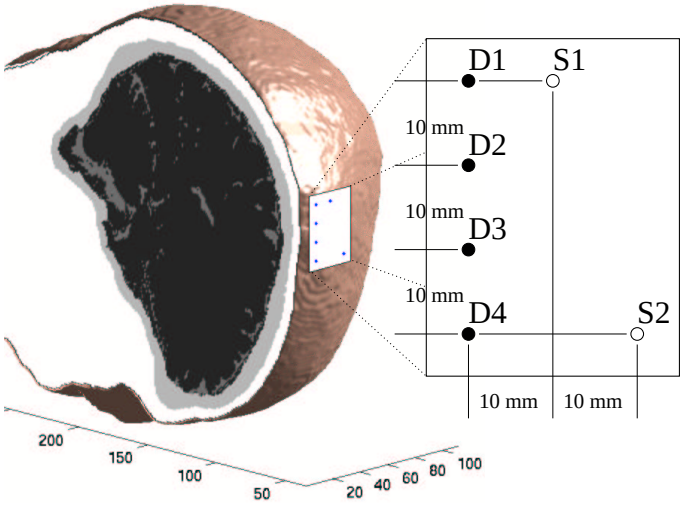

FIG. 2: Simulated optode arrangement and placement on the head (the face, pointing downwards, is mostly hidden). 3D MR-segmented head geometry is exposed by a sagittal slice. The tissue type color coding on this slice, is, from lightest to darkest: scalp, skull, CSF, brain.

III. FORWARD AND NOISE MODEL

A. The head system

We start with a human head segmentation geometry obtained at a 1 mm by 1 mm by 1.5 mm voxel resolution from a MR structural image using the technique of Dale, Fischl and Sereno [59]. This data consists of an assignment of each voxel to one of the five categories: scalp, skull, CSF, brain, or 'not tissue' (air). (We did not distinguish between grey and white matter; grey matter will dominate the optical properties because of the rapid decay of sensitivity with depth). For the subject data used the scalp and skull thicknesses were typical, being 5–7 mm and 7–8 mm respectively, in the relevant region of the head. Fig 2 shows our head geometry. Unless otherwise stated, we will use as our 'standard' optical properties for the four tissue types the values given in Table I. These are believed to be typical, falling within the quite wide variation of published values [33, 60, 61].

The CSF falls into a special category. It is presumed to be about 10^2 times less scattering and less absorbing than other tissue types, and occupies a much smaller volume. Its structure is convoluted and uneven, varying from a thin layer of about 1–2 mm thickness hugging the dura, to folds and pockets of several mm in thickness and size, following the folds of the brain surface [62]. Once the CSF geometry is held fixed, realistic biological variation in either CSF μ_a or μ_s' causes negligible changes in the photon tranport, and hence detected signals. For μ_a, this is because the values and volume are too small to cause any significant absorption of light compared to that caused by surrounding tissues.

For μ_s', the reason is subtely different: the extremely long reduced scattering length of $1/\mu_s' \approx 100$ mm is much longer (by a factor of order 5) than maximal line-of-sight distances within the CSF void region. This is mainly due to the folded and irregular geometry, and is most likely enhanced by surface roughness [63] and the presence of vasculature in the CSF below the MR resolution. Thus typical free transport distances in the CSF are dominated by the length-scale of the irregular folds and vessels. We believe that photons in the CSF pass into other highly-scattering tissue types long before the particular value of μ_s' becomes relevant.

Therefore we will not try to infer CSF properties in our segmented model, and our unknowns will consist of μ_a and μ_s' for the remaining 3 tissues. We will compare two models:

- $\mathcal{H}_{\text{hom}}$: the interior of the head is assumed *homogeneous*, with $N = 2$ unknown parameters $\mu_{a,\text{hom}}$ and $\mu'_{s,\text{hom}}$ describing a single tissue type
- $\mathcal{H}_{\text{seg}}$: the interior of the head is segmented as described above, with $N = 6$ unknown parameters $\mu_{a,\text{scalp}}$, $\mu'_{s,\text{scalp}}$, $\mu_{a,\text{skull}}$, $\mu'_{s,\text{skull}}$, $\mu_{a,\text{brain}}$ and $\mu'_{s,\text{brain}}$. Note that the CSF is present in this model but has fixed parameters.

The unknown (or fitting) parameters were given bounds $[x_{n,\min}, x_{n,\max}]$ of $[0.004, 0.04]$ mm^{-1} for all absorption, and $[0.4, 4.0]$ mm^{-1} for all reduced scattering coefficents.

tissue	μ_a (mm^{-1})	μ_s' (mm^{-1})
scalp	0.0149	0.8
skull	0.01	1.0
CSF	0.0004	0.01
brain	0.0178	1.25

TABLE I: Standard set of optical properties of human head tissue types used in this study. The tissue type 'brain' includes grey and white matter but is optically dominated by grey matter.

B. Optodes and detection model

We have chosen to model a 2-source ($N_s = 2$) and 4-detector ($N_d = 4$) fiber-coupled time-resolved system contacting a small region of the scalp at the top of the head. The arrangement, shown in Fig 2, was chosen to cover a range of source-detector distances 10–36 mm. We assume the system counts photons, and bins these counts according to their arrival time, building up a histogram of the temporal point spread function (TPSF). We assume no convolution of the TPSF due to the measurement system (*i.e.* instrument response function), but this would be easy to include in the forward model.

For simplicity, we assume optode locations are known accurately, and that the system has been calibrated to provide absolute measurements of amplitude and of time-of-flight. (We do not model the variable attenuation which is often used [7, 64] to prevent detector saturation at small optode separations).

One TPSF is available for each of the $N_s \times N_d = 8$ source-detector pairs. We represent each TPSF by 18 numbers, giving the photons collected in each of the 100 ps wide time-gates in the time-of-flight range 0.2–2.0 ns. We found that for time-of-flights below 0.2 ns, diffusion forward model errors were too large to be useful [65]. Beyond 2.0 ns the signals generally fall below the one-photon level. An example simulated signal vector $\mathbf{y}$ is shown in Fig. 3. Note that our signals are presented in units of detected photons.

C. Diffusion forward model

Photon transport in scattering media at macroscopic scales, where interference effects can be ignored, is described by the transport equation [15, 66]. In media where $\mu_a \ll \mu_s'$, and length-scales of interest are much

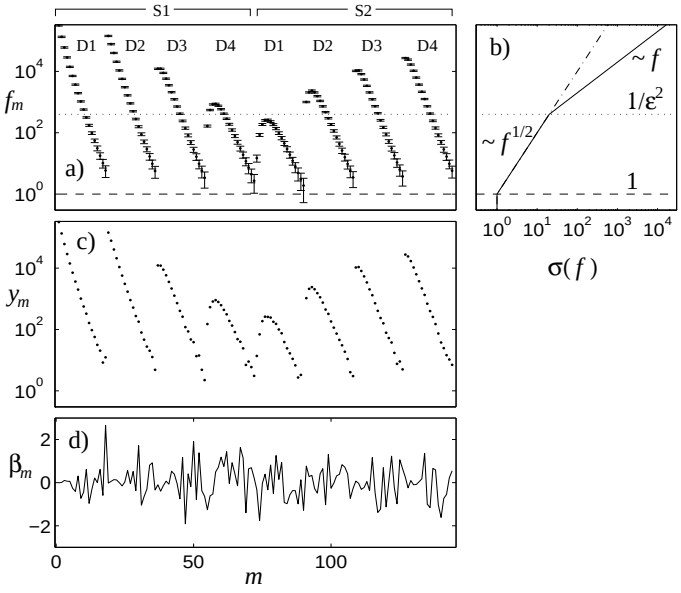

FIG. 3: Signals and noise models. a) shows typical signal expectation vector $\mathbf{f} \equiv \{f_m\}$ for $m = 1 \cdots M$, corresponding to $N_p = 1.1 \times 10^6$ detected photons, with errorbars representing the noise standard deviation σ used for inference. The 8 time-courses correspond to the interval 0.2–2.0 ns, and are labelled according to source and detector number. The one-photon level is shown dashed; the level for cross-over to fractional-error dominated noise at $1/\epsilon^2$ is shown dotted, for $\epsilon = 0.05$. b) shows (solid line) the inference noise $\sigma(f)$ given in Eq. 6, and (dash-dotted) the simulated experimental noise $\sigma_{\rm sim}(f)$ given in Eq. 7, plotted horizontally in order to share the same vertical scale as a). c) shows a simulated noisy signal vector $\mathbf{y}$, generated by adding Gaussian noise of size $\sigma_{\rm sim}(f)$ to the expectation $\mathbf{f}$, on same horizontal axis as a). d) shows the residual β_m (in standard deviation units, see Eq. A3) which would result from the $\mathbf{y}$ and $\mathbf{f}$ shown, on same horizontal axis as a).

larger than $1/\mu_s'$, transport can be well described [15] by the diffusion approximation (DA),

$$\frac{1}{v}\frac{\partial}{\partial t}\phi \;=\; \nabla \cdot (\kappa \nabla \phi) - \mu_a \phi + q, \qquad (5)$$

where $\phi(\mathbf{r}, t)$ is the fluence, $q(\mathbf{r}, t)$ is the source term, $\kappa(\mathbf{r})$ is the local diffusion constant, $\mu(\mathbf{r})$ is the local absorption, and $v(\mathbf{r})$ is the local speed of light in the medium. We will use the good approximation [67, 68] $\kappa = 1/(3\mu_s')$, and assume $v(\mathbf{r}) = c$ everywhere corresponding to unity refractive index. Our segmented map of μ_s' is given by assigning $\mu_s'(\mathbf{r}) = \mu_{s,\text{tissue}(\mathbf{r})}'$ where 'tissue($\mathbf{r}$)' represents the tissue type of the voxel within which $\mathbf{r}$ falls. Note that the resulting discontinuities in optical properties can cause discontinuities in $\nabla\phi$, but not in the value of ϕ itself (this would not be the case had we allowed refractive index changes). For tissue-air boundary conditions we have found the Dirichlet ($\phi = 0$) approximation to be sufficiently close to the more accurate Robin boundary condition [15], because κ is much smaller than the thicknesses of interest.

We use a Finite-Difference Time-Domain (FDTD) [69] method to compute the evolution of Eq. 5, from which the signal expectation vector $\mathbf{f}(\mathbf{x})$ is extracted using the detector locations. Details are given in Appendix B. The method represents the smooth function $\phi(\mathbf{r}, t)$ as discrete values on a cubical lattice of nodes, at a sequence of time steps. We resample the segmentation voxel map to the required lattice resolution; in this work either 2 mm or 1 mm. For a 2 mm voxel size, with 3.3×10^4 nodes in the required head volume, our current FDTD method takes about 8 seconds per source (on a 1GHz CPU) to simulate 2 ns of propagation time. We have not optimized the forward model, and there exist refinements of FDTD that can be much faster (see Appendix B). Typical error sizes at 2 mm can be judged from Fig. 9 (keep in mind that these signals span close to 6 orders of magnitude).

A note is necessary to explain how we treat the CSF. In recent years, work has been done which shows the large fluence errors which can result when the extremely low μ_s' value in void-like regions is fed directly into the DA [67]. (These errors were measured by comparison with transport equation solutions). We have chosen a different tactic, similar to that of Ripoll *et al.* [63]: by giving the CSF an *effective* μ_s' for use within the DA, we have been able to approximate the physics much better than possible using the true μ_s'. As discussed in Section III A, line-of-sight distances in the CSF are small, we believe of order $l \sim 10$ mm, due to highly irregular geometry and vasculature. We believe the optimal DA choice (which may vary subject to subject) is a μ_s' of order l^{-1}. Our preliminary results suggest that the fluence field is not very sensitive to the exact μ_s' chosen, when the full 3D MR head geometry is modelled. In contrast, most previous comparisons have used idealized 2D CSF geometries with long lines of sight [60, 67, 70], or 2D models taken from a single MR slice [67]. In this paper we fix $\mu_{s,\text{CSF}}' = 0.4$ mm^{-1}. The largeness of this choice is in part influenced by numerical efficiency: the CPU time for our current FDTD scales inversely with the smallest μ_s' in the system, which is always $\mu_{s,\text{CSF}}'$ in our case.

D. Noise model for inference

The forward model provides the signal expectation value vector $\mathbf{f}(\mathbf{x})$. Bayesian inverse methods always require a noise model, which captures our uncertainties about the difference between the numerically-generated $\mathbf{f}(\mathbf{x})$ and the real-world signals $\mathbf{y}$, given that the parameters $\mathbf{x}$ were actually correct.

Recall that in Eq. 3 we have assigned Gaussian distributions to each of the components $m = 1 \cdots M$ independently, each with mean f_m and standard deviation $\sigma(f_m)$. The Gaussian approximation to the Poisson photon detection statistics [71] would give $\sigma^2(f) = f$. Remember that signal units are in (effective) photons per time-gate. This approximation is good for $f \gg 1$, becoming bad (especially in the tails) as f approaches 1. To handle

sensibly the case $f < 1$ we place a lower limit of 1 on σ. (This limit could be increased to account for additive electronic noise).

However there is a problem at large signals: if some f_m is say 10^6, giving $\sigma_m = 10^3$, then we cannot be expected to believe that the fractional error σ_m/f_m, and therefore our trust in our forward model's closeness to reality, is 0.1% ! Rather we choose a fraction ϵ which is a lower limit on our fractional error, and takes account of

1. modeling of physics of photon migration in a given tissue geometry (*e.g.* transport corrections to the DA),
2. deviations from the assumption of uniform optical properties within a given tissue type,
3. incorrect tissue-type assignments from automatic MR segmentation,
4. the accuracy with which absolute amplitude (and time-offset) calibration can be carried out, including effects of optode-skin coupling variations due to *e.g.* skin pigmentation.

We typically choose $\epsilon = 0.05$–0.2, corresponding to 5%–20% model and calibration error.

Combining the above, the noise level $\sigma(f)$ is piecewise power-law:

$$\sigma(f) \;=\; \begin{cases} 1, & f \le 1, \\ f^{1/2}, & 1 < f \le 1/\epsilon^2, \\ \epsilon f, & f > 1/\epsilon^2. \end{cases} \qquad (6)$$

This is shown by the solid line in in Fig 3b. Typically the peaks in f are fractional-error dominated, while the tails are Poisson-statistics dominated.

E. Simulating experimental signals

In this proof-of-concept study, we generate *simulated experimental signals* using numerical forward models of the same type as used for the inference (inverse) problem. Given a true optical parameter vector $\mathbf{x}^{(0)}$ we generate a noise-free signal expectation vector $\mathbf{f}(\mathbf{x}^{(0)})$, to which we apply a *simulated* noise model.

Our simulated noise is identical to our inference noise (as described in the previous Section), except it only includes the detection noise component. We do not explicitly add noise to simulate model error; rather in Section IV E we will investigate the effects of model error by changing the forward model used to calculate $\mathbf{f}(\mathbf{x}^{(0)})$.

This gives the simulated noise level,

$$\sigma_{\rm sim}(f) \;=\; \begin{cases} 1, & f \le 1, \\ f^{1/2}, & f > 1. \end{cases} \qquad (7)$$

This is shown by the dash-dotted line in Fig 3b. Note that the difference between simulated and inference noise models manifests itself in Fig 3d as a residual β_m (see Eq. A3) which has a variance $\ll 1$ for the m values where $f_m \gg 1/\epsilon^2$.

The recipe for simulating experimental signals is,

$$y_m \;=\; \max\bigl[f_m(\mathbf{x}_0) + n_m, 0\bigr], \qquad m = 1 \cdots M, \quad (8)$$

where each n_m is independently sampled from a univariate Gaussian distribution with zero mean and variance $\sigma^2_{\rm sim}(f_m(\mathbf{x}_0))$. The maximum value operation removes unphysical negative signals.

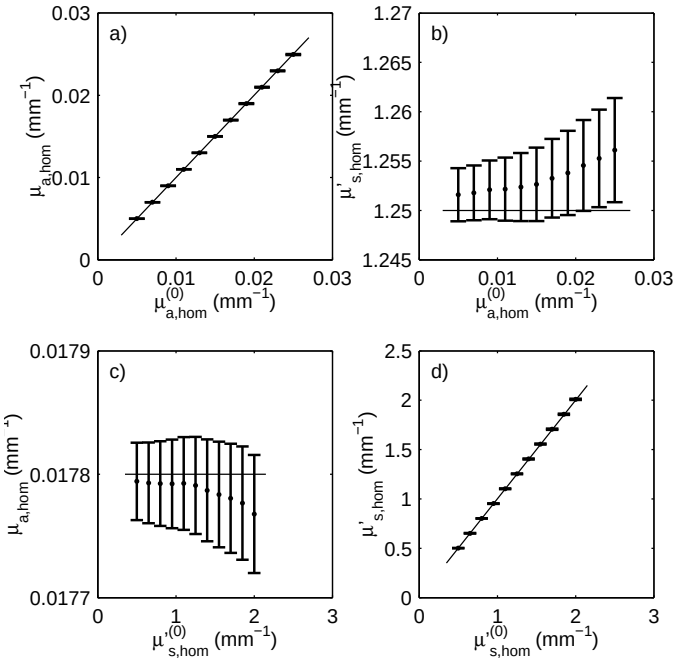

FIG. 4: Tracking of inferred optical parameters of a homogeneous head model, using noisy signals derived from the same model. a) and b) show the effect that changing the true absorption $\mu^{(0)}_{a,\rm hom}$ has on inferred $\mu_{a,\rm hom}$ and $\mu'_{s,\rm hom}$ respectively. The range of total detected photons N_p is from 3.4×10^6 at the smallest $\mu^{(0)}_{a,\rm hom}$ to 3.6×10^5 at the largest. c) and d) show the same, except changing the true reduced scattering $\mu'^{(0)}_{s,\rm hom}$, over which N_p varies from 1.8×10^6 to 3.5×10^5. The errorbars show $\pm 1\sigma$ about the MAP ('best-guess') value, for the (marginal) posterior distribution of inferred values, computed using the methods of Appendix A. For comparison, the thin lines show the true values. The fractional model-error is $\epsilon = 0.05$.

IV. RESULTS AND DISCUSSION

A. Homogeneous vs segmented head models

We are interested in finding out what improvement is possible in fitting baseline optical parameters when anatomical segmentation information becomes available. Therefore in this section we compare the models $\mathcal{H}_{\rm hom}$

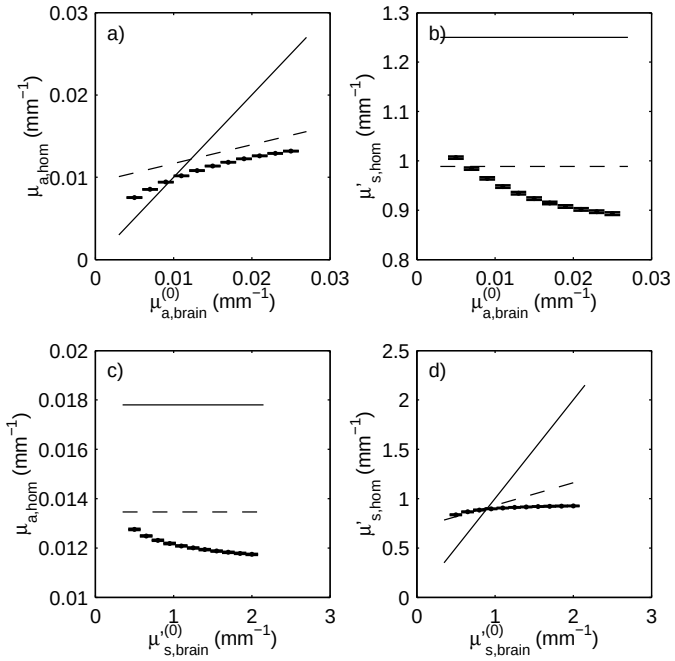

FIG. 5: Same as Fig. 4, except now the noisy signals have been generated using the segmented forward model, sweeping over a range of brain parameters. (The scalp and skull are fixed at the standard parameters in Table I). The inference is still performed using the homogeneous model. The dashed line shows true parameters volume-averaged over a depth of 17 mm. (We chose a constant effective thickness of 4 mm for the brain's volume contribution, a typical time-independent fluence decay length in a semi-infinite brain). N_p spans 2.4×10^6 to 2.1×10^6 over the range of $\mu^{(0)}_{a,\mathrm{brain}}$, and 2.1×10^6 to 2.2×10^6 over the range of $\mu'^{(0)}_{s,\mathrm{brain}}$. The fractional model-error is still $\epsilon = 0.05$.

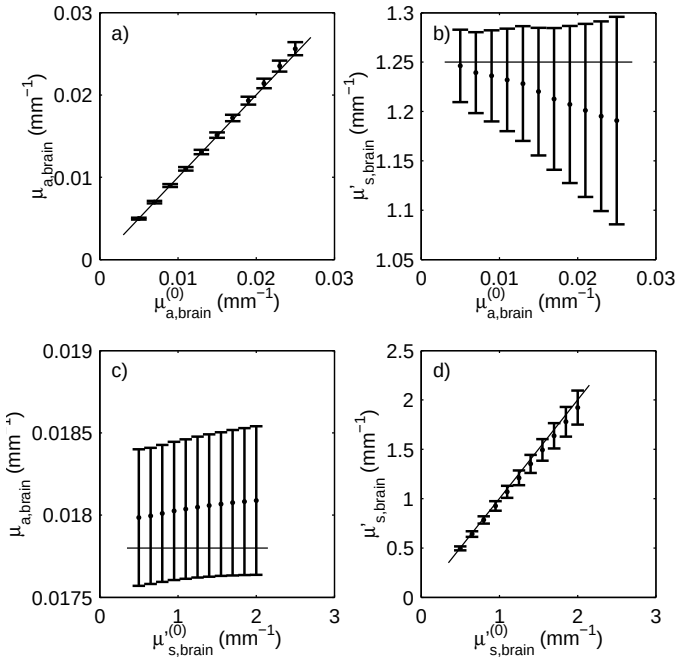

FIG. 6: Tracking of inferred brain optical parameters in a segmented model, using noisy signals also derived from the segmented model. (The scalp and skull are fixed at the standard parameters in Table I, and inferred scalp and skull parameters are not shown). N_p and ϵ are the same as for Fig. 5.

and $\mathcal{H}_{\mathrm{seg}}$ in their ability to infer these parameters, using simulated data also generated from $\mathcal{H}_{\mathrm{hom}}$ or $\mathcal{H}_{\mathrm{seg}}$. We choose a total detected photon number (which includes both sources),

$$N_p \equiv \sum_{m=1}^{M} y_m, \tag{9}$$

at approximately 2×10^6, typical for photon counting DOT systems. As we vary the head optical properties, we have decided to hold the collection time constant, rather than fix N_p. Note that some advantage, in terms of smaller N_p values, could be gained if variable detector attentuation [64] were used.

See figure captions for the N_p range for each experiment. In this Section we fix $\epsilon = 0.05$, and use a 2 mm forward model lattice. Recall that all simulated signals include realistic detection noise.

Homogeneous-homogeneous fit: Using signals from $\mathcal{H}_{\mathrm{hom}}$, we fit using $\mathcal{H}_{\mathrm{hom}}$, and sweep the true optical parameters. The results are presented in Fig. 4. This shows that if the head were indeed homogeneous, then very accurate inference of its baseline optical parameters would be possible. The errorbars (representing the expected measurement error of the parameters) are 0.5% or less. The true values fall within, or just outside, the errorbars.

Segmented-homogeneous fit: Using signals from the more realistic $\mathcal{H}_{\mathrm{seg}}$, we fit using $\mathcal{H}_{\mathrm{hom}}$ (see Fig. 5). This shows that the homogeneous model does a very poor job of measuring the brain's properties: changes in absorption are only 20% of the true change, and for scattering only 5%. This is due to detected photons spending a large fraction of time in the scalp and skull layers. To take this into account we also compare against local volume-averaged properties (dashed lines in Fig. 5); as expected, the homogeneous model reflects these better than the brain's properties. However there are two major failures of the homogeneous model: crosstalk and non-linearity. Neither of these failures can be compensated for by any (linear) volume-averaging correction. The absorption-scattering crosstalk dominates the fitted $\mu'_{s,\mathrm{hom}}$. (There is of course large crosstalk with scalp and skull parameters too, which we do not plot). The tracking is highly non-linear, meaning that even relative changes in brain parameters could not be quantitatively assessed. Note that $\mathcal{H}_{\mathrm{hom}}$ is a more physically accurate model than the homogeneous semi-infinite slab model commonly used to fit background properties because it incorporates the correct head surface shape. The fact that the true values are hundreds of standard de-

viations outside the errorbars implies that the inference noise model grossly under-represents the large true model errors.

Segmented-segmented fit: We now fit these same $\mathcal{H}_{\rm seg}$ signals with the $\mathcal{H}_{\rm seg}$ model (see Fig. 6). Measurement errors are 2%–2.5% in $\mu_{a,\rm brain}$ and 5%–10% in $\mu'_{s,\rm brain}$, across the complete range of biologically-relevant true parameter values. The linearity is excellent, and crosstalk causes a 0.5% change in $\mu_{a,\rm brain}$, and a 5% change in $\mu'_{s,\rm brain}$, across the parameter range. Importantly, although errorbars are larger than with $\mathcal{H}_{\rm hom}$, true values always fall within the errorbars. This shows that we believe reliable measurements of the brain's baseline properties can be made when anatomical information is provided.

Recalling that we are simultaneously fitting for scalp, skull and brain parameters, we can ask how accurately the scalp and skull have been characterised. At the standard values of Table I, expected errors are about 8% for $\mu_{a,\rm scalp}$, 3% for $\mu'_{s,\rm scalp}$, 4% for $\mu_{a,\rm skull}$, and 2% for $\mu'_{s,\rm skull}$.

We believe that in this fitting (non-imaging) example, the total photon number N_p will be more important in determining accuracy than the number of sources and detectors. Only when N is increased, as in an imaging application, or when unknown optode calibration parameters are included, do we expect that increasing the number of sources and detectors will significantly improve accuracy. Note that generally in order to optimize optode location and number, and TPSF parametrization, it would be important to know which components m of the signal $\mathbf{y}$ are most important in determining errorbar size for $\mu_{a,\rm brain}$ and $\mu'_{s,\rm brain}$. Within the Gaussian posterior approximation, all of this information is contained in the singular value decomposition of $\bar{J}$ (defined in Eq. A6). We will present these results in future publications.

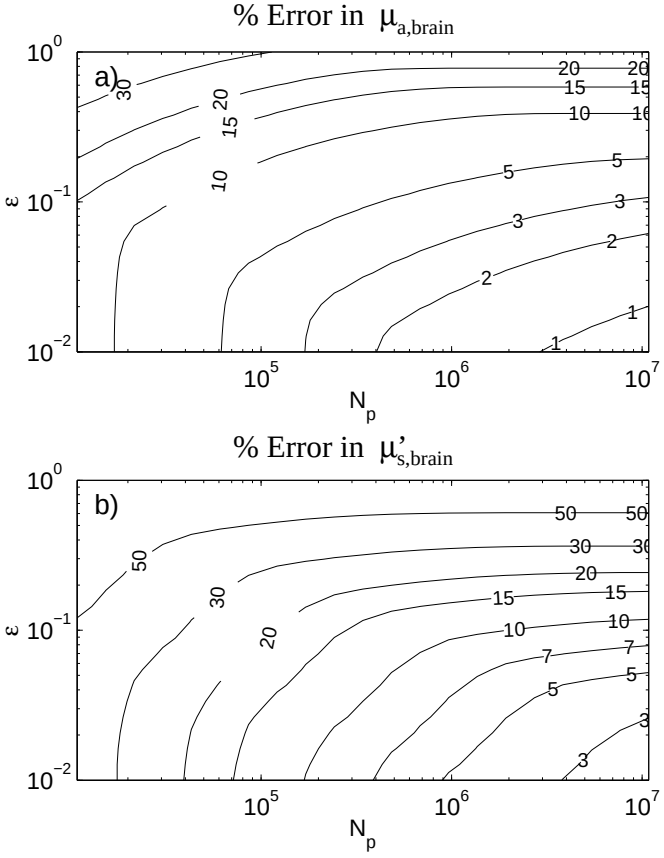

FIG. 7: Expected percentage marginal error as a function of noise model N_p (total detected photons) and ϵ (fractional model-error) for parameters a) $\mu_{a,\rm brain}$ and b) $\mu'_{s,\rm brain}$. The standard set of optical properties are used, for inference using $\mathcal{H}_{\rm seg}$ and simulated noisy signals from the same model. Note that the contour lines bend quite sharply at the transition from Poisson-statistics-limited error (lower left) to model-limited error (upper right).

B. Systematic deviations

In this subsection we discuss a detail of the inference procedure. (This can be skipped on first reading).

The reason why there are systematic deviations between MAP and true parameter values in Figs. 4 and 6 is not immediately intuitive. One might suspect that, since exactly the same models are used for simulation as for inference, zero error would result, averaging over the simulated noise. This effect is worth discussing further. In essence the systematic difference comes from the mismatch between the simulated and the inference noise models: $\sigma_{\rm sim}(f) \neq \sigma(f)$ for large f (see Eqs. 7 and 6). If the inference noise model is 'correct' (*i.e.* reflects the population from which noisy signals are sampled), Bayesian inference provides unbiased estimators [52] (in this case $\mathbf{x}_{\rm MAP}$ values). We have checked this: if $\sigma_{\rm sim}(f) = \sigma(f)$ for all f then the deviations change from systematic to random. However model error is *not* random; we prefer to investigate its effect via particular instances of realistic model error (Section IV E).

Mathematically, the systematic deviation can be seen to arise from the weak $\mathbf{x}$-dependence of $\sigma(f(\mathbf{x}))$ (giving σ' terms in Eq. A2). It would be misleading to attempt to correct this deviation, since this would be tantamount to benefitting from more information about the system than we have formalized in the noise model Eqs. 3 and 6. The deviation is always of a size consistent with the errorbars; the correct way to reduce this deviation is by building more accurate forward models whose smaller known model-error we can then use for inference.

C. Noise model effect on brain parameter errorbars

It is natural to wonder what the effects of the number of detected photons N_p, and the fractional model-error ϵ, are on the measurement accuracy. This allows the experiment designer to know in advance the detection time needed to reach a certain expected error in brain optical properties. In Fig. 7 we plot contours of expected

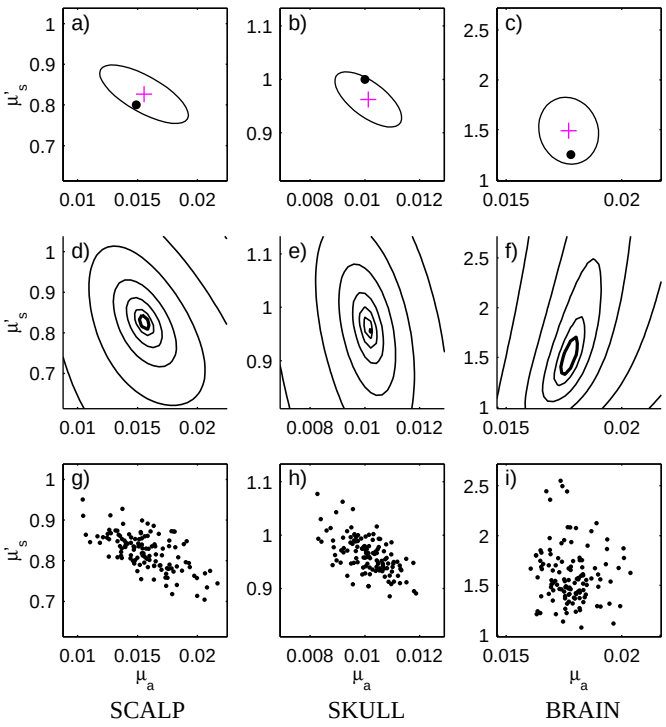

FIG. 8: Views of the posterior distribution $p(\mathbf{x}|\mathbf{y},\mathcal{H}_{\rm seg})$ with signals $\mathbf{y}$ from the same forward model (2 mm lattice), using our standard optical parameters in Table I. Only 3.2×10^5 photons have been collected, and $\epsilon = 0.1$. The three columns of graphs show the μ_a–μ_s' plane, separately for each tissue type. a), b), c) show true $\mathbf{x}_0$ (point), $\mathbf{x}_{\rm MAP}$ (cross), and the *marginal* PDF as an elliptical contour enclosing 63% of the probability mass in the Gaussian posterior approximation. The contour is at e^{-1} times the peak density. d), e), f) show for comparison, on the same axes, *conditional* distributions (slices through the PDF with other components of $\mathbf{x}$ fixed at the $\mathbf{x}_{\rm MAP}$ values), at contours of e^{-1} (shown with thicker line), $e^{-3}, e^{-10}, e^{-30}, e^{-100}, e^{-300}$ and e^{-1000} times the peak density. g), h), i) show 123 independent samples from the posterior obtained using Markov chain Monte Carlo. This displays the true marginal posterior PDF as a 'density cloud'.

percentage error in the two brain parameters, across the N_p–ϵ plane. We chose the standard set of optical parameters, using $\mathcal{H}_{\rm seg}$ for inference on a noisy signal from the same segmented model. The results show that even assuming fractional model-error as large as $\epsilon = 0.2$, expected errors on brain parameters can be 5% for absorption and 15%–20% for reduced scattering.

To generate this figure, solving for $\Sigma_{\rm MAP}$ by locating $\mathbf{x}_{\rm MAP}$ for each point in the plane in this figure would have been a very costly operation. Instead we observe that since $\mathbf{f}(\mathbf{x})$ is locally linear, its Jacobean J at $\mathbf{x}_{\rm MAP}$ will be similar to that at $\mathbf{x}^{(0)}$, when $\mathbf{x}_{\rm MAP}$ is close to $\mathbf{x}^{(0)}$. (In our case, it becomes a bad approximation only when marginal expected errors are greater than roughly 20%, when nonlinearity of $\mathbf{f}(\mathbf{x})$ becomes relevant). Therefore we can use J measured once at $\mathbf{x}^{(0)}$ to approximate $\Sigma_{\rm MAP}$ for any noise model parameters, according to Eqs. A6, A5 and A7. This procedure makes investigation of the noise model a rapid process.

D. Visualization of full posterior distribution

In Section IV A, for inference with $\mathcal{H}_{\rm seg}$, the errorbars (marginal expected errors) in the six parameters were of the same order of magnitude, 2%-10%. However, the (hyper-)ellipsoid defining the full posterior is actually very narrow in some directions; these directions happen *not* be aligned with the parameter axes. The ratio of the largest to smallest eigenvalue of $\Sigma_{\rm MAP}$ is 3×10^3, implying that the shortest principal axis is only about 2% of the longest. The distribution is in fact very 'pancake-like'. This phenomenon of both very well- and very poorly-constrained parameter directions is a feature of ill-posed inverse problems.

Combined with the nonlinear nature of $\mathbf{f}(\mathbf{x})$, this gives us some motivation to be suspicious of a Gaussian approximation to the posterior. Fig. 8 shows that even when our posterior is much wider (5% in $\mu_{a,\rm brain}$ and 14% in $\mu'_{s,\rm brain}$), the Gaussian approximations to the marginal distributions are adequate, and would give adequate confidence intervals on each x_n. However, the non-Gaussian nature has become relevant in the tails of the distribution (*e.g.* at large $\mu'_{s,\rm brain}$, Fig. 8i). Samples from the true posterior have been generated via MCMC, using the Metropolis method [55], with the stepping distribution chosen to be uniform within a hyper-ellipsoid of shape given by the Hessian matrix (Eq. A5). By optimizing the hyper-ellipsoid size, one independent sample could be extracted from the Markov chain roughly every 40 forward model evaluations. The total 'cloud' took 20 hours of CPU time to generate. Faster MCMC methods exist [55].

Note that in Fig. 8 the *conditional* distributions (*i.e.* slices though the posterior—these are much easier to generate than true marginal distributions), do a very poor job of telling us the width of the marginal distributions. This results from the pancake-like nature of the PDF.

E. Robustness to forward model error

If we run our forward model at a higher lattice resolution, we can generate simulated signals $\mathbf{y}$ that are more accurate, given the DA physics and the tissue segmentation model. In this way we can see how the inference procedure (based on a lower lattice resolution) handles realistic model error, which is highly correlated. In other words, we avoid committing the 'inverse crime' of using identical forward and inverse models, of which we were guilty in Section IV A.

In Fig. 9a we show that changing from 2 mm to 1 mm lattice voxel size in the forward model causes changes of roughly 20%. Most are smaller, and some are much

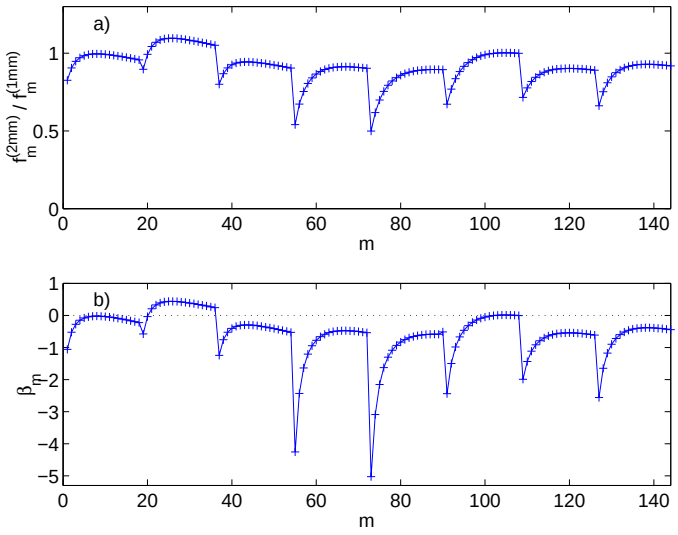

FIG. 9: a) shows approximate error of the 2 mm lattice forward-model signal expectation, expressed as a ratio against its (more accurate) 1 mm lattice equivalent. The standard set of optical properties are used, for the segmented head model. b) shows the normalized residual β_m which results when the 1 mm lattice signal is compared against the 2 mm lattice signal, using noise model parameters $y_p = 10^{-4}$ and $\epsilon = 0.2$.

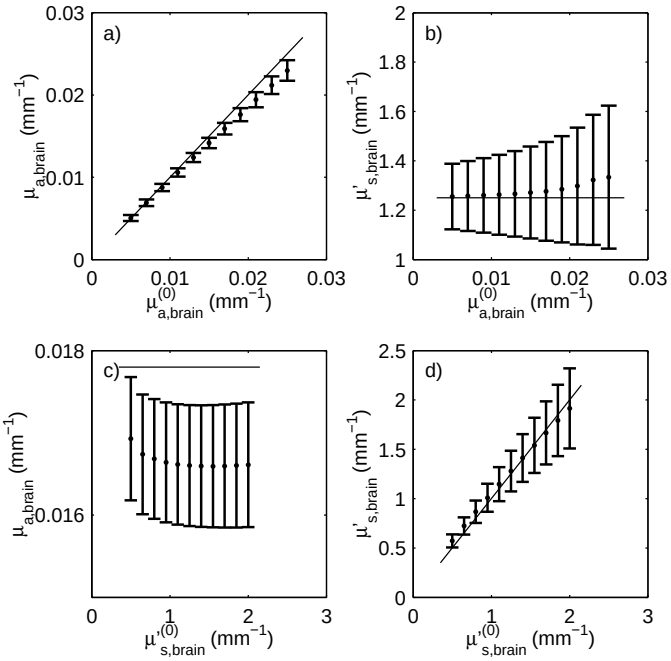

FIG. 10: Inference using the 2 mm lattice forward model $\mathcal{H}_{\rm seg}$, on simulated noisy signals generated from a 1 mm lattice forward model. The detection time was larger than in Fig. 6, with N_p spanning 1.18×10^7 to 1.35×10^7 over the range of $\mu^{(0)}_{a,\rm brain}$ and $\mu'^{(0)}_{s,\rm brain}$. The fractional model-error of $\epsilon = 0.2$ has been chosen to reflect our knowledge of the 2 mm lattice model errors.

larger, approaching 50% at early times. This latter early-time error is due to the non-zero depth below the surface at which the optode couples to the fluence gradient (see Appendix B). Smaller late-time error is due to the voxellated (rough) surface representation local to each optode.

We will perform inference using the usual 2 mm lattice model $\mathcal{H}_{\rm seg}$, using the choice $\epsilon = 0.2$ in the inference noise-model, motivated by the 20% observed model error. We increase the collection time to give roughly 1.2×10^7 photons. We find (Fig. 9b) that the variance of the resulting normalised residuals β_m (see Eq. A3) is close to 1, indicating a rough match of the noise model to true forward model errors.

Fig. 10 shows the results of sweeping the true brain optical parameters. The accuracy of measurement of $\mu_{a,\rm brain}$ is about 4%, and for $\mu'_{s,\rm brain}$ it is 10%–20%. The linearity is good, and maximum crosstalk is 6% (only significantly affecting $\mu'_{s,\rm brain}$). The systematic errors are certainly larger than those achieved in the more artificial situation shown in Fig. 6, however the errorbar still does a good job of indicating the size of this error. Therefore model error is being handled in a robust fashion.

Clearly incorporating the fact that model errors are large only at early times would be advantageous. We discuss ideas for model error improvement below.

V. CONCLUSIONS

We have demonstrated that an accurate determination of the baseline optical properties (both absorption μ_a and scattering μ'_s) of the human head could be made using a small time-resolved DOT apparatus using high-resolution 3D anatomical information from magnetic resonance imaging. We use an automatic MR segmentation into the basic tissue types (scalp, skull, CSF, and brain) and fit for μ_a and μ'_s of scalp, skull and brain (holding the CSF fixed), assuming each tissue is optically homogeneous. We perform simultaneous non-linear model-based fitting (inference) of the six optical parameters, and demonstrate the superiority of this approach over a homogeneous-head model in the biologically-relevant parameter ranges $0.005\,\mathrm{mm}^{-1} < \mu_{a,\rm brain} < 0.025\,\mathrm{mm}^{-1}$ and $0.5\,\mathrm{mm}^{-1} < \mu'_{s,\rm brain} < 2.0\,\mathrm{mm}^{-1}$. We emphasise that even the homogeneous-head model is more realistic than the commonly-used homogeneous semi-infinite slab model. For our work we use numerically-simulated noisy measured signals. Our finite-difference forward model can simulate the time-domain diffusion approximation in arbitrary 3D head geometries; we use about 3×10^4 nodes. We believe that this approximation is adequate given the convoluted CSF geometry, although this an important area of active research for us and others in the field. Fitting takes about 10 minutes on a standard 1 GHz CPU.

We employ a realistic but novel Gaussian noise model incorporating both shot-noise, parametrized by the total detected photons N_p, and model error (including calibration), parametrized by a fraction ϵ. In the Bayesian

inference framework, with a flat prior, we find the baseline measurement accuracy (errorbars) by examining the marginal width of the posterior distribution, for each parameter. We note that all previous demonstrations of Bayesian DOT have either been in 2D or simple cuboid geometries, and have not considered more than the single best-fit (MAP) solution. We find that 2×10^6 detected photons gives 2.5% errorbar in $\mu_{a,\mathrm{brain}}$ and 5%–10% in $\mu'_{s,\mathrm{brain}}$, assuming model errors of 5%. The crosstalk between surface layers and the cortex, and between μ_a and μ'_s, present with the homogeneous model, is eliminated.

By using a finer computational lattice, we tested realistic forward model error and find that even with 20% model error the errorbars can still be adequate: 4% in $\mu_{a,\mathrm{brain}}$ and 10%–20% in $\mu'_{s,\mathrm{brain}}$, for $N_p \approx 10^7$. (All N_p values could be reduced experimentally by the use of per-detector variable attenuation). Robustness is demonstrated by the fact that errorbars always reflect the size of the true deviations. We have also demonstrated the power of Markov chain Monte Carlo for posterior sampling when the Gaussian MAP approximation is no longer good.

The Bayesian framework allows, without fuss, use of improved noise models: we believe it will be important to encode into such models both measurement properties of DOT apparatus, and physical and numerical human head forward model errors. Noise models should also provide robustness to measurement outliers by replacing the Gaussian with a heavier-tailed distribution. In future work we plan to include unknown optode amplitude calibration parameters [49]: these can then be marginalized over naturally within the Bayesian framework. (We note that Oh *et al.* [45] have recently simulated Bayesian optode calibration in a 3D cubical geometry, at least in terms of a single MAP solution, with success). By increasing the number of parameters from 6 to 10^2 or more, our Bayesian approach could give errorbars and correlations of voxel parameters in a true 'imaging' (tomography) context, something we intend to address in future work. In general the value of the Bayesian paradigm in neuroimaging lies with its statistical roots in what is becoming an increasingly statistical practice. On a lower level, information from other modalities can optimally be included as a prior. On a higher level, the full posterior PDF, containing all error correlations, could be fed to dynamical models of neural activation.

Our study provides realistic expectations for achievable baseline accuracy in human hemodynamic parameters. An experimental implementation could have rapid clinical benefit in the stroke and neonatal fields. Our work also expands the possibility of simultaneous DOT and MR for functional neuroimaging: quantitative optical imaging of hemodynamic changes will require accurate baseline properties, which can best be achieved through the type of anatomical modeling presented here.

Acknowledgements

We have benefitted from discussions with Jon Stott, Quan Zhang, Tina Chaves, Greg Boverman, Radford Neal, and Jari Kaipio, among many others. We acknowledge funding by the Courant Institute at New York University, by grants NIH 2P01-NS35611-06, NIH R29-NS38842, NIH P41-RR14075 and from the Center for Innovative Minimally Invasive Therapies. This research was funded in part by the US Army, under Cooperative Agreement No. DAMD17-99-2-9001. The material presented does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

Appendix A. Finding the Gaussian approximation to the posterior

Maximizing Eq. 4 is equivalent to minimizing $L(\mathbf{y};\mathbf{x})$ over $\mathbf{x}$ within the given bounds. We rescale the components of $\mathbf{x}$ to have the same typical size by multiplying absorption coefficients by a constant $C = 10^2$, giving for instance in the fully-segmented case $\mathbf{x} \equiv \{C\mu_{a,\mathrm{scalp}},\ \mu'_{s,\mathrm{scalp}},\ C\mu_{a,\mathrm{skull}},\ \mu'_{s,\mathrm{skull}},\ C\mu_{a,\mathrm{brain}},\ \mu'_{s,\mathrm{brain}}\}$. This elementary preconditioning enhances performance of the inference methods presented below.

Because of the nature of the numerical solution of the forward model (Appendix B), the dependence of L on $\mathbf{x}$ is not completely smooth nor differentiable. This jitter in L, although estimated to be on a scale of less than 10^{-3}, means that care has to be taken in choosing an optimization algorithm. We have found Newton's method [72] most successful, the i^{th} iteration step being

$$\mathbf{x}_{i+1} \ = \ \mathbf{x}_i \ - \ H^{-1}(\mathbf{x}_i) \cdot \nabla L(\mathbf{x}_i), \qquad \text{(A1)}$$

where the exact gradient of L has components,

$$(\nabla L(\mathbf{x}))_n \ = \ \sum_{m=1}^{M} \frac{J_{mn}}{\sigma_m} \left[\beta_m + (1-\beta_m^2)\sigma'_m\right], \qquad \text{(A2)}$$

which follows from Eq. 3. Here the normalized residual is

$$\beta_m \ \equiv \ \frac{f_m(\mathbf{x}) - y_m}{\sigma_m}, \qquad \text{(A3)}$$

and the Jacobean (Frechet) derivative of the forward model is $J_{mn} \equiv \partial f_m / \partial x_n$. The symbol σ'_m is an abbreviation for $d\sigma/df|_{f_m}$, which in our case can be found in closed form by differentiating Eq. 6. All quantities are evaluated at $\mathbf{x}$. Since the number of unknowns N is small, we evaluate J using crude finite differencing along the axes in $\mathbf{x}$ space:

$$J_{mn}(\mathbf{x}) \ \approx \ \frac{1}{\delta x}\left[f_m(\mathbf{x} + \delta x\, \mathbf{e}_n) - f_m(\mathbf{x})\right], \qquad \text{(A4)}$$

where $\mathbf{e}_n$ is the n^{th} unit vector, requiring $N+1$ forward evaluations. With adjoint differentiation it would be possible to get J with effort roughly equal to one forward evaluation [40, 41, 57]; this would be worthwhile only if N were larger.

It is simple and efficient to use an approximate Hessian matrix,

$$H \;\approx\; H_{\text{approx}} \;\equiv\; \bar{J}^T \bar{J}, \tag{A5}$$

where

$$\bar{J}_{mn} \;\equiv\; \frac{\partial \beta_m}{\partial x_n} \;=\; \frac{J_{mn}}{\sigma_m}\left[1 - \beta_m \sigma'_m\right] \tag{A6}$$

is the Jacobean of the normalized residual. This requires no extra function evaluations to be performed. The stopping criterion for locating $\mathbf{x}_{\text{MAP}}$ is $|\mathbf{x}_{i+1} - \mathbf{x}_i| < \epsilon$ where ϵ is of order 10^{-3}. The initial value $\mathbf{x}_1$ is chosen randomly in the optical parameter range of interest, and has no effect on the $\mathbf{x}_{\text{MAP}}$ found. (Occasionally if $\mathbf{x}_{\text{MAP}}$ was very far from $\mathbf{x}_1$ then an initial rough Nelder-Mead simplex optimization [72] was required). Typically 3–8 Newton iterations are required. Given that a 2 mm lattice evaluation of $\mathbf{f}(\mathbf{x})$ takes about 16 seconds, this means 5–15 minutes are required to find $\mathbf{x}_{\text{MAP}}$ and Σ_{MAP}.

Because the derivatives of log terms from Eq. 3 are small, the above form of the Hessian usually agrees to within a few percent with a numerically-estimated Hessian, found by finite differencing $N(N+1)/2$ samples of L in $\mathbf{x}$-space. We have experimented with procedures for this estimation, but the combination of jitter on L, the fact that L is orders of magnitude more sensitive in some directions than others (*i.e.* the Hessian is close to singular), and the nonlinearity of $\mathbf{f}(\mathbf{x})$ and hence $L(\mathbf{x})$, causes a tendency for large errors and non-positive-definiteness. Nothing as reliable as the above (guaranteed positive definite) approximate Hessian form has been found. Therefore, our multivariate Gaussian covariance matrix estimate is

$$\Sigma_{\text{MAP}} \;=\; H^{-1}_{\text{approx}}. \tag{A7}$$

In order to calculate confidence intervals on single parameters, and the lower-dimensional marginal Gaussian distributions shown in Fig. 8a–c, we need the recipe for marginalizing a multivariate Gaussian onto a subspace. We split $\mathbf{x}$ into subspaces $\mathbf{a}$ and $\mathbf{b}$. By reordering $\{x_n\}$ this can without loss of generality be written $\mathbf{x}^T \equiv [\mathbf{a}^T, \mathbf{b}^T]$. Any covariance Σ in $\mathbf{x}$-space can then be written in block form,

$$\Sigma \equiv \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ab}^T & \Sigma_{bb} \end{pmatrix}. \tag{A8}$$

By analytically integrating over $\mathbf{b}$, the inverse covariance of the marginal distribution in subspace $\mathbf{a}$ can be shown to be [73]

$$\Sigma^{-1}_{(a)} \;=\; \Sigma_{aa} - \Sigma_{ab}\Sigma^{-1}_{bb}\Sigma^T_{ab}. \tag{A9}$$

In the case where $\mathbf{a}$ has a single dimension $a \equiv x_n$ then $\Sigma_{(a)} \equiv \sigma_{x_n}$ is a scalar giving the marginal standard deviation of parameter x_n.

Appendix B. Finite difference method for the diffusion equation

Here we provide details of the method used to approximate the time-evolution of the 3D parabolic Eq. 5, with constant $v(\mathbf{r}) = c$. There are many issues, most of which we discuss only cursorily.

Since $q(\mathbf{r}, t)$ is nonzero only at $t = 0$ this is a initial-value problem, and we use a FDTD method with a regular cubical lattice of nodes with spacing Δx, and a time-step Δt. We chose this over finite-element methods for simplicity and the fact that our segmentation map is also based on a cuboid lattice. We chose a node-based representation of fluence ϕ, and discretized the spatial derivative at node location $\mathbf{r}_{i,j,k}$ in the standard fashion [72, 74, 75],

$$\left.\frac{\partial}{\partial x}\kappa\frac{\partial \phi}{\partial x}\right|_{\mathbf{r}_{i,j,k}} \approx \tag{A10}$$
$$\frac{\kappa_{i+\frac{1}{2},j,k}(\phi_{i+1,j,k} - \phi_{i,j,k}) - \kappa_{i-\frac{1}{2},j,k}(\phi_{i,j,k} - \phi_{i-1,j,k})}{\Delta x^2}$$

with equivalent expressions in the y and z directions, where i, j, k are integer node labels in x, y, z axes respectively. This is $O(\Delta x^2)$ accurate. The material property κ is voxel-based: we locally average four voxels to give edge-based values $\kappa_{i\pm\frac{1}{2},j,k}$.

The initial fluence $q(\mathbf{r}, 0)$ is discontinuous, comprising a delta-function (or convolved delta-function) near the source; we found that Crank-Nicolson type $O(\Delta t^2)$ methods do *not* handle the non-smooth fluence well, due to the slow decay of non-physical oscillatory modes excited by the discontinuity. We note that these methods (specifically the alternating-direction implicit [ADI] method) are commonly used [15, 40, 41, 57] for simulation of Eq. 5 without mention of this issue. Instead we use the explicit Euler $O(\Delta t)$ method, which is less efficient and requires $\Delta t < \Delta x^2/(6\kappa c)$ at every point in the lattice for stability [75]. Evolution is perfomed by repeated sparse-matrix multiplication. Our computational effort is proportional to the number of nodes times the number of time-steps: this scales like $1/(\Delta x^3 \Delta t)$ which is proportional to $1/(\Delta x^5 \mu'_{s,min})$ where $\mu'_{s,min}$ is the smallest value of μ'_s in the system. Changing Δx from 2 mm to 1 mm therefore increases the effort by a factor of 32. Implicit methods would be more efficient; we are developing an optimal way to combine explicit and Douglas-Gunn [75] evolution to handle the discontinuous initial-condition.

We emulated Dirichlet boundary conditions by removing degrees of freedom $\phi_{i,j,k}$ for nodes on or outside the tissue-air boundary. (The issue of FDTD stability with the more correct Robin boundary conditions is an area

for research; we found resulting differences to be slight in our simulations for reasons given in Section III C). We simulate only a fraction of the human head volume, corresponding to about 3×10^4 nodes for $\Delta x = 2$ mm. Detector signal relative changes were less than 1% compared against the full head simulation of about 6×10^5 nodes.

In order to inject and detect fluence signals we developed lattice versions of standard optode models which couple to the fluence normal gradient $\hat{\mathbf{n}} \cdot \nabla \phi$ at the tissue-air boundary [15]. In order to reduce dependence on the detailed local voxel representation of the head surface, each optode was coupled to the fluence derivative in the beam direction, averaged over a beam of 2 mm radius at the rather large fixed depth of 1.5 voxels below the local surface along the beam direction. This large depth is responsible for the large early-time errors in *e.g.* Fig. 9, but is necessary to maximize immunity to voxellated surface roughness. The source opdotes give $q(\mathbf{r}, 0)$; the time-dependent detector signals are then integrated over the time-gates to give the components of the signal expectation vector $\mathbf{f}(\mathbf{x})$. Notice that this gradient-based source model is equivalent to more usual point-source models, but is more immune to voxellated surface roughness. Exploration of other optode models is an area for future work.

The procedure was validated against known analytic semi-infinite slab solutions ($< 5\%$ errors beyond 200 ps), and against a Monte Carlo calculation of a heterogeneous cuboid containing an absorbing and scattering inclusion whose nearest approach to the surface was as close as 6 mm ($\sim 20\%$ errors). We reserve validation against Monte Carlo in the full head model for a future publication.

[1] D. A. Boas, D. H. Brooks, E. L. Miller, C. .A. DiMarzio, M. Kilmer, R. J. Gaudette, and Q. Zhang, "Imaging the body with diffuse optical tomography", IEEE Signal Processing Magazine **18**(6), 57–75 (2001).

[2] B. Chance, D. T. Delpy, C. E. Cooper and E. O. R. Reynolds, eds., "Near-infrared spectroscopy and imaging of living systems", Phil. Trans. R. Soc. London B **352** (1997).

[3] D. A. Boas, M. A. Franceschini, A. K. Dunn, and G. Strangman, "Noninvasive imaging of cerebral activation with diffuse optical tomography", in *In vivo optical imaging of brain function*, R. D. Frostig, ed. (CRC Press, Boca Raton, 2002), pp. 193–221.

[4] A. Villringer and B. Chance, "Non-invasive optical spectroscopy and imaging of human brain function", Trends Neurosci. **20**, 435–442 (1997).

[5] S. Fantini, S. A. Walker, M. A. Franceschini, K. T. Moesta, P. M. Schlag, M. Kaschke, and E. Gratton, "Assessment of the size, position, and optical properties of breast tumors in vivo by non-invasive optical methods", Appl. Opt. **37**, 1982–1989 (1998).

[6] B. W. Pogue, T. O. McBride, S. Osterman, S. Poplack, U. Osterberg, and K. D. Paulsen, "Quantitative hemoglobin tomography with diffuse near-infrared spectroscopy: pilot results in the breast", Radiology, **218**(1), 261–266 (2001).

[7] J. C. Hebden, H. Veenstra, H. Dehghani, E. M. C. Hillman, M. Schweiger, S. R. Arridge, and D. T. Delpy, "Three dimensional time-resolved optical tomography of a conical breast phantom", Appl. Opt. **40**, 3278–87 (2001).

[8] N. Shah, A. Cerussi, C. Eker, J. Espinoza, J. Butler, J. Fishkin, R. Hornung, and B. Tromberg, "Noninvasive functional optical spectroscopy of human breast tissue", Proc. Natl. Acad. Sci. USA **98** 4420–25 (2001).

[9] S. Fantini, M. A. Franceschini-Fantini, J. S. Maier, S. A. Walker, B. Barbieri, and E. Gratton, "Frequency-domain multichannel optical detector for noninvasive tissue spectroscopy and oximetry", Opt. Eng. **34**, 32–42 (1995).

[10] A. Klose, A. H. Hielscher, K. M. Hanson, and J. Beuthan, "Three-dimensional optical tomography of a finger joint model for diagnostic of rheumatoid arthritis", in *Photon Propagation in Tissue IV*, D. A. Benaron, B. Chance, M. Ferrari, and M. Kohl, eds., Proc. SPIE **3566**, 151–160 (1998).

[11] G. Gratton, M. Fabiani, P. M. Corballis, D. C. Hood, M. R. Goodman-Wood, J. Hirsch, K. Kim, D. Friedman, and E. Gratton, "Fast and localized event-related optical signals (EROS) in the human occipital cortex: comparisons with the visual evoked potential and fMRI", Neuroimage **6** 168–180 (1997).

[12] J. Steinbrink, M. Kohl, H. Obrig, G. Curio, F. Syre, F. Thomas, H. Wabnitz, H. Rinneberg, and A. Villringer, "Somatosensory evoked fast optical intensity changes detected non-invasively in the adult human head", Neuroscience Letters **291**, 105–108 (2000).

[13] V. Ntziachristos, A .G. Yodh, M. Schnall, and B. Chance, "Concurrent MRI and diffuse optical tomography of breast after indocyanine green enhancement", Proc. Natl. Acad. Sci. USA, **97** 2767-72 (2000).

[14] G. Strangman, J. .P. Culver, J. H. Thompson, and D. A. Boas, "A quantitative comparison of simlutaneous BOLD fMRI and NIRS recordings during functional brain activation", NeuroImage, to appear.

[15] S. R. Arridge, "Optical tomography in medical imaging", Inverse Problems **15** R41–R93 (1999).

[16] D. A. Boas, M. A. O'Leary, B. Chance, and A. G. Yodh, "Detection and characterization of optical inhomogeneities with diffuse photon density waves: a signal-to-noise analysis", Appl. Opt. **36**, 75–92 (1997).

[17] S. Wray, M. Cope, and D. T. Delpy, "Characteristics of the near infrared absorption spectra of cytochrome aa3 and hemoglobin for the noninvasive monitoring of cerebral oxygenation", Biochim. Biophys. Acta **933** 184–192 (1988).

[18] W. D. Heiss, "Ischemic penumbra: Evidence from functional imaging in man", Journal of Cerebral Blood Flow and Metabolism **20** 1276–93 (2000).

[19] S. R. Hintz, W.-F. Cheong, J. P. van Houten, D. K. Stevenson, D. A. Benaron "Bedside imaging of intracranial hemorrhage in the neonate using light: comparison with ultrasound, computed tomography, and magnetic resonance imaging", Pediatr. Res. **45**, 54–59

(1999).
[20] S. R. Hintz, D. A. Benaron, A. M. Siegel, A. Zourabian, D. K. Stevenson and D. A. Boas, "Bedside functional imaging of the premature infant brain during passive motor activation", J. Perinat. Med. **29** 335–343 (2001).
[21] R. Barbour, H. Graber, Y. Wang, J. Chang, and R. Aronson, "A perturbation approach for optical diffusion tomography using continuous-wave and time-resolved data", in *Medical optical tomography: Functional imaging and monitoring, vol. IS 11* (SPIE Press, 1993), pp. 87–120.
[22] A. J. Devaney, "Reconstruction tomography with diffractive wave-fields", Inverse Problems **2**, 161-183 (1986).
[23] S. R. Arridge and J. C. Hebden, "Optical Imaging in Medicine: II. Modelling and reconstruction", Phys. Med. Biol. **42**, 841-854 (1997).
[24] X. Cheng and D. A. Boas, "Systematic diffuse optical image errors resulting from uncertainty in the background optical properties", Opt. Express **4**, 299–307 (1999).
[25] V. Ntziachristos, A. G. Yodh, M. D. Schnall, and B. Chance, "MRI-Guided Diffuse Optical Spectroscopy of Malignant and Benign Breast Lesions", Neoplasia **4**, 347–354 (2002).
[26] S. R. Arridge and M. Schweiger, "Reconstruction in optical tomography using MRI based prior knowledge", in *Information Processing in Medical Imaging*, eds. Y. Bizais, C. Barillot, and R. di Paola (Kluwer, 1995), pp. 77–88.
[27] M. Schweiger and S. .R. Arridge, "Optical tomographic reconstruction in a complex head model using a priori region boundary information", Phys. Med. Biol. **44**, 2703–22 (1999).
[28] M. Hämäläinen, R. Hari, R. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa, "Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain", Rev. Mod. Phys., **65**, 413–497 (1993).
[29] J. D. Oakley, "Magnetic resonance imaging based correction and reconstruction of positron emission tomography images", Ph. D. thesis, Service Hospitalier Frederic Joliot, CEA, Orsay, France (2000).
[30] V. Kolehmainen, M. Vauhkonen, J. P. Kaipio, and S. R. Arridge, "Recovery of piecewise constant coefficients in optical diffusion tomography", Opt. Express **7**, 468–481 (2000).
[31] M. Kilmer, E. Miller, D. A. Boas, and D. Brooks, "A shape-based reconstruction technique for DPDW data", Opt. Express **7**, 481–491 (2000).
[32] M. S. Patterson, B. Chance, and B. C. Wilson, "Time resolved reflectance and transmittance for the noninvasive measurement of tissue optical properties", Appl. Opt. **28**, 2331–36 (1989).
[33] R. Cubeddu, A. Pifferi, P. Taroni, A. Torricelli, and G. Valentini, "Time-resolved imaging on a realistic tissue phantom: μ_s' and μ_a images versus time-integrated images", Appl. Opt. **35**, 4533–40 (1996).
[34] A. Torricelli, A. Pifferi, P. Taroni, E. Giambattistelli, and R. Cubeddu, "In vivo optical characterization of human tissues from 610 to 1010 nm by time-resolved reflectance spectroscopy", Phys. Med. Biol. **46**, 2227–2237 (2001).
[35] A. Kienle, M. S. Patterson, N. Dögnitz, R. Bays, G. Wagnières, and H. van den Bergh, "Noninvasive determination of the optical properties of two-layered media", Appl. Opt. **37**, 779–791 (1998).
[36] A. Kienle, T. Glanzmann, G. Wagnières, H. and van den Bergh, "Investigation of two-layered turbid media with time-resolved reflectance", Appl. Opt. **37**, 6852–62 (1998).
[37] A. Pifferi, A. Torricelli, P. Taroni, and R. Cubeddu, "Reconstruction of absorber concentrations in a two-layer structure by use of multidistance time-resolved reflectance spectroscopy", Opt. Lett. **26**, 1963–65 (2001).
[38] C. K. Hayakawa, J. Spanier, F. Bevilacqua, A. K. Dunn, J. S. You, B. J. Tromberg, and V. Venugopalan, "Perturbation Monte Carlo methods to solve inverse photon migration problems in heterogeneous tissues", Opt. Lett. **26**, 1335–37 (2001).
[39] A. Kienle and T. Glanzmann, "In vivo determination of the optical properties of muscle with time-resolved reflectance using a layered model", Phys. Med. Biol. **44**, 2689–2702 (1999).
[40] K. M. Hanson, G. S. Cunningham, and S. S. Saquib, "Inversion based on computational simulations", in *Maximum Entropy and Bayesian Methods*, G. J. Erickson, J. T. Rychert, and C. R. Smith, eds. (Kluwer Academic, Dordrecht, 1998), pp. 121–135.
[41] S. S. Saquib, K. M. Hanson, and G. S. Cunningham, "Model-based image reconstruction from time-resolved diffusion data", in *Medical Imaging: Image Processing*, Proc. SPIE **3034**, 369–380 (1997).
[42] G. Nicholls and C. Fox, "Prior modelling and posterior sampling in impedance imaging", in *Bayesian Inference for Inverse Problems*, A. Mohammad-Djafari, ed., Proc. SPIE **3459**, 116–127 (1998).
[43] J. P. Kaipio, V. Kolehmainen, E. Somersalo, and M. Vauhkonen, "Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography", Inverse Problems **16**, 1487–1522 (2000).
[44] D. M. Schmidt, J. S. George, and C. C. Wood, "Bayesian inference applied to the electromagnetic inverse problem", Human Brain Mapping **7**, 195–212 (1999).
[45] S. Oh, A. B. Milstein, R. P. Millane, C. A. Bouman, and K. J. Webb, "Source-detector calibration in three-dimensional Bayesian optical diffusion tomography", J. Opt. Soc. Am. A, to appear.
[46] I. Kwee, "Towards a Bayesian framework for optical tomography", Ph. D. thesis, Department of Medical Physics and Bioengineering, University College London (1999).
[47] M. J. Eppstein, D. E. Dougherty, T. L. Troy, and E. M. Sevick-Muraca, "Biomedical optical tomography using dynamic parameterization and Bayesian conditioning on photon migration measurements", Appl. Opt. **38** 2138–50 (1999).
[48] K. J. Friston, "Bayesian Estimation of Dynamical Systems: An Application to fMRI", NeuroImage **16**, 513–530 (2002).
[49] D. A. Boas, T. J. Gaudette, and S. R. Arridge, "Simultaneous imaging and optode calibration with diffuse optical tomography", Opt. Express **8**, 263–270 (2001).
[50] David J.C. MacKay, *Information Theory, Inference, and Learning Algorithms*, Ch. 3, available at http://www.inference.phy.cam.ac.uk/mackay/itprnn/book.html
[51] J. Berger, *Statistical Decision theory and Bayesian Analysis* (Springer, 1985).
[52] S. James Press, *Bayesian Statistics: Principles, Models, and Applications*, Wiley Series in Probability and Statistics (Wiley, 1989).
[53] D. S. Sivia, *Data Analysis: a Bayesian tutorial* (Oxford

University Press, 1996).

[54] S. F. Gull, Bayesian inductive inference and maximum entropy, in *Maximum entropy and Bayesian methods in science and engineering, vol 1: Foundations*, G. R. Erickson and C. R. Smith, eds., (Kluwer, 1988).

[55] Neal, R. M., *Probabilistic inference using Markov chain Monte Carlo methods*, Technical Report CRG-TR-93-1, Dept. of Computer Science, University of Toronto (1993). Available at http://www.cs.toronto.edu/∼radford/review.abstract.html

[56] S. R. Arridge and M. Schweiger, "A gradient-based optimisation scheme for Optical Tomography", Opt. Express **2**, 213–226 (1998).

[57] A. H. Hielscher, A. D. Klose, and K. M. Hanson, "Gradient-based iterative image reconstruction scheme for time-resolved optical tomography", IEEE Transactions on Medical Imaging **18**, 262–271 (1999).

[58] F. Gao, H. Zhao, and Y. Yamada, "Improvement of image quality in diffuse optical tomography by use of full time-resolved data", Appl. Opt. **41** 778–791 (2002).

[59] A. M. Dale, B. Fischl, and M. I. Sereno, "Cortical surface-based analysis I. Segmentation and surface reconstruction", NeuroImage **9**, 179–194 (1999).

[60] E. Okada, M. Firbank, M. Schweiger, S. R. Arridge, M. Cope, and D. T. Delpy, "Theoretical and experimental investigation of near-infrared light propagation in a model of the adult head", Appl. Opt. **36**, 21–31 (1997).

[61] F. Bevilacqua, D. Piguet, P. Marquet, J. D. Gross, B. J. Tromberg, C. Depeursinge, "In vivo local determination of tissue optical properties: applications to human brain", Appl. Opt. **38**, 4939–50 (1999).

[62] J. C. Tamraz and Y. G. Comair, *Atlas of Regional Anatomy of the Brain Uisng MRI: with functional correlations*, (Springer 2000).

[63] J. Ripoll, S. R. Arridge, and M. Nieto-Vesperinas, "Effect of roughness in non-diffusive regions within diffusive media", J. Opt. Soc. Am. A. **18**, 940–947 (2001).

[64] F. E. Schmidt, M. E. Fry, E. M. C. Hillman, J. C. Hebden, and D. T. Delpy, "A 32-channel time-resolved instrument for medical optical tomography", Rev. Sci. Instruments **71**, 256–265 (2000).

[65] K. M. Yoo, F. Liu, and R. R. Alfano, "When does the diffusion approximation fail to describe photon transport in random media?", Phys. Rev. Lett. **64**, 2647 (1990).

[66] A. Ishimaru, *Wave Propagation and Scattering in Random Media, vol. 1* (Academic, New York, 1978).

[67] A. H. Hielscher, R. E. Alcouffe, and R. L. Barbour, "Comparison of finite-difference transport and diffusion calculations for photon migration in homogeneous and heterogeneous tissues", Phys. Med. Biol. **43**, 1285–1302 (1998).

[68] T. B. Durduran, B. Chance, A. G. Yodh, and D. A. Boas, "Does the photon diffusion coefficient depend on absorption?", J. Opt. Soc. Am. A **14**, 3358–65 (1997).

[69] Note that for us this acronym does not imply association with hyperbolic equations. We are evolving a parabolic equation.

[70] H. Dehghani, S. R. Arridge, M. Schweiger, and D. T. Delpy, "Optical tomography in the presence of void regions", J. Opt. Soc. Am. A **17**, 1659–70 (2000).

[71] S. R. Arridge, M. Hiraoka, and M. Schweiger, "Statistical basis for the determination of optical pathlength in tissue", Phys. Med. Biol. **40**, 1539–58 (1995).

[72] W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge University Press, 1992), available at http://lib-www.lanl.gov/numerical/bookcpdf.html

[73] S. J. Press, *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference*, 2nd Ed. (Krieger Publishing, Florida, 1982).

[74] R. D. Richtmeyer and K. W. Morton, *Difference Methods for Initial-Value Problems*, (John Wiley & Sons, NY, 1967).

[75] J. W. Thomas, *Numerical Partial Differential Equations: Finite Difference Methods*, (Springer-Verlag, 1995).