
https://doi.org/10.1093/gji/ggab457
Adjoint envelope tomography for scattering and absorption using radiative transfer theory

Tuo Zhang\textsuperscript{1,2} and Christoph Sens-Schönfelder\textsuperscript{1}

\textsuperscript{1}GFZ German Research Centre for Geosciences, 14473 Potsdam, Germany. E-mail: tuo@gfz-potsdam.de
\textsuperscript{2}Institute of Geological Sciences, Freie Universität Berlin, 12249 Berlin, Germany

SUMMARY
To investigate the small-scale elastic structure of the subsurface at length scales below the resolution limits of waveform tomography, envelopes of high-frequency scattered seismic waveforms have been used with a variety of approaches. However, a rigorous framework for the iterative inversion of seismogram envelopes to image heterogeneity and high-frequency attenuation comparable to full waveform inversion (FWI) is missing. We present the mathematical framework for an iterative full envelope inversion using forward and adjoint simulations of the radiative transfer equations, in full analogy to FWI that is based on the wave equation. The forward and adjoint problems are solved by modelling 2-D multiple non-isotropic scattering in a random elastic medium with spatially variable heterogeneity and attenuation using the Monte Carlo method. Sensitivity kernels are derived for the squared difference between the full observed and modelled envelopes which is iteratively minimized with the L-BFGS method. We apply this algorithm in numerical tests in the acoustic approximation and show that it is possible to image the spatial distribution of small-scale heterogeneity and attenuation in iterative inversions. Our analysis shows that the relative importance of scattering and attenuation anomalies needs to be considered when the model resolution is assessed. The inversions confirm that the early coda is important for imaging the distribution of heterogeneity while later coda waves are more sensitive to intrinsic attenuation and we show that this dependency can be used to cope with the trade-off that exists between both material properties.

Key words: Inverse theory; Seismic attenuation; Seismic tomography; Wave scattering and diffraction.

1 INTRODUCTION
Tomography as the method of inferring the spatial distribution of medium parameters from measurements of some physical observable has been successfully applied in different fields, like medicine, civil engineering and geophysics. In particular, seismic tomography uses seismograms to explore the structure of the Earth’s interior (Dziewonski & Anderson 1981). The different types of observations that can be made in seismic records offer various approaches to invert for physical properties of the Earth, leading to different classes of tomography. Since the traveltime of seismic waves is the most intuitive observation in seismic data, traveltime tomography was early developed to image the velocity or slowness structure (Aki & Lee 1976; Thurber 1983; Zhao et al. 1992; Li & Van Der Hilst 2010). However, the obvious limitation of the traveltime tomography is that only the information of first arriving waves is used (Rawlinson et al. 2003) and information from later reflected or refracted wave arrivals is neglected. Exploiting all information contained in a seismogram including shape and amplitude of all arriving waves is the goal of full-waveform inversion (FWI) also referred to as adjoint tomography (Tarantola 1984; Tromp et al. 2005; Fichtner et al. 2006; Fichtner 2010). The inversion problem in FWI is solved iteratively by calculating the Fréchet derivative of the misfit function with respect to the model parameters that shows how the current model can be improved. The term adjoint tomography indicates that the interaction of the forward field with the adjoint field that satisfies the adjoint equation with the adjoint source derived from the misfit between the observed and the synthetic data is used to calculate the model gradient (Fichtner 2010). With several decades of development, FWI has become a mature tool to reveal the velocity and attenuation structure of the medium (Tape et al. 2009; Fichtner et al. 2010; Zhu et al. 2012; Tao et al. 2018). Although FWI was successfully applied in global and regional imaging on multiple scales, its resolution is inherently limited by the wavelength of the observed wavefield. Structure on a length scale below this resolution limit cannot be resolved and causes misfit between modelled and observed wavefields that cannot be explained (Mancinelli et al. 2016). Therefore minor phases in the seismograms besides the main
body wave or surface wave arrivals are often excluded from FWI (Liu & Gu 2012).

The lithosphere and Earth’s mantle, however, contain multiscale heterogeneity ranging from the large-scale that can be interrogated deterministically to smallest scales far below the wavelength of seismic waves (Kennett et al. 2017). Well-log data provides the most direct evidence of this small-scale velocity heterogeneity (Holliger 1996). Improving the resolution limits of FWI to image smaller scales faces the following limitations: (i) computation costs for wavefield modelling increase drastically with frequency and (ii) higher resolution requires an increase of the station and event density (Chauris 2021).

For many applications, knowledge of the precise deterministic structure of the small-scale heterogeneity might not be required. Disregarding the geometric arrangement, the small-scale structure can be characterized by its statistical properties. In random medium theory it is assumed that heterogeneity can be regarded as a realization of a random process described by the amplitude of fluctuations and a spatial correlation structure (Batchelor 1953; Sato 1982). Different from the velocities and density used to describe the elastic properties in deterministic structure, the random medium heterogeneity is characterized by the strength ε and the spatial autocorrelation function of the fluctuations. Different correlation functions have been used in the literature but all contain a characteristic length scale called correlation length a. The statistical parameters ε and a of the small scale heterogeneity provide a description of the material that is complementary to the classical characterization of Earth materials based on macroscopic velocity (Sato et al. 2012).

The secondary phases resulting from the interaction of the wavefield with the small-scale heterogeneity form a continuous wave train that usually follows the first-arriving direct waves. These waves were first analysed by Aki (1969) who coined the term coda waves and suggested that these are backscattered waves from the lateral heterogeneity. Aki & Chouet (1975) then proposed the single backscattering model and the diffusion model to interpret the transfer of seismic energy by the random heterogeneities. The scattering during seismic wave propagating from the source to the station causes a redistribution of seismic energy which involves loss of direct wave energy and the excitation of coda waves as a consequence of energy conservation (Sato et al. 2012). Another process that causes a decay of the seismic wave amplitude is intrinsic attenuation or absorption that converts seismic energy into other forms of energy. As a result, scattering attenuation and intrinsic attenuation are collectively referred to as seismic attenuation. Both intrinsic and scattering attenuation lead to an exponential decay of wave amplitude with travelled distance as described by the respective quality factors $Q_i$ and $Q_a$. The temporal decay of coda wave amplitude can be described with the coda-Q (Aki 1980; Yoshimoto et al. 1993). However, the relationship between coda-Q and scattering or intrinsic attenuation is variable, which naturally leads to the question of how to distinguish between the simultaneous influences of scattering and intrinsic attenuation on seismic waveforms.

In order to separate scattering from the effect of absorption, the multiple scattering model was proposed by Wu (1985). For the first time Wu (1985) introduced the radiative transfer theory (RTT) to seismology. RTT describes the multiple scattering process including back and forward scattering, which has the advantage of solving much more complicated problems (Wegler et al. 2006). To numerically solve the radiative transfer equation, the Monte Carlo method was used by Gusev & Abubakirov (1987) and Hoshiba (1991). Following Wu (1985) and Hoshiba (1991), Fehler et al. (1992) proposed the multiple lapse-time window analysis method (MLTWA) to separate absorption and scattering, in which the seismic energy is integrated over different time windows since each attenuation parameter has a different influence on the individual lapse times. RTT is proved as an efficient tool to investigate absorption and seismic wave scattering of small-scale heterogeneity. Especially with the elastic radiative transfer equations derived by Weaver (1990) and Ryzhik et al. (1996) and the Monte Carlo method applied in the elastic case (Margerin et al. 2000; Yoshimoto 2000), RTT was widely used to synthesize seismogram envelopes (Wegler et al. 2006; Przybilla et al. 2006) and investigate scattering and intrinsic attenuation (Sens-Schönfelder & Wegler 2006; Padhy et al. 2007; Eulenfeld & Wegler 2016).

The approaches mentioned above are all based on the assumption of a uniform distribution of heterogeneity and constant intrinsic attenuation. Imaging the spatial distribution of non-uniform scattering and absorption properties has thus been a challenge. Using the MLTWA and following the work of Hoshiba (1993), Carcòle & Sato (2010) mapped the scattering loss parameter $Q^{-1}_s$, intrinsic absorption $Q^{-1}_i$ and the seismic albedo $B_i$ (i.e. the ratio $Q^{-1}_i/(Q^{-1}_s + Q^{-1}_i)$) of scattering attenuation to total attenuation) in Japan. A block with stronger scattering in a more homogeneous medium was used to model the lateral differences in the propagation of Lg-waves in the western Pyrenean crust with a Monte Carlo simulation algorithm for the 3-D elastic radiative transfer equations (Sens-Schönfelder et al. 2009). Another study of the spatially variable attenuation and scattering structure in the western Pyrenees was conducted by Calvet et al. (2013) based on the large lapse time coda decay as a proxy for intrinsic attenuation factor and peak delay time as the indicator of scattering.

Scattering in volcanoes has been studied extensively as their internal structure is very heterogeneous and wave scattering ubiquitous. Spatial variations of heterogeneity have been investigated by De Siena et al. (2013) who used a rim model with reflective diffusive boundary conditions to describe the distribution of heterogeneity at Campi Flegrei caldera, Italy. Prudencio et al. (2013) studied scattering and attenuation in Tenerife island and Mount St Helens was investigated by De Siena et al. (2014) and De Siena et al. (2016).

Compared to the FWI, these approaches for imaging heterogeneity and absorption have so far been rather simple. The above-mentioned approaches either prescribed a certain geometry of the anomalous regions (Sens-Schönfelder et al. 2009; De Siena et al. 2013) or estimated the parameters for a certain combination of source and station that is attributed to the straight line connecting source and receiver (Calvet et al. 2013; De Siena et al. 2014). Although more complicated space-weighting functions like Gaussian weighting function (Prudencio et al. 2013) or functions based on scattering and absorption sensitivity kernels (Del Pezzo et al. 2016; Del Pezzo & Ibáñez 2020) have been used, this approach is mostly a heuristic regionalization, that is in some cases applied to the non-physical medium parameter coda-Q that combines the effects of absorption and scattering (Bianco et al. 2002; Jin & Aki 2005; Sørgel et al. 2020). It is not an inversion of the nonlinear tomographic problem for the spatial distribution of heterogeneity and absorption that best describes the observations.

Takeuchi (2016) developed a differential Monte Carlo method for simulating the perturbations of seismogram envelopes due to the spatial changes in scattering and intrinsic attenuation parameters, which offers a direct inversion method for the structure. This method approaches some ideas of the conventional waveform inversion and was applied by Ogiso (2019) for the 3-D estimation of scattering and intrinsic attenuation in southwestern Japan.
This long-lasting quest for a reliable imaging method for attenuation and scattering on the one hand and the mature imaging technology of the full waveform inversion for velocity on the other hand inspired us to develop a new tomography method for the small-scale heterogeneity and high frequency absorption. The application of the adjoint method in FWI first benefited from the precise numerical solution of the wave equation (Komatitsch & Tromp 1999; Fichtner & Igel 2008), which allows simulating the seismic waves propagation in arbitrary models of velocities, density or other elastic properties.

The statistical properties that describe the small-heterogeneity have no direct expression in the wave equation which limits the applicability of FWI for spatial imaging of absorption and scattering. In our study, the forward problem is solved by modelling the multiple non-isotropic scattering in a random elastic medium based on the Radiative Transfer Equation using the Monte Carlo method (Zhang et al. 2021). The spatial variability of scattering and absorption is described by the spatial distribution of fluctuation strength $\varepsilon$ and intrinsic quality factors $Q_{\varepsilon}^{-1}$ and $Q_{r}^{-1}$ in the random medium. Key to the success of FWI is the efficient calculation of the gradient of the misfit function with respect to changes in the model parameters (Fréchet derivative) which is used to iteratively update the model for minimization of the misfit function. Although the choice of the model fluctuation strength in FWI can differ (Tromp et al. 2021). The spatial variability of scattering and absorption is described by the spatial distribution of fluctuation strength $\varepsilon$ and intrinsic quality factors $Q_{\varepsilon}^{-1}$ and $Q_{r}^{-1}$ in the random medium. Key to the success of FWI is the efficient calculation of the gradient of the misfit function with respect to changes in the model parameters (Fréchet derivative) which is used to iteratively update the model for minimization of the misfit function. Although the choice of the model fluctuation strength in FWI can differ (Tromp et al. 2005; Yuan et al. 2016; Tao et al. 2017), it is in general used as the adjoint source to generate the adjoint wavefield which in turn is used to obtain the gradient for model updates. In the present paper we follow the same approach and use the least-squares misfit between the observed and modelled envelopes as adjoint source for the adjoint energy field to obtain the Fréchet derivatives.

Here we present the adjoint tomography method for the radiative transfer equation with the misfit derived from the seismogram envelope to invert for scattering and intrinsic attenuation properties both in elastic and acoustic cases (Section 2) and describe an iterative inversion algorithm to update the model. Numerical tests are conducted in the 2-D acoustic case in Section 3 where we show the inversion results for scattering and intrinsic attenuation, individually and simultaneously. In Section 4, we discuss the trade-off between fluctuation strength $\varepsilon$ and intrinsic quality factor $Q^{-1}$ and the influence of the initial model and the inversion time window of the envelope. We finish with some concluding remarks in Section 5.

2 METHODOLOGY

2.1 Adjoint tomography with the radiative transfer equation

FWI uses the discrepancy between observed seismograms and synthetic waveforms calculated for a particular model to iteratively update the model until it fits the observations (Fichtner 2010). The model usually describes the spatial distribution of elastic parameters like seismic velocities and the synthetics are obtained by solving the wave equation in this model. Various possibilities exist to quantify the discrepancy between the synthetics and the observation. Mathematically the discrepancy is described by the misfit function that maps the observed and synthetic waveforms to a scalar misfit value. Finding the gradient of the misfit with respect to all the different model parameters is the core of the adjoint method and allows to obtain an image of the elastic parameters in the subsurface.

Here we aim to formulate a similar approach for the inversion of seismogram envelopes using the adjoint method with the radiative transfer equation (RTE). The model in our case does not describe the macroscopic seismic velocities as in FWI but the scattering and attenuation properties caused by the small-scale heterogeneity. Instead of comparing the waveforms, envelope inversion compares the seismogram envelopes with synthetic envelopes calculated on the basis of the RTE.

In scattering and absorbing media, spatial heterogeneity and intrinsic attenuation influence the energy propagation and absorption, which finally results in the energy density observed as seismogram envelope. To investigate these properties, we utilize the envelope of the full waveform including the ballistic wave and the scattered coda wave. We use the least-squares misfit function:

$$\chi_{YX}(\mathbf{m}) = \sum_i \sum_j \frac{1}{2} \int_0^T \left| E_{YX}(\mathbf{r}_i, t; \mathbf{r}_j, \mathbf{m}) - D_{YX}(\mathbf{r}_i, t; \mathbf{r}_j) \right|^2 dt. \quad (1)$$

$E_{YX}(\mathbf{r}_i, t; \mathbf{r}_j, \mathbf{m})$ is the synthetic energy density simulated in the current model $\mathbf{m}$, where $\mathbf{r}_i$ and $\mathbf{r}_j$ respectively represent the positions of the $i$th source and the $j$th receiver. In what follows, the subscript $YX$ of the energy density always indicates the wave mode $X$ of emission at the source and the mode of detection $Y$ at the receiver. For the $S$-wave energy we use the total $S$-energy as the sum of the energies of the two $S$-wave polarizations. This means that we disregard the specific influence of the $S$-wave polarization on the sensitivity kernels. $D_{YX}(\mathbf{r}_i, t; \mathbf{r}_j)$ designates the observation data. In most practical applications it is impossible to disentangle the recorded energy with respect to its mode and the mode of excitation. But in order to facilitate the derivation, $\chi_{YX}(\mathbf{m})$ is used to represent each component of the total misfit function $\chi(\mathbf{m})$. After the derivation we will discuss how to extend this formulation to applications. The time window of the integral is somewhat arbitrary and indicates the range of lapse times that is to be chosen in the inversion but is assumed to start at time $0$. Then the differential of misfit function reads:

$$\delta \chi_{YX}(\mathbf{m}) = \sum_i \sum_j \int_0^T \left[ E_{YX}(\mathbf{r}_i, t; \mathbf{r}_j, \mathbf{m}) - D_{YX}(\mathbf{r}_i, t; \mathbf{r}_j) \right]$$

$$\times \delta E_{YX}(\mathbf{r}_i, t; \mathbf{r}_j, \mathbf{m}) dt, \quad (2)$$

where $\delta E_{YX}(\mathbf{r}_i, t; \mathbf{r}_j, \mathbf{m})$ is the perturbation of energy density due to the perturbation of the model $\mathbf{m}$. We assume here that there are no changes in the macroscopic velocity structure and the shape of the power spectral density function that describes the small scale fluctuations (see Section 2.3). Under these assumptions, perturbations of the recorded energy density are either caused by changes in the amplitude of the small-scale fluctuations around the macroscopic velocity (e.g. due to the opening of cracks) that lead to changes of the scattering strength or due to changes of intrinsic attenuation described in transfer theory by the parameters fluctuation strength and intrinsic quality factors, respectively. A similar parametrization has been used for example by Takeuchi (2016). These dependencies are discussed in detail in Zhang et al. (2021) and can be expressed as

$$\delta E_{YX}(\mathbf{r}_i, t; \mathbf{r}_j, \mathbf{m}) =$$

$$\delta E_{YX}^{\varepsilon}(\mathbf{r}_i, t; \mathbf{r}_j, \varepsilon(\mathbf{r}')) + \delta E_{YX}^{Q}(\mathbf{r}_i, t; \mathbf{r}_j, Q(\mathbf{r}')) \quad (3)$$

where

$$\delta E_{YX}^{\varepsilon}(\mathbf{r}_i, t; \mathbf{r}_j, \varepsilon(\mathbf{r}')) =$$

$$\frac{1}{\varepsilon_0} \int_{\mathcal{V}} \left( \delta E^{\varepsilon}(\mathbf{r}') \cdot K_{YX}^{\varepsilon}(\mathbf{r}', t; \mathbf{r}_i, \mathbf{r}_j) \right) dV(\mathbf{r}') \quad (4)$$
and

\[ \delta E_{T,XY}(r_j, t; r_i, Q(r')) = - \omega \int_{V'} \left[ \delta Q_{\rho}^{\pm}(r') \delta E_{T,XY}(r', t; r_j, r_i) + \delta Q_\varepsilon^{\pm}(r') \delta E_{T,XY}(r', t; r_j, r_i) \right] dV'(r') \]  

(5)

Here \( \varepsilon_0 \) is the background value of the fluctuation strength. \( Q_\rho \) and \( Q_\varepsilon \) are the intrinsic quality factors of \( P \) and \( S \) waves, respectively. \( \omega \) is the angular frequency. \( \delta \varepsilon^{\pm}(r') \), \( \delta Q_{\rho}^{\pm}(r') \) and \( \delta Q_\varepsilon^{\pm}(r') \) indicate the perturbation of \( \varepsilon^2, 1/\rho Q_\rho \) and \( 1/Q_\varepsilon \) at the position \( r' \) respectively. \( V' \) is the full space in dimension \( d \). According to Zhang et al. (2021), the scattering sensitivity kernel and the attenuation sensitivity kernels read:

\[ \hat{T} E_{T,XY}(r', t; r_j, r_i) = \sum_{r} \int_{V'} \int_{0}^{T} G_{XY}(r_j, t - t'; r', \mathbf{n}) \left[ \int_{V'} c_v \left( g_{V-W}(\varepsilon_0^2) f_{V-W}(\mathbf{n}, \mathbf{n}') \right) E_{XY}(r', t', \mathbf{n}; r_j) d\mathbf{n}' \right] d\mathbf{n} \]  

and

\[ \delta T E_{T,XY}(r', t; r_j, r_i) = \int_{V'} \int_{0}^{T} G_{XY}(r_j, t - t'; r', \mathbf{n}) \left[ \int_{V'} c_v \left( g_{V-W}(\varepsilon_0^2) f_{V-W}(\mathbf{n}, \mathbf{n}') \right) E_{XY}(r', t', \mathbf{n}; r_j) d\mathbf{n}' \right] d\mathbf{n} \]  

(6)

(7)

The notation of the specific energy density \( E_{XY}(r', t', \mathbf{n}') \) is used to indicate the dependence on the propagation direction \( \mathbf{n}' \). \( G_{XY}(r_j, t - t'; r', \mathbf{n}) \) is Green’s function of the radiative transfer equation with the unit source at the position \( r' \) and an initial direction \( \mathbf{n} \). \( g_{V-W}(\varepsilon_0^2) f_{V-W}(\mathbf{n}, \mathbf{n}') \) is the total scattering coefficient that indicates the scattering probability of wave scattered from mode \( W \) to \( V \) with the background fluctuation strength \( \varepsilon_0^2 \). \( f_{V-W}(\mathbf{n}, \mathbf{n}') \) is the normalized differential scattering cross section. \( c_v \) is the macroscopic velocity of wave mode \( V \). \( S^d \) denotes the unit sphere in space dimension \( d \).

The iterative inversion requires Fréchet derivatives of the misfit function with respect to the model changes that can be obtained with the adjoint method (Tromp et al. 2005). We first discuss the misfit caused by variations of the scattering properties, that is the first term on the right-hand side of eq. (3):

\[ \delta \chi_{T,XY}^s(m) = \sum_{i} \sum_{j} \int_{V'} \int_{0}^{T} [E_{T,XY}(r_j, t; r_i, m) - D_{T,XY}(r_j, t; r_i)] dV'(r') dt \times \int_{V'} \frac{1}{\varepsilon_0^2} \delta \varepsilon^2(r') \hat{T} E_{T,XY}(r', t; r_j, r_i) dV'(r') \]  

(9)

Inserting eq. (6) into eq. (9), exchanging the order of the integrals and the integration bounds we rewrite eq. (9) as:

\[ \delta \chi_{T,XY}^s(m) = \sum_{i} \sum_{j} \int_{V'} \int_{0}^{T} \int_{V'} \left[ E_{T,XY}(r_j, t; r_i, m) - D_{T,XY}(r_j, t; r_i) \right] \times G_{T,XY}(r_j, t - t'; r', n) \times \sum_{V} \left[ \int_{V'} c_v \left( g_{V-W}(\varepsilon_0^2) f_{V-W}(n', n) \right) \times E_{T,XY}(r', t', n'; r_j) dV'(r') \right] \times \delta \varepsilon^2(r') dV'(r') \]  

(10)

This equation allows us to estimate the change of the misfit function due to a local change \( \delta \varepsilon^2(r') \). However, it involves the Green’s function \( G_{T,XY}(r_j, t - t'; r', n) \) which in practical terms requires a simulation with a source at \( r' \) which is every point in space where a perturbation \( \delta \varepsilon^2(r') \) can occur. Using the adjoint method we can circumvent this problem and generate the Fréchet derivative with one simulation of the adjoint equation that has its source at the location of the original receiver. We omit the derivation that is almost the same as shown in Tromp et al. (2005) for the wave equation. Note that the reciprocity relationship of the elastic radiative transfer equations is (Zhang et al. 2021):

\[ \frac{1}{S^d} G_{W,W}(r_j, t - t'; r', n) = G_{W,Y}^\dagger(r', t - t', -n; r_j), \]  

(11)

where \( G_{W,Y}^\dagger(r', t - t', -n; r_j) \) is the Green function of an elastic radiative transfer equations with the new scattering coefficient \( g_{W->V}^\dagger \) that obeys \( c_v g_{W->V}^\dagger = c_w g_{W->V} \). With this Green’s function, the adjoint wavefield is written as:

\[ E_{W,Y}^\dagger(r', t', n; r_j) = \int_{V'} \int_{0}^{T} G_{W,Y}^\dagger(r', t - t, n; r_j) \times F_{T,XY}(t, r') \delta V'(r') \]  

(12)

with the adjoint source:

\[ F_{T,XY}(t, r') = [E_{T,XY}(r_j, T - t; r_i, m) - D_{T,XY}(r_j, T - t; r_i)] \delta (r' - r_j), \]  

(13)

where \( \delta (r' - r_j) \) is the Dirac function. Note that the \( \delta (r' - r_j) \) in the expression of \( F_{T,XY}(t, r') \) indicates that the source position of the adjoint wavefield is the receiver position \( r_j \) of the forward wavefield. The expression of eq. (13) shows that the adjoint source is generated from the forward wavefield excited at \( r_j \) via the data misfit. Eq. (10) now can be written in a simplified form:

\[ \delta \chi_{T,XY}^s(m) = \int_{V'} \hat{T} E_{T,XY}(r') \delta \varepsilon^2(r') dV'(r') \]  

(14)
where \( K_{TX}^d(r') \) is the misfit kernel with respect to fluctuation strength \( \varepsilon \) which is:

\[
K_{TX}^d(r') = \sum_i \sum_j S^d_i \sum_w \int_0^T \int_{S^d_i} E_{W_X}^{\dagger}(r', T - t', -n; r_j) \times \\
\sum_{r'} \left[ \int_{S^d_i} c_V g^{W_{X} \to W}(\varepsilon_{0}^2) \right] f^{\to W}(n, n') \times \\
E_{V_X}(r', t', n', r_j) \alpha_d \Delta r \cdot d\Omega (n, n').
\] (15)

Note the difference to the energy sensitivity kernel \( K_{TX}^e(r') \) of eq. 6 which does not depend on the data.

Similarly, the misfit caused by the intrinsic attenuation is obtained. Eq. (2) is finally written as:

\[
\delta \chi_{X,Y}(m) = \int_{y_0}^{y} K_{X,Y}(r') \delta \varepsilon^2(r') dV(r') + \\
\int_{y_0}^{y} \left[ \int_{y_0}^{y} K_{X,Y}(r') \delta Q^p_{X}(r') \right] dV(r')
\] (16)

where the intrinsic attenuation misfit kernels are:

\[
\delta Q^p_{X,Y}(r') = - \sum_i \sum_j S^d_i \int_0^T \int_{S^d_i} E_{P_X}^{\dagger}(r', T - t', -n; r_j) \times \\
E_{P_X}(r', t', n', r_j) \alpha_d \Delta r \cdot d\Omega (n, n').
\] (17)

\[
\delta Q^p_{X,Y}(r') = - \sum_i \sum_j S^d_i \int_0^T \int_{S^d_i} E_{E_X}^{\dagger}(r', T - t', -n; r_j) \times \\
E_{E_X}(r', t', n', r_j) \alpha_d \Delta r \cdot d\Omega (n, n').
\] (18)

Eqs (15), (17) and (18) describe the kernels with respect to the misfit function \( \chi_{X,Y}(m) \) which compares the observed and simulated energy of recorded mode \( Y \) due to excitation of mode \( X \). These expressions are hard to apply since in most cases only the total energy \( D(r_j, t; r) \) is accessible. However, with some common assumptions about the composition of the wavefield we can solve this problem.

Without considering the S-wave polarization, we split the total energy into two parts as:

\[
D(r_j, t; r_j) = D_P(r_j, t; r_k) + D_S(r_j, t; r_k).
\] (19)

The subscript of the observable indicates the recording mode of the energy. Zhang et al. (2021) discussed the combination of the sensitivity kernels in eqs (6), (7) and (8) based on the ratio of the two types of energies \( R_{SP} = D_S/D_P \). This ratio can be reasonably approximated under the assumption of equipartition or it can be estimated using energy transfer simulations in the best available model. When the lapse time is large enough to reach the equipartition state \( R_{SP} = 2y^1 \) in 3-D, where \( y = c_p/c_s \) (Sato et al. 2012). Moreover, a general source simultaneously emits \( P \) and \( S \) energy in the proportions \( S_P \) and \( S_S \) with \( S_P + S_S = 1 \) for a unit source. The ratio between \( S_P \) and \( S_S \) that is required here, can either be estimated from the data or based on the assumption of a theoretical source model. So we can rewrite the misfit function:

\[
\chi(m) = \\
\sum_i \sum_j \sum_y \left[ \frac{1}{2} \int_{y_0}^{y} \left( \sum_X S_X E_{P_X}(r_j, t; r_k, m) - D_P(r_j, t; r_k) \right)^2 dt \right].
\] (20)

The differential of the misfit function becomes:

\[
\delta \chi_{X}(m) = \int_{y_0}^{y} \left[ \int_{y_0}^{y} K_{X,Y}(r') \delta \varepsilon^2(r') dV(r') + \\
\int_{y_0}^{y} \left[ \int_{y_0}^{y} K_{X,Y}(r') \delta Q^p_{X}(r') \right] dV(r')
\]

where the scattering total misfit kernel is:

\[
K_{TX}^d(r') = \sum_i \sum_j S^d_i \int_0^T \int_{S^d_i} E_{P_X}^{\dagger}(r', T - t', -n; r_j) \times \\
\left\{ \sum_X S_X \int_{y_0}^{y} E_{E_X}(r', t', n', r_j) \alpha_d \Delta r \cdot d\Omega (n, n') \right\} \alpha_d \Delta r \cdot d\Omega (n, n').
\] (22)

Note the adjoint wavefield in this kernel is still based on eq. (12) but the adjoint source now becomes:

\[
F^d_{\ell}(r, r') = \\
\left\{ \sum_X S_X E_{E_X}(r_j, T - t; r_k, m) - D_P(r_j, T - t; r_k) \right\} \times \\
\delta(r' - r).
\] (23)

With same adjoint source applied, the total misfit kernels with respect to intrinsic attenuation are easy to obtain:

\[
K_{TX}^d(r') = \\
- \sum_i \sum_j S^d_i \int_0^T \int_{S^d_i} E_{P_X}^{\dagger}(r', T - t', -n; r_j) \times \\
E_{E_X}(r', t', n', r_j) \alpha_d \Delta r \cdot d\Omega (n, n'),
\] (25)
2.2 Acoustic wave propagation

For the propagation of a single wave mode in the acoustic case, the misfit function simplifies to:

\[ \chi(m) = \sum_i \sum_j \frac{1}{2} \int_0^T \| E(r_j, t; r_i, m) - D(r_j, t; r_i) \|^2 \, dt . \] (26)

We rewrite the Fréchet derivative as in Section 2.1:

\[ \delta \chi(m) = \int_0^T \int_{\mathcal{S}_d} K(r') \delta \tau(r') \, dV(r') + \]

\[ \int_0^T \int_{\mathcal{S}_d} \partial_K r' \delta Q^{-1}(r') \, dV(r') , \] (27)

where the misfit kernels are:

\[ 'K'(r') = \sum_i \sum_j S^\ell \alpha_0 \frac{g_0(\tau^2)}{\epsilon_0} \int_0^T \int_{\mathcal{S}_d} E'(r', T - t', -n; r_j) \]

\[ \times \left[ \int_{\mathcal{S}_d} f(n, n') E(r', t', n'; r_j) \, d\text{nn}' \right] \, dt' \]

\[ -E(r', t', n; r_j) \] (28)

and

\[ \partial_K r'(r') = -\sum_i \sum_j S^\ell \omega \int_0^T \int_{\mathcal{S}_d} E'(r', T - t', -n; r_j) \]

\[ \times E(r', t', n; r_j) \, d\text{nn}' . \] (29)

The adjoint wavefield is written as:

\[ E'(r', t', n; r_i) = \int_0^T \int_{\mathcal{S}_d} G(r', t' - t, n; r_i) F(t, r') \, d\text{dV}(r') \] (30)

with the adjoint source:

\[ F(t, r') = [E(r_j, T - t; r_i, m) - D(r_j, T - t; r_i)] \delta(r' - r_i) . \] (31)

Note that the reciprocity relationship in the acoustic case is simpler than the elastic case without changing the scattering coefficient (Margerin et al. 2016):

\[ \frac{1}{2} \int_{\mathcal{S}_d} G(r_j, t; r', n) = G(r_j, t, -n; r_i) . \] (32)

2.3 Forward modelling

The application of adjoint envelope tomography for scattering and absorption requires modelling seismic energy propagating in a random medium with spatially variable heterogeneity and attenuation. Considering the form of the misfit kernels, the specific energy density of the wavefield containing the information of propagation direction needs to be calculated. Zhang et al. (2021) describe a complete simulation algorithm for the 2-D elastic multiple non-isotropic scattering with the Monte Carlo method. This simulation can be applied in elastic inversion as described in Section 2.1. To introduce the adjoint envelope tomography, we focus on the simpler acoustic case in the following. We use the acoustic radiative transfer equation in 2-D which is written as:

\[ \left( \frac{\partial}{\partial t} + \mathbf{n} \alpha_0 \cdot \mathbf{V} \right) E(r, n, t) = -\left( \alpha_0 g_0(\tau^2) + \frac{\omega}{Q(r)} \right) E(r, n, t) \]

\[ + \int\alpha_0 g(r, n', \tau^2) \, d\text{nn}' . \] (33)

The acoustic scattering coefficient is given by Wegler et al. (2006):

\[ g(\theta, \tau^2) = k_0 \Phi \left( \frac{2k_0 \sin \frac{\theta}{2} \tau^2}{2} \right) . \] (34)

where \( k_0 \) is the wavenumber and \( \theta \) is the scattering angle between the directions \( n \) and \( n' \). \( \alpha_0 \) indicates the mean background velocity. \( \Phi \) is the local power spectral density function (PSDF) of the parameter fluctuations in the random medium which is of exponential type in this paper:

\[ \Phi(k, \tau^2) = \frac{2\pi a^2 \tau^2}{(1 + a^2 k^2)^{3/2}} . \] (35)

with the wave vector \( k \). This algorithm allows us to simulate the acoustic wave propagating in the spatially variable medium with spatially variable heterogeneity and attenuation.

2.4 Iterative inversion

The target of the inversion is the minimization of the misfit function. This is achieved by iterative updates of the model converging toward the true model. The iterative inversion starts with the initial model \((m_i^0, m_i^j)\) and uses the following initial search directions to update the model:

\[ m_{i+1} = m_i^0 + \eta_i^0 \mathbf{h}_i \] (36)

\[ m_{i+1}^j = m_i^j + \eta_i^j \mathbf{h}_i^j . \] (37)

Here, \( m_i^0 \) and \( m_i^j \) are the models used in the simulation of the \((k+1)\)th iteration so that \( k = 0, 1, 2, 3, \ldots \) \( \mathbf{h}_i \) and \( \mathbf{h}_i^j \) are the search directions derived from the gradients of misfit function, that is the misfit kernels \( 'K'_i \) and \( \partial K_i^j \). To derive the search directions from the gradients we use the quasi-Newton L-BFGS method (Liu & Nocedal 1989). The benefit of this method is that we obtain an approximation of the Hessian matrix of misfit function to improve the convergence in comparison to a steepest decent method. More details are discussed in Appendix A. \( \eta_i^0 \) and \( \eta_i^j \) are the step lengths of the model updates in the resulting search directions. The step lengths are determined under the assumption that the misfit is a quadratic function of the step length (Fichtner 2010). Using two more simulations with different trial step lengths in the \( \mathbf{h}_i \) and \( \mathbf{h}_i^j \) directions we approximate these quadratic functions and estimate the step lengths to reach their minima.

3 NUMERICAL EXAMPLES

In the previous section we have outlined the mathematical formalism for an iterative adjoint tomography of the radiative transfer equation to infer the distribution of heterogeneity in the subsurface with scattered seismic waves. In this section we use numerical examples to demonstrate the performance of this approach. We use three synthetic examples in a 2-D acoustic setting to illustrate how to realize the adjoint tomography with the envelope for scattering and intrinsic attenuation. Two individual inversions are performed separately for the scattering and intrinsic attenuation structures separately with a uniform distribution of the other property. The trade-off between heterogeneity and attenuation structure is investigated in a third inversion in which the seisogram envelopes are inverted for the spatial distribution of both parameters simultaneously. The
density is 2.7 g·cm⁻³. The intrinsic attenuation $Q^{-1}$ is constant throughout the domain.

The observation data in this synthetic test is obtained with a Monte Carlo simulation using eq. (33) in the true model that is shown in Fig. 1(a). The intrinsic attenuation $Q^{-1}(r)$ is uniform in space (Fig. 1b). For the application of FWI, a starting model is required that is a fair representation of the target (Fichtner 2010) and has to be inferred using a complementary method. In our case we assume that the diffusion approximation (Wegler 2004) or the multiple isotropic scattering approximation (Sens-Schönfelder & Wegler 2006) can be used to infer the background values of heterogeneity and attenuation.

Inspired by the scale of local to regional seismological experiments (Gaebler et al. 2019; Izgi et al. 2020; van Dinther et al. 2021), the model extends over a 50 km by 50 km area and is discretized in a 100 by 100 cells grid (Fig. 1). There are two square anomaly areas in the $\varepsilon$ model with a side length of 10 km. Twenty-four numbered locations around the anomaly areas in Fig. 1 represent the colocated positions of the sources and receivers. When the source is at one location, all other locations are the receivers, a layout inspired by seismic interferometry (Curtis et al. 2006) that can turn receivers into virtual sources and the use of transducers in acoustic experiments that can act either as source or receiver (Niederleithinger et al. 2015).

The macroscopic velocity is spatially uniform at 6 km·s⁻¹ and the density is 2.7 g·cm⁻³. The correlation length $a$ is constant 0.3 km. With an angular frequency $\omega = 30$ rad·s⁻¹ (i.e. $f = 4.78$ Hz) representing high frequency seismic waves, the normalized wavevector $\mathbf{k}_0 = 1.5$. Since the forward modelling is conducted with the Monte Carlo method, 100 million particles are used to simulate the energy transport in the forward simulation for the true model to create the synthetic data. Although the propagation of particles is free and there is no boundary of the model, a $100 \times 100 \times 150$ grid of cell size $0.5 \times 0.5 \times 0.1$ s is used for recording the energy and the propagation direction is recorded in 72 non-overlapping angular bins of $5^\circ$ width.

The iterative inversion starts with a uniform initial model of $\varepsilon = 0.1$ that has to be inferred independently. The corresponding total scattering coefficient $g_{01}$ is 0.157 km⁻¹ which is equivalent to an inverse scattering quality factor $Q^{-1} = 0.0314$. The misfit obtained from the differences between the simulations in the initial model and the observation (here it is the synthetic data simulated in the true model) is used in the adjoint method that we introduced in Section 2.2 to obtain the gradient of the misfit function (sensitivity kernel) for each source–receiver couple. All kernels for every combination are then summed to compute the misfit kernel for updating the model. The energy at the sources and the stations (sources of the adjoint field) is much higher than elsewhere in the model resulting in singularities of the gradients at these positions. In FWI this influence can be removed by a pre-conditioner that was designed to approximate the Hessian matrix or its diagonal terms (Zhu et al. 2015). In this study, we use a simple approach: a normalized Gaussian distribution subtracted from 1 ($1 - \exp[-(r - r_j)^2 / 2\sigma^2]$) is multiplied with the kernel for each source/sensor location $r_j$ to suppress the singularities of the gradient. Additionally, we also apply a Gaussian smoothing filter to the kernel, resulting in final misfit gradient. This step is usually regarded as a regularization procedure that guarantees the stability of inversion (Zhu et al. 2015; Tao et al. 2018). For details about the regularization which we applied here, please refer to Appendix B.

Using the L-BFGS method the search direction of the $\varepsilon^2$ model is obtained. In each iteration, two tentative step lengths are tested by separate forward simulations and calculation of the misfit. The best step length is determined as the minimum of a quadratic function fitted to the three misfit values obtained for the initial model of the current iteration (step length zero) and the two tested step lengths. With this step length, the initial model is updated each iteration. This workflow is repeated several times and the model is iteratively improved until the misfit converges. Fig. 2 shows the evolution of the normalized misfit over 8 iterations which shows that after about five iterations the misfit starts to converge to 36 per cent. The slight increase in the misfit curve at the 6th and 8th iteration is due to changes in the details of the model during the model update and will be discussed later. The rather rapid convergence results from the use of the L-BFGS method, the testing of the step length, and the good ray coverage (24 sources × 23 receivers). But it is also clear that a significant improvement of the model is achieved after the first iteration, documenting the non-linearity of the problem and the benefit of the iterative inversion over the single-step kernel based approaches (e.g. Obermann et al. 2013; Sánchez-Pastor et al. 2019).

Fig. 3 illustrates the inversion result of $\varepsilon$ after eight iterations. Although the initial model is uniform which means that there is no prior information given about the perturbations, the two anomaly areas of the true model (Fig. 3a) are successfully retrieved in location, shape and amplitude. To quantitatively evaluate the results, three profiles across the anomaly at $x = 12.5$ km, $x = 37.5$ km, $y = 25$ km are shown in Fig. 3(b). The evolution of the model in each iteration is shown with differently coloured curves ranging from the initial model (light yellow) to the final result (dark purple) compared with the true model (red).

The location of the anomalies is picked up already in the 1st iteration while the amplitudes are gradually fitted with successive iterations. This process is not monotonic but can include some oscillations as seen for the negative anomaly where some overshoot in the central part after the fifth iteration is corrected in subsequent iterations. For this reason the misfit curve slightly fluctuates around 36 per cent after the fifth iteration. Smaller step lengths can improve this behaviour but will increase the computational costs and are more likely to result in local minima. Both in Figs 3(a) and (b), the final result shows that the anomaly areas are smooth and do not exactly match the true model. This limitation is partially due to the Gaussian smoothing we use to in the preconditioning of the gradient, but also due to the setup of sources and receivers around the target area. Gaussian smoothing is chosen here as a simple method for demonstration and should be replaced in further applications by more advanced regularization techniques that are, for example, more suitable to resolve sharp-sided anomalies (Akcelik et al. 2002).

In a realistic inversion scenario, one cannot compare the inverted model with the true model. The only possibility to assess the final result, is the comparison of the modelled and measured envelopes.
Figure 1. The illustrations of the scattering anomaly model. The background of \( \varepsilon \) (left-hand panel) is 0.1. The white and the dark blue are two anomaly areas, respectively with the value 0.05 and 0.15. The attenuation model (right-hand panel) is homogeneous with the value 0.005. The red points represent both the sources and receivers. Note that the model we demonstrated here is the partial space and there is no boundary for the simulation.

Figure 2. The misfit in the scattering inversion experiment alters with the number of iterations. The values are normalized by the misfit between the uniform initial model and the true model shown in Fig. 1.

Fig. 4 shows envelopes of four source–receiver combinations and the adjoint sources (misfit) in the different iteration from the initial (light yellow) to the final (dark purple) compared with the observation (red). Since the initial model is homogeneous, the synthetic energy density of direct wave is larger than the observations in Fig. 4(a) that shows a source–receiver combination in which the ballistic wave passes through the stronger scattering anomaly. This results in increased energy loss of the ballistic wave. The amplitude of the adjoint source gradually decreases indicating convergence of the modelled envelope towards the data obtained in the true model. The same but with opposite sign can be observed in Fig. 4(b) that shows a source–receiver configuration for which the ballistic wave passes through the anomaly with reduced heterogeneity. The energy of the ballistic peak is underestimated in the initial model and gradually increases during the inversion.

Fig. 4(c) shows a source–receiver combination that is similarly affected by both anomalies as the ballistic wave passes through both anomaly areas. Since the effects of both anomalies compensate to some extent the absolute amplitude of the adjoint source is smaller than in Figs 4(a) and (b). However, as the effect of the increased heterogeneity on \( \varepsilon^2 \) is larger than the effect of the reduced heterogeneity for equal \( |\Delta \varepsilon| \) the adjoint source is positive and decreases in amplitude during the inversion.

The absolute amplitude of the adjoint source for the source–receiver combination for which the direct wave does not pass through any anomalous region (Fig. 4d) is even smaller. It has positive and negative segments that are improved during the inversion but it is devoid of a signal associated to the ballistic peak that is correctly represented by the initial model. The fluctuation of the adjoint source in Fig. 4(d) mainly represents the statistical fluctuations of the simulations with the Monte Carlo method and also contains some scattered waves from distant scatterers.

3.2 Intrinsic attenuation inversion

In this section, we conduct an inversion for the intrinsic attenuation only and fix the strength of the heterogeneity \( \varepsilon \). We call it intrinsic attenuation inversion. Fig. 5 shows the new model in which there are two anomaly areas in the \( Q^{-1} \) model with values of 0.0075 and 0.0025, respectively whereas \( \varepsilon \) is uniform in space. The parameters of the background and the setup of the grid are the same as in Section 3.1.

Again the iterative inversion starts form a uniform initial model and the workflow follows Section 3.1 but now the model of \( Q^{-1} \) is updated in each iteration. Fig. 6 shows the evolution of the model misfit during the inversion. It converges to a value of about 40 per cent after the 7th iteration. The resulting model of the 9th iteration is shown in Fig. 7. Both anomaly areas are recovered (Fig. 7a) but their shape is smoother compared to the inversion result for heterogeneity in Section 3.1.

The envelope fits are shown in Fig. 8 together with the adjoint sources. Figs 8(a) and (b) show the source–receiver configurations with ballistic wave passing through the areas of smaller and larger intrinsic attenuation, respectively. Both envelopes are fitted in a similar way over the course of the iterations leading to a vanishing adjoint source of the ballistic waves and its early coda. The two source–receiver combinations shown in Figs 8(c) and (d) that are affected by both anomaly areas again have significantly smaller adjoint sources fluctuating around zero. No significant changes occur here during the course of the inversion. This is to be expected since both anomalous areas have similar influence on the envelope as a shift of \( Q^{-1} \) in both anomaly areas is 0.0025.
3.3 Simultaneous inversion

In the previous sections we demonstrated that the iterative adjoint envelope tomography allows to recover anomalies in the distribution of small-scale heterogeneity as well as in intrinsic attenuation. However, in both cases we used a uniform model for the other property and did not update the respective part of the model. This is not possible in a real application in which it is a priori unknown whether an observed misfit between model and data is caused by attenuation.
Adjoint envelope tomography

Figure 5. The illustrations of the intrinsic attenuation anomaly model. The background of $Q^{-1}$ (right-hand panel) is 0.005. The white and the dark blue are two anomaly areas, respectively with the value 0.0075 and 0.0025. The $\varepsilon$ model (left-hand panel) is homogeneous with the value 0.1.

Figure 6. The misfit in the intrinsic attenuation inversion experiment alters with the number of iterations. The values are normalized by the misfit between the uniform initial model and the true model shown in Fig. 5.

or scattering. Consequently both parameters of the model have to be updated simultaneously.

This simultaneous inversion of scattering and intrinsic attenuation is much more complicated than the separate inversions. In this section, we combine the $\varepsilon$ model that we used in Section 3.1 for scattering inversion and the $Q^{-1}$ model from Section 3.2 for the inversion of intrinsic attenuation. Both scattering and intrinsic attenuation of the true model have two anomaly areas as shown in Fig. 9. The workflow of the simultaneous inversion is almost the same as the separate inversion. The only difference is that both $\varepsilon$ and $Q^{-1}$ are updated in each iteration. As a result, the simultaneous update of the models requires two different step lengths, respectively, for $\varepsilon$ and $Q^{-1}$. We have introduced in Section 2.4 the assumption that the relationship between the misfit and each step fits the quadratic function. Here, we estimate the step length independently using two trial step lengths for each parameter to determine the best step length. Finally, four test simulations are required for each iteration.

As shown in Fig. 10 the normalized misfit decreases to 24.81 per cent after eight iterations which is lower than in the separated inversions. Although the absolute value of the misfit in each numerical test is different, the faster decrease illustrates more significant changes in the envelopes due to the recovery of the anomalies. The inversion results are shown in Fig. 11. Fig. 11(a) shows the final inverted models after eight iterations. The anomaly areas in the true model of $\varepsilon$ and $Q^{-1}$ are located in the north–south and east–west directions, respectively. The model of $\varepsilon$ is well recovered—similar to the result in Section 3.1. Profiles along the four lines shown in Fig. 11(b) and (d) illustrate that the inverted model agrees well with the true $\varepsilon$ model.

However, in the inverted model of $Q^{-1}$ one can observe four anomaly areas. Two of them are located in the east and west which agrees with the locations in the true model. The profiles across these two areas in Fig. 11(c) show that the amplitudes of these anomalies are underestimated. Normally the underestimation of the amplitudes is reflected by the misfit and improved by more iterations. Yet, in the joint inversion this process seems to be impaired by the presence of the other two anomaly areas in the north and south. Since these areas with anomalies in $\varepsilon$ introduce changes in the envelopes the misfit can be reduced also by adapting the $Q^{-1}$. Profiles through the $Q^{-1}$ model across the locations of the $\varepsilon$ anomalies in Fig. 11(e) show these misplaced anomalies. The trade-off between improvements of the $Q^{-1}$ and $\varepsilon$ models will be discussed in the next section in detail.

The envelope fit and the adjoint sources in the simultaneous inversion are shown in Fig. 12. Both the redistribution of energy due to the changes in heterogeneity and the energy loss caused by intrinsic attenuation anomaly contribute the changes in the envelopes. Disentangling these influences is the challenge of the joint inversion which complicates the model updates.

Let us first consider the source–receiver combinations that are affected strongest by the anomalies in heterogeneity because of the ballistic path traversing through the $\varepsilon$ anomalies in Figs 12(a), (b) and (c). The convergence of these envelopes is similar to the separate inversion for $\varepsilon$ in Section 3.1.

As expected from the inverted model the evolution of envelopes affected by the attenuation anomalies is different. The source–receiver combination with ballistic path across the stronger intrinsic attenuation anomaly (Fig. 12d) shows that the adjoint source improves in the first few iterations but then grows again (with opposite sign) in the later iterations. In contrast the envelope across the weak attenuation area in Fig. 12(f) nicely converges and the adjoint source
remains minimal in the last iterations. In general the amplitudes of the adjoint sources of source–receiver combinations with ballistic path through the $Q^{-1}$ anomalies in the east and west are smaller than those affected by the $\varepsilon$ anomalies.

4 DISCUSSION

To demonstrate that the mathematical framework of adjoint envelope tomography with the radiative transfer equation which we have presented in Section 2 is able to recover material properties,
we performed different synthetic tests. We want to remark first, that we used the radiative transfer MC-algorithm for the inversion and for the simulation of the synthetic data. We chose this strategy over the more realistic scenario of generating the synthetic data with wavefield simulations to avoid the additional fluctuations that result from the specific realization of the medium heterogeneity that has to be generated for the wave field simulation. This effect will be investigated in the future, but in this first application we want to focus on the methodological aspects of the adjoint tomography. For the same reason we restrict ourselves to the acoustic case in two dimensions.

In the three numerical experiments we assume that first order estimates of the large scale average material properties are available from separate investigations, a prerequisite shared with FWI. In the first two experiments we show that the adjoint tomography successfully recovers the heterogeneity and absorption structures if the other material property is known and fixed. Significant improvements of the model are achieved following the first iteration. These improvements cannot be achieved with existing single-step inversion approaches where sensitivity kernels are calculated in a uniform model. In both experiments with the separate inversion for either heterogeneity or absorption the misfit converges after a few iterations and the remaining misfit can be attributed to the enforced smoothness of the retrieved model that prohibits further improvements. Successively lowering the smoothness constraint imposed by the filtering of the gradient could be a strategy to reduce the misfit further. Such a strategy would be similar to successively raising the high frequency limit in FWI.

The situation is different in the third experiment in which we aim to invert the heterogeneity and attenuation structures simultaneously. Fig. 11(a) gives the impression that the intrinsic attenuation is more difficult to constrain than heterogeneity. We will show in the following that this difficulty is due to the intrinsic trade-off between \( \varepsilon \) and \( Q^{-1} \) in the inversion process and depends on the specific parameter distribution that is to be imaged.

### 4.1 Trade-off

The fundamental problem we meet in the simultaneous inversion for heterogeneity and attenuation comes from the trade-off between changes in \( \varepsilon \) and \( Q^{-1} \). Eq. (28) shows that the scattering kernel is composed of two terms: (i) the active kernel that describes the energy gained by scattering from another directions and (ii) the passive kernel that describes the loss of energy due to scattering in other directions (Margerin et al. 2016; Zhang et al. 2021). The absorption kernel (eq. 29) has the shape of the passive kernel. Separating the effects of \( \varepsilon \) and \( Q^{-1} \) in the inversion is complicated by the passive kernel that is part of both kernels.

In order to quantify the influence of \( \varepsilon \) and \( Q^{-1} \) on the misfit, we design a test to calculate the misfit caused by the changes of \( \Delta \varepsilon \) or \( \Delta Q^{-1} \). Note that the expression (28) is a function of \( \varepsilon^2 \) while we use \( \varepsilon \) for illustration. The model is shown in Fig. 13. It contains only one anomaly area that is used for \( \varepsilon \) and \( Q^{-1} \) separately, while the other parameter is kept uniform. Background values are as before (\( \varepsilon = 0.1 \) and \( Q^{-1} = 0.005 \)). We compare the adjoint sources caused by the two types of changes in the anomaly area (\( \Delta \varepsilon^2 = 0.02 \) or \( \Delta Q^{-1} = 0.02 \)). Two source–receiver combinations are shown for which the ballistic wave crosses the anomaly or travels far away from the anomaly area as shown in Figs 14(a) and (b), respectively. Although the difference between the two perturbations (the blue and red curves) is clearly observed in the later coda waves, it is impossible to estimate the character of the anomaly from the ballistic wave. In the case of distant sensors in Fig. 14(b) the wavefield obviously does not encounter the anomaly prior to approximately 9 s making it impossible to learn anything about the anomaly. But even in the case of ballistic propagation through the anomaly (Fig. 14(a))

---

**Figure 9.** The illustrations of the scattering and intrinsic attenuation anomaly model in the simultaneous inversion. The background of \( \varepsilon \) and \( Q^{-1} \) is 0.1 and 0.005, respectively. The white and the dark blue are two anomaly areas, with the value of 0.05 and 0.15 in the scattering model and the value of 0.0075 and 0.0025 in the intrinsic attenuation model, respectively.

**Figure 10.** The misfit in the simultaneous inversion experiment alters with the number of iterations. The values are normalized by the misfit between the uniform initial model and the true model shown in Fig. 9.
Figure 11. The simultaneous inversion results: (a) the final model of $\varepsilon$ and $Q^{-1}$ after 8 iterations; (b)–(d) the profiles across the anomalous areas (illustrated in the right-hand column) of the updated model in each iteration from the initial model (light yellow) to the final model (dark purple) compared with the true model (red).
Figure 12. The comparison of observed (red) and synthetic envelopes in each iteration from the initial (light yellow) to the final (dark purple) as well as the corresponding adjoint sources in the simultaneous inversion.

Figure 13. Model used for trade-off test. The anomaly area is separately perturbed by changes in $\varepsilon$ and $Q^{-1}$. The other simulation parameters are the same as before.

the effect of increased scattering and increased attenuation on the ballistic wave and early coda is approximately the same as both perturbations decrease the coherent energy.

The time window chosen in the simultaneous inversion test includes the whole period from 0 to 15 s so the misfit is dominated by the large amplitude ballistic waves. Since the scattering kernel for...
the ballistic wave is dominated by the passive kernel and attenuation is only influenced by the passive kernel variation of both scattering and attenuation can each explain misfit of the ballistic wave resulting in a trade-off. This can only be resolved if sufficient weight is given to the late coda in which the active kernel contributes to the scattering kernel.

Fig. 14(c) shows the misfit for varying amplitudes of the anomalies. It allows to estimate the values of $\varepsilon$ and $Q^{-1}$ at which both parameters have comparable influence on the misfit. The relative influence on the misfit can be illustrated in the $\varepsilon - Q^{-1}$ domain as shown in Fig. 15 where the red line represents the relationship $\text{Misfit}(|\Delta \varepsilon^2|) = \text{Misfit}(|\Delta Q^{-1}|)$ which splits the domain into two parts. The green and pink areas represent combinations of $\varepsilon - Q^{-1}$ in which the misfit is dominated by the scattering and intrinsic attenuation, respectively. A similar discussion of this trade-off on the fit to the envelopes was conducted by Cormier & Sanborn (2019). Different from the adjoint tomography focused on the misfit function, their work compared the influence of the two parameters on the measurements with MLTW A which is affected by the choice of time windows. We will discuss the effects of the time window in our approach in Section 4.3.

The values of $|\Delta \varepsilon^2|$ and $|\Delta Q^{-1}|$ that we used in the numerical experiment for the simultaneous inversion in Section 3.3 are illustrated in Fig. 15 by the blue dot and the red triangle for the positive and negative anomalies, respectively. It is clear that the misfit in this situation is dominated by the anomalies in $\varepsilon$. To illustrate how this dominance affects the inversion Fig. 16(a) shows the kernels in the first iteration in this simultaneous inversion experiment. While the $\varepsilon$ kernel correctly highlights the areas of the $\varepsilon$ anomalies in the north and south, the $Q^{-1}$ kernels also highlights the $\varepsilon$ anomalies supplemented by moderate values at the locations of the $Q^{-1}$ anomalies in the east and west. As a consequence the $Q^{-1}$ update starts off in the wrong direction. A similar effect but to a much smaller extend can be observed for the $\varepsilon$ kernel that shows a slight increase at the locations of the $Q^{-1}$ anomalies in the east and west. It is obvious that the misfit is dominated by the heterogeneity anomalies which results in a bias of the absorption kernels and leads to the inverted model in Fig. 11(a).

On the contrary we can try to recover a model in which the anomalies of intrinsic attenuation dominate the inversion as shown in left column of Fig. 16(b). The values of the $|\Delta \varepsilon^2|$ anomaly are 0.0019 and 0.0021 while that for $|\Delta Q^{-1}|$ is 0.005 as indicated by the green square and orange diamond in Fig. 15. As expected the resulting kernels of the first iteration shown in Fig. 16(b) reflect the dominance of the $Q^{-1}$ anomalies. The anomaly areas of $Q^{-1}$ are correctly recognized while the scattering kernel is biased by the anomalies in absorption.

### 4.2 Influence of the Initial model

In the two individual inversions and the simultaneous inversion above, we assumed that the initial models are all uniform meaning that there is no priori information about the location, shape or amplitude of the anomaly. In the separate inversions for scattering and intrinsic attenuation the respectively other parameter distribution is uniform, equals the true model and is not updated, leading to a good recovery of each parameter as shown in Figs 3(a) and 7(a). If updates of both parameters are required, the joint calculation may project the data misfit into updates of both $\varepsilon$ and $Q^{-1}$ and it is instructive to investigate how the knowledge about one parameter influences the recovery of the other.

We assume that we have additional priori information about the $\varepsilon$ structure such that we can start the inversion with a non-uniform but correct $\varepsilon$ model. The kernels in the first iteration under this condition are shown in Fig. 17(a). Now the intrinsic attenuation kernel highlights the correct anomaly areas because there is no misfit from a wrong $\varepsilon$ model. However, the $Q^{-1}$ anomalies also show up in the $\varepsilon$ model, meaning that the correct $\varepsilon$ model would be altered in a simultaneous inversion for both parameters. The same applies when the correct $Q^{-1}$ model but a uniform $\varepsilon$ model is used to obtain the kernels as shown in Fig. 17(b). The $\varepsilon$ kernel highlights the correct anomalies in the north and south while the $Q^{-1}$ kernel would alter the correct model of absorption in a simultaneous inversion for both parameters.

These tests show that the kernels of absorption and heterogeneity are both correctly recovered also if the other parameter has a complicated spatial structure that is well described already. However, if the structure of one or both parameters is not known and both parameters are inverted for, the contributions to the data misfit of both parameters show up in both kernels to a variable degree.

### 4.3 Time window for full envelope inversion

In the experiments described above we use a time window that contains the whole envelope ranging from the source time until the late coda at more than twice the traveltine of the ballistic wave. This choice corresponds to setting $T$ in eq. (28) and (29) to the end of the time window. Although the full information of wave propagation through the medium is utilized in this time window, we show in the following how absorption and heterogeneity have different imprints on the envelopes at different lapse times.

Scattering reduces the energy of the ballistic wave and increases the coda wave while the intrinsic attenuation absorbs the energy during the whole lapse time including both the ballistic wave and coda waves. Since the effect of absorption and heterogeneity is the same for the ballistic wave one can anticipate that the coda is crucial to distinguish between both effects which has been illustrated in
Figure 16. The kernels of the first iteration in two simultaneous inversion cases: (a) the misfit is dominated by $\varepsilon$ which is the case conducted in Section 3.3 and (b) the misfit is dominated by $Q^{-1}$.

Figure 17. The kernels of the first iteration in the simultaneous inversion conducted in Section 3.3 with different initial models: (a) $\varepsilon$ is the same as the true model while $Q$ is homogeneous and (b) $Q^{-1}$ is the same as the true model while $\varepsilon$ is homogeneous.
Figure 18. The kernels of the first iteration in the simultaneous inversion whose misfit is dominated by $\varepsilon$ with different time windows: (a) ballistic wave and (b) coda wave. The results with the time window containing both ballistic wave and coda wave is shown in Figs 16(a).

Figure 19. The kernels of the first iteration in the simultaneous inversion whose misfit is dominated by $Q^{-1}$ with different time windows: (a) ballistic wave and (b) coda wave. The results with the time window containing both ballistic wave and coda wave is shown in Figs 16(b).
Figs 14(a) and (b). In this section we use the models discussed in Section 4.1 with anomalies in both heterogeneity and absorption and uniform initial models to study the effects of the ballistic wave and the coda wave by separating them in different time windows.

Figs 18 and 19 show results under the condition that the misfit is dominated by heterogeneity and absorption, respectively. Figs 18(a) and (b) show the misfit kernels in the first iteration of the inversion with time windows restricted to the ballistic waves and coda waves, respectively. The results for the ballistic time window shown in Fig. 18(a) are similar to the ones in Fig. 16(a) with the full lapse time range and show the strong bias of the absorption kernel from the anomalies in heterogeneity. This is reasonable since the energy peak of the ballistic wave dominates the inversion when the full envelope is inverted and the effect on the ballistic wave energy is the same for scattering and intrinsic attenuation. A significant improvement of the intrinsic attenuation kernel is achieved by restricting the time window to the coda wave as shown in Fig. 18(b). The effect of the heterogeneity anomaly on the absorption kernel is drastically reduced and reversed in polarity compared to Fig. 18(a). This reflects the different effects of anomalies in heterogeneity and absorption on coda waves and indicates a possibility to steer the inversion by a lapse time dependent weighting of the misfit in the calculation of the misfit kernels.

Figs 19(a) and (b) show the absorption dominated misfit kernels restricted to ballistic and coda wave time window, respectively. Here the absorption kernel is rather independent of the time window used due to its dominant effect. The scattering kernel, however, is very different for the two time windows. The kernels derived from both the coda and ballistic time windows are biased by the absorption anomalies, but the bias has opposite polarity for the ballistic and coda windows while the polarity at the location of the scattering anomalies is the same. An appropriate weighting of both
time windows can thus improve the inversion also in this situation. In conclusion, separating the ballistic wave and coda wave into separate time windows indicates a potential strategy to improve the trade-off.

4.4 The effect of noise

The Monte Carlo simulation which we use for the forward modelling is a stochastic method that introduces fluctuations in the synthetic data. These random fluctuations can be regarded as the noise of observation. However, the use of a sufficient amount of particles ensures convergence of the simulation suppressing this random noise to a desired degree. For more details refer to appendix B of Zhang et al. (2021). To study a more realistic situation, we artificially add random noise to the synthetic data in our numerical tests. Gaussian noise is generated with a specific signal-to-noise ratio (SNR) defined as:

$$\text{SNR} = 10 \times \log_{10} \frac{E_{\text{MS}}}{N_{\text{MS}}},$$

where $E_{\text{MS}}$ indicates the mean square of the noise free signal envelope $E$ and the noise amplitude $N$. The addition of Gaussian noise can lead to negative values of the envelope which seems unrealistic. However, it is common practice in the analysis of scattered waves to estimate the noise energy from time windows prior to the arrival of direct waves and subtract this noise energy from the signal. This process does also lead to partially negative values in observed energy time-series, consistent with our approach.

Fig. 20 shows examples of the kernels in the first iteration for different SNR. The left-hand column illustrates the energy density and the adjoint source of one source–receiver couple. The black and red curves indicate the original data and that with added noise, respectively. From top to bottom the SNR is 20, 10 and 1, respectively. At the lowest level of noise (SNR = 20) shown in Fig. 20(a), the noise has rather little influence on the energy density. Also the adjoint source is still clearly visible and there is almost no difference between Figs 20(a) and 16(a) that shows the result without any noise. With increasing noise (SNR = 10), the coda part of the adjoint source is drowned in the noise but the ballistic wave can be still recognized. Resulting changes in the scattering kernel shown in Fig. 20(b) are still minor but the intrinsic attenuation kernel starts to degrade. At the high level of noise (SNR = 1), the additional fluctuations are comparable to the signal level. Although it is almost impossible to recognize the adjoint source in the data of a single source-station combination the scattering kernel shown in Fig. 20(c) still reflects the anomaly areas in $\varepsilon$ to some degree. Apparently information of the numerous records that are combined in the kernel still facilitates recovery of a somewhat degraded image even under the influence of strong noise.

5 CONCLUSION

We have presented a new physically rigorous tomography method for heterogeneity and attenuation. The iterative tomography combines forward modelling of the radiative transfer equations with simulations of the adjoint equations to obtain the Fréchet derivatives that guide the iterative model improvements. This full envelope inversion is thus the radiative transfer counterpart to full waveform tomography but allows us to obtain information about the structure of the propagation medium on length scales below the resolution limits of waveform tomography. We have demonstrated the success of this approach in a synthetic test of 2-D acoustic scattering.

The small-scale structure is statistically described by the spatially variable scattering and intrinsic attenuation properties of the medium described by the fluctuation strength $\varepsilon$ and intrinsic quality factor $Q^{-1}$. Seismic energy propagation in this medium is described by the radiative transfer equation with uniform macroscopic velocity and density. The presence of nonuniform scattering and intrinsic attenuation causes spatially variable seismic energy redistribution and loss with complex influence on the observed seismic waveform envelopes. We therefore use the full envelopes of the observed seismic waveforms for the inversion.

As objective function we use the squared difference between observed/synthetic and modelled data. Inspired by FWI we use the adjoint method to derive the Fréchet derivatives of the least-square misfit function. The radiative transfer equation is solved with the Monte Carlo method in 2-D with multiple non-isotropic scattering in an acoustic medium with spatially variable heterogeneity and attenuation. This algorithm is used to generate the synthetic data as well as for the forward and adjoint simulations. The synthetic data could have been generated also with wavefield simulations which would have been a more realistic experiment. However, we decided to use the same approach for generation of the synthetic data that we use in the inversion to focus on the performance of the adjoint formalism that we present here and leave the test against wavefield data for a later stage.

The numerical experiments demonstrate the success of the presented inversion concept. Fréchet derivatives of the two experiments with separate inversions for either $\varepsilon$ or $Q^{-1}$ reliably guide the inversion towards lower values of the misfit function. Models which better reproduce the synthetic envelopes as expressed by lower misfit also converge towards the true model in the model space. We have also shown that inversion is well behaved with respect to the addition of noise.

The iterative inversion correctly recovers the locations and amplitudes of the anomalies in each parameter, if the other parameter is fixed at the correct value. If both parameters are free in the inversion we observe a trade-off that is common to most inverse problems with multiple parameters. In this case the inversion process is dominated by the parameter that has the stronger anomalies, that is dominates the misfit function.

Although the trade-off between scattering and absorption cannot be avoided, some strategies exist to improve the simultaneous inversion. At first, correct knowledge of one of the parameters improves the recovery of the other. This means that independent knowledge of one parameter can be incorporated to stabilize the inversion. Secondly, the time window used to construct the adjoint source affects the recovery of the two parameters differently. While the late coda is essential to infer the distribution of absorption, the ballistic wave and early coda is important to locate anomalies in the small-scale heterogeneity. Applying a weighting between the influence of early and late coda in the inversion can help to improve the recovery of both parameters.

Future developments of adjoint envelope tomography for scattering and absorption include the application to real data (a laboratory study on a concrete test specimen is on the way), an extension to 3-D and the treatment of elastic wave scattering. We hope that the approach presented here allows for more detailed analysis of heterogeneous media, providing more accurate knowledge of heterogeneity and attenuation structures of targets such as volcanoes, the Earth’s crust on a regional scale, and even the deep Earth.
ACKNOWLEDGEMENTS

We thank Frederik Tilmann for suggestions on the discussion of the results. TZ likes to thank Ya-Jian Gao for inspiring discussion on the adjoint method and acknowledges financial support from the China Scholarship Council (CSC).

DATA AVAILABILITY

This work used synthetic data, which is available on request.

REFERENCES


Aki, K., 1980. Attenuation of shear-waves in the lithosphere for frequencies from 0.05 to 25 Hz, Phys. Earth planet. Inter., 21(1), 50–60.


**APPENDIX A: L-BFGS METHOD**

The L-BFGS method is applied to approximate the Hessian matrix that allows us to generate the search direction with the gradient (kernel). For the search direction $h_k$ of one parameter, we define the start model as $m_0$ with the the gradient $K_k$ in the $(k+1)$th iteration. The L-BFGS algorithm is shown in Algorithm 1 (Nocecul & Wright 2006). The number of most recent iterations $n$ is chosen for saving memory in this quasi-Newton method. In our tests, it is set between three and the maximum since the number of iteration is not huge. Note that $k$ in this algorithm must more than 0 so it does not work for the first iteration. We use the steepest descent method to calculate the first search direction $h_0$.

**APPENDIX B: CONDITIONING OF THE MISFIT KERNELS**

The regularization in our inversion has two parts, which are (i) removing the singularities at the sources/stations and (ii) smoothing the results.

The source–receiver singularities are an intrinsic problem of the adjoint method. Their effect is significant and it is almost impossible to recover the anomaly without removing them. We multiply the kernels with a 2-D Gaussian notch centred on each station and truncated at 2 km distance. The kernels before and after removing the singularities are shown in Fig. A1 bottom left and upper middle, respectively. Actually the influence of this step on kernels is mainly
Algorithm 1 L-BFGS

\( q_k \leftarrow K_k \),

do the number of most recent iterations \( n, n \leq k \)

\( \text{for } i = k - 1 \text{ to } k - n \text{ do} \)

\( s_i = m_{i+1} - m_i, \)

\( y_i = K_{i+1} - K_i, \)

\( \rho_i = 1/(y_i^T s_i), \)

\( \alpha_i = \rho_i y_i^T q_{i+1}, \)

\( q_i = q_{i+1} - \alpha_i y_i, \)

end for

\( r_{k-n} = [(s_{k-1}^T y_{k-1})/(y_{k-1}^T y_{k-1})] q_{k-n}, \)

\( \text{for } i = k - n \text{ to } k - 1 \text{ do} \)

\( \beta_i = \rho_i y_i^T r_i, \)

\( r_{i+1} = r_i + (\alpha_i - \beta_i)s_i, \)

end for

\( h_k \leftarrow r_k. \)

determined by the setup of sources and stations, which can be reduced with high enough station coverage.

Moreover, the kernels are smoothed by a 2-D Gaussian filter with the same \( x \)- and \( y \)-axis standard deviations \( \sigma_x = \sigma_y = \sigma \). The length of \( \sigma \) is normally determined by the wavelength and the structure in FWI and decreases with the model improving (Zhu et al. 2015).

Our approach is based on RTE that is beyond the limitation of the wavelength. But the model is discretized in a 100 by 100 cells grid with the size of 0.5 km \( \times \) 0.5 km. The Figs A1 shows the Gaussian filters with different \( \sigma \) (the upper left) and the corresponding kernels after smoothing (the right two columns). The one with \( \sigma = 0 \) km is the original kernel after removing the singularities at the sources and receivers. In the beginning, we set the \( \sigma \) as 3 km and after more iterations reduce it to 2 km. The \( \sigma \) less than 2 km has never been used which can result in the inversion unstable. In order to demonstrate the effects from smoothing with different \( \sigma \) on the adjoint sources, five simulations are conducted in these five models including the initial model and four updated models based on four different smoothed kernels in Figs A1 using the steepest descend. The resulting envelopes are shown in Fig. A2. The comparison shows, that the differences originating from the different smoothing of the kernels are insignificant compared to the change of the envelopes resulting from the model update. This indicates that the precise choice of the smoothing filter has minor influence on the result and our choice of the filter size of the order of the station distance is acceptable.
Figure A2. Comparison of observed (red) and synthetic envelopes with the corresponding adjoint sources in the initial model (blue) and four updated models. The four updated models are obtained with the steepest descend method using the four differently smoothed kernels shown in the two rightmost columns of Fig. A1. Source–receiver configurations are shown in the insets.