

MARGINALIZATION OF UNINTERESTING DISTRIBUTED PARAMETERS IN INVERSE PROBLEMS —APPLICATION TO DIFFUSE OPTICAL TOMOGRAPHY

Ville Kolehmainen,¹ Tanja Tarvainen,¹ Simon R. Arridge,² & Jari P. Kaipio^{1,3,*}

¹Department of Physics and Mathematics, University of Eastern Finland, P.O. Box 1627, 70211 Kuopio, Finland

²Department of Computer Science, University College London, Gower Street, London WC1E 6BT, UK

³Department of Mathematics, University of Auckland, Private Bag 92019, Auckland Mail Centre, Auckland 1142, New Zealand

Original Manuscript Submitted: 01/10/2010; Final Draft Received: 04/04/2010

With inverse problems there are often several unknown distributed parameters of which only one may be of interest. Since assigning incorrect fixed values to the uninteresting parameters usually leads to a severely erroneous model, one is forced to estimate all distributed parameters simultaneously. This may increase the computational complexity of the problem significantly. In the Bayesian framework, all unknowns are generally treated as random variables and estimated simultaneously and all uncertainties can be modeled systematically. Recently, the approximation error approach has been proposed for handling uncertainty and model-reduction-related errors in the models. In this approach approximate marginalization of these errors is carried out before the estimation of the interesting variables. In this paper we discuss the adaptation of the approximation error approach to the marginalization of uninteresting distributed parameters. As an example, we consider the marginalization of scattering coefficient in diffuse optical tomography.

KEY WORDS: *inverse problems, Bayesian inference, parameter estimation, spatial uncertainty, diffuse optical tomography*

1. INTRODUCTION

There are several inverse problems in which there are many unknown distributed parameters. Often, only one or a subset of the unknowns is of main interest. For example, in hydrogeophysics the unknown distributed parameters may include permittivity, capillarity, and diffusivity [1, 2]. In diffuse optical tomography (DOT), the most important distributed parameters are the scattering and absorption coefficients [3]. Of these, at least in biomedical applications, the absorption coefficient is the one of interest since it is related to the oxygenization level of tissues [4]. In these applications, the scattering coefficient is considered as a nuisance parameter. The scattering coefficient, however, has to be estimated simultaneously due to the so-called crosstalk of the coefficients [3].

The simultaneous estimation of two distributed parameters is naturally a more unstable problem than estimating either of these if the other were known. In addition, with nonlinear problems the convergence of algorithms is a further problem. With applications which are eventually meant to be almost real time ones, such as biomedical optical tomog-

*Correspond to Jari P. Kaipio, E-mail: jari@math.auckland.ac.nz, URL: <http://www.math.auckland.ac.nz/~kaipio/>

raphy, it is of major interest to reduce the computational times as much as possible. Thus, in addition to estimating the interesting parameters only, there is usually pressure to also use otherwise heavily reduced computational models.

The approximation error approach was introduced in [5, 6] originally to handle pure model reduction errors. For example, in electrical impedance (resistance) tomography (EIT, ERT) and deconvolution problems, it was shown that significant model reduction is possible without essentially sacrificing the quality of estimates. With EIT, for example, this means that very low dimensional finite element approximations can be used. Later, the approach was also applied to handle other kinds of approximation and modeling errors as well as other inverse problems. Model reduction, domain truncation, and unknown anisotropy structures in diffuse optical tomography were treated in [7–10]. Missing boundary data in the case of image processing and geophysical ERT/EIT were considered in [11] and [12], respectively. In [13–15] the problem of recovery from simultaneous geometry errors and model reduction was found to be possible.

The approximation error approach was extended to nonstationary inverse problems in [16] in which linear nonstationary (heat transfer) problems were considered, and in [17] and [18] in which nonlinear problems and state space identification problems were considered, respectively. The earliest similar but partial treatment within the framework of nonstationary inverse problems was considered in [19], in which the the boundary data that is related to stochastic convection diffusion models was partially unknown. A modification in which the approximation error statistics can be updated with accumulating information was proposed in [20] and an application to hydrogeophysical monitoring in [21].

From pure model reduction and unknown (nondistributed) parameters or boundary data, a step forward was recently considered in [22] in which the physical forward model itself was replaced with a (computationally) much simpler model. In [22], the radiative transfer model (Boltzmann transfer equation), which is considered to be the most accurate model for light transfer in (turbid) media, was replaced with the diffusion approximation. It was found that also in this kind of case, the statistical structure of the approximation errors enabled the use of a significantly less complex model, again simultaneously with significant model reduction for the diffusion approximation. But also here, both the absorption and scattering coefficients were estimated simultaneously.

The approximation error approach relies on the Bayesian framework of inverse problems, in which all unknowns are explicitly modeled as random variables [5, 23, 24]. The uncertainty in the unknowns is given in the models and measurements is reflected in the posterior (probability) distribution. In the Bayesian framework, all unknowns are subject to inference simultaneously, which often results in excessively heavy computational loads. Generally, Markov chain Monte Carlo algorithms have to be used to obtain a representative set of samples from the posterior distribution. Then, after a set of samples has been computed, marginalization over the uninteresting unknowns is trivial. Only in a few special but important cases, such as the additive error model, some of the uninteresting unknowns can be eliminated before inference. We refer to such elimination as premarginalization.

In the present paper we consider the approximation error approach in the context of approximate premarginalization of uninteresting distributed parameters. Furthermore, we also consider the simultaneous treatment of the errors that are related to model reduction. As a computational example, we consider the approximate premarginalization of the scattering coefficient in diffuse optical tomography. This example shows that at least in this case it is possible to premarginalize over one distributed parameter and successfully estimate another.

The rest of the paper is structured as follows. In Section 2 we give a brief account of the approximation error approach and its formulation for the case of several distributed parameters. In Section 3 we describe the diffuse optical tomography problem. In Section 4, numerical examples of reconstructing the scattering coefficient in optical tomography with different degrees of severity are treated.

2. APPROXIMATION ERROR APPROACH

In the Bayesian framework for inverse problems, all unknowns are treated and modeled as random variables [5, 23, 24]. Once the probabilistic models for the unknowns and the measurement process have been constructed, the *posterior distribution* $\pi(x | y)$ is accessed, which reflects the uncertainty of the interesting unknowns x given the measurements y . This distribution can then be explored to answer all questions which can be expressed in terms of probabilities. For general discussion of Bayesian inference (see, for example, [25, 26]).

Bayesian inverse problems are a special class of problems in Bayesian inference. Usually, the dimension of a feasible representation of the unknowns is significantly larger than the number of measurements. Thus, for example, a maximum likelihood estimate is impossible to compute. Even in cases in which the number of unknowns would be significantly smaller than the number of measurements, the structure of the forward problem is such that maximum likelihood estimates would still be unstable. In addition to the instability, the variances of the likelihood model are almost invariably much smaller than the variances of the prior models. The posterior density is often extremely narrow and, in addition, may be a nonlinear manifold.

2.1 Marginalization Over Additive Errors

In the approximation error approach, the modeling and other errors are treated as additive errors. Therefore, we review briefly how the additive errors are formally premarginalized [5]. Let the observation model be

$$y = \bar{A}(x) + e \quad (1)$$

where e are the additive errors and $x \mapsto \bar{A}(x)$ is the deterministic forward model. With deterministic we mean that the model \bar{A} does not contain any uncertainties or other model errors. Let the joint *prior distribution* of the unknowns x and e be $\pi(x, e)$. Using the Bayes' theorem repeatedly, we can decompose the joint distribution of all associated random variables as

$$\pi(y, x, e) = \pi(y | x, e)\pi(e | x)\pi(x) \quad (2)$$

$$= \pi(y, e | x)\pi(x) \quad (3)$$

In the case of the additive model (1), the conditional distribution $\pi(y | x, e)$ is formally given by

$$\pi(y | x, e) = \delta(y - \bar{A}(x) - e)$$

which yields the *likelihood distribution*

$$\pi(y | x) = \int \pi(y, e | x) de = \quad (4)$$

$$= \int \delta(y - \bar{A}(x) - e)\pi(e | x) de \quad (5)$$

$$= \pi_{e|x}(y - \bar{A}(x) | x) \quad (6)$$

and further,¹ noting that once the measurements have been obtained, $\pi(y) > 0$ is a fixed normalization constant, we have the posterior distribution

$$\pi(x | y) \propto \pi(y | x)\pi(x) \quad (7)$$

$$= \pi_{e|x}(y - \bar{A}(x) | x)\pi(x) \quad (8)$$

In the quite common case of mutually independent x and e , we have $\pi_{e|x}(e | x) = \pi_e(e)$. Furthermore, if e and x are normal, we can write $\pi(e) = \mathcal{N}(e_*, \Gamma_e)$ and $\pi(x) = \mathcal{N}(x_*, \Gamma_x)$ and we have the familiar form

$$\pi(x | y) \propto \exp\left(-\frac{1}{2} (\|L_e(y - \bar{A}(x) - e_*)\|^2 + \|L_x(x - x_*)\|^2)\right) \quad (9)$$

where $L_e^T L_e = \Gamma_e^{-1}$ and $L_x^T L_x = \Gamma_x^{-1}$, for the posterior distribution. In the above, the unknown (uninteresting) additive error e was *pre-marginalized*, that is, marginalized before the inference procedure, and is not present in (8) or (9).

¹The subscripts, such as $e | x$ in $\pi_{e|x}$, are used to determine the actual probability density function. If the arguments, however, coincide with the density, we drop the subscripts. For example, we write $\pi_x(x) = \pi(x)$ and $\pi_{e|x}(e | x) = \pi(e | x)$, but retain the subscript in $\pi_{e|x}(y - \bar{A}(x) | x)$. Furthermore, we use the terms density and distribution interchangeably.

2.2 Approximate Premarginalization Over Model Reduction Related Errors and Other Uncertainties

The problem that is generally related to uninteresting auxiliary unknowns ξ is that we usually cannot perform premarginalization such as in Eqs. (5) and (6). In most cases we have to estimate both x and ξ , which may be a considerably more demanding undertaking than estimating just x when ξ were known. For example, if a Markov chain Monte Carlo (MCMC) approach were used, the marginalization over ξ can only be done after running the chain for both x and ξ . Once this is carried out, however, the marginalization over ξ is trivial. For MCMC methods in general (see, for example, [27, 28]). For MCMC and inverse problems, see [29–31] for applications to EIT. In this section we discuss the computational procedure in more detail in the case in which there are two distributed parameters of which premarginalization over the other one is to be carried out.

Now let the unknowns be (x, z, ξ, e) , where again e represents additive errors and ξ represents auxiliary uncertainties such as unknown boundary data, and (x, z) are two distributed parameters of which x is of interest only. The accurate forward model

$$(x, z, \xi) \mapsto \bar{A}(x, z, \xi) \quad (10)$$

is usually a nonlinear one. The uncertainties ξ can sometimes be modeled to be mutually dependent with (x, z) , especially when ξ represents boundary data on the computational domain boundary and (x, z) are modeled as random fields. On the other hand, if ξ represents an unknown boundary shape, ξ might be modeled as mutually independent with (x, z) . In the following we consider the case in which the noise e is additive and the unknowns (x, z, ξ) are not necessarily mutually independent.

Let

$$y = \bar{A}(\bar{x}, z, \xi) + e \in \mathbb{R}^m$$

denote an accurate model for the relation between the measurements and the unknowns,² and let e be mutually independent with (x, z, ξ) .

In the following we approximate the accurate representation of the primary unknown \bar{x} by $x = P\bar{x}$, where P is typically a projection operator. Let $\pi(x, z, \xi, e)$ be a feasible model for the joint distribution of the unknowns. We identify $x = P\bar{x}$ with its coordinates in the associated basis when applicable.

In the approximation error approach, we proceed as follows. Instead of using the accurate forward model $(\bar{x}, z, \xi) \mapsto \bar{A}(\bar{x}, z, \xi)$ with (\bar{x}, z, ξ) as the unknowns, we fix the random variables $(z, \xi) \leftarrow (z_0, \xi_0)$ and use a computationally (possibly drastically reduced) approximative model

$$x \mapsto A(x, z_0, \xi_0)$$

Thus, we write the measurement model in the form

$$y = \bar{A}(\bar{x}, z, \xi) + e \quad (11)$$

$$= A(x, z_0, \xi_0) + [\bar{A}(\bar{x}, z, \xi) - A(x, z_0, \xi_0)] + e \quad (12)$$

$$= A(x, z_0, \xi_0) + \varepsilon + e \quad (13)$$

where we define the *approximation error* $\varepsilon = \varphi(\bar{x}, z, \xi) = \bar{A}(\bar{x}, z, \xi) - A(x, z_0, \xi_0)$. Thus, the approximation error is the discrepancy of predictions of the measurements (given the unknowns) when using the accurate model $\bar{A}(\bar{x}, z, \xi)$ and the approximate model $A(x, z_0, \xi_0)$. Note that (13) is exact.

Formally, after the models \bar{A} and A are fixed, we have $\pi(\varepsilon | \bar{x}, z, \xi) = \delta[\varepsilon - \varphi(\bar{x}, z, \xi)]$. We will later, however, employ approximative joint distributions and therefore consider $\pi(\varepsilon, \bar{x}, z, \xi)$ without any special structure. As the first approximation, we approximate $\varphi(\bar{x}, z, \xi) \approx \varphi(Px, z, \xi)$ and thus $\pi(\varepsilon | \bar{x}, z, \xi) \approx \pi(\varepsilon | Px, z, \xi)$. This means that we assume that the model predictions and thus the approximation error is essentially the same for \bar{x} as $x = P\bar{x}$. This assumption holds for inverse problems in general and for such projections in particular.

²If there are no additive errors, we write $e = 0$ and consider the other types of errors to be included in ξ .

Proceeding as in Section 2.1, we use the Bayes' formula repeatedly

$$\begin{aligned}\pi(y, x, z, \xi, e, \varepsilon) &= \pi(y | x, z, \xi, e, \varepsilon)\pi(x, z, \xi, e, \varepsilon) = \delta[y - A(x, z_0, \xi_0) - e - \varepsilon]\pi(e, \varepsilon | x, z, \xi)\pi(z, \xi | x)\pi(x) \\ &= \pi(y, z, \xi, e, \varepsilon | x)\pi(x)\end{aligned}$$

Hence,

$$\begin{aligned}\pi(y | x) &= \iiint \pi(y, z, \xi, e, \varepsilon | x)de d\varepsilon dz d\xi = \iint \delta[y - A(x, z_0, \xi_0) - e - \varepsilon] \\ &\cdot \left[\iint \pi(e, \varepsilon | x, z, \xi)\pi(z, \xi | x)dz d\xi \right] de d\varepsilon = \iint \delta[y - A(x, z_0, \xi_0) - e - \varepsilon]\pi(e, \varepsilon | x)de d\varepsilon \\ &= \int \pi_e[y - A(x, z_0, \xi_0) - \varepsilon]\pi_{\varepsilon|x}(\varepsilon | x) d\varepsilon\end{aligned}\quad (14)$$

since e and x are mutually independent, and (14) is a convolution integral with respect to ε . In particular, since e is mutually independent with (x, z, ξ) , e and ε are also mutually independent.

At this stage, in the approximation error approach, both π_e and $\pi_{\varepsilon|x}$ are approximated with normal distributions. Let the normal approximation for the joint density $\pi(\varepsilon, x)$ be

$$\pi(\varepsilon, x) \propto \exp \left\{ -\frac{1}{2} \begin{pmatrix} \varepsilon - \varepsilon_* \\ x - x_* \end{pmatrix}^T \begin{pmatrix} \Gamma_{\varepsilon\varepsilon} & \Gamma_{\varepsilon x} \\ \Gamma_{x\varepsilon} & \Gamma_{xx} \end{pmatrix}^{-1} \begin{pmatrix} \varepsilon - \varepsilon_* \\ x - x_* \end{pmatrix} \right\} \quad (15)$$

Thus we write

$$e \sim \mathcal{N}(e_*, \Gamma_e), \quad \varepsilon | x \sim \mathcal{N}(\varepsilon_{*,x}, \Gamma_{\varepsilon|x})$$

where

$$\varepsilon_{*,x} = \varepsilon_* + \Gamma_{\varepsilon x} \Gamma_{xx}^{-1} (x - x_*) \quad (16)$$

$$\Gamma_{\varepsilon|x} = \Gamma_{\varepsilon\varepsilon} - \Gamma_{\varepsilon x} \Gamma_{xx}^{-1} \Gamma_{x\varepsilon} \quad (17)$$

Define the normal random variable ν so that³ $\nu | x = e + \varepsilon | x$

$$\nu | x \sim \mathcal{N}(\nu_{*,|x}, \Gamma_{\nu|x})$$

where

$$\nu_{*,|x} = e_* + \varepsilon_* + \Gamma_{\varepsilon x} \Gamma_x^{-1} (x - x_*) \quad (18)$$

$$\Gamma_{\nu|x} = \Gamma_e + \Gamma_\varepsilon - \Gamma_{\varepsilon x} \Gamma_x^{-1} \Gamma_{x\varepsilon} \quad (19)$$

Thus, we obtain for the approximate likelihood distribution

$$y | x \sim \mathcal{N}[y - A(x, z_0, \xi_0) - \nu_{*,|x}, \Gamma_{\nu|x}]$$

Since we are after computational efficiency, a normal approximation for the prior model is also conventionally used:

$$x \sim \mathcal{N}(x_*, \Gamma_x)$$

Thus, we obtain the approximation for the posterior distribution

$$\pi(x | y) \propto \pi(y | x)\pi(x) \propto \exp \left[-\frac{1}{2} V(x) \right]$$

³With autocovariances, we may notate $\Gamma_{xx} = \Gamma_x$ below.

where $V(x)$ is the posterior potential

$$V(x) = [y - A(x, z_0, \xi_0) - \nu_{*|x}]^T \Gamma_{\nu|x}^{-1} [y - A(x, z_0, \xi_0) - \nu_{*|x}] + (x - x_*)^T \Gamma_x^{-1} (x - x_*) \quad (20)$$

$$= \|L_{\nu|x}[y - A(x, z_0, \xi_0) - \nu_{*|x}]\|^2 + \|L_x(x - x_*)\|^2 \quad (21)$$

where $\Gamma_{\nu|x}^{-1} = L_{\nu|x}^T L_{\nu|x}$ and $\Gamma_x^{-1} = L_x^T L_x$.

2.3 Computational Considerations

In Section 2.2, we wrote the normal approximation (15) for the joint distribution of (x, ε) . Generally, this approximation is done to make an efficient computation of the maximum a posteriori (MAP) estimate feasible. If the actual prior model is normal, the marginal distribution of x induced by (15) coincides with the actual prior model. The prior model $\pi(\bar{x}, z, \xi)$ does not, however, have to be jointly normal and neither, in particular, does the marginal prior model $\pi(\bar{x})$. In practice, whatever the prior model $\pi(\bar{x}, z, \xi)$ is, a set of samples $(\bar{x}^{(\ell)}, z^{(\ell)}, \xi^{(\ell)})$ is usually to be drawn and the approximation errors

$$\varepsilon^{(\ell)} = \varphi(\bar{x}^{(\ell)}, z^{(\ell)}, \xi^{(\ell)}) = \bar{A}(x^{(\ell)}, z^{(\ell)}, \xi^{(\ell)}) - \bar{A}(x^{(\ell)}, z_0, \xi_0), \quad \ell = 1 \dots n_{\text{samp}}$$

are then to be computed, where n_{samp} is the number of draws. The normal approximation for $\pi(\varepsilon, x)$ is then formed by setting $x^{(\ell)} = P\bar{x}^{(\ell)}$ and computing the mean and joint covariance as sample averages over the ensemble.

In the *enhanced error model*, one neglects the cross covariance and sets $\Gamma_{\varepsilon x} = 0$ (see, for example, [5]). With the enhanced error model and nonlinear forward problems, we need to estimate the covariance Γ_ε practically always by simulations and sample averages. If the prior model $\pi(\bar{x})$ is Gaussian, however, the covariance $\Gamma_{\bar{x}}$ is available in the first place and in principle, would, not have to be computed as a sample average.

Irrespective of what the (original) prior model for the primary unknown is, we note the following: When the cross covariances $\Gamma_{\varepsilon x}$ are employed, the sample average has to be used in practice also for Γ_x . Although the prior model covariance $\Gamma_{\bar{x}\bar{x}}$ would yield a “better estimate” for the covariance of the reduced order covariance Γ_x than a sample covariance, we might be forced to use the latter because when the sample covariance is computed, the term $\Gamma_{\varepsilon\varepsilon} - \Gamma_{\varepsilon x} \Gamma_x^{-1} \Gamma_{x\varepsilon}$ is guaranteed to be non-negative definite. If, in addition, Γ_ε has full rank, $\Gamma_{\nu|x}$ is guaranteed to be positive definite and the Cholesky factor $L_{\nu|x}$ exists. But if the prior model covariance $\Gamma_x = P\Gamma_{\bar{x}\bar{x}}P^T$ is used, this condition is not generally met. By the law of large numbers, the condition is met asymptotically but it is impossible to specify a safe sample size. From the point of view of numerical stability, this is a problem especially when the approximation errors clearly dominate the additive errors, that is, the case for which the approximation error approach is targeted in the first place. It is thus advised to use the sample covariance estimate also for Γ_x .

It is difficult to predict how many draws are needed to compute the mean and joint covariance for (x, ε) . Loosely speaking, this depends generally on (the covariances of) the model $\pi(x, z, \xi)$ and the degree of nonlinearity of \bar{A} . With relatively small covariances, few draws seem to be enough (see, for example, [12]). In the approximation error approach, the bottleneck is the computation of the solutions of the full accurate forward model $\bar{A}(\bar{x}, z, \xi)$. As for using the full accurate forward model in the inversion with nonlinear problems, we, of course, have to compute this model along the iteration, but also typically compute the related Jacobian mapping. Thus, the overhead that is related to the computation of the approximation error statistics often corresponds to the computation of a few MAP estimates with the full accurate model.

3. DIFFUSE OPTICAL TOMOGRAPHY

Diffuse optical tomography (DOT) is a noninvasive imaging modality in which images of the optical absorption and scattering within turbid media are derived based on measurements of near-infrared light on the surface of the body (for reviews see [3, 32]). The DOT problem is an exceptionally challenging inverse problem due to the diffuse nature of the forward model, and also since the measurements can span 10 orders of magnitude. Furthermore, there are several unknown distributed parameters involved, of which the *absorption coefficient* and the (*reduced*) *scattering coefficient* are usually reconstructed. Both coefficients are usually measured in mm^{-1} .

The applications of DOT include the detection and classification of tumors from breast tissue, monitoring of infant brain tissue oxygenation level, and functional brain activation studies. For reviews on clinical applications (see [4, 33]). The absorption coefficient, which is related to the oxygenation level of blood, is usually the interesting parameter. In most applications, the scattering coefficient is considered a nuisance parameter. In this paper, our task is thus to employ the approximation error approach for approximate premarginalization over the inhomogeneous scattering coefficient $z(\vec{r})$ and reconstruct only the inhomogeneous absorption coefficient $x(\vec{r})$.

In the numerical examples below, we consider cases in which there are no auxiliary unknowns ξ of any type. Thus, we only have the uninteresting distributed parameter z and additive measurement noise to deal with.

3.1 Measurements in Optical Tomography

In the experimental setup of DOT, m_s optical fibers are placed on the source positions (surface patches) $\partial\Omega_{s,k} \subset \partial\Omega$ on the boundary of the body Ω . The measurements are obtained through m_d optical fibers that are placed in the detector positions (surface patches) $\partial\Omega_{m,i} \subset \partial\Omega$. A collection of measurements is formed by turning on the sources one at a time and measuring the light intensity at all measurement locations (for each source). The measurements in DOT may consist of direct intensity measurements, frequency-modulated amplitude and phase shift measurements, or time-resolved impulse response measurements.

In this paper, we consider the frequency domain measurement system. In the frequency domain measurements light from a sinusoidally modulated laser source is guided via the optic fibers to one of the source locations $\partial\Omega_{s,k}$ at a time, and the amplitudes and phase shifts of the transmitted light are measured on all the detector locations $\partial\Omega_{m,i}, i = 1, \dots, m_d$. The measurement vector is thus $y \in \mathbb{R}^m$ with $m = m_s m_d$. The inverse problem in DOT is to estimate a pair of functions (x, z) , representing the (spatially inhomogeneous) optical absorption coefficient $x(\vec{r})$ and the scattering coefficients $z(\vec{r})$ of the tissues in Ω , given the measurements y and the forward model for the measurement process and noise.

3.2 Forward Model

In the context of inverse problems, the model $x \mapsto A(x)$ is referred to as the *forward model*. We consider DOT simulations in a diffusive regime where the body Ω consists of turbid, scattering dominated media. In such cases the light transport is commonly modeled with the diffusion approximation (DA) of the radiative transfer equation (RTE) [3]. The diffusion approximation is also used as transport model in this paper. For further details on the derivation and properties of the transport models and boundary conditions (see [3, 34, 35]).

We consider the frequency domain system below. Let the light source $\partial\Omega_{s,k}$ be on and $\Phi_k(\vec{r}, \omega)$ be the induced photon density at \vec{r} , where ω is the modulation frequency of the light source. The frequency domain version of the diffusion approximation with the Robin boundary condition is of the form [3, 35]

$$-\nabla \cdot D(\vec{r}) \nabla \Phi_k(\vec{r}, \omega) + x(\vec{r}) \Phi_k(\vec{r}, \omega) + \frac{i\omega}{c} \Phi_k(\vec{r}, \omega) = 0, \quad \vec{r} \in \Omega \quad (22)$$

$$\Phi_k(\vec{r}, \omega) + 2\zeta D(\vec{r}) \frac{\partial \Phi_k(\vec{r}, \omega)}{\partial \mathbf{v}} = g_k(\vec{r}, \omega) \quad \vec{r} \in \partial\Omega \quad (23)$$

where

$$D(\vec{r}) = \{3[x(\vec{r}) + z(\vec{r})]\}^{-1}$$

where D (units [mm]) is the diffusion coefficient, c is the speed of light in the medium, ζ is a parameter that describes reflection on the boundary, \mathbf{v} is the outward normal vector at $\partial\Omega$, and $g_k(\vec{r}, \omega)$ is the boundary source term for source at $\partial\Omega_{s,k}$,

$$g_k(\vec{r}, \omega) = \begin{cases} I & \text{on } \partial\Omega_{s,k} \\ 0 & \text{on } \partial\Omega \setminus \partial\Omega_{s,k} \end{cases} \quad (24)$$

where I is the intensity of the source. The complex-valued flux $\rho_{i,k}(\omega)$ at the measurement site $\partial\Omega_{m,i}$ can be written as the surface integral

$$\rho_{i,k}(\omega) = \int_{\partial\Omega_{m,i}} \frac{1}{2\zeta} \Phi_k(\vec{r}, \omega) dS \quad (25)$$

The collection $\{\rho_{i,k}\}$, $k = 1, \dots, m_s$, $i = 1, \dots, m_d$ is the *raw* data for the experiment. For numerical reasons, primarily for the range of measurements, transformed data and the correspondingly transformed forward model are typically used. Furthermore, the measurement systems are constructed according to these transformations. The experimental systems for frequency domain optical tomography export the log-amplitude and phase shift of the complex valued (demodulated) raw data. Thus, the measurements can be written

$$y = \begin{pmatrix} \text{Re} \log(\rho) \\ \text{Im} \log(\rho) \end{pmatrix} \in \mathbb{R}^{2m} \quad (26)$$

and the forward model is transformed accordingly. For different end uses, data with different frequencies ω may be acquired.

For the numerical realization of the diffusion approximation model (22)–(24), the finite element method (FEM) is typically used (see, for example, [3, 7]). In the FEM approximation, photon density is approximated in a finite dimensional basis as

$$\Phi^h(\vec{r}) = \sum_{i=1}^{N_n} \alpha_i \varphi_i(\vec{r}) \quad (27)$$

where $\varphi_i(\vec{r})$ are the nodal basis functions of the finite element mesh and N_n is the number of nodes in the FEM mesh, and h is mesh element size parameter.

The absorption and scattering coefficients are written as finite dimensional approximations

$$x(\vec{r}) = \sum_{j=1}^{n_p} x_j \chi_j(\vec{r}), \quad z(\vec{r}) = \sum_{j=1}^{n_p} z_j \chi_j(\vec{r}) \quad (28)$$

where χ_j denotes characteristic functions of disjoint elements in the reconstruction mesh. In the following, we identify the absorption and scattering coefficients and their representations as the coordinates

$$x = (x_1, \dots, x_{n_p})^T \in \mathbb{R}^{n_p}, \quad z = (z_1, \dots, z_{n_p})^T \in \mathbb{R}^{n_p}$$

Thus, the solution of the forward problem amounts to the solution of m_s complex valued $N_n \times N_n$ systems of equations for one DOT experiment. The FEM-based forward model is thus of the form

$$y = A_h(x, z) \quad (29)$$

where h refers to the discretization level parameter in (27).

3.3 Gaussian MRF Prior Model for Scattering and Absorption Coefficients

In the following, we model (x, z) as mutually independent. As the prior model for both $\pi(x)$ and $\pi(z)$, we used a proper Gaussian smoothness prior, constructed similarly as in [5, 7, 8]. In this construction, the distributed parameter, say x , is considered in the form

$$x(\vec{r}) = x_{\text{in}}(\vec{r}) + x_{\text{bg}}(\vec{r})$$

where $x_{\text{in}}(\vec{r})$ is a spatially inhomogeneous (absorption) coefficient⁴ with zero mean, and $x_{\text{bg}}(\vec{r})$ is a spatially homogeneous (background) absorption coefficient with nonzero mean. For the latter, we can write $x_{\text{bg}}(\vec{r}) = q\mathbb{I}$, where \mathbb{I} is a vector of ones and q is a scalar random variable with distribution $q \sim \mathcal{N}(x_*, \sigma_{\text{bg},x}^2)$. With respect to the basis for x , we have the coordinates $x_{\text{in}} \in \mathbb{R}^{n_p}$, $\mathbb{I} \in \mathbb{R}^{n_p}$, and set $x_{\text{in}} \sim \mathcal{N}(0, \Gamma_{\text{in}})$. We model the spatial distributions x_{in} and

⁴In the sequel, “in” refers to inhomogeneous, “bg” to background.

$q\mathbb{I}$ as mutually independent, that is, the background is mutually independent with the inhomogeneities. An equivalent construction for z was considered.

Thus, we have the $\Gamma_x = \Gamma_{\text{in},x} + \sigma_{\text{bg},x}^2 \mathbb{I}\mathbb{I}^T$, $\Gamma_z = \Gamma_{\text{in},z} + \sigma_{\text{bg},z}^2 \mathbb{I}\mathbb{I}^T$ and

$$\pi(x) = \mathcal{N}(x_*\mathbb{I}, \Gamma_x), \quad \pi(z) = \mathcal{N}(z_*\mathbb{I}, \Gamma_z)$$

In the construction of $\Gamma_{\text{in},x}$ and $\Gamma_{\text{in},z}$, the approximate correlation lengths can be adjusted to match the size of the expected inhomogeneities. (See [5, 7, 8] for details.)

This prior model is a proper distribution, that is, the covariance exists. Traditional smoothness prior models are improper and samples cannot be drawn from such distributions. The approximation error approach, on the other hand, is based on computing the statistics of ε over the prior distribution. This is not possible with a prior of unbounded variances.

In [7] it was found that such a specific construction for the prior model for scattering and absorption coefficients was exceptionally suitable. This was the case even without the variable background. In [8], the prior described above with the variable background was shown to be feasible also for real data.

4. NUMERICAL STUDIES

We evaluate the approximate premarginalization by the approximation error approach with three two-dimensional numerical examples. In the first one, we study only the errors that are related to marginalization over the scattering coefficient using an otherwise accurate forward model $\bar{A}(x, z_0)$. Thus, numerical model reduction errors are not present. The second example is similar, but the prior model for the scattering coefficient is somewhat off in the sense that the actual background of scattering differs from the modeled one in $\bar{A}(x, z_0)$. In the third example, the numerical model reduction is included, that is, we use the model $A(x, z_0)$.

In the following we explain the common details for the numerical examples.

4.1 Computational Forward Models and Prior Model

In the numerical studies, the domain $\Omega \subset \mathbb{R}^2$ is a circle with radius 25 mm. The measurement setup consists of $m_s = 32$ sources and $m_d = 32$ detectors, located at equispaced intervals on the boundary $\partial\Omega$. With this setup, the number of log-amplitude and phase measurements is $2m = 2048$.

The simulated measurement data is computed with the FEM approximation of the diffusion approximation model in a mesh \mathcal{M}_1 which is dense enough to consider the solution as the solution of the problem (22)–(24). Two other finite element meshes and models are constructed, the (relatively) accurate $[\mathcal{M}_2, \bar{A}(x, z)]$ and the radically reduced one $[\mathcal{M}_3, A(x, z)]$ (see Table 1). The actual simulated measurement data was obtained by adding mutually independent (non-identically distributed) noise $\sigma_{e,j} = \delta|y_{*,j}|/100$ with $\delta = 0.5$, that is, the error level was 0.5% of the respective noiseless measurement. Thus, the (instrument) noise is additive and independent, identically distributed, but the individual variances are not equal, that is, the covariance Γ_e is diagonal but the diagonal entries are not equal.

For the reconstruction basis in (28) for the coefficients (x, z) in the inverse problem, we divide the domain Ω into $n_p = 1904$ square pixels for both the accurate and approximate models. Thus we have $P = I$, $\bar{x} = Px = x$, and the parameter vectors $\bar{x}, z \in \mathbb{R}^{1904}$. Having $P : \mathbb{R}^n \mapsto \mathbb{R}^m$ where $m \ll n$ has an impact mainly when m is very small,

TABLE 1: The FEM meshes used in the simulations: N_n is the number of nodes and N_e is the number of triangular elements in the mesh.

Mesh	Use	N_n	N_e
\mathcal{M}_1	Simulation of measurements	12,853	25,326
\mathcal{M}_2	Accurate forward model	11,329	22,302
\mathcal{M}_3	Reduced forward model	703	1326

and less with having $m \ll n$. An exception is poor (initial) modeling, that is, if the accurate representation \bar{x} is a poor approximation for the reality.

The prior model was constructed as in Section 3.3. The parameters for the prior distribution are given in Table 2. Elementwise, this means that the values of absorption and scatter coefficients are expected to lie within the two standard deviation intervals $x_* \pm 2\sigma_x$ and $z_* \pm 2\sigma_z$ with probability of 95%, with respect to the marginal prior distributions.

Two draws from the prior models $\pi(x)$ and $\pi(z)$ are shown in Fig. 1. In the draws, the variation of the background that was part of the construction of the prior model is clearly visible.

4.2 Estimates and Approximation Error Statistics

We compute MAP estimates only by minimizing the respective posterior potentials. The following particular estimates are computed in the three test cases:

TABLE 2: The parameters for the prior model: means and the standard deviations σ for the homogeneous background and inhomogeneities. The correlation length for both coefficients was set as 8 mm.

x_*	$10.0 \cdot 10^{-3} \text{ mm}^{-1}$
$\sigma_{\text{bg},x}$	$1.2 \cdot 10^{-3} \text{ mm}^{-1}$
$\sigma_{\text{in},x}$	$3.0 \cdot 10^{-3} \text{ mm}^{-1}$
z_*	$1000 \cdot 10^{-3} \text{ mm}^{-1}$
$\sigma_{\text{bg},z}$	$120 \cdot 10^{-3} \text{ mm}^{-1}$
$\sigma_{\text{in},z}$	$300 \cdot 10^{-3} \text{ mm}^{-1}$

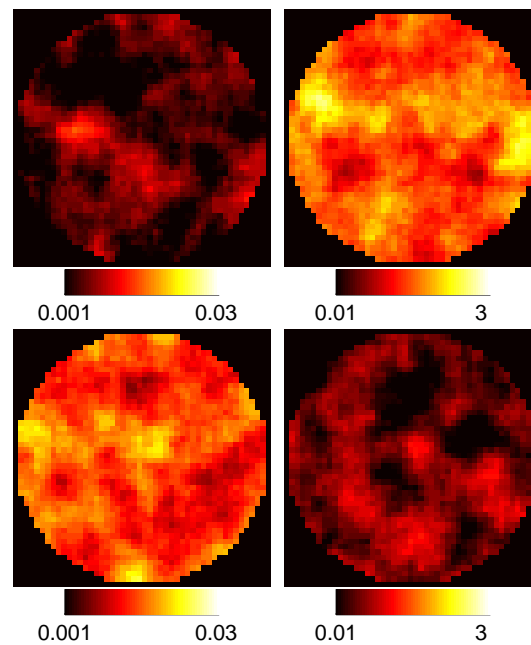


FIG. 1: Two draws from the prior distribution: (left) absorption coefficients $x(\vec{r})$ and (right) scattering coefficient $z(\vec{r})$.

1. MAP-REF—Maximum a posteriori estimate for both parameters (x, z) with the conventional error model $y = \bar{A}(x, z) + e$. This estimate is obtained by computing

$$\min_{x,z} \{ \|L_e[y - \bar{A}(x, z) - e_*]\|^2 + \|L_x(x - x_*)\|^2 + \|L_z(z - z_*)\|^2 \} \quad (30)$$

and can be considered as a reference estimate in which both distributed parameters are estimated simultaneously.

2. MAP-CEM—Maximum a posteriori estimate for the primary unknown x using fixed $z = z_*$ and conventional error model $y = \bar{A}(x, z_*) + e$, corresponding to

$$\min_x \{ \|L_e[y - \bar{A}(x, z_*) - e_*]\|^2 + \|L_x(x - x_*)\|^2 \} \quad (31)$$

3. MAP-AEM—Maximum a posteriori estimate for x using fixed $z = z_*$ and the approximation error model, corresponding to

$$\min_x \{ \|L_{\nu|x}[y - \bar{A}(x, z_*) - \nu_{*|x}]\|^2 + \|L_x(x - x_*)\|^2 \} \quad (32)$$

In the above functionals, when the reduced-order model is used, the model $\bar{A}(x, \cdot)$ is to be substituted by $A(x, \cdot)$ [see the posterior potential $V(x)$ in (21)].

In the following numerical examples, the realization z_* is the mean of the prior model $\pi(z)$. The estimates (30)–(32) are computed with the Gauss–Newton optimization method with an explicit line search [36].

For the construction of the approximation error statistics, we proceed as follows. The statistics were used only with MAP with the approximation error model (MAP–AEM) and were computed to correspond to the employed forward model. The means and covariances for (x, ε) in the approximation error model (15) were estimated by Monte Carlo simulation. For this, we draw the sets of samples $\{x^{(\ell)}, \ell = 1, \dots, n_{\text{samp}}\}$ and $\{z^{(\ell)}, \ell = 1, \dots, n_{\text{samp}}\}$ from the prior models $\pi(x)$ and $\pi(z)$, respectively. Using the sets of samples, the realizations of the approximation error are computed as

$$\varepsilon^{(\ell)} = \bar{A}(x^{(\ell)}, z^{(\ell)}) - \bar{A}(x^{(\ell)}, z_0), \quad \ell = 1 \dots n_{\text{samp}}$$

for the case where ε is due to using fixed $z = z_0$ only, that is, no model reduction errors are present, and

$$\varepsilon^{(\ell)} = \bar{A}(x^{(\ell)}, z^{(\ell)}) - A(x^{(\ell)}, z_0), \quad \ell = 1 \dots n_{\text{samp}}$$

for the cases in which both errors from model reduction and using fixed $z = z_0$ are present. The means x_* and $\nu_{*|x}$ and the covariances Γ_x and $\Gamma_{\nu|x}$ are then estimated as sample averages using the samples $\{x^{(\ell)}, \varepsilon^{(\ell)}, \ell = 1 \dots n_{\text{samp}}\}$. In the following examples, we use sample size $n_{\text{samp}} = 20,000$.

4.3 Reconstructions

We have three different parameters z : $z_{\text{bg,actual}}$ is the actual (unknown distributed) parameter, z_* is the mean of the modeled prior distribution, and z_0 is the fixed value of z used in the posterior model. Both z_0 and z_* can be chosen separately and we do not need to have $z_0 \neq z_*$. Technically, it is possible to optimize z_0 , but this is not considered here.

Three reconstruction cases are considered: case 1—marginalization over z only with $z_0 = z_*$, and furthermore, $z_0 = z_{\text{bg,actual}}$; case 2—as in Case 1 but with $z_0 \neq z_{\text{bg,actual}}$ to assess the robustness toward the poor choice of z_0 ; and case 3—both marginalization over z and numerical model reduction errors are present. The actual spatial distributions for x and z are blocky targets in a homogeneous background. The probability of the actual x and z with respect to the prior models is relatively low since they are discontinuous (see top row of any of Figs. 2–4). This is one of the ways to check the robustness of the approximation error approach against prior model design.

Case 1: Modeling errors caused by using a fixed value for z .—The results for case 1 are shown in Fig. 2. The actual coefficients (x, z) are shown in the top row. The background values of the target distributions coincide with the prior means. In particular, $z_{\text{bg,actual}} = z_0$.

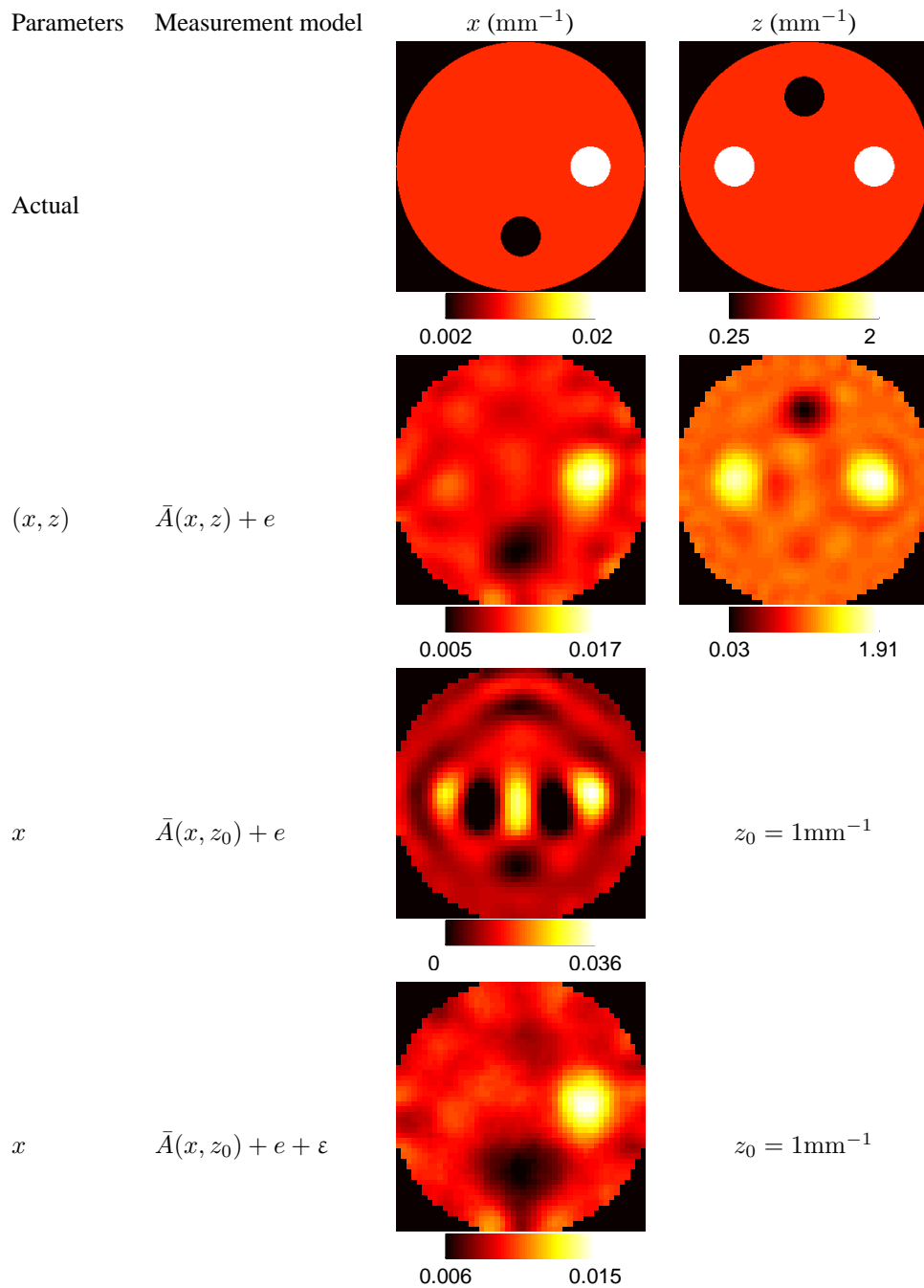


FIG. 2: Case 1. Rows from top to bottom: actual distributions, MAP-REF estimate (30) for (x, z) , MAP-CEM estimate (31) for x using fixed $z_0 = z_*$ and conventional noise model, and MAP-AEM estimate (32) using fixed $z_0 = z_*$ and the approximation error model. The modeling error in MAP-CEM and MAP-AEM is caused solely by using the (incorrect) fixed value $z_0 = z_*$. Here $z_{\text{bg,actual}} = z_* = z_0$.

The reference estimates MAP-REF for (x, z) , computed by minimization of (30), are shown in the second row. The estimates are computed using the accurate forward model $\bar{A}(x, z)$, meaning that there are no discretization-related errors present. The MAP estimate with the conventional error model (MAP-CEM) estimate (31) for x , using the fixed

value $z = z_*$ for the scattering and the conventional noise model $y = \bar{A}(x, z_*) + e$, is shown in the third row. As can be seen, the use of incorrect value z_* (which is here equivalent to the background scatter) has effectively destroyed the reconstruction of the absorption coefficient x , although the actual z differs from the modeled z_0 only in three small subdomains.

The fourth row shows the MAP-AEM estimate (32) for x using the same fixed $z = z_*$ and the approximation error model $y = \bar{A}(x, z_*) + \varepsilon + e$. The estimate for the absorption is similar to the MAP-REF estimate, but the circular targets are slightly more blurred and the values of the inhomogeneous targets are slightly more off the actual values than with the MAP-REF. This is unavoidable in most cases, since the inclusion of the statistics of the approximation errors increases the variances of their likelihood, which in turn drives the estimates toward the mean of the prior model.

Case 2: Tolerance against poor choice of z_0 .— The results for case 2 are shown in Fig. 3. The estimates are computed and arranged as in Fig. 2, the only difference being that in Fig. 3 the actual background value of the scattering target in $z_{\text{bg,actual}} = 1.5z_0$, the discrepancy of which corresponds to $1.5\sigma_z$ with respect to the prior model.

The decrease of the quality of the MAP-REF is due to using a poor model for the prior (background) mean only. This is still a reasonable estimate showing all five objects of the actual spatial distributions. The MAP-CEM estimate is completely useless.

The reconstruction quality in the MAP-AEM estimate is similar to that in Fig. 2. This means that a $1.5\sigma_z$ underestimation of the scattering coefficient in the model $A(x, z_0)$ is tolerated well. We also tested a case where the background of the scattering target was $-1.5\sigma_z$ and $\pm 2\sigma_z$ away from the mean z_0 . The MAP-AEM estimate of x remained similar to the estimate shown in bottom row in Fig. 3, except in the case $z_{\text{bg,actual}} = z_0 - 2\sigma_z$. In this case, the quality of the MAP-AEM estimate decreased significantly, producing a nearly useless estimate. Such misspecification of the background is not expected in practice, since the estimation of the best spatially homogeneous estimates for x and z can be done readily if the approximate background values are not known (see [8]). Based on these numerical tests, the approximate premarginalization with respect to the unknown scattering parameter z is reasonably tolerant against a discrepancy between the prior model and the fixed $z = z_0$.

Case 3: Combined approximation error caused by fixed $z = z_*$ and model reduction.— The results for case 3 are shown in Fig. 4. The estimates are computed from the same data that was used in Fig. 2. The only difference between case 1 and case 3 is that in case 3 the estimates are computed using the reduced forward model $A(x, z_0)$, with the number of nodes in the FEM approximation being $N_n = 1326$, instead of $N_n = 22,302$ in the accurate model $\bar{A}(x, z_0)$.

The reference estimate MAP-REF for both unknowns (x, z) was destroyed by the unaccounted for approximation error caused by use of the reduced FEM model. As with case 2, the MAP-CEM estimate is again useless. On the other hand, the MAP-AEM estimate with the proposed approach remained similar to case 1, showing that simultaneous recovery from the use of an incorrect fixed value of z and model-reduction-related errors is feasible with the proposed approach.

We did not consider any auxiliary uncertainties ξ in the above example. In DOT, the principal candidate for ξ is poorly known exterior geometry, but there are also other topics. In clinical applications of optical tomography in particular, the actual optode locations might be slightly off the modeled locations. Furthermore, the channel amplifications can only be calibrated up to a constant with, which constant is difficult to estimate.

5. CONCLUSIONS

In this paper we applied the recently proposed approximation error approach for approximate premarginalization of uninteresting distributed parameters. The approximation error approach is based on the Bayesian framework for inverse problems. The approximation error approach has earlier been shown to be able to deal with diverse types of approximation and modeling errors and uncertainties, such as those related to pure model reduction, unknown boundary data, mismodeled geometry, and use of approximative physical models.

We considered the special case of approximate premarginalization over the inhomogeneous scattering coefficient in diffuse optical tomography. This is an example of a problem in which there are two or more unknown distributed parameters, of which only one is of interest. The results for this example problem suggest that the approximation error

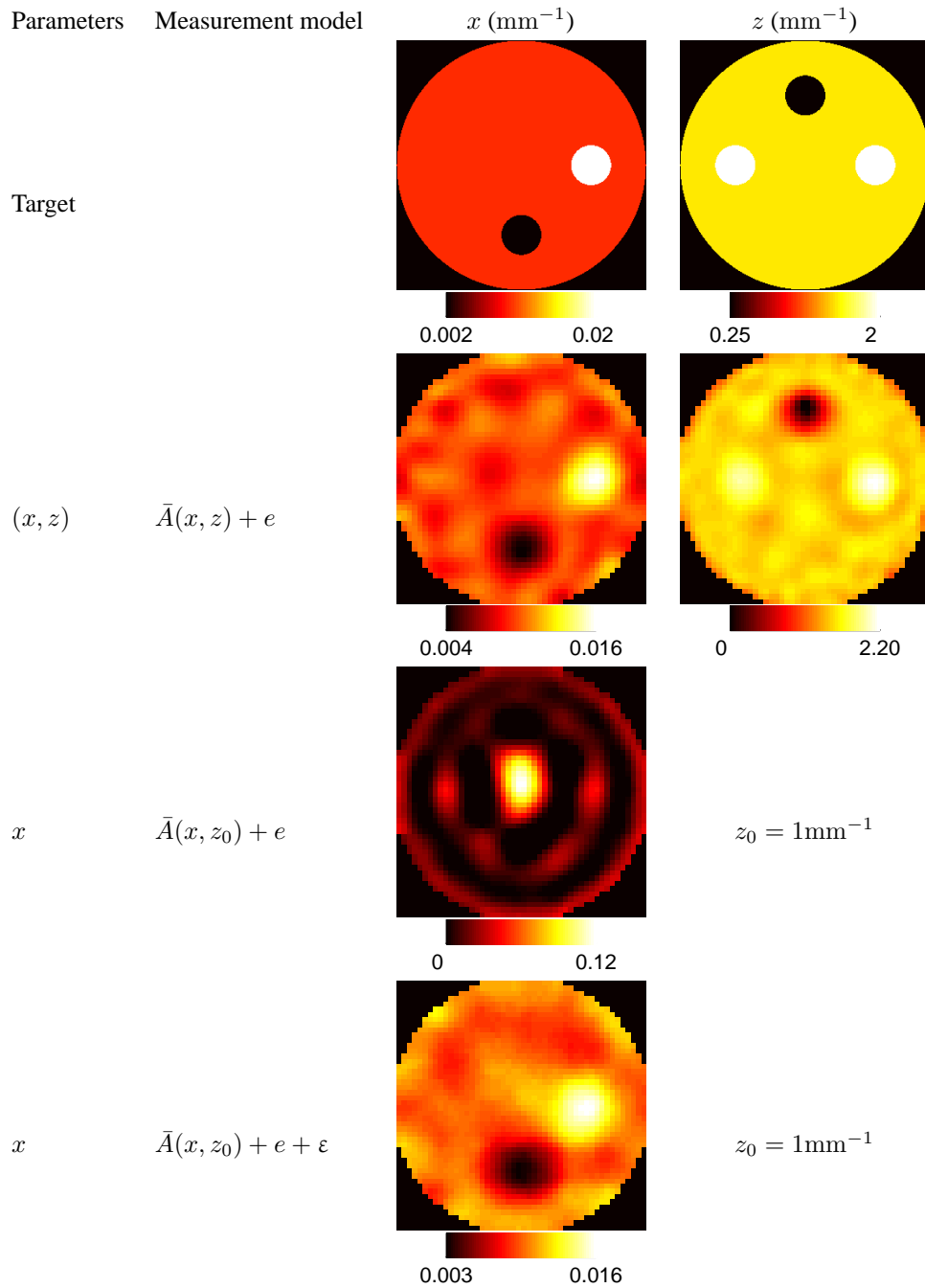


FIG. 3: Case 2. Misspecification of z_0 . Rows as in Fig. 2. The modeling error in MAP-CEM and MAP-AEM is caused solely by using a fixed value $z = z_0$. Here $z_{\text{bg,actual}} = z_* + 1.5\sigma_z = z_0 + 1.5\sigma_z$.

approach is able to compensate for using an incorrect fixed value for the uninteresting distributed parameter. In this particular example, the premarginalization was carried over the scattering coefficient, as well as the errors related to simultaneous reduction of the computational forward model. It was also shown that at least in the studied cases, the approach is tolerant to a reasonable misspecification of the fixed scattering coefficient.

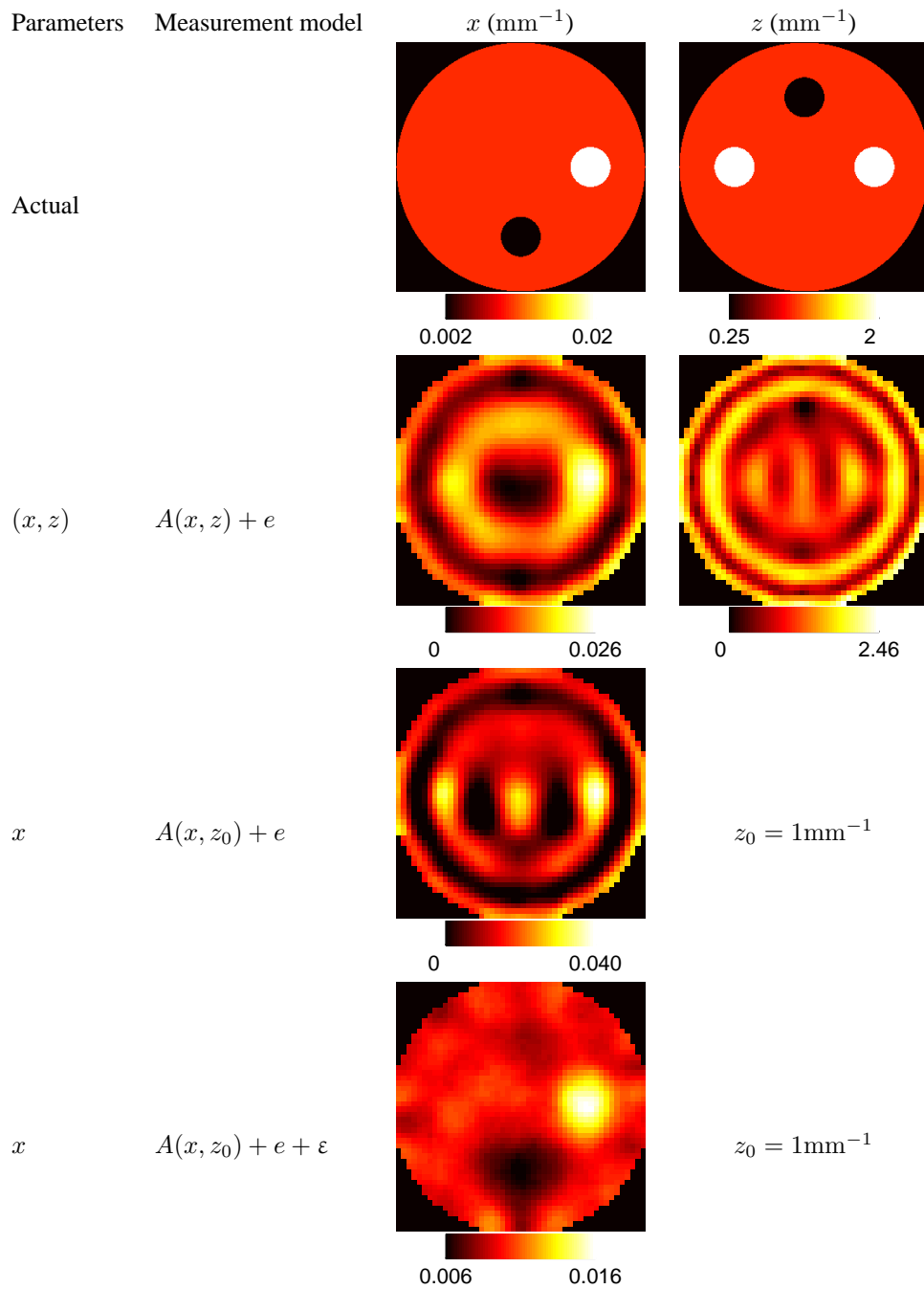


FIG. 4: Case 3. Rows as in Fig. 2. The modeling error in MAP-CEM and MAP-AEM is caused by using a fixed value $z = z_0$ and the reduced-order model $A(x, z_0)$ for the forward problem. Here $z_{\text{bg,actual}} = z_* = z_0$.

ACKNOWLEDGMENTS

This work was supported by the Academy of Finland, projects 119270, 122499, 218183, 140731 and 213476, the Finnish Centre of Excellence in Inverse Problems Research 2006-2011, and by EPSRC project EP/E034950/1.

REFERENCES

1. Daily, W. D., Ramirez, A. L., LaBrecque, D. J., and Nitao, J., Electrical resistivity tomography of vadose water movement, *Water Resour. Res.*, 28:1429–1442, 1992.
2. Daily, W. D. Ramirez, A. L., Electrical resistivity tomography, *The Leading Edge*, 23(1):438–442, 2004.
3. Arridge, S. R., Optical tomography in medical imaging, *Inverse Probl.*, 15:R41–R93, 1999.
4. Gibson, A. P., Hebden, J. C., and Arridge, S. R., Recent advances in diffuse optical imaging, *Phys. Med. Biol.*, 50:R1–R43, 2005.
5. Kaipio, J. and Somersalo, E., *Statistical and Computational Inverse Problems*, Springer, New York, 2005.
6. Kaipio, J. and Somersalo, E., Statistical inverse problems: Discretization, model reduction and inverse crimes, *J. Comput. Appl. Math.*, 198:493–504, 2007.
7. Arridge, S., Kaipio, J., Kolehmainen, V., Schweiger, M., Somersalo, E., Tarvainen, T., and Vauhkonen, M., Approximation errors and model reduction with an application in optical diffusion tomography, *Inverse Probl.*, 22:175–195, 2006.
8. Kolehmainen, V., Schweiger, M., Nissilä, I., Tarvainen, T., Arridge, S., and Kaipio, J., Approximation errors and model reduction in three-dimensional optical tomography, *J. Opt. Soc. Am. A*, 26:2257–2268, 2009.
9. Heino, J. and Somersalo, E., A modelling error approach for the estimation of optical absorption in the presence of anisotropies, *Phys. Med. Biol.*, 49:4785–4798, 2004.
10. Heino, J., Somersalo, E., and Kaipio, J., Compensation for geometric mismodelling by anisotropies in optical tomography, *Opt. Express*, 13(1):296–308, 2005.
11. Calvetti, D., Kaipio, J. P., and Somersalo, E., Aristotelian prior boundary conditions, *Int. J. Math.*, 1:63–81, 2006.
12. Lehtikoinen, A., Finsterle, S., Voutilainen, A., Heikkinen, L., Vauhkonen, M., and Kaipio, J., Approximation errors and truncation of computational domains with application to geophysical tomography, *Inverse Probl. Imag.*, 1:371–389, 2007.
13. Nissinen, A., Heikkinen, L., and Kaipio, J. P., Approximation errors in electrical impedance tomography—An experimental study, *Meas. Sci. Technol.*, 19:015501, 2008.
14. Nissinen, A., Heikkinen, L., Kolehmainen, V., and Kaipio, J. P., Compensation of errors due to discretization, domain truncation and unknown contact impedances in electrical impedance tomography, *Meas. Sci. Technol.*, 20:105504, 2009.
15. Nissinen, A., Kolehmainen, V., and Kaipio, J. P., Compensation of modelling errors due to unknown domain boundary in electrical impedance tomography, *IEEE Trans. Med. Imag.*, 2010, in press.
16. Huttunen, J. and Kaipio, J., Approximation errors in nonstationary inverse problems, *Inverse Probl. Imag.*, 1(1):77–93, 2007.
17. Huttunen, J. and Kaipio, J., Approximation error analysis in nonlinear state estimation with an application to state-space identification, *Inverse Probl.*, 23:2141–2157, 2007.
18. Huttunen, J. and Kaipio, J., Model reduction in state identification problems with an application to determination of thermal parameters, *Appl. Numer. Math.*, 59:877–890, 2009.
19. Seppänen, A., Vauhkonen, M., Vauhkonen, P., Somersalo, E., and Kaipio, J., State estimation with fluid dynamical evolution models in process tomography — An application to impedance tomography, *Inverse Probl.*, 17:467–484, 2001.
20. Huttunen, J., Lehtikoinen, A., Hämäläinen, J., and Kaipio, J., Importance filtering approach for the nonstationary approximation error method, *Inverse Probl.*, 2010, in press.
21. Lehtikoinen, A., Huttunen, J., Voutilainen, A., Finsterle, S., Kowalsky, M., and Kaipio, J. P., Dynamic inversion for hydrological process monitoring with electrical resistance tomography under model uncertainties, *Water Resour. Res.*, 46:W04513, 2010.
22. Tarvainen, T., Kolehmainen, V., Pulkkinen, A., Vauhkonen, M., Schweiger, M., Arridge, S. R., and Kaipio, J. P., Approximation error approach for compensating for modelling errors between the radiative transfer equation and the diffusion approximation in diffuse optical tomography, *Inverse Probl.*, 26:015005, 2010.
23. Tarantola, A., *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, Philadelphia, 2004.
24. Calvetti, D. and Somersalo, E., *An Introduction to Bayesian Scientific Computing, Ten Lectures on Subjective Computing*, ISBN 978-0-387-73393-7, Springer, 2007.
25. Berger, J., *Statistical Decision Theory and Bayesian Analysis*, Springer, 2006.
26. Robert, C., *The Bayesian Choice*, Springer, 2007.

27. Chen, M.-H., Shao, Q.-M., and Ibrahim, J., *Monte Carlo Methods in Bayesian Computation*, Springer, 2000.
28. J. Liu, *Monte Carlo Strategies in Scientific Computing*, ISBN 0-387-95230-6 in Springer Series in Statistics, Springer, 2005.
29. Somersalo, E., Kaipio, J., Vauhkonen, M., Baroudi, D., and Järvenpää, S., Impedance Imaging and Markov Chain Monte Carlo Methods, In R. Barbour, M. Carvlin, and M. Fiddy (Eds.), *Proc SPIE's 42nd Annual Meeting, Computational, Experimental and Numerical Methods for Solving Ill-Posed Inverse Imaging Problems: Medical and Nonmedical Applications*, pp. 175–185, San Diego, USA, June 27-Aug 1, 1997.
30. Fox, C. and Nicholls, G., Sampling conductivity images via MCMC, In K. V. Mardia, C. A. Gill, and R. G. Aykroyd (Eds.), *The Art and Science of Bayesian Image Analysis, Proceedings of the Leeds Annual Statistics Research Workshop*, pp. 91–100, Leeds University Press, Leeds, UK, 1–4 July 1997.
31. Kaipio, J. P., Kolehmainen, V., Somersalo, E., and Vauhkonen, M., Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography, *Inverse Probl.*, 16:1487–1522, 2000.
32. Arridge, S. and Schotland, J., Optical tomography: Forward and inverse problems, *Inverse Probl.*, 25:123010, 2009.
33. Tromberg, B. J., Pogue, B. W., Paulsen, K. D., Yodh, A. G., Boas, D. A., and Cerussi, A. E., Assessing the future of diffuse imaging technologies for breast cancer management, *Med. Phys.*, 35:2443–2451, 2008.
34. Schweiger, M., Arridge, S. R., Hiraoka, M., and Delpy, D. T., The finite element model for the propagation of light in scattering media: Boundary and source conditions, *Med. Phys.*, 22(11):1779–1792, 1995.
35. Heino, J. and Somersalo, E., Estimation of optical absorption in anisotropic background, *Inverse Probl.*, 18:559–573, 2002.
36. Nocedal, J. and Wright, S. J., *Numerical Optimization*, 2nd ed., Springer, New York, 2006.