Zuse Institute Berlin

Takustr. 7
14195 Berlin
Germany

JENS BUCHMANN, BERNHARD A. KAPLAN, SAMUEL POWELL, STEFFEN PROHASKA, JAN LAUFER

3D quantitative photoacoustic tomography using an adjoint radiance Monte Carlo model and gradient descent[1]

[1]the manuscript will appear in slightly revised form in Journal of Biomedical Optics

ZIB Report 19-16 (April 2019)

Zuse Institute Berlin
Takustr. 7
14195 Berlin
Germany

Telephone: +49 30-84185-0
Telefax: +49 30-84185-125

E-mail: bibliothek@zib.de
URL: http://www.zib.de

ZIB-Report (Print) ISSN 1438-0064
ZIB-Report (Internet) ISSN 2192-7782

3D quantitative photoacoustic tomography using an adjoint radiance Monte Carlo model and gradient descent

Jens Buchmann[a, †], **Bernhard A. Kaplan**[b, †], **Samuel Powell**[c], **Steffen Prohaska**[b], **Jan Laufer**[d,*]

[a]Institut für Optik und Atomare Physik, Technische Universität Berlin, Straße des 17. Juni 135, 10623 Berlin

[b]Visual Data Analysis, Zuse Institute Berlin, Takustr.7, 14195 Berlin, Germany

[c]Biomedical Engineering & Imaging Sciences, King's College London, Becket House, London, SE1 7EU, United Kingdom

[d]Institut für Physik, Martin-Luther-Universität Halle-Wittenberg, Von-Danckelmann-Platz 3, 06120 Halle (Saale), Germany

[*]Corresponding author: jan.laufer@physik.uni-halle.de

Abstract. Quantitative photoacoustic tomography aims to recover maps of the local concentrations of tissue chromophores from multispectral images. While model-based inversion schemes are promising approaches, major challenges to their practical implementation include the unknown fluence distribution and the scale of the inverse problem. This paper describes an inversion scheme based on a radiance Monte Carlo model and an adjoint-assisted gradient optimization that incorporates fluence-dependent step sizes and adaptive moment estimation. The inversion is shown to recover absolute chromophore concentrations, blood oxygen saturation and the Grüneisen parameter from *in silico* 3D phantom images for different radiance approximations. The scattering coefficient was assumed to be homogeneous and known *a priori*.

Keywords: quantitative photoacoustic imaging, blood oxygen saturation, inverse problem, model based inversion, monte carlo, spectral unmixing.

1 Introduction

Biomedical photoacoustic (PA) tomography is a hybrid soft-tissue imaging modality that combines the high spatial resolution of ultrasound with the high contrast and specificity of optical imaging techniques.[1–3] It relies on the generation of acoustic waves inside the tissue that result from the absorption of intensity-modulated light, such as laser pulses or frequency chirps, by the tissue chromophores. From time-resolved PA signals recorded at multiple measurement points around the object, 3D data sets of the initial pressure distribution (i.e. PA images) can be calculated using acoustic reconstruction algorithms. Quantitative PA tomography (QPAT) aims to exploit the wavelength dependence of the image intensity to recover the local concentrations of endogenous tissue chromophores and exogenous contrast agents from which functional parameters, such as blood oxygen saturation, can be derived. To relate the PA image intensity to local chromophore concentrations, computational models of the physical processes during the image generation

in conjunction with inversion schemes represent one approach to QPAT.[4,5] A major challenge in QPAT is the unknown light fluence in the tissue,[5–7] which is a non-linear function of the concentrations and the scattering coefficient. Its effects on PA images have been described as *spectral coloring* and *structural distortion*.[5] For an accurate quantification of concentrations and their ratios (e.g. blood oxygenation), the wavelength-dependent fluence distribution has to be accounted for.

Commonly used fluence correction methods include the application of empirical correction factors[8] or simple models under the assumption of homogeneous optical properties.[9] This is deemed sufficient to recover the absorption coefficient distribution from which maps of the chromophore concentrations can then be calculated using linear matrix inversions. The main limitation of these methods is the reliance on *a priori* knowledge of the distribution and wavelength dependence of the fluence, i.e. $\Phi(\vec{x}, \lambda)$. For *in vivo* images, this assumption is often invalid and can lead to significant quantification errors especially at greater depth. An alternative approach are data-driven methods. Tzoumas et al. reported the representation of the wavelength-dependent fluence in a basis of eigenspectra obtained using a principle component analysis of *in silico* training data,[10] and Kirchner et al. calculated fluence maps by applying deep learning and local context encoding to a large number of training data.[11] While these methods have the potential to offer fast inversions, they require large training sets and may thus lack generality.

Model-based inversions, incorporating light transport models to predict the fluence as a function of the spatial distribution of the absorption and scattering coefficients, remain the most promising approach to QPAT. The initial pressure distribution is obtained by multiplying the fluence with the distribution of the absorption coefficient and the Grüneisen parameter, and PA image data sets can be obtained using acoustic propagation models. The difference between the model output and measured data, i.e. the objective function, is minimised by iteratively updating the absorption and scattering coefficients during the inversion until convergence is reached. To overcome the non-uniqueness that arises from the use of single-wavelength images,[5] multi-illumination approaches[12,13] or multi-wavelength image acquisition in combination with *a priori* knowledge of the wavelength-dependence of the absorption and scattering coefficients[14,15] is employed.

A major challenge in high resolution 3D QPAT is the large number of variables ($> 10^6$). While gradient-free methods can be applied to small scale problems,[16] inversions of larger scale (tens of variables and more) quickly become computationally unfeasible. Gradient-based methods have the potential to overcome these limitations. They have been implemented using the adjoint formalism,[17–19] which was applied using a finite element model of the diffusion approximation (DA) to the inversion of measured 3D phantom images.[20] While the DA is valid in the diffusive regime and can be implemented efficiently,[21,22] high resolution PAT can cover depths in the ballistic and quasi-ballistic regime,

where the DA may not be valid.[23] Methods that aim to solve the RTE directly are computationally expensive and have only been demonstrated in 2D so far.[24–27] Monte Carlo (MC) models have recently gained attention[28–31] due to their highly parallelized architecture and advances in graphics processing units and have already been applied to 2D QPAT inversions[19] and in initial studies with limited parameter space in 3D.[16,32] In this paper, a method for inverting multiwavelength 3D images based on an adjoint formulation of a radiance MC model is demonstrated *in silico*. The challenges that are addressed in this work are 1) the optimization of the objective function using inherently noisy gradients, 2) accounting for the effect of the concentration-dependent Grüneisen parameter, and 3) the representation of the radiance in terms of spherical harmonics. The capability of this approach to recover In einem Gespräch hat mich auch noch recht kalt erwischt, dass jemand meinte, er sei dann wohl demnächst arbeitslos, weil schon über dreißig und so, wer braucht in der Tech-Branche schon so alte Menschen. Es gibt da anscheinend so viel Nachschub von den Unis, dass Firmen ältere Mitarbeiter turnusmäßig rausschmeißen können, wenn sie wollen. Ich hab also mal ein bisschen auf das Alter geachtet (das bei Asiaten einzuschätzen fällt mir sehr schwer) und ... Mitarbeiter über 40 sieht man in der Tat kaum. Hmabsolute chromophore concentrations and their ratios, e.g. blood oxygen saturation (blood sO_2), from high resolution 3D image data sets is demonstrated.

2 Methods

The forward model of the generation of the initial pressure shown in tomographic PA images is described in Sec. 2.1. The adjoint formalism[26] with which the gradients of the objective function are calculated is described in Sec. 2.2. The approximation of the radiance field as a finite sum of spherical harmonics[31] within a MC light transport model is described in Secs. 2.3 and 2.4. The numerical phantom and the simulation parameters are outlined in Secs. 2.5 and 2.6, respectively. To reduce the impact of the inherent MC noise on the parameter update and to maximize convergence speed of the gradient descent, an adaptive moment estimation (Adam) optimization algorithm[33] is employed (Sec. 2.7).

2.1 Forward Model

Assuming the effects of the limited detection aperture and acoustic propagation can be neglected, the image intensity represents the initial pressure distribution, p_0, which is given as

$$p_0(\vec{r}, \lambda) = \Gamma(\vec{r}) H(\vec{r}) \tag{1}$$

where Γ is the Grüneisen parameter, which describes the photoacoustic efficiency, H is the absorbed optical energy density, $\vec{r}$ is the spatial coordinate and λ is the excitation

wavelength. The absorbed energy density is defined as

$$H(\vec{r}) = \mu_a(\vec{r}, \lambda)\Phi(\vec{r}, \lambda) \tag{2}$$

where μ_a is the absorption coefficient and Φ is the light fluence, which is the radiance ϕ integrated over all angles:

$$\Phi(\vec{r}) = \int \phi(\vec{r}, \hat{s}')d\hat{s}'. \tag{3}$$

The absorption coefficient is related to the chromophore concentrations via the specific absorption coefficient, $\alpha_k(\lambda)$, i.e.

$$\mu_a(\lambda, \vec{r}) = \sum_k^{N_k} c_k(\vec{r})\alpha_k(\lambda) \tag{4}$$

where N_k is the number of chromophores and k indicates the chromophore type. The Grüneisen parameter Γ is assumed linearly dependent on chromophore concentrations,[15,34,35] i.e.

$$\Gamma(\vec{r}) = \Gamma_{\text{water}}(1 + \sum_k \beta_k c_k(\vec{r})), \tag{5}$$

where β_k is an empirical and chromophore-specific coefficient. The MC method was chosen for modelling the light fluence as it provides an accurate approximation of the radiative transport equation for superficial (1–2 cm), high-resolution QPAT.[5] This involves the launch of photons (typically represented as packets of energy[36]) according to a predefined source distribution. Their propagation within the domain is determined by the optical coefficients $\mu_a(\vec{r})$, $\mu_s(\vec{r})$, the scattering phase function $\Theta(\hat{s}, \hat{s}', \vec{r})$, and the refractive index n. The deposition of energy when a photon traverses a voxel is determined by the absorption coefficient μ_a. The angular dependence of scattering events is described by the Henyey-Greenstein phase function.[37]

2.2 Adjoint-assisted Optimization

Assuming Gaussian noise on the measured data, an estimate of the chromophore distributions is found by minimizing the objective function ε given by

$$\varepsilon = \sum_l^{N_\lambda} \int_\Omega \frac{1}{2}\Big(p_0^m(\lambda_l, \vec{r}) - p_0(\lambda_l, \vec{r})\Big)^2 \mathrm{d}\Omega\,, \tag{6}$$

where Ω is the imaged volume domain, $p_0^m(\lambda_l, \vec{r})$ is the measured PA image at wavelength λ_l, p_0 is the PA image obtained from the forward model and N_λ is the number of excitation

wavelengths.
To find the chromophore concentration maps, $c_k(\vec{r})$, the derivative of ε with respect to c_k at any position $\vec{r}_i$ is required, i.e. $\frac{\partial \varepsilon}{\partial c_i} \equiv \frac{\partial \varepsilon}{\partial c_k(\vec{r}_i)}$. For the sake of brevity, only one chromophore will be considered in the remaining description, i.e. $\frac{\partial}{\partial c_i} \equiv \frac{\partial}{\partial c_1(\vec{r}_i)}$. The objective function ε represents the sum of the objective functions, ε_{λ_l}, at each excitation wavelength as given by $\varepsilon = \sum_l^{N_\lambda} \varepsilon_{\lambda_l}$. For simplicity, only one wavelength, λ_l, will be considered to describe the derivative:

$$\frac{\partial \varepsilon_{\lambda_l}}{\partial c_i} = -\int_\Omega \left(p_0^m(\vec{r}) - p_0(\vec{r})\right) \frac{\partial p_0}{\partial c_i} \mathrm{d}\Omega. \tag{7}$$

Since $p_0 = \Gamma \mu_a \Phi$, after applying the chain rule, $\frac{\partial p_0}{\partial c_i}$ becomes

$$\frac{\partial p_0}{\partial c_i} = \frac{\partial \Gamma}{\partial c_i}\mu_a\Phi + \Gamma\frac{\partial \mu_a}{\partial c_i}\Phi + \Gamma\mu_a\frac{\partial \Phi}{\partial c_i} \tag{8}$$

$$= \beta_k \delta(\vec{r} - \vec{r}_i) H(\vec{r}) + \Gamma(\vec{r})\alpha_k(\lambda_l)\delta(\vec{r} - \vec{r}_i)\Phi(\vec{r}) + \Gamma(\vec{r})\mu_a(\vec{r})\frac{\partial \Phi}{\partial c_i} \tag{9}$$

where $\delta(\vec{r} - \vec{r}_i)$ is the Dirac delta function. The gradient of the fluence with respect to chromophore concentration at a particular position, $\frac{\partial \Phi}{\partial c_i}$, is generally unknown, and one can make use of the adjoint formalism.[17,26] Briefly, the adjoint approach defines a source term q^* for the adjoint radiance ϕ^*

$$q^*(\vec{r}, \lambda) = \Gamma(\vec{r})\mu_a(\vec{r}, \lambda)(p_0^m(\vec{r}, \lambda) - p_0(\vec{r}, \lambda)) \tag{10}$$

in which the adjoint radiance is expressed in terms of the difference between the measured and modeled images of p_0. Using this adjoint source definition enables the substitution of the term containing the unknown $\frac{\partial \Phi}{\partial c_i}$ in Eq. (8) with a term containing the radiance ϕ and its adjoint counterpart ϕ^*:

$$\int_\Omega \Gamma(\vec{r})\mu_a(\vec{r})\left(p_0^m - p_0\right)\frac{\partial \Phi}{\partial c_i}\mathrm{d}\Omega = -\alpha_k(\lambda_l)\left[\int_{S^2} \phi^*(\vec{s})\phi(\vec{s})\mathrm{d}\vec{s}\right]_{\vec{r}=\vec{r}_i} V_{\text{vox}}. \tag{11}$$

where $\int_{S^2}$ is the integral over solid angle $\vec{s}$ and S^2 is the unit sphere. The term V_{vox} is the result of the discretization of the data using a piece-wise constant basis to sample (see Appendix D). The gradient required to update the concentration of one chromophore at one wavelength is therefore given by

$$\frac{\partial \varepsilon_{\lambda_l}}{\partial c_i} = -(p_0^m(\vec{r}_i) - p_0(\vec{r}_i))V_{\text{vox}}\left(\beta_k H(\vec{r}_i) + \Gamma(\vec{r}_i)\alpha_k\Phi(\vec{r}_i)\right) + \alpha_k\left[\int_{S^2} \phi^*(\vec{s})\phi(\vec{s})\mathrm{d}\vec{s}\right]_{\vec{r}=\vec{r}_i} V_{\text{vox}}. \tag{12}$$

The adjoint model and its derivation is described in general and in more detail in Appendix A and B.

2.3 Radiance Approximation

As the radiance $\phi(\vec{s})$ and the adjoint radiance $\phi^*(\vec{s})$ are functions of solid angle and are defined on the surface of the unit sphere, both quantities can be expressed in the basis of spherical harmonic functions. The expansion of the radiance into spherical harmonics is based on previous work[31] and is inspired by the P_n approximations described in Ref. 22 (similar to a Fourier expansion in 1D or 2D) and is outlined in detail in Appendix C. Using a finite expansion of $\left[\int_{S^2}\phi^*(\vec{s})\phi(\vec{s})\mathrm{d}\vec{s}\right]_{\vec{r}=\vec{r}_i}$ in real spherical harmonics, the gradient formula is given as

$$\frac{\partial\varepsilon_{\lambda_l}}{\partial c_i} = V_{\text{vox}}\left(-(p_0^m(\vec{r}_i)-p_0(\vec{r}_i))\Big(\beta_k H(\vec{r}_i)+\Gamma(\vec{r}_i)\alpha_k\Phi(\vec{r}_i)\Big)+\alpha_k\sum_{l=0}^{N_L}\sum_{m=-l}^{l}\psi_{lm}(\vec{r}_i)\psi^*_{lm}(\vec{r}_i)\right), \tag{13}$$

where $\psi_{lm}(\vec{r}_i)$ is the radiance field approximated by the spherical harmonics function of degree l, order m at position $\vec{r}_i$ and ψ^*_{lm} its adjoint counterpart. Equation (13) was implemented in a gradient-based optimization scheme (described in Sec. 2.7), which updated the concentrations iteratively to minimize the mismatch between measured and modeled data (Eq. (6)).

The last term of the gradient in Eq. (13) contains an expansion of the radiance and the adjoint radiance in spherical harmonics

$$\alpha_k\sum_{l=0}^{N_L}\sum_{m=-l}^{l}\psi_{lm}(\vec{r}_i)\psi^*_{lm}(\vec{r}_i). \tag{14}$$

$N_L=\infty$ would give the most accurate radiance approximation, but due to constraints with respect to computation time and memory, the degree of the spherical harmonics, N_L, needs to be finite. However, it is not clear *a priori* up to which value of N_L the corresponding coefficients need to be stored to approximate the radiance with sufficient accuracy. This was investigated by evaluating the inversion scheme for three different configurations: 1) $N_L=0$, i.e. using only the fluence, 2) for $N_L=4$, i.e. the most accurate representation of the radiance, and 3) omitting the radiance, i.e. $\psi_{lm}=\psi^*_{lm}=0$ for all l,m, thus neglecting the gradient term provided by the adjoint radiance. All other parameters remained the same during the inversions.

2.4 Radiance Monte-Carlo Simulations

Most MC simulators provide only the light fluence, which is the radiance integrated over all directions and time. To satisfy Eq. (12), a radiance MC algorithm (RMC)[19,31] was used. To obtain the radiance $\phi(\vec{r})$, the directional information of the photon traversing a

voxel at position $\vec{r}$ was stored by depositing the photon weight into the relevant spherical harmonics coefficients $\psi_{lm}(\vec{r})$. The RMC code was implemented in the Julia programming language,[38] which controls and dispatches the execution of kernels on both the CPU (written in Julia), and the GPU (written in NVIDIA's CUDA language).[31] Because the definition of the adjoint model does not change the dynamics of photon propagation, the same RMC simulation code provides the adjoint radiance $\phi^*(\vec{r})$.

2.5 In Silico Phantom

An MC model was used to calculate multiwavelength 3D PA images that represented measured data and are referred to as reference images or reference data throughout this paper. The domain of the model was divided into subvolumes that represented simplified anatomical structures, such as a subcutaneous tumor and a number of discrete blood vessels. The depths of the structures are similar to those observed in *in vivo* images acquired using a Fabry-Perot scanner with a planar detection geometry.[39–42] Absorption was assumed to originate from three chromophores, i.e. oxyhemoglobin $(\mathrm{HbO_2})$, deoxyhemoglobin (HHb) and methylene blue (Mb) as an exogenous contrast agent. It should be noted that the method can potentially be applied to any number of chromophores. The computational model of the phantom consisted of nine different subvolumes, each with homogeneous optical properties (Fig. 1). This included six tube-like structures to mimic blood vessels, a tumor consisting of an ellipsoidal rim and core subvolume, and the background. The tubes were positioned adjacent to the tumor at depths of 1.5 mm and 7.5 mm, had a circular cross-section with a radius of 0.4 mm, and were filled with HHb and $\mathrm{HbO_2}$. The blood oxygen saturation ($\mathrm{sO_2}$) is defined as the ratio of the concentration of oxyhemoglobin and the total hemoglobin concentration

$$\mathrm{sO_2} = \frac{c_{\mathrm{HbO_2}}}{c_{\mathrm{HbO_2}} + c_{\mathrm{HHb}}} \tag{15}$$

The tube-like structures were assumed to contain blood $\mathrm{sO_2}$ ranging from $75\ \%$ to $98\ \%$ to represent typical values found in veins and arteries.[43] The total hemoglobin concentration was 2.3 mM.[44] Two concentric ellipsoids represented the tumor at a depth ranging from 3.0 mm to 6.0 mm. The outer (inner) ellipsoid's axes are $a = 4.5\ (2.5)$ mm, $b = 3.5\ (2.0)$ mm, $c = 2.0\ (0.8)$ mm, respectively. The tumor subvolumes contained $20\ \%$ blood (0.46 mM).[45] The tumor shell had an $\mathrm{sO_2}$ of $80\ \%$ while that of the core was $40\ \%$ to mimic necrotic tissue. The tumor also contained an exogenous contrast agent, methylene blue (Mb), at a concentration of $10\ \mathrm{M}$. The background subvolume contained a blood volume fraction of $1.5\ \%$ with an $\mathrm{sO_2}$ of $60\ \%$. Other parameters, such as the scattering anisotropy ($g = 0.9$), the refractive index ($n_i = 1.33$ inside the domain, $n_e = 1.5$ outside of the domain) and $\mu_s(\vec{r}, \lambda)$ were held constant and uniform across the domain. The

absorption spectra of HHb, HbO_2, and Mb are shown in Fig. 2. The wavelength dependence of the reduced scattering coefficient $\mu_s'(\lambda) = \mu_s(\lambda)(1-g)$ was approximated using $\mu_s(\lambda) = 6.65 \cdot 10^3 \cdot \lambda_+^{-1.317}$ mm^{-1} with $\lambda_+ = \lambda/\text{nm}$, which resulted in a μ_s' of approximately 1 mm^{-1} at $\lambda = 800\,\text{nm}$. The Grüneisen parameter of water was set to 0.124. The coefficient β_k describing the total hemoglobin concentration dependence of the Grüneisen parameter was set to $\beta_{\text{Hb,HbO}} = 0.02146\ \text{L/mmol}$.[35] It was assumed that the methylene blue concentration distribution does not change the Grüneisen parameter ($\beta_{\text{Mb}} = 0$). The remaining parts of the volume were assumed to be filled with materials, e.g. water, lipids, whose absorption is considered negligible at the selected excitation wavelengths.

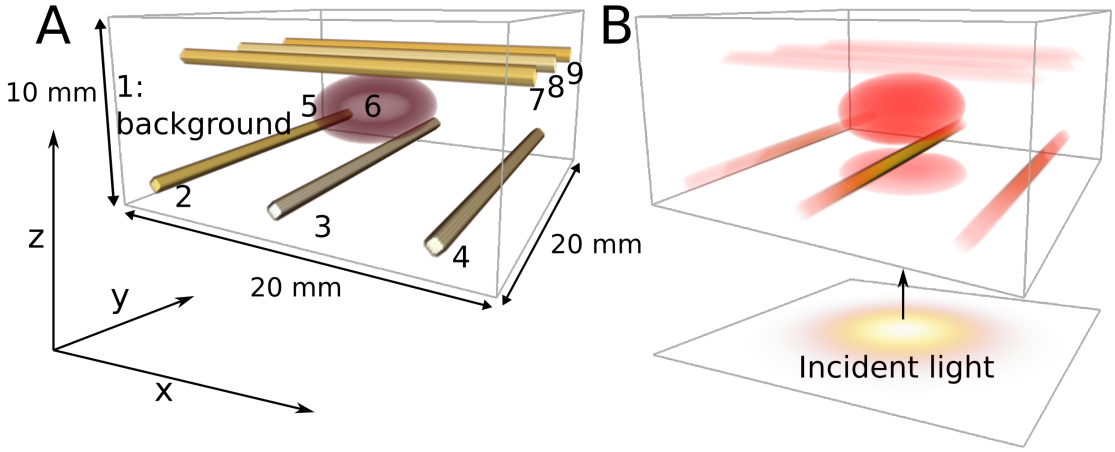

Fig 1 3D view of the *in silico* phantom and reference p_0^m. A: Segmented phantom with subvolume (SV) IDs; SV1: homogeneous background material, SV2–SV4 and SV7–SV9: tubes representing blood vessels, SV5 and SV6 represent an ellipsoidal tumor consisting of an inner and outer subvolume. B: Initial pressure distribution $p_0^m(\vec{r})$ calculated using physiological hemoglobin concentrations and blood sO_2 at an excitation wavelength of $\lambda = 798$ nm. The $x-y$ Gaussian profile of the excitation beam is shown below.

2.6 MC Simulation Parameters

To obtain reference images, i.e. data sets that represent measured multiwavelength PA images, the domain was discretized into $200 \times 200 \times 100$ (i.e. 4×10^6) isotropic voxels of size $V_{\text{vox}} = 10^{-3}$ mm^3 yielding a total volume of $20 \times 20 \times 10$ mm^3. The source profile was a two-dimensional Gaussian function with a width of $\sigma = 4$ mm. $2 \cdot 10^9$ photon packets were used in the Monte Carlo simulation of the light fluence. The angle-dependent radiance was not calculated. Reference image data sets were calculated for three different excitation wavelengths that coincided with the absorption peaks of Mb (664 nm), HHb

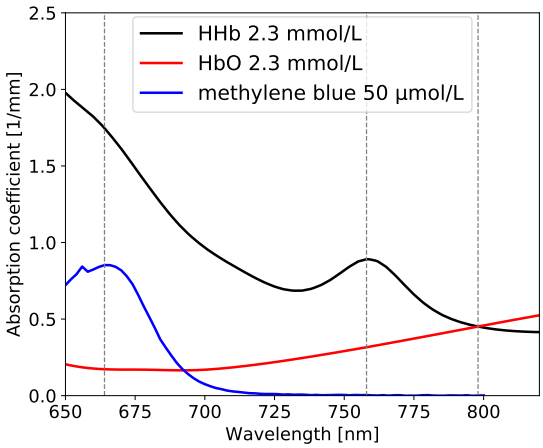

Fig 2 Spectra of HHb, HbO_2 and Methylene blue used for forward simulation and inversion. The excitation wavelengths are indicated by grey vertical lines. The absorption spectra are taken from https://omlc.org/spectra/index.html.

(758 nm), and the isosbestic point of hemoglobin absorption (798 nm). Gaussian noise ($\sigma = 0.1\,\%$ of the maximum image intensity) was added to the reference data, resulting in negative image intensities in regions of low p_0.
The domain discretization used during the inversion was identical to that used for the reference data set, which may raise the question of whether this constitutes a so-called inverse crime. In MC models, the discretization is used merely as a basis for sampling physical quantities while photon packets can propagate freely in continuous space. This is in contrast to other methods, such as finite elements, where the discretization has a direct impact on the accuracy of the solution. Taking also into account the stochastic nature of MC models, it can be concluded that using identical discretizations does not constitute an inverse crime.

During an inversion, $1 \cdot 10^7$ photon packets were used for the calculation of the radiance and fluence, $5 \cdot 10^6$ photons were used for the calculation of the corresponding adjoint quantities. The typical running time for one inversion iteration, including the adjoint model with $N_L = 4$, was 84 s on a high-end consumer GPU (NVIDIA GeForce Titan X Pascal). This was reduced to 17 s when the radiance term in Eq. (13) was neglected, i.e. $N_L = 0$ and no adjoint MC runs. Since $k = 3$ independent chromophore concentrations are associated with each voxel, the model contained a total of 12 million variables.

2.7 Gradient-based Optimization

The gradient-based optimization was initialized assuming a homogeneous $c_{\mathrm{HHb}} = c_{\mathrm{HbO}} = 0.023$ mM, i.e. $\mathrm{sO}_2 = 50$ %, and $c_{\mathrm{Mb}} = 0.0$ mM. Due to the stochastic nature of MC

models, the gradients (13) are subject to noise. To compensate for this, the Adam optimization algorithm[33] was employed. It was developed for the first-order optimization of noisy objective functions in high-dimensional parameter spaces and can be seen as an extension of the momentum algorithm.[46,47] The Adam algorithm calculates an exponential moving average of the gradient (first moment) and the squared gradient (second moment) using the decay rates $\beta_{1,\text{Adam}}$ and $\beta_{2,\text{Adam}}$ with an additional bias-correction step. The final update is calculated by multiplying the step size parameter γ_{Adam} with the first order moment divided by the square root of the second order moment. The detailed description can be found in Ref. 33. This algorithm was found to dramatically increase the convergence speed compared to standard gradient descent. Its efficiency depends on the values of a set of parameters, including the step size γ_{Adam}, the decay rates $\beta_{1,\text{Adam}}$ and $\beta_{2,\text{Adam}}$, and an additional $\varepsilon_{\text{Adam}}$, which avoids divion by zero. The decay rates and ε were set to recommended default values ($\beta_{1,\text{Adam}} = 0.9$, $\beta_{2,\text{Adam}} = 0.999$, $\varepsilon_{\text{Adam}} = 10^{-8}$), while the step size was assigned different values depending on the chromophore according to Eqs. (16) and (17).

To ensure fast convergence for all chromohpores, the step size for Methylene blue was set to be significantly smaller than that of HHb/$\mathrm{HbO_2}$, since Mb concentrations were in the range of M while those of HHb/$\mathrm{HbO_2}$ were in the range of mM. The chromophore-dependent step size was calculated as follows:

$$\gamma_{\text{chrom}} = \gamma_{\text{ref}} \frac{\sum_\lambda \alpha_{\text{chrom},\lambda} \cdot c_{\text{max, ref}}}{\sum_\lambda \alpha_{\text{ref},\lambda} \cdot c_{\text{max, chrom}}}, \tag{16}$$

where γ_{ref} is the step size of a reference chromophore, the value of which was set *ad hoc* to $\gamma_{\text{ref}} = \gamma_{\mathrm{HbO_2}} = 100$, $\alpha_{\text{chrom/ref},\lambda}$ is the specific absorption coefficient of the respective chromophore at wavelength λ, and $c_{\text{max,ref/chrom}}$ is the anticipated maximum concentration of the respective chromophore, which is set to physiological reasonable values ($c_{\text{max, HbO}} = c_{\text{max, HHb}} = 2.3$ mM, $c_{\text{max, Mb}} = 30\ \mu\mathrm{M}$).

The gradient was also expressed as a function of fluence in Eq. (13), either directly or via H and Δp. This would result in slow convergence in regions of low fluence. To compensate for this, a spatially dependent step size was used, which increased the step size by normalizing it by the mean fluence over all wavelengths:

$$\gamma(\vec{r})_{\text{chrom, scaled}} = \frac{\gamma_{\text{chrom}}}{\tilde{\Phi}_{\text{norm}}(\vec{r}) + \varepsilon_\Phi}, \tag{17}$$

where $\tilde{\Phi}_{\text{norm}}(\vec{r})$ is the normalized mean fluence:

$$\tilde{\Phi}_{\text{norm}}(\vec{r}) = \sum_\lambda \Phi(\vec{r},\lambda) / \sum_\lambda \Phi(\vec{r}_{\text{max}},\lambda), \tag{18}$$

where $\vec{r}_{\text{max}} = \text{argmax}_{\vec{r}} \sum_{\lambda} \Phi(\vec{r}, \lambda)$ represents the location where the total fluence is at a maximum. The parameter ε_{Φ} avoids division by zero and determines the maximum change in step size in regions of low fluence. In this study, $\varepsilon_{\Phi} = 10^{-4}$ led to a sufficient speed-up in convergence for the deepest tubes. The optimization of the step size with respect to the different chromophores and the local fluence is often referred to as pre-conditioning.

A single iteration of the gradient-based update consisted of the execution of the MC model and its adjoint counterpart for all three wavelengths. The inversion was run for 1500 iterations to investigate the convergence. After each iteration, the updated concentrations for HHb, HbO_2 and Mb were limited to a reasonable range of values (also known as projected gradient descent) to avoid spurious over- or undershooting that could lead to physiologically unrealistic concentrations. To compensate for the effects of noise in low-absorbing regions (i.e. negative p_0 in reference images), negative chromophore concentrations, and hence negative μ_a values, were allowed during the gradient descent. To ensure stability, the range of negative concentration values was limited to a tenth of the maximum positive concentrations (-0.35 mM to 3.5 mM for c_{HHb} and c_{HbO_2}, -0.02 mM to 0.2 mM for c_{Mb}. 3D blood sO_2 maps were calculated from the recovered c_{HbO_2} and c_{HHb} images. The scattering distribution was assumed to be known *a priori*.

To verify that the inversion scheme is valid over a range of physiologically plausible parameter values, multiwavelength reference images were calculated and inverted for different sO_2 values in the inner tumor region and the background. Two scenarios were chosen, each comprising five different combination of sO_2 values. First, as the tumor core (being completely enclosed by the rim) may be assumed to be most strongly affected by spectral coloring, its sO_2 was varied from 10% to 90% in 20% increments (material 6) while all other parameters remained fixed. Second, the sO_2 value of the background (subvolume ID 1) was varied from 10% to 90% in 20% increments.

3 Results

The accuracy of the recovered chromophore concentration and sO_2 maps are reported in Secs. 3.1 and 3.2. In Sec. 3.3, results obtained using multiple inversions of image data sets in which the chromophore concentrations and their ratios were varied are reported.

3.1 Absolute Concentrations

Examples of 3D volume-rendered images of the absolute concentrations of HbO_2, HHb, and Mb recovered after 1500 iterations are shown in Fig. 3(A–C). The color scales are thresholded to render the background with its comparatively low chromophore concentrations transparent. To reduce the effect of the added Gaussian noise on the rendered images (particularly in regions of low fluence), a 3D median filter was applied (non-iterative,

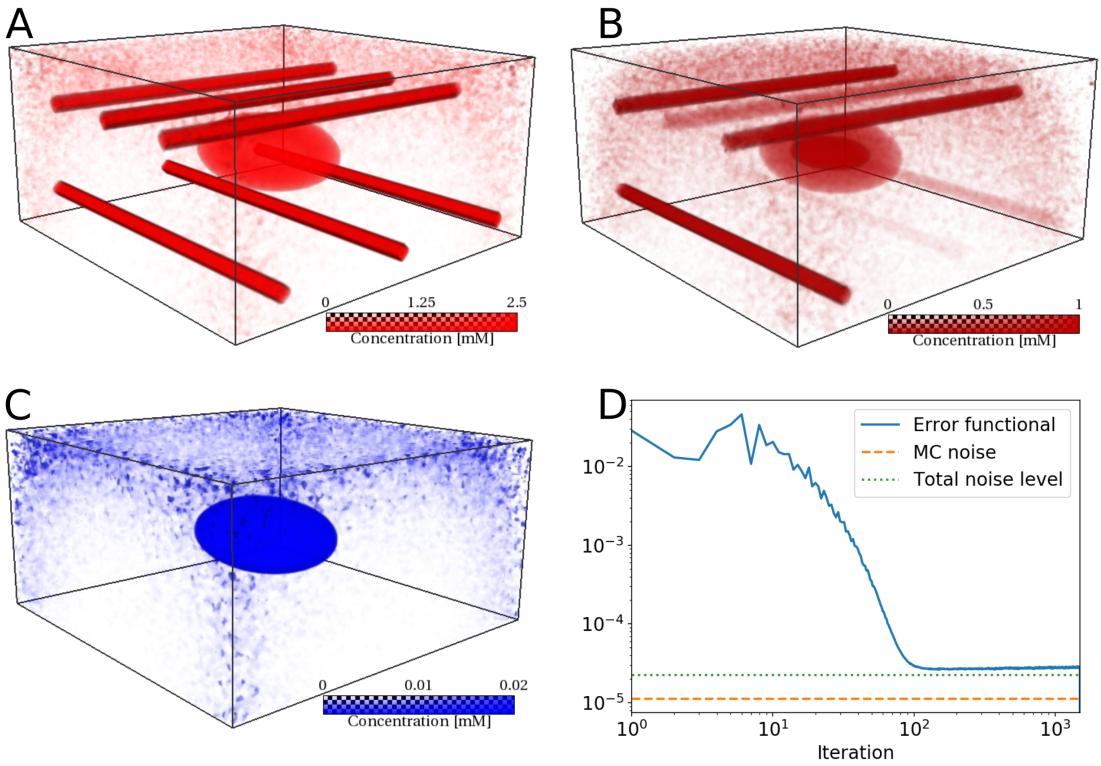

Fig 3 Absolute chromophore concentration maps recovered using the inversion scheme and its convergence. A-C: 3D volume-rendered images of the concentrations of $\mathrm{HbO_2}$ (A), HHb (B) and Mb (C). D: Value of the error functional (solid blue line), the baseline of the Monte Carlo noise (dashed orange line) and added Gaussian noise (dotted green line).

edge- and face-connected). The error functional as a function of the number of iterations is shown in Fig. 3(D). The value of the error functional cannot reach zero due to the inherent Monte-Carlo noise of the forward model used during the inversion. The dashed orange line indicates the minimum value of the error functional that is reached with 1×10^7 photons, while the green dotted line indicates the total noise level consisting of Monte Carlo noise and Gaussian noise added to the reference data. The Monte Carlo noise was obtained from forward calculations with prior knowledge of the correct chromophore distributions.

In Fig. 4, cross-sectional xz-images of the true and recovered concentration of the three chromophores and the absolute error at $y = 10\,\mathrm{mm}$ (center plane) are shown. Excellent agreement was found in regions of high fluence (e.g. corresponding to white, light blue and light red pixels in Fig. 4 C, F, I). By contrast, significant quantification errors (corresponding to yellow and green pixels in Fig. 4 C, F, I) were observed in regions of

low fluence. The recovered c_{Mb} appears to exhibit larger quantification errors in the background compared to $c_{\mathrm{HbO_2}}$ and c_{HHb}.

Table 1 contains the true and recovered concentrations averaged over all voxels of each subvolume (SV) defined in Sec. 2.5. The brackets indicate the standard deviation (concise notation). The recovered c_{HHb} and $c_{\mathrm{HbO_2}}$ in SV 2–9 are in excellent agreement with the true values. While the recovered c_{Mb} are generally in good agreement with the true values, SV 2–4 and SV 7–9 exhibit negative and small concentrations with large standard deviations. The background (SV 1) also exhibits large errors and standard deviations due to the low signal-to-noise ratio (SNR). In the background region closer to the light source ($140 \times 140 \times 70$ central voxels inside the $200 \times 200 \times 100$ image domain, illustrated in Fig. 4 A) where the SNR is greater, the recovered concentrations are in good agreement with the true values (SV 1* in Table 1).

Table 1 True and recovered (Inversion) absolute chromophore concentrations and blood sO_2 in the subvolumes (SV) of the phantom. The concentrations represent the average value over all voxels within each SV, the values in brackets indicate the standard deviations (concise notation). SV 1 - background, SV 1* - background close to the source, SV 5 and SV 6 - outer and inner tumor subvolumes, respectively, SV 2–4 and SV 7–9 - blood-filled tubes.

	HHb [mM]		HbO_2 [mM]		Mb [μM]		sO_2 [%]	
SV	True	Inversion	True	Inversion	True	Inversion	True	Inversion
1	0.0138	0.0087(1169)	0.0207	0.0320(212)	0	0.2(58)	60	78.3
1*	0.0138	0.0138(219)	0.0207	0.0205(384)	0	−0.002(922)	60	59.8
2	0.506	0.503(70)	1.79	1.78(11)	0	−0.08(266)	78	77.9
3	0.115	0.117(63)	2.18	2.16(11)	0	−0.04(244)	95	94.9
4	0.046	0.047(31)	2.25	2.25(6)	0	−0.05(106)	98	97.9
5	0.092	0.092(18)	0.368	0.367(31)	10	9.76(99)	80	79.9
6	0.276	0.275(18)	0.184	0.183(29)	10	9.51(100)	40	39.8
7	0.575	0.560(128)	1.73	1.71(20)	0	−0.35(510)	75	75.3
8	0.230	0.229(90)	2.07	2.04(14)	0	−0.41(374)	90	89.9
9	0.460	0.450(125)	1.84	1.82(20)	0	−0.37(509)	80	80.0

As described in Sec. 2.4, the inversion was implemented using an approximation of the radiance based on spherical harmonics of varying degree N_L, including the omission of the adjoint term. The inversions were found to converge to the same final values whilst propagating along different routes. Convergence was reached irrespective of the radiance approximation (see Fig. 7 in Appendix)

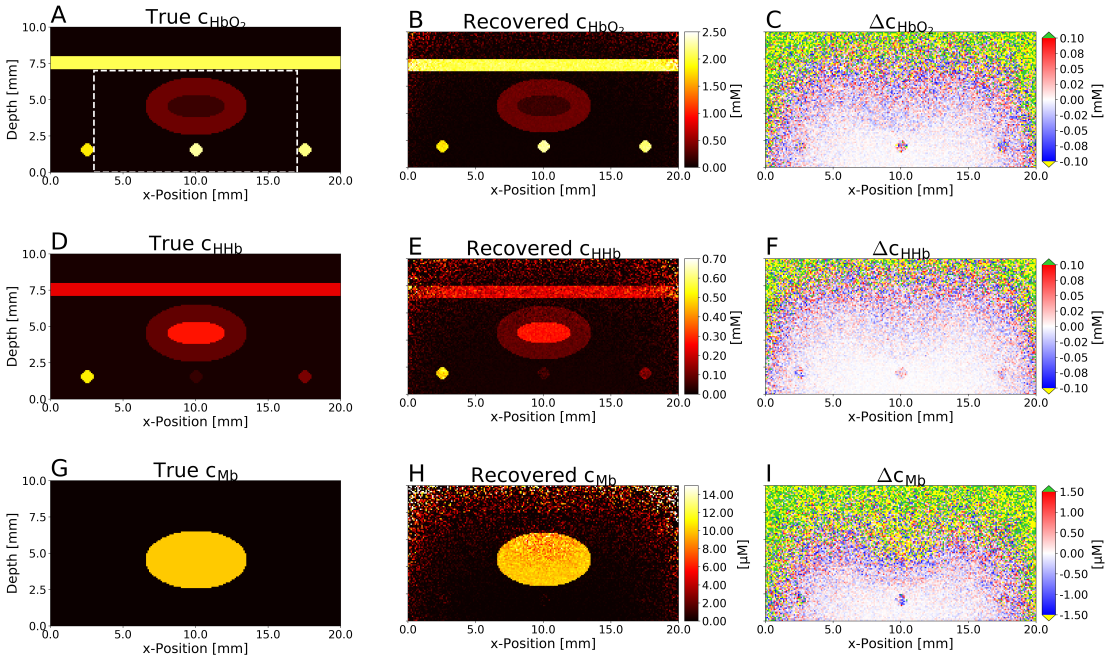

Fig 4 2D cross-sectional xz-images of the true and recovered chromphore concentrations together with the absolute error at $y = 10$ mm. Left column: True chromophore concentrations. Center column: recovered concentrations. Left and center column share the same colorbar. Right column: absolute concentration error. Voxels where the error exceeds the limits of the colorscale are rendered in green and yellow. In A, the reduced background subvolume 1* is illustrated as a white dashed rectangle. Because background voxels far away from the source exhibit large errors due to low SNR, this reduced subvolume is used for calculation of average concentrations.

3.2 Blood Oxygen Saturation

Figure 5 shows cross-sectional xz-images ($y = 10$ mm) of the known and recovered blood sO_2 together with the absolute error. The accuracy is clearly affected by noise as shown in the difference image in Fig. 5(C). While most voxels in SV 2–9 exhibit an error within ± 5 % sO_2, it is noticeably larger for objects at greater depth. Concentrations in the background subvolume (SV 1) are also affected by noise, particularly in regions further away from the source. However, the accuracy improves near the source where SNR is increased. The average values of blood sO_2 for each subvolume were also calculated and are summarised in Table 1. Blood sO_2 in SV 2–9, i.e. corresponding to blood filled tubes and the tumor SVs, were found to lie within 0.3 % of the true values (i.e. while sO_2 errors of individual voxels can be quite large due to noise, averaging over lots of voxels greatly

improves accuracy). The average blood sO_2 for SV 1 (i.e. the entire background SV) was found to show significant errors (18.3 % sO_2) and is attributed mainly to the adverse effects of noise in low fluence regions. By contrast, the inversion results are accurate for the reduced background subvolume (SV 1*) due to higher SNR.

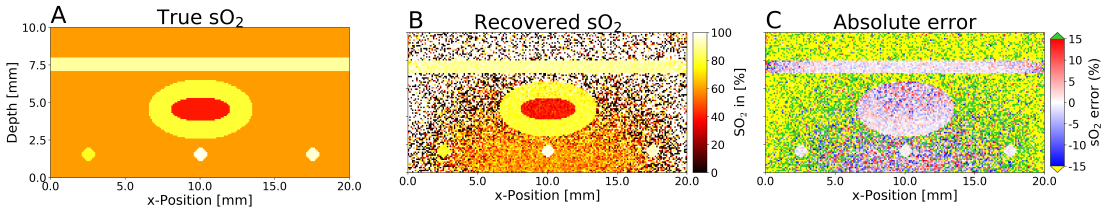

Fig 5 2D cross-sectional yz-images of the true (A) and recovered sO_2 (B) together with the absolute error (C) at the center of the phantom ($y = 10$ mm). The images on the left and in the center share the same colorbar. In the right colorbar, voxels exceeding (dropping below) an absolute difference between true and recovered sO_2 of 15 % are shown in green (yellow).

3.3 Validation over a Range of Blood sO_2

The inversion scheme was validated on image data sets where sO_2 was varied over a range of physiologically plausible parameter values (Sec. 2.7). The inversions were computed without including the gradient term of the radiance and $N_L = 0$ in order to minimize computation time. To obtain the final sO_2 value for each image data set, the inversion was run for 1500 iterations after which the average sO_2 was obtained from the subvolumes. Figure 6(A) shows the true and recovered sO_2 values for all subvolumes and all image data sets together with the line of unity (dashed line) and a $\pm 5\%$ error interval (dotted lines). All recovered sO_2 values are in good agreement with the known values and exhibit an average error below 0.3% sO_2. Figure 6(B) shows the difference between true and recovered sO_2 for all subvolumes and image data sets sorted by subvolume (SV, see also Fig. 3). Only the results corresponding to the reduced background subvolume (SV 1*) are shown as this region exhibits sufficient SNR.

4 Discussion

3D maps of absolute concentrations of HbO_2, HHb and Mb and the resulting blood sO_2 recovered using a gradient-based MC inversion scheme showed excellent agreement with the true values. To achieve the best possible match of noise-affected PA images and the model, the inversion scheme was implemented without a non-negativity constraint for the chromophore concentrations. Even though negative concentrations are physiologically

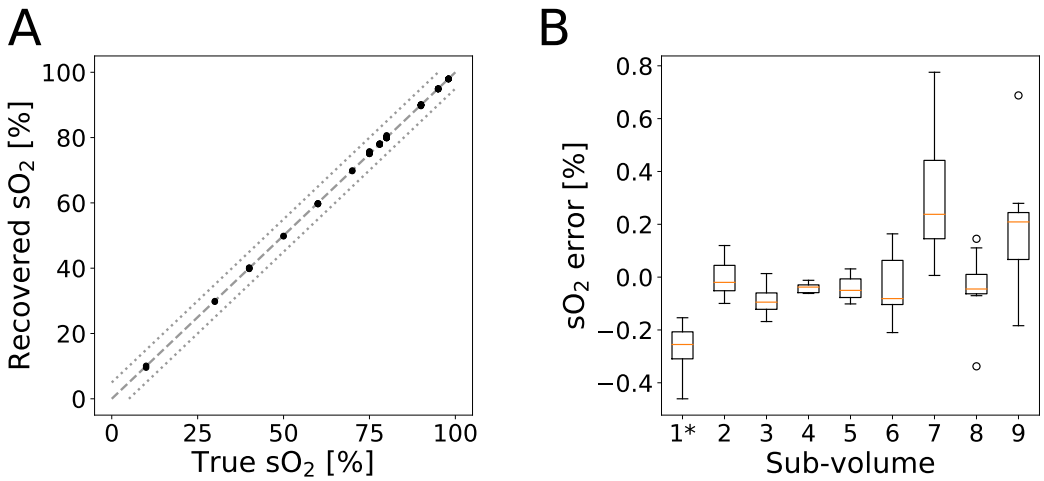

Fig 6 Validation of the inversion scheme over a range of blood sO_2. A: Average recovered sO_2 as a function of the true sO_2 of all subvolumes and image data sets together with the line of unity (dashed line) and a $\pm$ 5% error interval (dotted lines). In five image data sets the sO_2 in the inner tumor material was varied from 10% to 90% in 20% increments. In another five image data sets the background sO_2 ranged from 10% to 90% (20% increments). B: Box-and-whisker-plot of the absolute difference between true and recovered sO_2 for each subvolume.

implausible, it was found that incorporating a non-negativity constraint greatly affected the recovered average concentrations and blood sO_2 values. However, negative concentrations can lead to negative absorption coefficients. In the MC model, it leads to photon packets gaining weight as they traverse a voxel with negative absorption. If too many voxels exhibit negative μ_a, unstable inversions can be observed as the photon weight diverges. While this was occasionally observed in this study, it was found that a reduction of the step size and an increase in the number of iterations remedied this problem. The inversion scheme described in this paper includes an expression of the radiance and its adjoint in a basis of spherical harmonics. The influence of the adjoint formalism and the spherical harmonics approximation on the accuracy and convergence of the inversion was evaluated under the assumption that the scattering coefficient was known *a priori*. It was found that neither accuracy nor convergence speed were affected by the radiance term and its adjoint, i.e. the last term in Eq. (13). This was also observed when the radiance term was omitted (Fig. 7) and was confirmed by the relative magnitudes of the individual gradient terms. The radiance term (irrespective of the spherical harmonic approximations) was always significantly smaller than the remaining terms of Eq. (13). Omitting the computation of the adjoint radiance resulted in a major increase in computational speed. However, from

the limited investigation presented here, it can only be concluded that the adjoint term may be neglected if the scattering coefficient is known. If the recovery of the scattering coefficient is of interest, a radiance approximation to a minimum of $N_L = 1$ degrees may be be necessary.[31]

The gradient-based inversion was found to benefit greatly from optimization algorithms in which parameters, such as the step sizes and the exponential decay rates of the Adam algorithm, are predefined to enable a fast and accurate convergence. A potential drawback of such methods is the need to test several sets of these parameters prior to an inversion to assess whether they have a positive impact on the convergence speed. Within the scope of this study, only minor and *ad hoc* parameter tuning was conducted. A more thorough investigation, including the development of automated parameter selection algorithms, may yield significantly faster convergence.

The chromophore-dependent step size and the fluence-dependent spatial step size scaling (Eqs. 16 and 17) proved to be vital to achieving convergence. Without chromophore-dependent step sizes, Mb concentrations diverged to the upper and lower fit limits. Similarly, the fluence-dependent spatial step size scaling was crucial to achieving fast convergence in regions of low fluence.

The selection of the excitation wavelengths could also be optimized further to improve inversion accuracy and convergence speed.[48] However, such a study would exceed the scope of this paper. Despite potentially sub-optimal excitation wavelengths, the inversion was shown to recover blood sO_2 over a wide range (Fig. 6(A)) with high accuracy ($< 1\%$ error in sO_2) across the domain.

Gradient-based methods do not guarantee convergence to a global minimum, especially when the inversion is adversely affected by a noisy gradient. While the Adam optimization algorithm (compared to for example standard or momentum gradient descent) has been shown to greatly reduce the likelihood of finishing the inversion in a local minimum or on a saddle point, such a result cannot be ruled out entirely. It should also be noted that the application of this method to measured PA images does not require their segmentation into sub-regions. While this makes the method generally valid, some form of image segmentation may still be advantageous as it would reduce the number of variables, the risk of convergence to a local minimum, and increase convergence speed. Moreover, explicit regularization of the objective function could further improve the accuracy and speed of the convergence.

While the general methodology of a gradient-based inversion using an adjoint formulation of a MC model has been demonstrated *in silico*, the application of this approach to experimental 3D PA images, especially those acquired *in vivo*, requires further investigation. One of the perhaps most critical points is the recovery of the scattering coefficient as

it is likely to have an impact on the importance of the radiance approximation using spherical harmonics. Other issues, such as the choice of inversion parameters, the selection of optimal excitation wavelengths are also important in the translation of QPAT methods towards applications in the medical and life sciences.

5 Conclusions

An inversion scheme based for recovering absolute chromophore concentrations and their ratios, such as blood sO_2 from 3D multiwavelength PA images was developed and validated *in silico*. The scheme was based on an adjoint formulation of an MC light transport model and allowed an approximation of the radiance using spherical harmonics. It was found that the adjoint radiance is not required to obtain accurate inversion results provided the scattering coefficient is constant. The speed of convergence was increased by incorporating the Adam optimization algorithm, chromophore-dependent step-sizes, and fluence-dependent step-size scaling. This work represents an important step in the development of robust and generally applicable methods for quantitative functional and molecular PA imaging.

Appendix A: Definition of the adjoint model

The idea of the adjoint formalism is to define non-physical quantities, an adjoint source $q^*(\vec{r},\lambda)$, adjoint radiance $\phi^*(\vec{r},\vec{s},\lambda)$, and an adjoint fluence $\Phi^*(\vec{r},\lambda) = \int_{S^2}\phi^*(\vec{r},\vec{s},\lambda)\mathrm{d}\vec{s}$, that help to replace the integral term containing the unknown $\frac{\partial\Phi}{\partial c_i}$ in the definition of the gradient. In our case the gradient equation is

$$\frac{\partial\varepsilon_{\lambda_l}}{\partial c_i} = -\Big(p_0^m(\vec{r}_i)-p_0(\vec{r}_i)\Big)V_{vox}\Big(\beta_k H(\vec{r}_i)+\Gamma(\vec{r}_i)\alpha_k\Phi(\vec{r}_i)\Big)+\int_\Omega\Big(p_0^m-p_0\Big)\Gamma(\vec{r})\mu_a(\vec{r})\frac{\partial\Phi}{\partial c_i}\mathrm{d}\Omega. \tag{19}$$

The adjoint approach has been used in the context of PAT earlier, see e.g. Refs. 17–19, 26, 27.

As a first step, the adjoint source term is defined as

$$q^*(\vec{r},\lambda) = (p_0^m(\vec{r},\lambda) - p_0(\vec{r},\lambda))\Gamma(\vec{r})\mu_a(\vec{r},\lambda). \tag{20}$$

Since the approach is targeted for multi-spectral QPAT, each wavelength requires its own definition of an adjoint source based on the difference between modeled $p_0(\vec{r},\lambda)$ and measured data $p_0^m(\vec{r},\lambda)$. We only denote one wavelength here and omit the dependence on λ for the sake of brevity.

The adjoint source is usually defined as the "pre-factor" of the unknown term that is to be replaced and contains the error between modeled and measured data. By defining

the behavior of the adjoint radiance ϕ^* and adjoint fluence Φ^*, a relationship between these and the desired $\frac{\partial \Phi}{\partial c_i}$ can be established, which is outlined in the following derivation. One important aspect of the definition of the adjoint quantity is to leave the equations underlying the development similar to the ones of their physical counterpart, which is in this context the time-independent radiative transfer equation (RTE).

The RTE is given by

$$(\vec{s} \cdot \nabla + \mu_a(\vec{r}) + \mu_s(\vec{r}))\phi(\vec{r}, \vec{s}) - \mu_s \int_{S^2} \Theta(\vec{s}, \vec{s}')\phi(\vec{r}, \vec{s}')\mathrm{d}\vec{s}' = q(\vec{r}, \vec{s}). \tag{21}$$

Similarly, we define the adjoint RTE (ARTE) as

$$(-\vec{s} \cdot \nabla + \mu_a(\vec{r}) + \mu_s(\vec{r}))\phi^*(\vec{r}, \vec{s}) - \mu_s \int_{S^2} \Theta(\vec{s}, \vec{s}')\phi^*(\vec{r}, \vec{s}')\mathrm{d}\vec{s}' = q^*(\vec{r}, \vec{s}). \tag{22}$$

One advantage of defining the adjoint radiance in that way is that the propagation dynamics are identical to that of the normal radiance defined by the RTE, since the left hand side of both the RTE and the ARTE are practically identical, the only difference being the negative sign in the ARTE indicating a change of direction in light propagation. One way to interpret the negative sign is to follow the propagation of photons in the opposite direction which does not affect the photon's movement and the final results in terms of energy deposit. Thus, the mechanisms for absorption and scattering remain the same as in the RTE. Because the light transport is dominated by scattering and absorption and not whether photons move in the forward or backward direction, light propagation can be seen as reciprocal and hence the numerical framework implementing the ARTE is unaffected by the additional negative sign. Thus, the same simulation code as for the RTE can be used for the ARTE. Only the difference in the source distributions needs to be taken into account, with the adjoint source being three dimensional, whereas normal source distributions are usually two dimensional.

It is important to note that the adjoint fluence and adjoint radiance have different physical units as their physical counterparts. The adjoint radiance has units $\mathrm{J}/(\mathrm{m}^3\mathrm{sr})$ and the adjoint fluence $\mathrm{J}/(\mathrm{m}^3)$.

Appendix B: Adjoint-assisted derivation of the gradient

Using the above definition of the adjoint source in Eq. (20), the substitution of the unknown term $\frac{\partial \Phi}{\partial c_i}$

$$\int_\Omega \left(p_0^m - p_0\right)\Gamma(\vec{r})\mu_a(\vec{r})\frac{\partial \Phi}{\partial c_i}\mathrm{d}\Omega = -\alpha_k \left[\int_{S^2} \phi^*(\vec{s})\phi(\vec{s})\mathrm{d}\vec{s}\right]_{\vec{r}=\vec{r}_i} V_{\text{vox}} \tag{23}$$

can be derived.

The derivation follows the ideas presented in previous works, in particular Refs. 27. In there, $\frac{\partial \text{RTE}}{\partial c_i}$ is combined with the ARTE

$$\phi^* \cdot \frac{\partial \text{RTE}}{\partial c_i} - \frac{\partial \phi}{\partial c_i} \cdot \text{ARTE}. \tag{24}$$

The term $\frac{\partial \text{RTE}}{\partial c_i}$ is

$$(\vec{s} \cdot \nabla + \mu_a(\vec{r}) + \mu_s(\vec{r})) \frac{\partial \phi(\vec{r}, \vec{s})}{\partial c_i} + \frac{\partial \mu_a}{\partial c_i} \phi(\vec{r}, \vec{s}) - \mu_s \int_{S^2} \Theta(\vec{s}, \vec{s}') \frac{\partial \phi(\vec{r}, \vec{s}')}{\partial c_i} \mathrm{d}\vec{s}' = 0 \tag{25}$$

as $\frac{\partial q}{\partial c_i} = 0$ since the external light source does not depend on chromophore concentrations. The basic idea underlying the following steps is to rearrange all terms in Eq. (24) so that all terms on the left-hand side can be set to zero after integrating over space and angles. First, we insert $\frac{\partial \mu_a}{\partial c_i} = \alpha \delta(\vec{r} - \vec{r}_i)$ and rearrange Eq. (25)

$$(\vec{s} \cdot \nabla + \mu_a(\vec{r}) + \mu_s(\vec{r})) \frac{\partial \phi(\vec{r}, \vec{s})}{\partial c_i} - \mu_s \int_{S^2} \Theta(\vec{s}, \vec{s}') \frac{\partial \phi(\vec{r}, \vec{s}')}{\partial c_i} \mathrm{d}\vec{s}' = -\alpha(\vec{r}) \phi(\vec{r}, \vec{s}) \delta(\vec{r} - \vec{r}_i) \tag{26}$$

The combination of the ARTE and the RTE from Eq. (24) is (for brevity we omit the dependency on $\vec{r}$ and $\vec{s}$ for the moment)

$$\begin{aligned} &\phi^* (\vec{s} \cdot \nabla + \mu_a + \mu_s) \frac{\partial \phi}{\partial c_i} - \phi^* \mu_s \int_{S^2} \Theta(\vec{s}, \vec{s}') \frac{\partial \phi}{\partial c_i} \mathrm{d}\vec{s}' \\ &- \frac{\partial \phi}{\partial c_i} (-\vec{s} \cdot \nabla + \mu_a + \mu_s) \phi^* + \frac{\partial \phi}{\partial c_i} \mu_s \int_{S^2} \Theta(\vec{s}, \vec{s}') \phi^* \mathrm{d}\vec{s}' \\ &= -\alpha \phi^* \phi \delta(\vec{r} - \vec{r}_i) - \frac{\partial \phi}{\partial c_i} q^* \end{aligned} \tag{27}$$

The left-hand side can be simplified, giving

$$\begin{aligned} &\phi^* (\vec{s} \cdot \nabla) \frac{\partial \phi}{\partial c_i} + \frac{\partial \phi}{\partial c_i} (\vec{s} \cdot \nabla) \phi^* - \phi^* \mu_s \int_{S^2} \Theta(\vec{s}, \vec{s}') \frac{\partial \phi}{\partial c_i} \mathrm{d}\vec{s}' + \frac{\partial \phi}{\partial c_i} \mu_s \int_{S^2} \Theta(\vec{s}, \vec{s}') \phi^* \mathrm{d}\vec{s}' \\ &= -\alpha \phi^* \phi \delta(\vec{r} - \vec{r}_i) - \frac{\partial \phi}{\partial c_i} q^*. \end{aligned} \tag{28}$$

The next steps of the derivation are from now on identical to Eqs. (4.11 ff) in Ref. 27. The left-hand side of Eq. (28) equates to zero, which can be seen after integrating first over all

angles $s \in S^{n-1}$ and over the volume Ω with surface $\partial\Omega$:

$$\begin{aligned}
&\int_\Omega \int_{S^2} \phi^*(\vec{s}\cdot\nabla)\frac{\partial\phi}{\partial c_i}\mathrm{d}\vec{s}\mathrm{d}\Omega + \int_\Omega \int_{S^2} \frac{\partial\phi}{\partial c_i}(\vec{s}\cdot\nabla)\phi^*\mathrm{d}\vec{s}\mathrm{d}\Omega \\
&- \int_\Omega \mu_s \int_{S^2} \phi^*(\vec{s}) \int_{S^2} \Theta(\vec{s},\vec{s}')\frac{\partial\phi(\vec{s}')}{\partial c_i}\mathrm{d}\vec{s}'\mathrm{d}\vec{s}\mathrm{d}\Omega \\
&+ \int_\Omega \mu_s \int_{S^2} \frac{\partial\phi(\vec{s})}{\partial c_i} \int_{S^2} \Theta(\vec{s},\vec{s}')\phi^*(\vec{s}')\mathrm{d}\vec{s}'\mathrm{d}\vec{s}\mathrm{d}\Omega \\
&= -\int_\Omega \alpha\delta(\vec{r}-\vec{r}_i)\int_{S^2}\phi^*\phi\mathrm{d}\vec{s}\mathrm{d}\Omega - \int_\Omega q^* \int_{S^2}\frac{\partial\phi}{\partial c_i}\mathrm{d}\vec{s}\mathrm{d}\Omega. \qquad (29)
\end{aligned}$$

To see that the left-hand side equates to zero the volume integral with the terms involving ∇ can be transformed into a surface integral using this form of the divergence theorem:

$$\int_\Omega ab\cdot\nabla c\mathrm{d}\Omega + \int_\Omega cb\cdot\nabla a\mathrm{d}\Omega = \int_{\partial\Omega} b\cdot\hat{n}ac\ \mathrm{d}\Omega, \qquad (30)$$

along with the following substitions:

$$a = \phi^*(\vec{s}), \qquad b = \vec{s} \qquad \text{and} \qquad c = \frac{\partial\phi(\vec{s})}{\partial c_i}. \qquad (31)$$

Hence, using the divergence theorem, the first two terms in Eq. (28) are replaced by a single term

$$\int_\Omega \int_{S^2} \phi^*(\vec{s}\cdot\nabla)\frac{\partial\phi}{\partial c_i}\mathrm{d}\vec{s}\mathrm{d}\Omega + \int_\Omega \int_{S^2} \frac{\partial\phi}{\partial c_i}(\vec{s}\cdot\nabla)\phi^*\mathrm{d}\vec{s}\mathrm{d}\Omega = \int_{\partial\Omega}\int_{S^2}(\vec{s}\cdot\hat{\vec{n}})\phi^*(\vec{s})\frac{\partial\phi(\vec{s})}{\partial c_i}\mathrm{d}\vec{s}\mathrm{d}\Omega.$$

By definition both $\phi^* \to 0$ and $\partial\Phi/\partial c_i \to 0$ on the boundary of the volume $\partial\Omega$. Thus, the integrand on the right-hand side and hence the integral equate to zero.

This reduces Eq. (28) to

$$\begin{aligned}
&- \int_\Omega \mu_s \int_{S^2} \phi^*(\vec{s}) \int_{S^2} \Theta(\vec{s},\vec{s}')\frac{\partial\phi(\vec{s}')}{\partial c_i}\mathrm{d}\vec{s}'\mathrm{d}\vec{s}\mathrm{d}\Omega \\
&+ \int_\Omega \mu_s \int_{S^2} \frac{\partial\phi(\vec{s})}{\partial c_i} \int_{S^2} \Theta(\vec{s},\vec{s}')\phi^*(\vec{s}')\mathrm{d}\vec{s}'\mathrm{d}\vec{s}\mathrm{d}\Omega \\
&= -\int_\Omega \alpha\delta(\vec{r}-\vec{r}_i)\int_{S^2}\phi^*\phi\mathrm{d}\vec{s}\mathrm{d}\Omega - \int_\Omega q^* \int_{S^2}\frac{\partial\phi}{\partial c_i}\mathrm{d}\vec{s}\mathrm{d}\Omega. \qquad (32)
\end{aligned}$$

Because we can assume that all functions are integrable, the terms on the left-hand side of Eq. (32) can be rearranged after changing the order of integration, yielding

$$\int_\Omega \mu_s \int_{S^2}\int_{S^2} \Theta(\vec{s},\vec{s}')\phi^*(\vec{s})\frac{\partial\phi(\vec{s}')}{\partial c_i}\mathrm{d}\vec{s}'\mathrm{d}\vec{s}\mathrm{d}\Omega$$
$$-\int_\Omega \mu_s \int_{S^2}\int_{S^2} \Theta(\vec{s},\vec{s}')\phi^*(\vec{s}')\frac{\partial\phi(\vec{s})}{\partial c_i}\mathrm{d}\vec{s}'\mathrm{d}\vec{s}\mathrm{d}\Omega.$$

Hence, the left-hand side of Eq. (32) equates to zero, which leaves us with

$$\int_\Omega q^* \int_{S^2}\frac{\partial\phi}{\partial c_i}\mathrm{d}\vec{s}\mathrm{d}\Omega = -\int_\Omega \alpha\delta(\vec{r}-\vec{r}_i)\int_{S^2}\phi^*\phi\mathrm{d}\vec{s}\mathrm{d}\Omega \tag{33}$$

The fluence Φ is by definition the integral of the time-integrated radiance $\phi(\vec{s})$ over all directions $\vec{s}$, that is

$$\Phi(\vec{r}) = \int_{S^2}\phi(\vec{r},\vec{s})\mathrm{d}\vec{s} \tag{34}$$

Applying the derivative with respect to c_i yields

$$\frac{\partial\Phi(\vec{r})}{\partial c_i} = \int_{S^2}\frac{\partial\phi(\vec{r},\vec{s})}{\partial c_i}\mathrm{d}\vec{s}, \tag{35}$$

which lets us write Eq. (33) as

$$\int_\Omega q^*\frac{\partial\Phi}{\partial c_i}\mathrm{d}\Omega = -\alpha(\vec{r}_i)\left[\int_{S^2}\phi^*(\vec{s})\phi(\vec{s})\mathrm{d}\vec{s}\right]_{\vec{r}=\vec{r}_i} V_{vox} \tag{36}$$

Because we have defined the adjoint source term as $q^* = (p_0^m - p_0)\Gamma\mu_a$ the left-hand side of Eq. (36) is exactly the last term in Eq. (19):

$$\int_\Omega (p_0^m - p_0)\Gamma\mu_a\frac{\partial\Phi}{\partial c_i}\mathrm{d}\Omega = -\alpha(\vec{r}_i)\left[\int_{S^2}\phi^*(\vec{s})\phi(\vec{s})\mathrm{d}\vec{s}\right]_{\vec{r}=\vec{r}_i} V_{vox}. \tag{37}$$

Inserting this result into the error-gradient Eq. (19) provides us with the sub-gradient term obtained using one wavelength

$$\frac{\partial\varepsilon_{\lambda_l}}{\partial c_i} = -(p_0^m(\vec{r}_i) - p_0(\vec{r}_i))\alpha(\lambda_l)\Phi(\vec{r}_i)V_{vox} + \alpha(\vec{r}_i)\left[\int_{S^2}\phi^*(\vec{s})\phi(\vec{s})\mathrm{d}\vec{s}\right]_{\vec{r}=\vec{r}_i} V_{vox}. \tag{38}$$

In summary, the adjoint formalism enables the update of the concentration distribution using only terms obtained from running the forward model implemented by the RMC algorithm. The following section will focus on the question as to how the term $\phi^*\phi$ can be computed and approximated.

Appendix C: Radiance approximations

In order to simulate the radiance, some discretization over angle is required. One option is to use a piecewise constant set of basis functions over angle. However, due to the importance of ballistic and quasi-ballistic light propagation in PAT, a high number of discretization orders would be required to capture the directionality of the radiance in regions near the source. This would result in very large memory demands in finite-element implementations, see Refs. 29,49,50 for more details on different approaches for estimating the radiance, their limitations, and suggested solutions. For a summary thereof, see Ref. 51. Inspired by the P_n approximation[22] and continuing the work presented by Refs. 19,51 we approximate the radiance in 3D using spherical harmonics as basis functions as in Ref. 31. Instead of discretising the angular domain into segments, the radiance field ϕ at any position $\vec{r}$ can be expanded using a series of spherical harmonics[52,53]

$$\phi(\vec{r},\vec{s}) = \sum_{l=0}^{N_L=\infty} \sum_{m=-l}^{l} \psi_{lm}(\vec{r}) Y_{lm}(\vec{r},\vec{s}), \tag{39}$$

where $\psi_{lm}(\vec{r})$ are the coefficients corresponding to the real spherical harmonics $Y_{lm}(\vec{r},\vec{s})$, expressed as

$$Y_{lm} = \begin{cases} \sqrt{2}\sqrt{\frac{2l+1}{4\pi}}\sqrt{\frac{(l-|m|)!}{(l+|m|)!}} P_l^{|m|}\cos(\theta)\sin(|m|\phi) & \text{if } m<0, \\ \sqrt{\frac{2l+1}{4\pi}} P_l^m(\cos(\theta)) & \text{if } m=0, \\ \sqrt{2}\sqrt{\frac{2l+1}{4\pi}}\sqrt{\frac{(l-m)!}{(l+m)!}} P_l^m\cos(\theta)\cos(m\phi) & \text{if } m>0, \end{cases} \tag{40}$$

where l is the degree of the spherical harmonic, m is the order, P_l^m are the associated Legendre polynomials. The coefficient $\psi_{lm}(\vec{r})$ scales the total weight deposited by all simulated photons at position (voxel) $\vec{r}$ for the associated spherical harmonic.

The advantage of expressing $\phi^*\phi$ in a spherical harmonics expansion lies in the fact that the Y_{lm} form an orthonormal basis, i.e.

$$\int_{S^2} Y_{lm}\cdot Y_{l'm'} d\vec{s} = \delta_{ll'}\delta_{mm'}. \tag{41}$$

Using this orthonormality condition greatly simplifies the term $\left[\int_{S^2}\phi^*(\vec{s})\phi(\vec{s})\mathrm{d}\vec{s}\right]_{\vec{r}=\vec{r}_i}$

from Eq. (38), when the radiance is expressed in spherical harmonics:

$$\left[\int_{S^2}\phi^*(\vec{s})\phi(\vec{s})\mathrm{d}\vec{s}\right]_{\vec{r}=\vec{r}_i} = \int_{S^2}\left(\sum_{l=0}^{N_L}\sum_{m=-l}^{l}\psi^*_{lm}(\vec{r}_i)Y^*_{lm}(\vec{r}_i,\vec{s})\right)\left(\sum_{l'=0}^{N_L}\sum_{m'=-l'}^{l'}\psi_{l'm'}(\vec{r}_i)Y_{l'm'}(\vec{r}_i,\vec{s})\right)\mathrm{d}\vec{s} \tag{42}$$

$$= \left(\sum_{l=0}^{N_L}\sum_{m=-l}^{l}\psi^*_{lm}(\vec{r}_i)\right)\left(\sum_{l'=0}^{N_L}\sum_{m'=-l'}^{l'}\psi_{l'm'}(\vec{r}_i)\right)\left(\underbrace{\int_{S^2}Y^*_{lm}(\vec{r}_i,\vec{s})Y_{l'm'}(\vec{r}_i,\vec{s})\mathrm{d}\vec{s}}_{\delta_{ll'}\delta_{mm'}}\right) \tag{43}$$

$$= \sum_{l=0}^{N_L}\sum_{m=-l}^{l}\psi_{lm}(\vec{r}_i)\psi^*_{lm}(\vec{r}_i). \tag{44}$$

Hence, with this approximation the gradient to update the distribution of chromophore k finally becomes

$$\frac{\partial\varepsilon_{\lambda_l}}{\partial c_i} = V_{vox}\left(-(p_0^m(\vec{r}_i)-p_0(\vec{r}_i))\Big(\beta_k H(\vec{r}_i)+\Gamma(\vec{r}_i)\alpha_k\Phi(\vec{r}_i)\Big)+\alpha_k\sum_{l=0}^{N_L}\sum_{m=-l}^{l}\psi_{lm}(\vec{r}_i)\psi^*_{lm}(\vec{r}_i)\right). \tag{45}$$

Appendix D: Discretization

To solve the given equations numerically, the bases in which the data and model output are represented must be defined. Assuming a sampling of continuous fields in a point-wise basis $\Psi_j(\vec{r}) = \delta(\vec{r}-\vec{r}_i)$ as in Ref. 18. Hence, the data projected onto this basis becomes a vector of coefficients $\vec{p}_0^m$,

$$h_j = \langle\Psi_j, p_0^m(\vec{r})\rangle = p_0^m(\vec{r}_j) \tag{46}$$

Equally, all other continuous fields are discretized. Transforming an integral of any integrable function $f(\vec{r})$ over the continuous domain Ω into discretized space introduces a volume element $\mathrm{d}\Omega = V_{\text{vox}}$:

$$\int_\Omega f(\vec{r})\mathrm{d}\Omega = \sum_j^{N_{\text{vox}}}\langle\Psi_j, f(\vec{r})\rangle V_{\text{vox}} = \sum_j^{N_{\text{vox}}} f(\vec{r}_j)V_{\text{vox}} \tag{47}$$

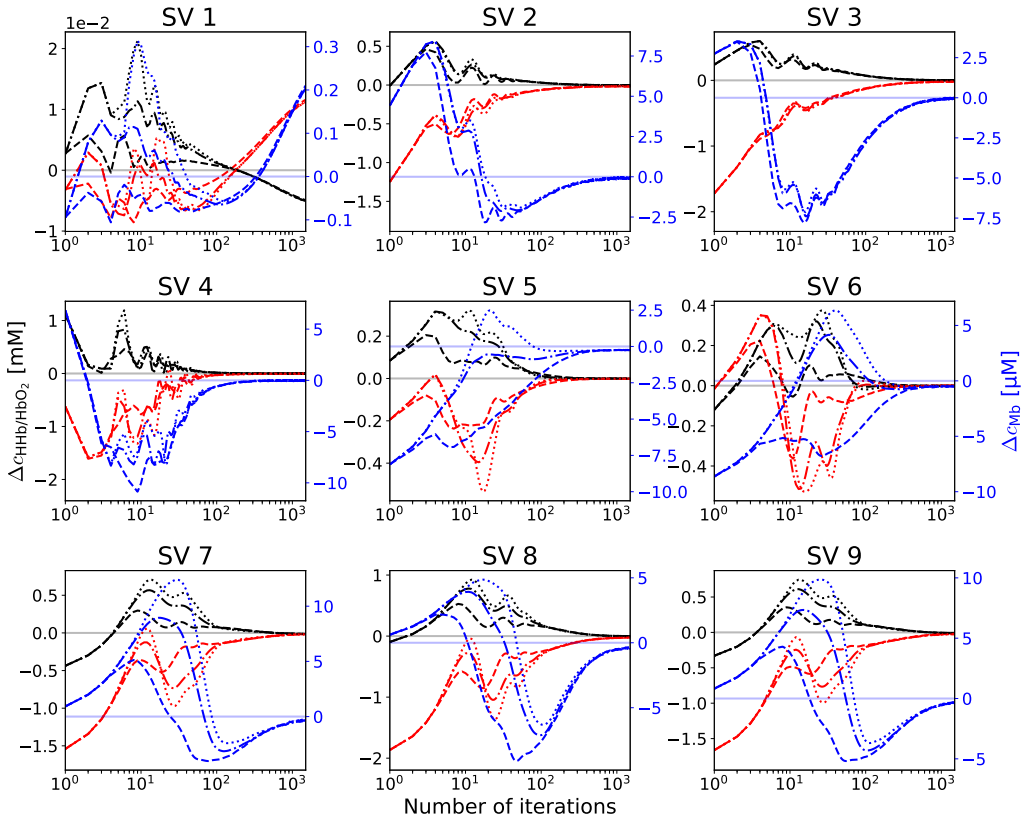

Fig 7 Convergence of the true and recovered concentrations for inversions incorporating the gradient term of the adjoint radiance for $N_L = 4$ (dotted line, $\cdots$), $N_L = 0$ (dashed-dotted line, $- \cdot -$), and without the adjoint term (dashed line, $- -$). The chromphores are shown in blue (Mb), red ($\mathrm{HbO_2}$) and black(HHb). Horizontal lines in light color depict perfect convergence, i.e. $\Delta c = 0$.

Funding

This project was funded by the Deutsche Forschungsgemeinschaft (DFG project grants LA3273/1-1, PR1226/5-1), the Royal Academy of Engineering (RAEng Fellowship RF1516\15\33), and the Engineering and Physical Sciences Research Council (EPSRC grant EP/N032055/1).

Disclosures

The authors declare that there are no conflicts of interest related to this article.

Acknowledgments

The authors would like to thank Simon Arridge and Ben Cox for valuable discussions.
†These authors contributed equally to this work.

References

1 P. Beard, "Biomedical photoacoustic imaging," *Interface Focus* (2011).

2 J. Xia, J. Yao, and L. Wang, "Photoacoustic tomography: principles and advances," *Electromagnetic waves* **147**, 1–22 (2014).

3 L. Wang and J. Yao, "A practical guide to photoacoustic tomography in the life sciences," *Nature Methods* **13:627638** (2016).

4 B. Cox, J. Laufer, and P. Beard, "The challenges for quantitative photoacoustic imaging," in *Photons Plus Ultrasound: Imaging and Sensing 2009*, *Proc.SPIE* **7177**, 717713, International Society for Optics and Photonics (2009).

5 B. Cox, J. G. Laufer, S. R. Arridge, *et al.*, "Quantitative spectroscopic photoacoustic imaging: a review," *Journal of Biomedical Optics* **17**(6), 0612021–06120222 (2012).

6 C. Lutzweiler and D. Razansky, "Optoacoustic imaging and tomography: Reconstruction approaches and outstanding challenges in image performance and quantification," *Sensors* **13**(6), 7345–7384 (2013).

7 M. Li, Y. Tang, and J. Yao, "Photoacoustic tomography of blood oxygenation: A mini review," *Photoacoustics* **10**, 65–73 (2018).

8 X. Wang, X. Xie, G. Ku, *et al.*, "Noninvasive imaging of hemoglobin concentration and oxygenation in the rat brain using high-resolution photoacoustic tomography," *Journal of Biomedical Optics* **11**(2), 024015 (2006).

9 D. Razansky, M. Distel, C. Vinegoni, *et al.*, "Multispectral opto-acoustic tomography of deep-seated fluorescent proteins in vivo," *Nature Photonics* **3**, 412 EP – (2009). Article.

10 S. Tzoumas, A. Nunes, I. Olefir, *et al.*, "Eigenspectra optoacoustic tomography achieves quantitative blood oxygenation imaging deep in tissues," *Nature communications* **7**, 12121 (2016).

11 T. Kirchner, J. Gröhl, and L. Maier-Hein, "Context encoding enables machine learning-based quantitative photoacoustics," *Journal of Biomedical Optics* **23**(5), 056008 (2018).

12 G. Bal and K. Ren, "On multi-spectral quantitative photoacoustic tomography in diffusive regime," *Inverse Problems* **28**(2), 025010 (2012).

13 R. J. Zemp, "Quantitative photoacoustic tomography with multiple optical sources," *Appl. Opt.* **49**, 3566–3572 (2010).

14 B. T. Cox, S. R. Arridge, and P. C. Beard, "Estimating chromophore distributions from multiwavelength photoacoustic images," *J. Opt. Soc. Am. A* **26**, 443–455 (2009).

15 J. Laufer, B. Cox, E. Zhang, *et al.*, "Quantitative determination of chromophore concentrations from 2d photoacoustic images using a nonlinear model-based inversion scheme," *Applied Optics* **49**(8), 1219–1233 (2010).

16 B. A. Kaplan, J. Buchmann, S. Prohaska, *et al.*, "Monte-Carlo-based inversion scheme for 3D quantitative photoacoustic tomography," in *Photons Plus Ultrasound: Imaging and Sensing 2017*, **10064**, 100645J, International Society for Optics and Photonics (2017).

17 B. Cox, S. Arridge, and P. Beard, "Gradient-based quantitative photoacoustic image reconstruction for molecular imaging," in *Photons Plus Ultrasound: Imaging and Sensing 2007*, **6437**, 64371T, International Society for Optics and Photonics (2007).

18 E. Malone, S. Powell, B. T. Cox, *et al.*, "Reconstruction-classification method for quantitative photoacoustic tomography," *Journal of Biomedical Optics* **20**(12), 126004–126004 (2015).

19 R. Hochuli, S. Powell, S. Arridge, *et al.*, "Quantitative photoacoustic tomography using forward and adjoint Monte Carlo models of radiance," *Journal of Biomedical Optics* **21**(12), 126004–126004 (2016).

20 M. Fonseca, E. Malone, F. Lucka, *et al.*, "Three-dimensional photoacoustic imaging and inversion for accurate quantification of chromophore distributions," in *Photons Plus Ultrasound: Imaging and Sensing 2017*, **10064**, 1006415, International Society for Optics and Photonics (2017).

21 G. Bal and G. Uhlmann, "Inverse diffusion theory of photoacoustics," *Inverse Problems* **26**(8), 085010 (2010).

22 S. R. Arridge, "Optical tomography in medical imaging," *Inverse Problems* **15**(2), R41 (1999).

23 T. Tarvainen, B. T. Cox, J. P. Kaipio, *et al.*, "Reconstructing absorption and scattering distributions in quantitative photoacoustic tomography," *Inverse Problems* **28**(8), 084009 (2012).

24 A. V. Mamonov and K. Ren, "Quantitative photoacoustic imaging in the radiative transport regime," *Communications in Mathematical Sciences* **12**(2), 201–234 (2014).

25 S. Rabanser, L. Neumann, and M. Haltmeier, "Stochastic proximal gradient algorithms for multi-source quantitative photoacoustic tomography," *Entropy* **20**(2), 121 (2018).

26 T. Saratoon, T. Tarvainen, B. Cox, *et al.*, "A gradient-based method for quantitative photoacoustic tomography using the radiative transfer equation," *Inverse Problems* **29**(7), 075006 (2013).

27 T. Soonthornsaratoon, *Gradient-based methods for quantitative photoacoustic tomography*. PhD thesis, UCL (University College London) (2014).

28 S. T. Flock, M. S. Patterson, B. C. Wilson, *et al.*, "Monte Carlo modeling of light propagation in highly scattering tissues. I. Model predictions and comparison with diffusion theory," *IEEE Transactions on Biomedical Engineering* **36**(12), 1162–1168 (1989).

29 L. Wang and S. L. Jacques, "Hybrid model of Monte Carlo simulation and diffusion theory for light reflectance by turbid media," *J. Opt. Soc. Am. A* **10**, 1746–1752 (1993).

30 Q. Fang and D. A. Boas, "Monte Carlo simulation of photon migration in 3D turbid media accelerated by graphics processing units," *Optics Express* **17**(22), 20178–20190 (2009).

31 S. Powell, R. Hochuli, and S. R. Arridge, "Radiance Monte-Carlo for application of the radiative transport equation in the inverse problem of diffuse optical tomography," in *Optical Tomography and Spectroscopy of Tissue XII*, **10059**, 100590W, International Society for Optics and Photonics (2017).

32 J. Buchmann, B. A. Kaplan, S. Prohaska, *et al.*, "Experimental validation of a Monte-Carlo-based inversion scheme for 3D quantitative photoacoustic tomography," in *Photons Plus Ultrasound: Imaging and Sensing 2017*, **10064**, 1006416, International Society for Optics and Photonics (2017).

33 D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980* (2014).

34 M. B. Fonseca, L. An, and B. T. Cox, "Sulfates as chromophores for multiwavelength photoacoustic imaging phantoms," *Journal of Biomedical Optics* **22**(12), 125007 (2017).

35 D.-K. Yao, C. Zhang, K. Maslov, *et al.*, "Photoacoustic measurement of the Grüneisen parameter of tissue," *J Biomed Opt* **19**, 017007 (2014). 130780R[PII].

36 S. A. Prahl, "A monte carlo model of light propagation in tissue," in *Dosimetry of laser radiation in medicine and biology*, **10305**, 1030509, International Society for Optics and Photonics (1989).

37 L. G. Henyey and J. L. Greenstein, "Diffuse radiation in the galaxy," *The Astrophysical Journal* **93**, 70–83 (1941).

38 J. Bezanson, A. Edelman, S. Karpinski, *et al.*, "Julia: A fresh approach to numerical computing," *SIAM review* **59**(1), 65–98 (2017).

39 J. Laufer, P. Johnson, E. Zhang, *et al.*, "In vivo preclinical photoacoustic imaging of tumor vasculature development and therapy," *Journal of Biomedical Optics* **17**(5), 0560161–0560168 (2012).

40 J. Märk, H. Dortay, A. Wagener, *et al.*, "Dual-wavelength 3D photoacoustic imaging of mammalian cells using a photoswitchable phytochrome reporter protein," *Communications Physics* **1**(1), 3 (2018).

41 E. Zhang, J. Laufer, and P. Beard, "Backward-mode multiwavelength photoacoustic scanner using a planar Fabry-Pérot polymer film ultrasound sensor for high-resolution three-dimensional imaging of biological tissues," *Applied Optics* **47**(4), 561–577 (2008).

42 J. Buchmann, J. Guggenheim, E. Zhang, *et al.*, "Characterization and modeling of fabry–perot ultrasound sensors with hard dielectric mirrors for photoacoustic imaging," *Applied optics* **56**(17), 5039–5046 (2017).

43 B. G. Barratt-Boyes and E. H. Wood, "The oxygen saturation of blood in the venae cavae, right-heart chambers, and pulmonary vessels of healthy subjects," *The Journal of Laboratory and Clinical Medicine* **50**(1), 93–106 (1957).

44 E. Beutler and J. Waalen, "The definition of anemia: what is the lower limit of normal of the blood hemoglobin concentration?," *Blood* **107**(5), 1747–1750 (2006).

45 G. Serŝa, K. Beravs, M. Ĉemazˆar, *et al.*, "Contrast enhanced mri assessment of tumor blood volume after application of electric pulses," *Electro- and Magnetobiology* **17**(2), 299–306 (1998).

46 B. T. Polyak, "Some methods of speeding up the convergence of iteration methods," *USSR Computational Mathematics and Mathematical Physics* **4**(5), 1–17 (1964).

47 G. Goh, "Why momentum really works," *Distill* (2017).

48 G. P. Luke, S. Y. Nam, and S. Y. Emelianov, "Optical wavelength selection for improved spectroscopic photoacoustic imaging," *Photoacoustics* **1**(2), 36–42 (2013).

49 T. Tarvainen, M. Vauhkonen, V. Kolehmainen, *et al.*, "Hybrid radiative-transfer–diffusion model for optical tomography," *Applied optics* **44**(6), 876–886 (2005).

50 T. Tarvainen, *Computational methods for light transport in optical tomography*. PhD thesis, Faculty of Natural and Environmental Sciences, Department of Physics, University of Kuopio (2006).

51 R. Hochuli, *Monte Carlo methods in quantitative photoacoustic tomography*. PhD thesis, UCL (University College London) (2016).

52 F. Inanc and A. Rohach, "A nodal solution of the multigroup neutron transport equation using spherical harmonics," *Annals of Nuclear Energy* **15**(10-11), 501–509 (1988).

53 A. G. Buchan, A. Calloo, M. G. Goffin, *et al.*, "A POD reduced order model for resolving angular direction in neutron/photon transport problems," *Journal of Computational Physics* **296**, 138–157 (2015).