PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Analyzing resolution and model uncertainties for ultrasound computed tomography using Hessian information

Ines E. Ulrich, Christian Boehm, Andrea Zunino, Andreas Fichtner

Ines E. Ulrich, Christian Boehm, Andrea Zunino, Andreas Fichtner, "Analyzing resolution and model uncertainties for ultrasound computed tomography using Hessian information," Proc. SPIE 12038, Medical Imaging 2022: Ultrasonic Imaging and Tomography, 120380A (4 April 2022); doi: 10.1117/12.2608546

SPIE.

Event: SPIE Medical Imaging, 2022, San Diego, California, United States

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Jun 2022 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Analyzing Resolution and Model Uncertainties for Ultrasound Computed Tomography using Hessian Information

Ines E. Ulrich[a], Christian Boehm[a], Andrea Zunino[a], and Andreas Fichtner[a]

[a]ETH Zurich, Sonneggstrasse 5, Zurich, Switzerland

ABSTRACT

Characterizing the spatial resolution and uncertainties related to a tomographic reconstruction are crucial to assess its quality and to assist with the decision-making process. Bayesian inference provides a general framework to compute conditional probability density functions of the model space. However, analytic expressions and closed-form solutions for the posterior probability density are limited to linear inverse problems such as straight-ray tomography under the assumption of a Gaussian prior and data noise. Resolution analysis and uncertainty quantification is significantly more complicated for non-linear inverse problems such as full-waveform inversion (FWI), and sampling-based approaches such as Markov-Chain Monte-Carlo are often impractical because of their tremendous computational cost. However, under the assumption of Gaussian priors in model and data space, we can exploit the machinery of linear resolution analysis and find a Gaussian approximation of the posterior probability density by using the Hessian of the regularized objective functional. This non-linear resolution analysis rests on (i) a quadratic approximation of the misfit functional in the vicinity of an optimal model; (ii) the idea that an approximation of the Hessian can be built efficiently by gradient information from a set of perturbed models around the optimal model. The inverse of the preconditioned Hessian serves as a proxy of the posterior covariance from which space-dependent uncertainties as well as correlations between parameters and inter-parameter trade-offs can be extracted. Moreover, the framework proposed here also allows for inter-comparison between different tomographic techniques. Specifically, we aim for a comparison between tissue models obtained from ray tomography and models obtained with FWI using ultrasound data.

Keywords: resolution analysis, uncertainty quantification, ultrasound-computed tomography, full-waveform inversion, inverse problem

1. INTRODUCTION

Tomographic models obtained from indirect measurements are one of the most important sources of information that medical diagnoses are based on. The specific nature of the acquired ultrasound data often dictates the tomographic method that can be used to infer tissue structure from the data. A crucial question that arises in the subsequent analysis of models obtained by different tomographic methods is how to compare their resolution, or more precisely, how much information on resolving qualities of the model can be gained from the specific mathematical description used in the tomographic method. In this study, we focus on the inter-comparison between the resolution quality of models recovered from straight-ray tomography (SRT) and reconstructions obtained with full-waveform inversion (FWI). For SRT, where the mapping from slowness to times-of-flight is linear, Bayes's rule provides a rigorous framework to compute the posterior distribution of the model parameters. More specifically, one obtains a closed-form solution of the posterior covariance matrix in the case of a Gaussian prior and data errors.

In FWI, where the parameter-to-solution map is non-linear, such an analytical assessment of model uncertainties is generally not possible. Therefore, the first part of this study will define a resolution framework for FWI by linearizing the inverse problem under the assumption of Gaussian uncertainties, which allows us to exploit the machinery of linear resolution analysis resulting in an approximation to the posterior covariance matrix. Tomographic images obtained with full-waveform inversion (FWI) use the entire waveform information including non-linear effects of wave propagation, which has the ability to provide a higher resolution compared

Further author information: (Send correspondence to I.E.U.)
I.E.U.: E-mail: ines.ulrich@erdw.ethz.ch, Telephone: +41 44 633 35 66

Medical Imaging 2022: Ultrasonic Imaging and Tomography, edited by Nick Bottenus, Nicole V. Ruiter,
Proc. of SPIE Vol. 12038, 120380A · © 2022 SPIE · 1605-7422 · doi: 10.1117/12.2608546

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Jun 2022
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

to reconstructions obtained with ray theory. The latter, although frequently used in medical ultrasound, is solely a valid description of wave propagation for a medium that is sufficiently smooth and does not vary significantly over a wavelength.[1] Recent applications of FWI to ultrasound computed tomography (USCT) data show great potential in significantly increasing the quality of recovered models.[2–6] However, a pitfall of visually analyzing the resolution of tomographic reconstructions is to mistake artifacts for small-scale features within the tissue structure.

Therefore, a rigorous resolution analysis in pixel and model space accounting for inter-parameter trade-offs is desirable to measure the quality of a reconstruction. For non-linear problems such as FWI, probabilistic methods that quantify uncertainties and approximate probability density functions by exploring the model space via random sampling strategies[7] are commonly employed. Although these methods provide statistical metrics for uncertainty quantification, they are computationally demanding for problems with high dimensional parameter spaces as for instance in USCT. In this work, we propose to stay within the deterministic inversion framework and to exploit the properties of the Hessian operator to obtain measures for resolution. Specifically, we show that the Hessian operator, which encodes the curvature of the misfit functional and arises naturally in a Newton-based inversion framework, can contribute in three ways to a resolution analysis: (1) it can be interpreted in terms of the inverse posterior covariance;[8–10] (2) probing its columns (resp. rows) enables measuring inter-parameter blurring and position- and direction-dependent resolution lengths[11] similar to common point-spread function tests[12] and (3) the Hessian can be used to determine the resolution gain during an inversion process.

2. METHODS

In order to treat different inversion strategies based on (straight-)ray tomography or full-waveform inversion within the same framework, we consider a generic data vector $\mathbf{d}^{\text{obs}}$ combining measurements from all emitters and receivers, as well as a model vector $\mathbf{m}$ describing the tissue structure. For instance, $\mathbf{d}^{\text{obs}}$ can encode traveltime data for ray tomography, or the entire time series of A-scans for FWI, respectively. Likewise, $\mathbf{m}$ could be slowness, speed-of-sound, density, attenuation or a combination thereof. We consider a Gaussian prior in model space with covariance $\mathbf{C}_M$ and mean $\mathbf{m}^{\text{prior}}$, as well as Gaussian data errors with covariance $\mathbf{C}_D$ and zero mean.

Solving the regularized inverse problem can then be formulated as an optimization problem minimizing a non-linear function $\chi(\mathbf{m})$, which is the sum of data misfit and regularization term

$$\chi(\mathbf{m}) = \frac{1}{2}(\mathbf{F}(\mathbf{m}) - \mathbf{d}^{\text{obs}})^T \mathbf{C}_D^{-1}(\mathbf{F}(\mathbf{m}) - \mathbf{d}^{\text{obs}}) + \frac{1}{2}(\mathbf{m} - \mathbf{m}^{\text{prior}})^T \mathbf{C}_M^{-1}(\mathbf{m} - \mathbf{m}^{\text{prior}}). \tag{1}$$

Here, $\mathbf{F}(\mathbf{m})$ describes the parameter-to-observable map, which can depend linearly or non-linearly on $\mathbf{m}$. A standard result from Bayesian inference[8,13] states that the posterior probability density function defined as the conditional probability of a model $\mathbf{m}$ given data $\mathbf{d}^{\text{obs}}$ satisfies

$$\rho(\mathbf{m}|\mathbf{d}^{\text{obs}}) \propto \exp(-\chi(\mathbf{m})). \tag{2}$$

In this work, we assess the resolution quality of tomographic images by exploiting curvature information of the objective function $\chi(\mathbf{m})$ contained in the Hessian $\mathbf{H}(\mathbf{m}) = \nabla^2\chi(\mathbf{m})$. From hereon, we suppose all uncertainties to be normally distributed. To set the stage, we assume that in the vicinity of the optimal model $\tilde{\mathbf{m}}$, the misfit functional can be approximated quadratically as

$$\chi(\mathbf{m}) \approx \chi(\tilde{\mathbf{m}}) + \frac{1}{2}(\mathbf{m} - \tilde{\mathbf{m}})^T \mathbf{H}(\tilde{\mathbf{m}})(\mathbf{m} - \tilde{\mathbf{m}}), \tag{3}$$

where the optimality condition $\nabla\chi(\tilde{\mathbf{m}}) = \mathbf{0}$ has been used.

2.1 Resolution analysis for linear straight-ray tomography

By construction, the linear least-squares problem together with the assumption of Gaussian priors results in a Gaussian posterior. In this case the misfit $\chi(\mathbf{m})$ is quadratic in $\mathbf{m}$ and eq. (3) is exact. This guarantees that

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Jun 2022
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

there exists a unique minimum at the optimal model $\tilde{\mathbf{m}}$, which coincides with the mean of the posterior Gaussian and therefore describes the posterior probability density $\rho(\mathbf{m}|\mathbf{d}^{\text{obs}})$ as[8]

$$\rho(\mathbf{m}|\mathbf{d}^{\text{obs}}) \propto \exp(-\chi(\mathbf{m})) = \exp\left(-\frac{1}{2}(\mathbf{m}-\tilde{\mathbf{m}})^T \tilde{\mathbf{C}}_M^{-1}(\mathbf{m}-\tilde{\mathbf{m}})\right), \tag{4}$$

$$\tilde{\mathbf{C}}_M^{-1} = \mathbf{F}^T \mathbf{C}_D^{-1} \mathbf{F} + \mathbf{C}_M^{-1} = \mathbf{H}. \tag{5}$$

Specifically, eqs. (4) and (5) allow us to compute the Hessian $\mathbf{H}$ as well as the solution of the inverse problem $\rho(\mathbf{m}|\mathbf{d}^{\text{obs}})$ analytically.

2.2 Full-waveform inversion

Before we describe how the linear resolution recipe can be adapted for non-linear problems, we first outline the mathematical framework of non-linear full-waveform inversion. FWI iteratively constructs a model by comparing simulated waveforms to observed ultrasound data. This defines an optimization problem constrained by the wave equation for modeling ultrasonic wave propagation in soft tissue

$$\mathbf{L}(\mathbf{p},\mathbf{m}) = \mathbf{f}(\mathbf{x},t), \quad \mathbf{x} \in \Omega \subset \mathbb{R}^3, \quad t \in [t_0, t_1] \subset \mathbb{R}. \tag{6}$$

Here, $\mathbf{L}$ is the discrete wave operator in N dimensions, $\mathbf{p}$ is a vector describing the forward pressure wavefield inside the domain $\Omega \subset \mathbb{R}^3$ from which synthetic measurements $\mathbf{d}$ can be extracted, and $\mathbf{f}(\mathbf{x},t)$ denotes a vector of external sources. Thus, FWI can exploit the full wave information and captures the relationship between model parameters and synthetic data in a non-linear forward operator $\mathbf{F}$, to infer a high-resolution model of the tissue structure. Quasi-Newton methods are commonly used to solve the non-convex FWI problem as given in eq. (1) by finding a model that fits the data in a least-squares sense using a scheme that iteratively updates a plausible starting model. An update of the model $\mathbf{s}_k$ at iteration k reduces the misfit if

$$\chi(\mathbf{m}_{k+1}) < \chi(\mathbf{m}_k) \quad \text{with} \quad \mathbf{m}_{k+1} = \mathbf{m}_k + \mathbf{s}_k. \tag{7}$$

An efficient method to obtain the model update iteratively is the trust-region L-BFGS algorithm.[14] To avoid the evaluation of the misfit at several points in model space, the trust-region method approximates $\chi(\mathbf{m})$ in the vicinity of the current model $\mathbf{m}_k$ by a second-order Taylor expansion, which leads to the subproblem

$$\min_{(\mathbf{m}-\mathbf{m}_k)} \chi(\mathbf{m}_k) + \nabla\chi(\mathbf{m}_k)^T(\mathbf{m}-\mathbf{m}_k) + \frac{1}{2}(\mathbf{m}-\mathbf{m}_k)^T \nabla^2\chi(\mathbf{m}_k)(\mathbf{m}-\mathbf{m}_k) \quad \text{s.t.} \quad ||\mathbf{m}-\mathbf{m}_k|| \leq \Delta_k, \tag{8}$$

that needs to be solved at each iteration k to obtain the model update $\mathbf{s}_k$. Equation (8) involves the gradient of the objective at the current model $\nabla\chi(\mathbf{m}_k) = \frac{\partial\chi(\mathbf{m}_k)}{\partial m_i}$ as well as the Hessian $\mathbf{H}(\mathbf{m}_k) = \nabla^2\chi(\mathbf{m}_k) = \frac{\partial^2\chi(\mathbf{m}_k)}{\partial m_i^2}$. The trust-region radius $\Delta_k > 0$ constrains the magnitude of change, denoted by the Euclidean distance $||\cdot||$, between subsequent models and is adjusted for each iteration depending on the performance of the local approximation around the current model. Storing the Hessian in memory is generally not feasible, but L-BFGS works matrix-free and approximates the Hessian or its inverse based on a series of updates of the model $\mathbf{s}_k$ and their corresponding change in gradients $\mathbf{y}_k = \nabla\chi(\mathbf{m}_{k+1}) - \nabla\chi(\mathbf{m}_k)$ within previous iterations. An efficient method to compute gradients is the adjoint method.[15] Here, $\mathbf{B}^{-1}(\mathbf{m}_k)$ is a $n \times n$ symmetric positive-definite matrix obtained by the L-BFGS approximation of the inverse of the Hessian $\mathbf{H}(\mathbf{m}_k)$ with n being the number of model parameters.

The mapping $\mathbf{B}^{-1}(\mathbf{m}_{k+1})$ explicitly uses the approximation of the curvature of the misfit at the update position $k+1$ derived from previous iterations. Therefore, during each iteration information from another region in model space is used to update the approximate Hessian. To avoid storing the whole matrix in memory, which is generally not feasible for large model spaces, L-BFGS solely operates on a set of p previous gradients and model updates and computes the matrix-vector product on the fly. Hence, $\{\mathbf{s}_i, \mathbf{y}_i\}$ with $i = 1, ..., p$ vector pairs, need to be stored. With L-BFGS we then also find an efficient way to better explore the approximate inverse Hessian in the vicinity of one specific model $\bar{\mathbf{m}}$ instead of relying on the previous model updates. Assuming Gaussian

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Jun 2022
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

priors, which effectively justifies the quadratic approximation of the misfit given in eq. (3), and choosing the vectors

$$\mathbf{y}_i = \nabla\chi(\mathbf{m}_i^{\text{pert}}) - \nabla\chi(\bar{\mathbf{m}}), \tag{9}$$

$$\mathbf{s}_i = \mathbf{m}_i^{\text{pert}} - \bar{\mathbf{m}}, \tag{10}$$

to denote the difference in gradients between a realization i of a perturbed model $\mathbf{m}_i^{\text{pert}}$ and the model $\bar{\mathbf{m}}$ and respectively the corresponding change in model space ties the Hessian approximation to a region around $\bar{\mathbf{m}}$. In the particular case of linearizing the misfit around the optimum as stated in eq. (3) it follows that the Hessian is assumed constant at the optimum and a consistent choice for $\bar{\mathbf{m}}$ is thus to choose it to be equal to the optimal model, thus $\bar{\mathbf{m}} = \tilde{\mathbf{m}}$. In this case, the set of perturbed models lie within a certain percentage of the optimal model and connect the information contained in the L-BFGS approximated Hessian to a region around around the optimum. Some additional constraints, which are typically obtained from the Wolfe condition are required to ensure that the resulting approximation is a positive-definite matrix.

2.3 Resolution analysis for FWI

Because of the non-linearity of the forward operator, the posterior is not normally distributed for FWI and there exist no closed-form solution for the posterior such as the one given by eq. (4) and eq. (5). However, as stated above, the inverse problem may be linearized by considering a Gaussian approximation of the posterior, which conincides with linearizing the forward operator. This implies that in the vicinity of the optimal model $\tilde{\mathbf{m}}$, the objective function can be approximated quadratically according to eq. (3). Then, using the L-BFGS method, one can construct the approximate inverse Hessian $\mathbf{B}^{-1}(\tilde{\mathbf{m}})$. Note that we use a set of perturbed models that deviate from the optimal model by a few percent to approximate the inverse Hessian with L-BFGS as described in the previous section. This results in a Gaussian approximation of the posterior $\rho(\mathbf{m}|\mathbf{d}^{\text{obs}})$ with mean $\tilde{\mathbf{m}}$ and covariance $\mathbf{H}^{-1}(\tilde{\mathbf{m}})$. The posterior distribution can then be written in a similar form as for the linear case:

$$\rho(\mathbf{m}|\mathbf{d}^{\text{obs}}) \propto \exp(-\chi(\mathbf{m})) = \exp\left(-\frac{1}{2}(\mathbf{m}-\tilde{\mathbf{m}})^T\mathbf{H}(\tilde{\mathbf{m}})(\mathbf{m}-\tilde{\mathbf{m}})\right). \tag{11}$$

Comparing eq. (11) with eq. (4), we see that the Hessian can be interpreted in the sense of the inverse of the posterior covariance matrix[8,10]

$$\mathbf{H}(\tilde{\mathbf{m}}) = \left((\mathbf{F}'(\tilde{\mathbf{m}}))^T\mathbf{C}_D^{-1}\mathbf{F}'(\tilde{\mathbf{m}}) + \mathbf{C}_M^{-1}\right)^{-1} \approx \tilde{\mathbf{C}}_M^{-1}, \tag{12}$$

where $\mathbf{F}'$ denotes the Fréchet derivative of $\mathbf{F}$. Since $\rho(\tilde{\mathbf{m}}|\mathbf{d}^{\text{obs}})$ is not Gaussian ($\mathbf{F}$ is a non-linear operator), there is no equality in eq. (12) and the Hessian solely serves as an approximation. Furthermore, we see that in the non-linear case, the Hessian is always related to a certain model, here the optimal model $\tilde{\mathbf{m}}$. Following the outlined L-BFGS strategy, one also notes that $\mathbf{B}^{-1}(\tilde{\mathbf{m}}) \approx \mathbf{H}^{-1}(\tilde{\mathbf{m}})$.

2.4 Hessian-Vector products

When assessing the resolution of a reconstructed image, one is generally interested in i) inter parameter trade-offs and ii) correlations between model parameters. Information on both of these aspects are contained within the approximate Hessian or its inverse. Defining a model vector $\mathbf{v}$ that points to a specific location in pixel space, which may for instance identify the location of a lesion, and given the vectors $\mathbf{y}_i$ and $\mathbf{s}_i$ for a set of perturbed models, the approximate Hessian $\mathbf{B}(\tilde{\mathbf{m}})$ can be constructed and the mapping

$$\mathbf{B}(\tilde{\mathbf{m}})\mathbf{v}, \tag{13}$$

then probes a certain row (res. column) of $\mathbf{B}(\tilde{\mathbf{m}})$. Whilst the Hessian itself is a square matrix with dimensions equal to the number of model parameters, one of its row (resp. column) can be reshaped to the dimensions of the model space, which illustrates the inter-parameter trade-offs with respect to the lesion location. In a linearized tomographic inversion, where the posterior can be expressed as in eq. (11), the point-localized perturbation at the lesion location is smeared into the surrounding pixels by the mapping (13), similar to what one observed in common point-spread function tests. The width of this spread can therefore serve as a measure of resolution.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Jun 2022
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

For ray tomography, the resolution in model space is commonly assessed with the so-called checkerboard test. The latter uses a 2D sinusoidal field to construct the perturbed model $\mathbf{m}^{\text{pert}}$ by adding small model perturbations $\delta\bar{\mathbf{m}}$ to a reference model $\bar{\mathbf{m}}$. Knowing the set of gradients for $\nabla\chi(\bar{\mathbf{m}} + \delta\bar{\mathbf{m}})$ and $\nabla\chi(\bar{\mathbf{m}})$, the Hessian vector product (13) may also be approximated directly as[11]

$$\mathbf{H}\delta\bar{\mathbf{m}} \approx \nabla\chi(\bar{\mathbf{m}} + \delta\bar{\mathbf{m}}) - \nabla\chi(\bar{\mathbf{m}}). \tag{14}$$

If $\delta\bar{\mathbf{m}}$ is equal to a single point perturbation, we recover the forward mapping (13). Assuming that the inversion converged, a visual representation of resolution for the entire final model can be obtained by choosing as a reference model the optimal model $\bar{\mathbf{m}} = \tilde{\mathbf{m}}$. If resolution in the inverted model and the linearization was perfect, application of eq. (14) should recover the original checkerboard pattern. If there are trade-offs between model parameters, the Hessian-vector product will uncover them, providing a proxy for estimating the spatial resolution length of the model.

3. RESULTS

3.1 Example: Fan array, one source

The goal of this study is to compare the resolution quality of models obtained with different tomographic methods, which each involve a specific forward operator. In this study, we consider i) the linear straight-ray tomographic framework, which leads to a linear forward operator and a closed-form solution for the posterior and ii) the non-linear FWI framework, which upon linearization allows for an approximation of the posterior covariance matrix. It is important to note that SRT uses substantially different physics to model wave propagation than FWI. In SRT, waves are assumed to propagate along straight ray paths with infinite frequency, whereas FWI per definition accounts for finite frequency effects. In order to make a fair comparison, we compute gradients in the FWI problem via cross-correlation time shifts, which account only for the first arrival.

We start with a simple, yet insightful example for a fan array setup with 1 source and 20 receivers as seen in Figure 1 (a). Note that we show the phantom slowness map, since this is the natural parametrization in the straight-ray framework. Synthetic observed data using a Ricker wavelet with a center frequency at 400 kHz is

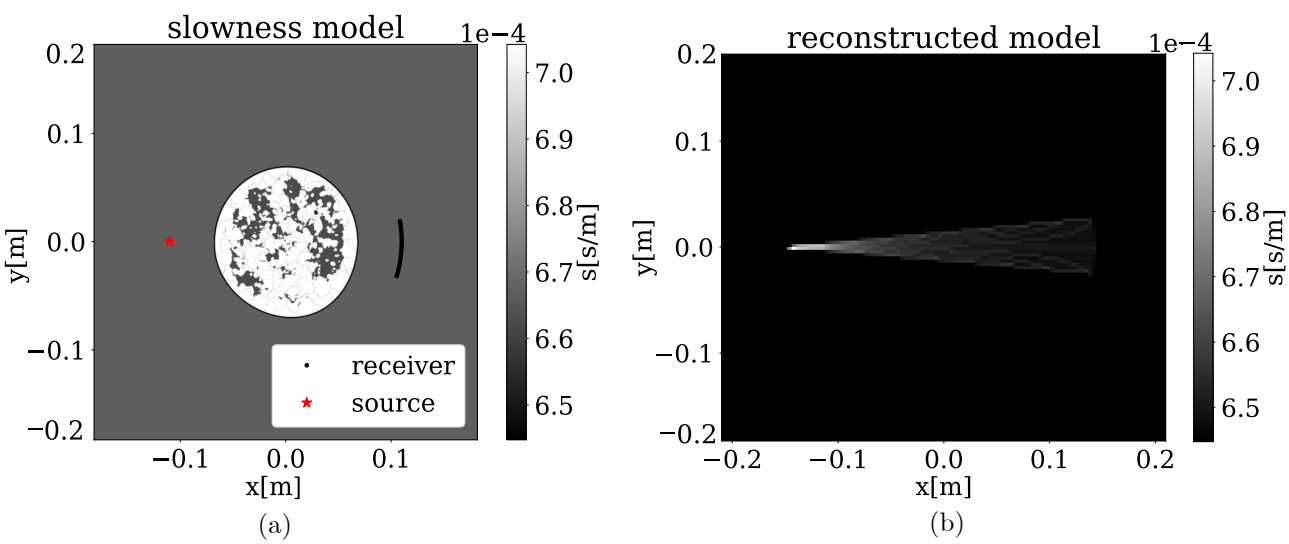


Figure 1: (a) Setup for the fan aperture and the slowness model. (b) The final slowness SRT model reconstructed with non-regularized least-squares.

modelled with the spectral-element software Salvus[16] and one of the 2D acoustical breast phantoms provided by Li et al.[17] as well as for the background medium (water with $s = 1/1500$ s/m). The observed data is then obtained as the set of cross-correlation traveltime shifts. In the objective, eq. (1), the data and model covariance matrices encode the variance that we attach a-priori to the measurements and the models. These variances reflect

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Jun 2022
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

(subjective) uncertainties about posterior distribution in model space. From the information gain of the data brought into the system by solving the inverse problem, uncertainties and hence the variance in model space should be reduced. We assume a standard deviation of $1e^{-7}$ s for the data (around 10% of the mean value of the traveltime shifts) and a standard deviation for the slowness parameters of 1/10 s/m. No correlations in model or data space are assumed in the prior covariance matrices, which means that both, $\mathbf{C}_D$ and $\mathbf{C}_M$ are diagonal matrices. As a prior, we use a constant slowness model with $s = 1/1500$ s/m. Solving the inverse problem (4) leads then to the maximum a-posteriori solution: the reconstruction shown in Figure 1 (b). It is clear that at such low frequencies and with such a sparse source-receiver setup, one cannot expect to resolve any phantom structure. However, what can be observed in Figure 1 (b) is that only within the area crossed by straight-rays from the source to the receivers, slowness values are updated with respect to the prior.

Using the forward matrix $\mathbf{F}$ obtained from straight-ray tracing, eq. (5) is solved to obtain the Hessian $\mathbf{H}$, which is inverted to obtain the posterior covariance matrix $\tilde{\mathbf{C}}_M$. The latter is the quantity of interest, since it encodes the the variance in the model parameters after information from the data has been brought into the system. By analyzing the posterior, we therefore obtain an estimate by how much the uncertainties in model space could be reduced. A comparison between the prior and the posterior is shown in Figure 2. Per

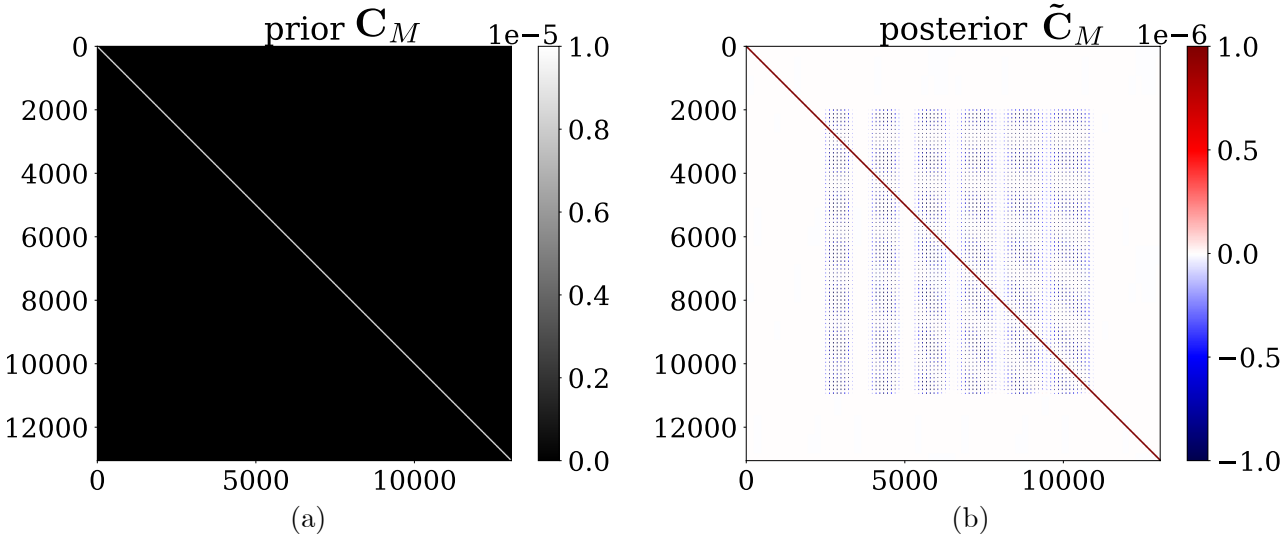
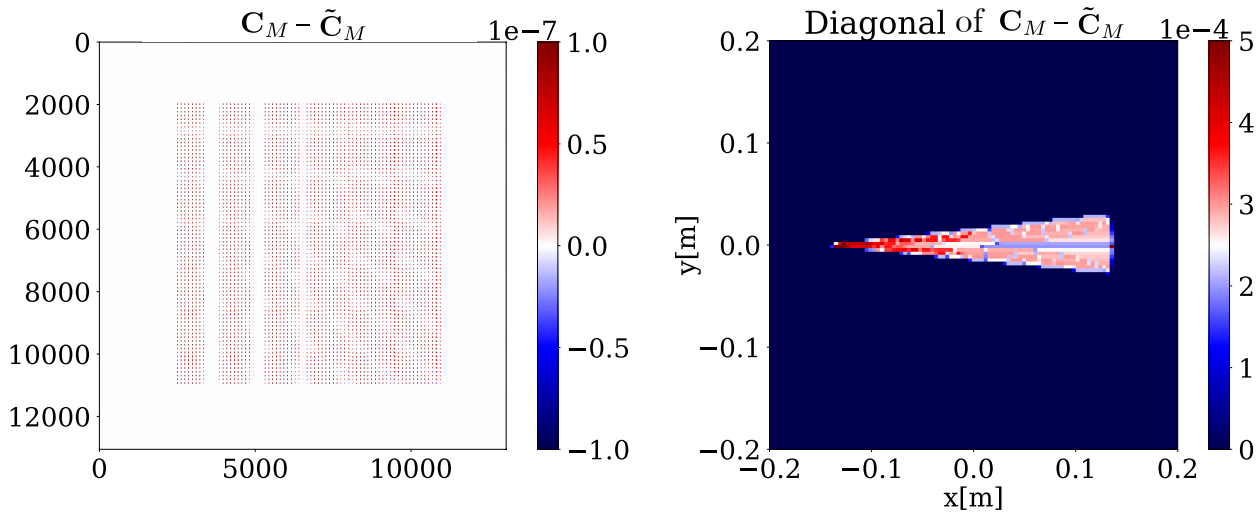


Figure 2: (a) Prior model covariance matrix and (b) posterior model covariance matrix for the fan array encoding the variance of the values in model space. No correlations in model space are assumed in the prior, hence $\mathbf{C}_M$ is diagonal. By solving the SRT inverse problem, information on the relation between the model parameters is brought into the system by the data, thus the posterior $\tilde{\mathbf{C}}_M$ has non-zero off-diagonal elements.

construction, the prior $\mathbf{C}_M$ is diagonal, but we observe that in the posterior off-diagonal elements appear that encode correlations between model parameters. If non-zero off-diagonal elements appear, this information must have been brought into the system by the data $\mathbf{d}^{\text{obs}}$ - we have learned from the data and consequently should observe a decrease of in the variance of the model parameters. This is also visualized in Figure 3 (a), where the difference between the prior and posterior clearly marks the off-diagonal elements and therefore displays the information obtained from the observed data. In Figure 3 (b) the diagonal elements of the difference plot display the information gain in pixel space. Diagonal elements that belong to pixels that are crossed by a ray segment are non-zero. In these areas the pixel values could be updated according to the observed data and the variance in model space could be reduced. Contrary for pixels that are not crossed by rays, no information from the data can be used to update the pixel values and the variance in model space does not change with respect to the prior. These observations concur with the straight-ray physical model, a high-frequency approximation of the wave equation. Approximating waves as infinitely narrow rays, traveltimes can be computed by integrating slowness over the ray path. Therefore the sensitivity of the computed traveltimes is restricted to the the path integral. If one value along a ray is changed, other pixel values need to be adjusted in order to still fit the data.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Jun 2022
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

For all pixels off the ray path, straight-ray tomography cannot provide information on model parameter values.



(a) (b)

Figure 3: (a) Differences between the prior and the posterior display the information gained from the observed data in model space. (b) The diagonal of (a) displays the information gain in pixel space.

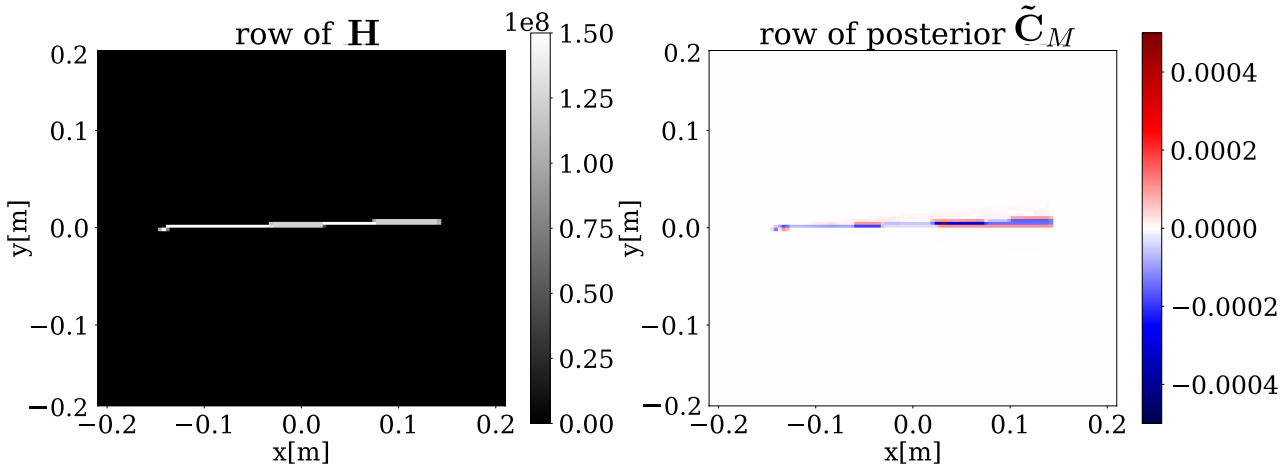
Lastly we analyze the the structure and content of the Hessian. Considering eq. (5), one notices that off-diagonal elements of the Hessian encode information on parameter correlations via the product of ray segment lengths in specific pixels (described by $\mathbf{F}^T\mathbf{F}$), weighted by the inverse of the data variance $\mathbf{C}_D^{-1}$. Diagonal elements on the other hand contain a measure of the ray segment lengths in a specific pixel, weighted again by the inverse of the data variance but with the additional a-priori information on model parameter uncertainties. The diagonal elements thus depict the inverted prior complemented by the additional information gained from the data.

Figure 4 (a) shows a specific row selected from the Hessian plotted in pixel space. One row of $\mathbf{H}$ extracts a particular part of the reconstructed image, which here picks out a specific ray from a source to one of the receivers. In Figure 4 (b), the same row has been selected in the posterior. We see that the posterior is solely sensitive to a very narrow region around the straight path, which adheres to the infinite frequency assumption taken in SRT. The algorithm is blind to everything that happens outside the ray path and assumes information for this data point can only come from the infinitely narrow region around the straight-ray path. Thus changing a value off the path will not affect the reconstruction.

In Figure 5, a particular row of the approximate Hessian $\mathbf{B}(\tilde{\mathbf{m}})$ obtained from the linearized FWI problem has been picked by performing a matrix-vector product with the vector $\mathbf{v}$ pointing to a central region in the phantom as described in section 2.4. Note that the size of the perturbation introduced by $\mathbf{v}$ is on the order of one wavelength, which is here approximately equal to 3.75 mm. The vector $\mathbf{v}$ can be thought of as a point-localized perturbation that is then smeared out in the spatial domain by the forward mapping $\mathbf{B}(\tilde{\mathbf{m}})\mathbf{v}$. Thus the application of $\mathbf{B}(\tilde{\mathbf{m}})$ on $\mathbf{v}$ blurs the information on the spatial location of the perturbation in space. However, since the forward mapping shows trade-offs between model parameters, we may also interpret Figure 5 in terms of the influence that the perturbation has on the other model parameters. Since FWI accounts for the finite frequency content of the waveforms, the sensitivity area of the entire model space for the values at the perturbed location extends to the first Fresnel zone. This includes areas outside of the direct straight-ray paths and notably also areas behind the direct source-receiver line. Consequently, with FWI, reconstructions account for non linear effects such as scattering and diffraction off the straight-ray path. This is especially interesting when considering reconstructions in a 2D plane that will be influenced to some extend by scattering effects off the plane.

To investigate the resolution gain that can be expected by using a more accurate physical description throughout the entire model instead of only a single point perturbation, we perform checkerboard tests by application

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Jun 2022
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use



(a) (b)

Figure 4: (a) A specific row of the Hessian $\mathbf{H}$ encodes the correlations and trade-offs for for one model parameter, i.e. one pixel location. (b) The same row has been selected from the posterior $\tilde{\mathbf{C}}_M$. As expected for SRT, uncertainties in model parameters are only influenced by the information in a narrow region around the straight-ray path from source to one receiver, which maps to one data point in data space.

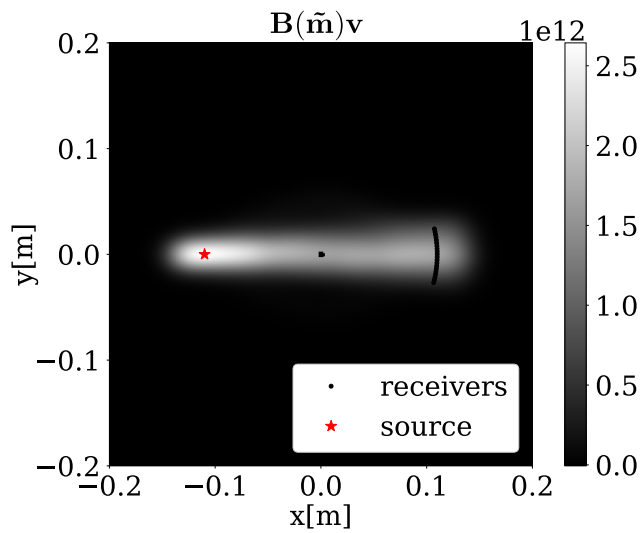


Figure 5: Hessian vector products with the L-BFGS approximated Hessian matrix $\mathbf{B}(\tilde{\mathbf{m}})$ and a vector $\mathbf{v}$ pointing to a certain region in the reconstruction plane. The vector can be interpreted as a point-localized perturbation, which is smeared in the spatial domain by the forward mapping $\mathbf{B}(\tilde{\mathbf{m}})\mathbf{v}$.

of eq. (14). The model parameters are perturbed by a checkerboard patter consisting of 2D sinusoids with variations of $\pm$ 0.066 s/m or 2% of the background slowness in a region of interest inside the source-receiver array covering the phantom as seen in a qualitative plot in Figure 6 (a). If the resolution is perfect, we recover the original checkerboard pattern. For the fan array setup this cannot be expected since the data coverage is too sparse. However, by using different misfit functionals for the calculation of the gradient in FWI, a resolution proxy for the expected region of influence on the model parameters is obtained. Choosing as a misfit functional cross-correlation time-shifts restricts the available data to a single data point (one traveltime shift for each source-receiver pair), therefore the data space equals the one in SRT. The resulting sensitivity kernel is displayed in Figure 6 (b). In comparison to SRT, the Fresnel zone, equal to the sensitivity area to model perturbations, is clearly visible. Although using the same data (cross-correlation traveltime shifts), FWI offers a significantly wider sensitivity imprint. Compared to Figure 6 (c), which shows the result for a L2 misfit functional, one notes

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Jun 2022
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

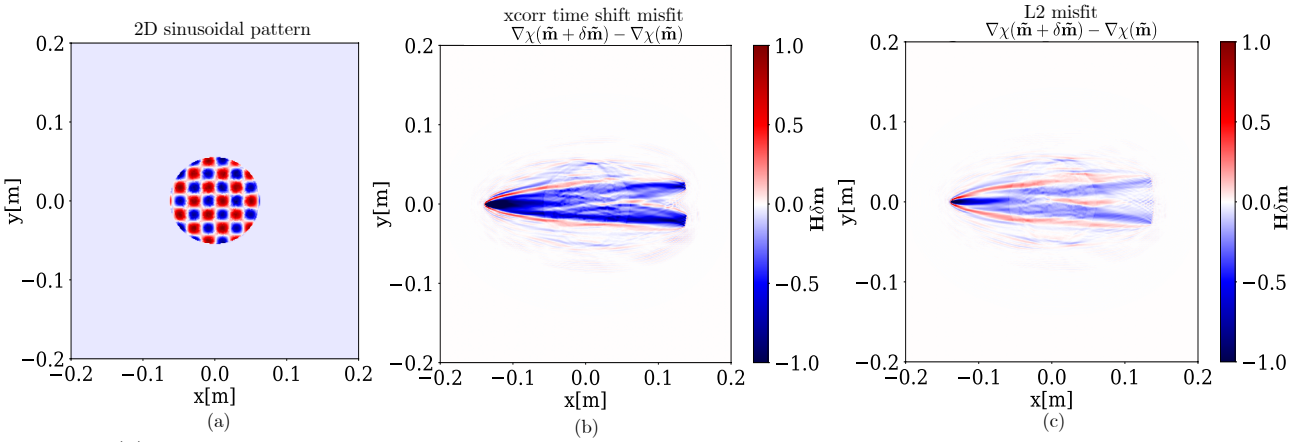


Figure 6: (a) Qualitative plot of the checkerboard pattern applied to a region of interest inside the source-receiver array covering the phantom. (a) Checkerboard test for the fan array setup using a cross-correlation time shift misfit functional and (b) using a L2-misfit functional.

however that the sensitivity pattern is more uniform and primarily aligned along the path that connects the source to the receivers. The L2 misfit functional, which is commonly employed in FWI, computes the mismatch between observed and synthetic data as the L2 norm of their difference taking into account the whole recorded signal. Therefore, the L2 misfit offers significantly more data points including also information from scattering for instance, which allows for an increase in the sensitivity area (more red areas) and a finer resolution of the checkerboard features.

3.2 Full source-receiver array

In this section, we consider a full aperture of 256 receivers and 100 sources equidistantly spaced on a circular array (see Figure 7 (a)) as is common in USCT setups. A set of cross-correlation traveltime shifts for each source-

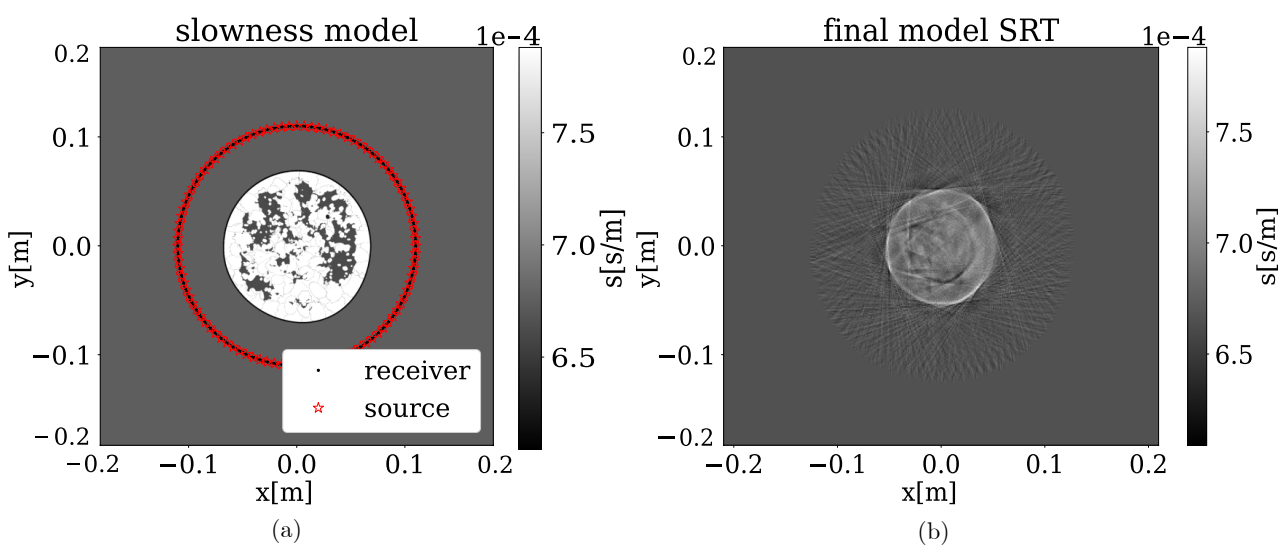


Figure 7: (a) Setup for the full aperture and the slowness model. (b) The final slowness SRT model reconstructed with non-regularized least-squares.

receiver combination is computed following the same recipe as in the previous section. The final SRT model reconstructed by non-regularized least-squares shown in Figure 7 (b) exposes severe ray artefacts, which may be mitigated by introducing damping. Furthermore, the reconstruction suffers from the relatively low number

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Jun 2022
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

of source-receiver pairs, which leads to a relatively sparse ray coverage that produces the circular artefacts in the reconstruction. However in this study, we focus on a statistical analysis of the problem, with the goal to evaluate the information gain in the posterior with respect to the prior, independent of the specific nature of the reconstruction.

Figure 8 shows the counterpart to Figure 4 but for the full array. On the left, the selected row of $\mathbf{H}$ corresponds to the central region of the reconstruction, that is densely crossed by rays. Therefore, changing the model parameters at these locations affects pixel values in the entire reconstruction plane. The value of a perturbation in this central region is smeared out radially and the width of this spread may be analyzed similar to common point-spread function tests for resolution.[18] In Figure 8 (b), the same row has been selected from the posterior.

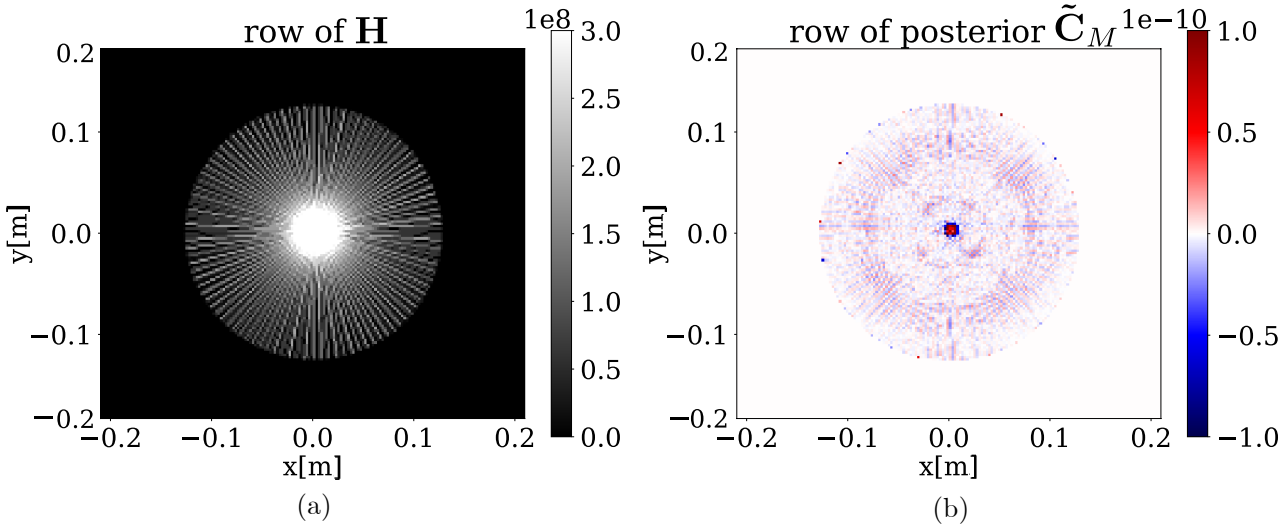


Figure 8: (a) A row of the Hessian and (b) the corresponding row of the posterior.

The posterior variances in the central region are, as expected, influenced the most by the central perturbation. Interestingly, we observe that whilst the variance directly at the center increases, the values just outside are decreased (blue circle around the center). Again, this adheres to the fact that the model still needs to fit the data. Furthermore, in the region where rays travel though the phantom, variances are adjusted on a local level, whilst outside the phantom, where rays only travel though water, the ray pattern that links sources to receivers reappears.

Figure 9 shows the result of a point-spread function test for FWI for using sources at 100 kHz and reveals the potential of the L2 misfit functional to provide a better estimate of the spatial extend of a central point perturbation compared to a cross-correlation misfit. The latter smears the perturbation substantially in space as seen in Figure 9 (a) and the resulting width of the point spread function is thus significantly broader than the result for the L2 misfit. Since point spread functions offer an objective proxy for resolution lengths in the reconstructed image, we conclude that the choice of the misfit functional in a FWI is essential in order to outweigh the increased computational costs by a more accurate reconstruction.

Analogously to the checkerboard test for the fan array, a proxy for the resolution throughout the model for the full array is obtained by performing the Hessian-vector product eq. (14). The results for different anomaly sizes are presented in Figure 10 and suggest a similar interpretation as for the fan array. The checkerboard pattern is far better resolved when using the L2 misfit functional. The resulting pattern for the cross-correlation traveltime shift is smeared along the single source-receiver ellipses. In fact, this pattern may be viewed as a version of Figure 6 (a) where the single source sensitivity kernels from different angles are overlain. Since one source-receiver pair only leads to one point in data space, the sole information obtained is an average slowness value over all checkerboard cells that the ray crosses, ignoring any information from scattering. This in principle prohibits the any spatial resolution of the checkerboard pattern for a single source-receiver pair. We note however that by sampling the domain from all angles and by adding the individual sensitivity kernels, some information

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Jun 2022
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

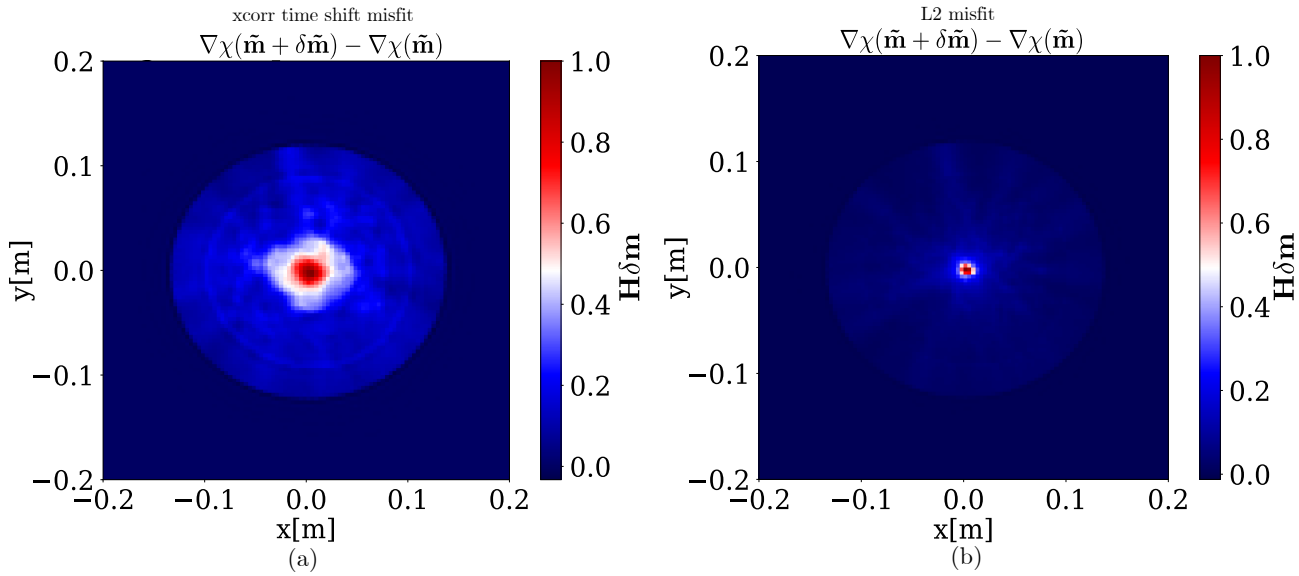


Figure 9: Point-spread function test for a central perturbation with (a) using a cross-correlation traveltime shift misfit functional and (b) using a L2 misfit functional. Since the L2 misfit is not limited to first arrival information only, the point perturbation is resolved more accurately in space in contrast to the cross-correlation misfit, which smears the perturbation substantially.

on the spatial distribution of the anomalies is obtained. By virtue of the even source-receiver coverage for the full array, the L2 misfit result recovers the checkerboard pattern fairly accurate and the strong directional dependency visible in the fan array result is significantly reduced. From Figure 10 one may further draw the conclusion that even in the case of using the same data space for FWI as for SRT when applying a cross-correlation traveltime shift misfit, the additional information from the more accurate physical model as well as the full coverage of the domain lead to an increase in spatial resolution. However, the full potential of FWI can only be leveraged when increasing the amount of data by virtue of using the L2 misfit functional. In Figure 10, the size of the checkerboard pattern is varied from left to right to increasingly finer features. Whilst for the first column, the pattern is well resolved especially for the L2 misfit, the resolution quality degrades as the pattern size decreases. For the cross-correlation misfit, resolution completely breaks down in the last column. Performing a series of checkerboard tests with varying pattern sizes therefore constitutes a proxy for a qualitative resolution measure.

4. CONCLUSIONS

Although the assessment of the resolution quality of tomographic models obtained with different tomographic techniques is fundamental in order to be able to compare the results, there exists no general framework that allows the comparison between linear and non-linear models. However, using the same parametrization of the model space, such an inter-comparison becomes possible. Using the assumption of Gaussian priors, which allows for a linearization around the optimal model, a consistent framework for non-linear models that follows the machinery of linear resolution analysis is obtained. By exploiting gradient information from a set of perturbed optimal models, an approximation for the Hessian of the misfit functional can be constructed. The Hessian encodes the curvature information of the misfit and thus describes by how much the misfit changes if model parameters are perturbed. For reconstructed models, this kind of information is extremely valuable since it allows us to infer parameter correlations and trade-offs. In this study, we have shown that the sensitivity area in which model uncertainties can be reduced through the information gained from the observed data is significantly smaller for straight-ray models than for models obtained with FWI. Choosing a cross-correlation traveltime shift functional, which forces FWI to use the same limited data as ray tomographic methods, the additional physics taken into account by finite frequency modelling and a complete angular data coverage of the domain increase the resolution of recovered features. In order to exploit the full potential of finite frequency models however,

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Jun 2022
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

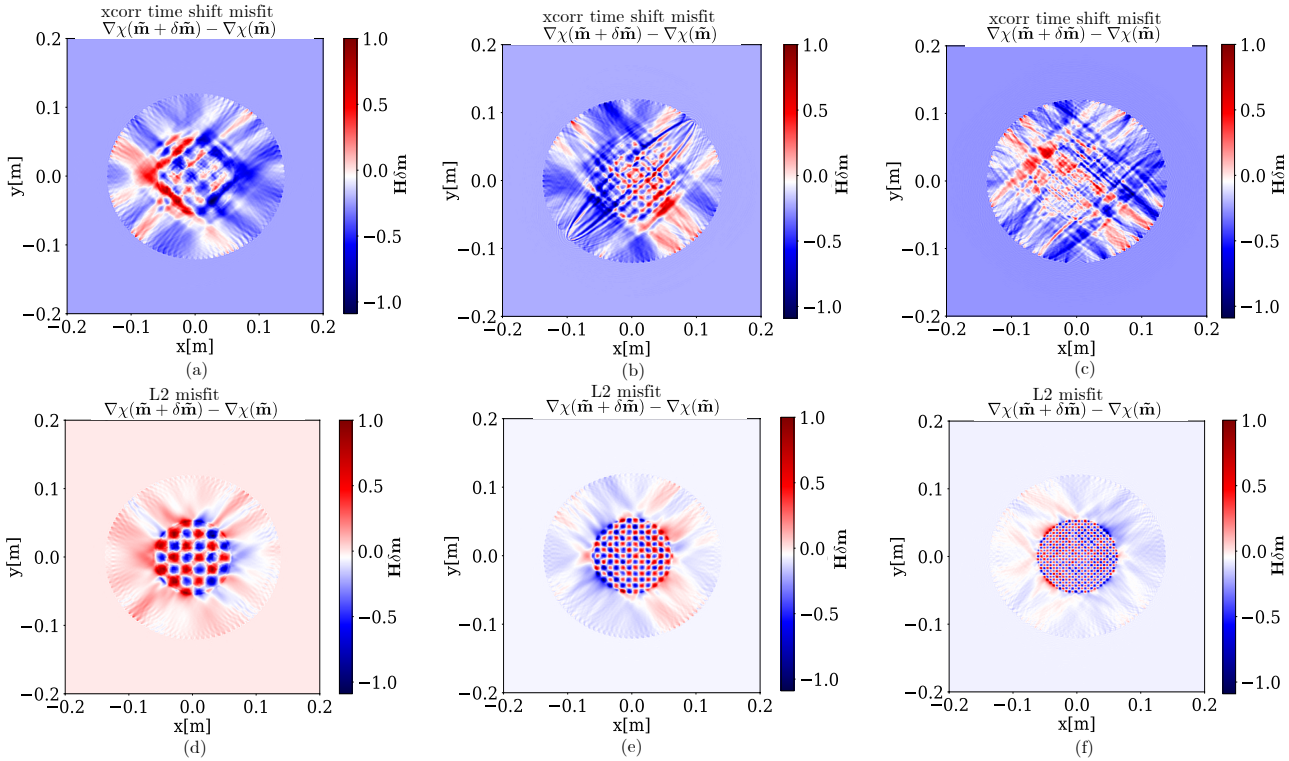


Figure 10: Checkerboard test for the full array setup. (a)-(c) Using a cross-correlation time shift misfit functional and (d)-(f) using a L2-misfit functional. The frequency of the 2D sinusoidal field is increased by a factor of 2 respectively between the columns. Especially for the cross-correlation time shift misfit, the resolution is significantly degraded for a finer checkerboard pattern.

additional information on wavefield interactions contained within the reflection record of the signal is needed. The L2 misfit functional considers this additional data naturally and thus leads to an increase in the sensitivity area with respect to changes in model parameters as well as to an increase in resolution of the recovered features. An interesting approach for a FWI would therefore be to start with the robust cross-correlation time shift misfit and to switch to the more accurate L2 misfit in the course of the inversion. Since the choice of the form of the misfit functional has no influence on the overall computational cost of an inversion procedure, this could be implemented without additional restrains.

ACKNOWLEDGMENTS

This work was supported by the Swiss National Supercomputing Centre (CSCS) under project ID sm59.

REFERENCES

[1] A. Sieminski, J.-J. Lévêque, and E. Debayle, "Can finite-frequency effects be accounted for in ray theory surface wave tomography?," *Geophysical Research Letters* **31**(24), 2004.

[2] L. Guasch, O. Calderon Agudo, M.-X. Tang, P. Nachev, and M. Warner, "Full-waveform inversion imaging of the human brain," *npj Digital Medicine* **3**, 12 2020.

[3] R. Pratt, L. Huang, N. Duric, and P. Littrup, "Sound-speed and attenuation imaging of breast tissue using waveform tomography of transmission ultrasound data," *Proceedings of SPIE - The International Society for Optical Engineering* **6510**, 03 2007.

[4] E. Bachmann and J. Tromp, "Source encoding for viscoacoustic ultrasound computed tomography," *The Journal of the Acoustical Society of America* **147**(5), pp. 3221–3235, 2020.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Jun 2022
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

[5] K. Wang, T. Matthews, F. Anis, C. Li, N. Duric, and M. A. Anastasio, "Waveform inversion with source encoding for breast sound speed reconstruction in ultrasound computed tomography," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* **62**(3), pp. 475–493, 2015.

[6] M. Pérez-Liva, J. L. Herraiz, J. M. Udías, E. Miller, B. T. Cox, and B. E. Treeby, "Time domain reconstruction of sound speed and attenuation in ultrasound computed tomography using full wave inversion," *The Journal of the Acoustical Society of America* **141**(3), pp. 1595–1604, 2017.

[7] L. Gebraad, C. Boehm, and A. Fichtner, "Bayesian elastic full-waveform inversion using hamiltonian monte carlo," 09 2019.

[8] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*, Society for Industrial and Applied Mathematics, 2005.

[9] A. Zunino and K. Mosegaard, "An efficient method to solve large linearizable inverse problems under gaussian and separability assumptions," *Computers & Geosciences* **122**, pp. 77–86, 10 2018.

[10] A. Fichtner and J. Trampert, "Hessian kernels of seismic data functionals based upon adjoint techniques," *Geophysical Journal International* **185**, pp. 775 – 798, 02 2011.

[11] A. Fichtner and T. v. Leeuwen, "Resolution analysis by random probing," *Journal of Geophysical Research: Solid Earth* **120**(8), pp. 5549–5573, 2015.

[12] N. Ruiter, R. Dapp, M. Zapf, and H. Gemmeke, "Evaluation of 3d point spread function of a semi-ellipsoidal ultrasound computer tomography system," in *2011 IEEE International Ultrasonics Symposium*, pp. 242–245, 2011.

[13] T. Bui-Thanh, O. Ghattas, J. Martin, and G. Stadler, "A computational framework for infinite-dimensional bayesian inverse problems part i: The linearized case, with application to global seismic inversion," *SIAM Journal on Scientific Computing* **35**(6), pp. A2494–A2523, 2013.

[14] C. Boehm, N. K. Martiartu, N. Vinard, I. J. Balic, and A. Fichtner, "Time-domain spectral-element ultrasound waveform tomography using a stochastic quasi-Newton method," in *Medical Imaging 2018: Ultrasonic Imaging and Tomography*, N. Duric and B. C. Byram, eds., **10580**, pp. 92 – 100, International Society for Optics and Photonics, SPIE, 2018.

[15] A. Tarantola, "Inversion of seismic reflection data in the acoustic approximation," *GEOPHYSICS* **49**(8), pp. 1259–1266, 1984.

[16] M. Afanasiev, C. Boehm, M. van Driel, L. Krischer, M. Rietmann, D. A. May, M. G. Knepley, and A. Fichtner, "Modular and flexible spectral-element waveform modelling in two and three dimensions," *Geophysical Journal International* **216**(3), pp. 1675–1692, 2019.

[17] F. Li, U. Villa, S. Park, and M. Anastasio, "2D Acoustic Numerical Breast Phantoms and USCT Measurement Data," 2021.

[18] A. Fichtner and J. Trampert, "Resolution analysis in full waveform inversion," *Geophysical Journal International* **187**, pp. 1604 – 1624, 12 2011.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Jun 2022
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use