A specific type of modeling error evaluation for viscoelastic full waveform inversion

Tianze Zhang, Kris Innanen, Daniel, Trad

ABSTRACT

In the traditional viscoelastic Full Waveform Inversion (FWI) approach based on the generalized standard linear solid (GSLS) model, the quality factor, denoted as Q, is not directly inverted. Instead, it is first converted into a set of relaxation variables. Thus, the viscoelastic FWI process involves inverting for both elastic and relaxation models and subsequently translating the relaxation variables back into the Q model to achieve the final inversion results. This indirect approach is necessitated partly because the relationship between the Q values and relaxation variables is complex, and the functions that map between these two domains are not straightforward inverses of each other. In this report, we propose a novel method where a Multi-Layer Perceptron (MLP) is pre-trained to learn the mapping from Q values to relaxation variables. This trained MLP is then integrated into a Recurrent Neural Network (RNN)-based GSLS viscoelastic FWI framework, creating a complete computational graph. This graph connects Q models, relaxation variables, and elastic models directly to the synthetic data, thereby enabling the direct inversion of the Q model. Additionally, we employ the Monte Carlo dropout technique within the neural network to quantify the uncertainty associated with the MLP's learning process for mapping Q values to relaxation variables. In our assumption of a constant Q model, this uncertainty quantification reflects the relaxation variables' limited capacity to accurately represent a constant Q model. The potential of extending this approach to variable Q models is straightforward. The impact of this limited capacity, which is effectively a modeling error, on the forward modeling data is also examined. For our experiments, we employ simple layered models as well as subsets of the Marmousi models. Our observations suggest that this modeling error primarily affects the attenuation models. The findings of this study have significant implications for improving the accuracy and reliability of viscoelastic FWI, particularly in how we understand and handle the inherent uncertainties in modeling complex subsurface properties.

STANDARD LINEAR SOLID VISCOELASTIC FORWARD MODELING

Numerous numerical methods have been developed for the time domain forward modeling of viscoelastic wavefields. Day and Minster (1984) explored the incorporation of elasticity into 2-D time-domain modeling methods using a Padé approximation. Meanwhile, the generalized standard linear solid (SLS) method was proposed by Emmerich and Korn (1987). This method, rooted in rheological modeling, utilizes the finite difference method to enhance computational efficiency. Robertsson et al. (1994) introduced a staggered grid finite difference method, focusing on a single layer of SLS, to simulate viscoelastic wavefields. Using the SLS for viscoelastic wavefield modeling, the quality factor Q is often converted into stress and strain relaxation times for forward modeling calculations. Each SLS layer possesses unique stress and strain relaxation times. Consequently, determining these times based on the corresponding Q value becomes a pivotal step in modeling viscoelastic wavefields. To solve this issue, Fichtner and van Driel (2014) introduced a method that Q can be directly used for the forward modeling of viscoelastic wavefields.

Besides the Standard Linear Solid (SLS) model, alternative methods are available for modeling viscoelastic wavefields. Many of these alternatives necessitate the use of fractional derivatives. For instance, Zhu and Carcione (2014) developed a time-domain differential equation to model seismic wave propagation in constant Q viscoelastic media. This was achieved using fractional spatial derivatives, focusing on Laplacian differential operators of fractional order. Building on this, Zhou et al. (2023) enhanced the fractional spatial derivative Q modeling methods. They introduced a wavenumber-time domain viscoelastic k-space equation, employing the eigenvalue decomposition method and solutions to matrix differential equations.

Utilizing fractional derivative spatial derivatives to model viscoelastic wavefields can be seen as a "memory effect" modeling approach. Such methods capture the hereditary properties intrinsic to viscoelastic materials, providing flexibility when modeling various wave propagation behaviours, especially within intricate geological structures. However, the computation of fractional derivatives can be intricate. And while they may describe certain behaviours mathematically, their physical interpretation isn't always straightforward. The quality factor Q for the GSLS model can be obtained as follows:

$$Q_{\mathrm{GSLS}}(\omega, \tau_{\varepsilon 1}, \tau_{\sigma 1}, \dots) = \frac{1 - L + \sum_{l=1}^{L} \frac{1+\omega^2 \tau_{\varepsilon l} \tau_{\sigma l}}{1+\omega^2 {\tau_{\sigma l}}^2}}{\sum_{l=1}^{L} \frac{\omega(\tau_{\varepsilon l} - \tau_{\sigma l})}{1+\omega^2 {\tau_{\sigma l}}^2}}, \tag{1}$$

where $Q_{\mathrm{GSLS}}(\omega, \tau_{\varepsilon 1}, \tau_{\sigma 1}, \dots)$ is the Q model as a function of the frequency and relaxation parameters: $\tau_{\varepsilon l}$, and $\tau_{\sigma l}$, representing the strain and stress relaxation times of the l^{th} SLS mechanism, where $l \in \{1, L\}$. L is the maximum number of SLS mechanisms. Several methods have been developed to evaluate strain and stress relaxation times to approximate a Q model, with a certain degree of accuracy, i.e., Blanch et al. (1995) introduced the τ-method, which reduces the number of the relaxation parameters. An empirical interpolation method has been introduced by Liu and Archuleta (2006) to model the constant Q model. Fan et al. (2016) introduce a global optimization method to optimize the following objective function to obtain the relaxation variables for the constant Q model:

$$\tau_{\varepsilon 1}, \tau_{\sigma 1}, \dots = \underset{\tau_{\varepsilon 1}, \tau_{\sigma 1}, \dots}{\operatorname{argmin}} \sqrt{\frac{1}{N-1} \sum_{n=1}^{N} \left[\frac{1}{Q_{\mathrm{GSLS}}\left(\omega_n, \tau_{\varepsilon 1}, \tau_{\sigma 1}, \dots\right)} - \frac{1}{Q_{\mathrm{true}}(\omega_n)} \right]^2}, \tag{2}$$

where the ω_n is the discrete angular frequency and $Q_{\mathrm{true}}(\omega_n)$ representing the true relationship of how Q varies with respect to frequency ω_n. Post optimization, the relaxation variables are saved in the table, which can be used for modeling whenever needed.

Mapping Q with relaxation variables with neural networks

Here, I train a neural network in which the input of the neural network is the Q value, and the outputs are the desired relaxation times.

$$\mathcal{N}(Q; \theta) = \tau_{\varepsilon_{nn_1}}, \tau_{\sigma_{nn_1}}, \tau_{\varepsilon_{nn_2}}, \tau_{\sigma_{nn_2}}, \cdots, \tau_{\varepsilon_{nn_L}}, \tau_{\sigma_{nn_L}}, \tag{3}$$

where $\mathcal{N}$ is the functional mapping Q value into the relaxation time in the form of the neural network parameterized by the neural network parameter θ, and $\tau_{\varepsilon_{nn_l}}$, and $\tau_{\sigma_{nn_l}}$ are Q corresponding strain and stress relaxation times for the l^{th} SLS mechanism. Then, according to equation 1, frequency-dependent Q modelled with the relaxation variables generated with the neural network $\mathcal{N}$ is demoted as Q_{nn} which has the formulation of:

$$Q_{nn}(\omega, \mathcal{N}(Q;\theta)) = \frac{1 - L + \sum_{l=1}^{L} \frac{1+\omega^2 \tau_{\varepsilon_{nn_l}} \tau_{\sigma_{nn_l}}}{1+\omega^2 \tau_{\sigma_{nn_l}}^2}}{\sum_{l=1}^{L} \frac{\omega\left(\tau_{\varepsilon_{nn_l}} - \tau_{\sigma_{nn_l}}\right)}{1+\omega^2 \tau_{\sigma_{nn_l}}^2}} \tag{4}$$

During the training of the neural network, similar to equation (2), I set the objective function for neural network training as:

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \sqrt{\frac{1}{N-1} \sum_{n=1}^{N} \left[\frac{1}{Q_{nn}\left(\omega_n, \mathcal{N}(Q;\theta)\right)} - \frac{1}{Q_{\text{true}}(\omega_n)} \right]^2}. \tag{5}$$

Unlike the method introduced by Fan et al. (2016), which stores the relaxation variables for each discrete Q value in a large table, a well-trained network provides a continuous mapping function relating the Q values to the relaxation variables. This means that the neural network can map any Q value on the real axis within the range of interest to its corresponding relaxation variables. Furthermore, the proposed method meets our expectation of directly updating the Q value during inversion. During the inversion process, Q can take any value on the real axis due to the fine step length used in optimization, and a well-trained network can readily generate the corresponding relaxation variables for modeling.

During training, I generate 2100 random Q values within the range of $1 \sim 1000$. Two thousand of these values are used as training label data, while the remaining one hundred are used as testing data. I employ a dense network with Tanh() activation functions for the hidden layers and a Sigmoid() activation functions for the output layers. I test the network to generate relaxation variables under 2, 3, and 4 SLS mechanisms to obtain the constant Q model, which implies that the Q values are constant with respect to the frequency bandwidth of exploration seismic. An example of the dense network used to generate the relaxation variables for 2 SLS mechanisms is described in Table **??**. The networks for training the 3 and 4 SLS mechanisms are obtained by increasing the dimension of the weight in the neural network. I use the Adam algorithm as the optimizer with a learning rate of 10^{-4}. The stopping condition is set such that if the testing loss decreases to 10^{-3} of its initial value, training will stop. The training and testing loss with respect to the training epoch is plotted in Figure 1. The prediction stopped after around 11300 epochs, and the convergence property is promising.

In Figure 2, I plot the predicted $Q_{nn}(\omega_n, \mathcal{N}(Q;\theta))$ for different Q values using two layers of the SLS mechanism and compare them with the constant Q model, where $Q_{\text{true}}(\omega_n) = Q$. The corresponding relaxation stress and strain times are generated with the well-trained neural network. In Figure 2(a), we compare $Q = 30$. I forward propagate the well-trained neural network 1000 times to generate different $Q_{nn}(\omega_n, \mathcal{N}(30;\theta))$ models with a dropout rate of 0.5 in the neural network. The predicted lines are plotted as black lines in Figure

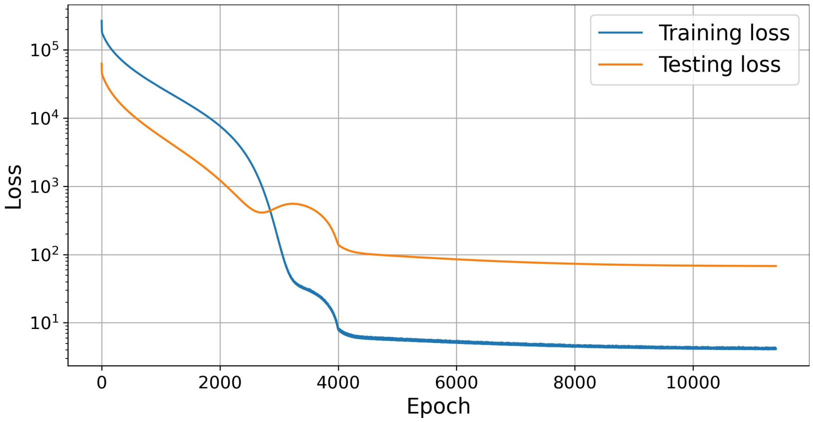

FIG. 1. Training and testing loss for the training the Dense network using two layers of SLS mechanisms for constant Q model.

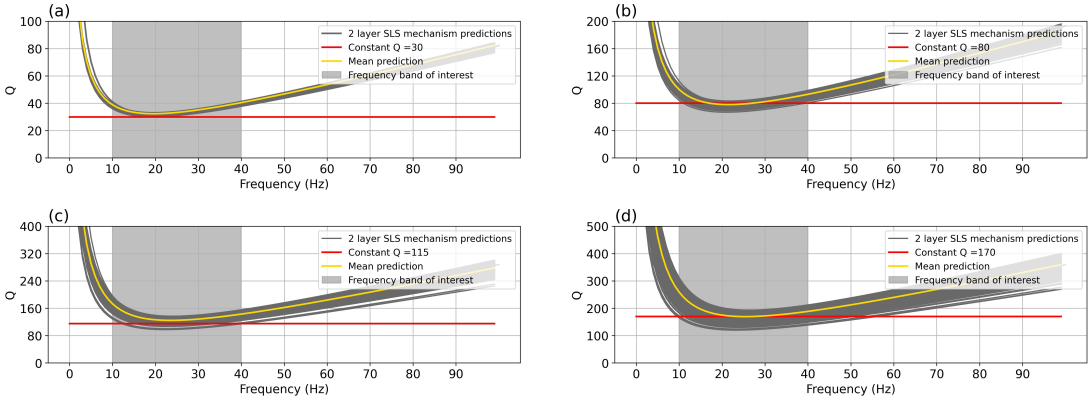

FIG. 2. Using two layers of SLS mechanism to predict constant Q models, where the relaxation stress and strain time are generated with the well-trained neural network.

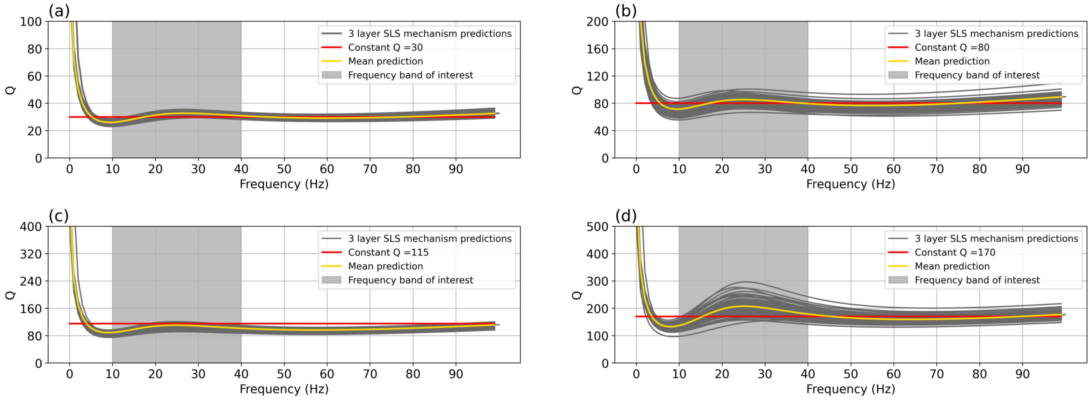

FIG. 3. Using three layers of SLS mechanism to predict constant Q models, where the relaxation stress and strain time are generated with the well-trained neural network.

2(a). The mean model of these lines is plotted as the yellow line, and the constant Q model, which is the target of training, is plotted as the red line. Within the exploration seismic frequency bandwidth of interest, i.e., $10\,\text{Hz} \sim 40\,\text{Hz}$, the mean predicted line matches the constant Q line with an acceptable error. However, for frequencies beyond our interests, the prediction lines deviate from the target constant Q model. Similar observations can be seen from the predictions of other Q values. Figure 2(b), (c), and (d) show the comparisons

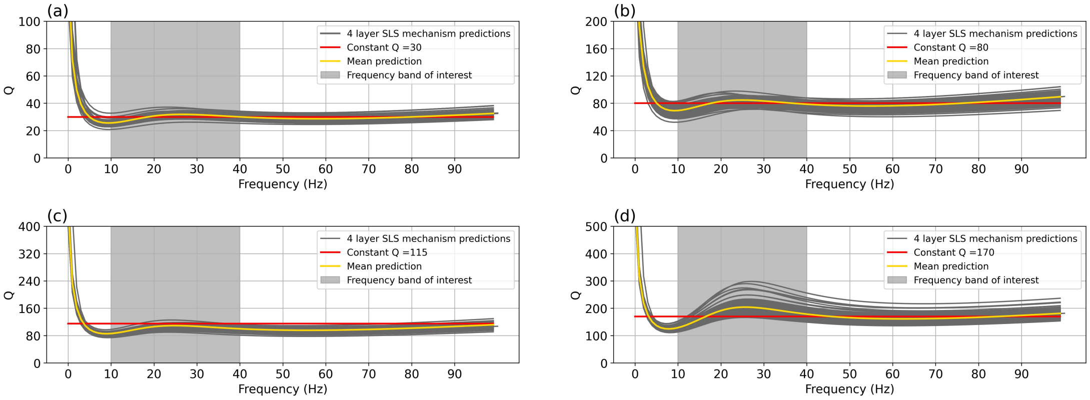

FIG. 4. Using four layers of SLS mechanism to predict constant Q models, where the relaxation stress and strain time are generated with the well-trained neural network.

for Q values of 80, 115, and 170, respectively. We can see that the predictions provide promising results within the seismic frequency bandwidth of interest.

In Figure 3, I plot the predicted $Q_{nn}(\omega_n, \mathcal{N}(Q;\theta))$ for different Q values using three layers of the SLS mechanism and compare them with the constant Q model. Figure 3(a) shows the prediction for $Q = 30$. Compared with Figure 2(a), we can see that with three layers of the SLS mechanism, we can better predict the constant Q model across all the frequencies used here. Within the frequency bandwidth of interest, the predicted lines align well with the constant Q model. Similar observations can be seen from the prediction results for other Q values, which are plotted in Figures 3(b), (c), and (d).

In Figure 4, I plot the prediction results for the 4 layers of SLS. The neural network can also generate relaxation variables that can be used to generate the constant Q model as described by equation (3). Compared with the predictions using 3 and 4 layers of SLS, I did not observe a significant improvement in predicting the constant Q model.

Phase velocity dispersion relationship

According to O'connell and Budiansky (1978), the elastic modulus can be written as the complex value function of frequency in which the real and imaginary part of the complex value is related by Q, with the form of:

$$Q(\omega) = \frac{\Re[M(\omega)]}{\Im[M(\omega)]}, \tag{6}$$

where the $\Re$ and $\Im$ represent the real and complex part of the complex modulus $M(\omega)$. With the well-trained neural network, which can generate the corresponding relaxation variables, we can determine the relaxed modulus M_P for a reference frequency f_r, a reference velocity v_0 at this reference frequency and density. According to Fan et al. (2016), the relaxed modulus can be formulated as:

$$M_R(\omega) = \rho v_0^2 \frac{\sqrt{\Phi_1^2 + \Phi_2^2} + \Phi_1}{2(\Phi_1^2 + \Phi_2^2)}, \tag{7}$$

where

$$\Phi_1 = 1 - L + \sum_{l=1}^{L} \frac{1+\omega^2\tau_{\varepsilon_l}\tau_{\sigma_l}}{1+\omega^2\tau_{\sigma_l}^2} \tag{8}$$

$$\Phi_2 = \sum_{l=1}^{L} \frac{\omega(\tau_{\varepsilon_l}-\tau_{\sigma_l})}{1+\omega^2\tau_{\sigma_l}^2} \tag{9}$$

The phase velocity dispersion formula can be calculated as:

$$v_p(\omega) = v_0\sqrt{\frac{\sqrt{\Phi_1^2+\Phi_2^2}+\Phi_1}{2(\Phi_1^2+\Phi_2^2)}}, \tag{10}$$

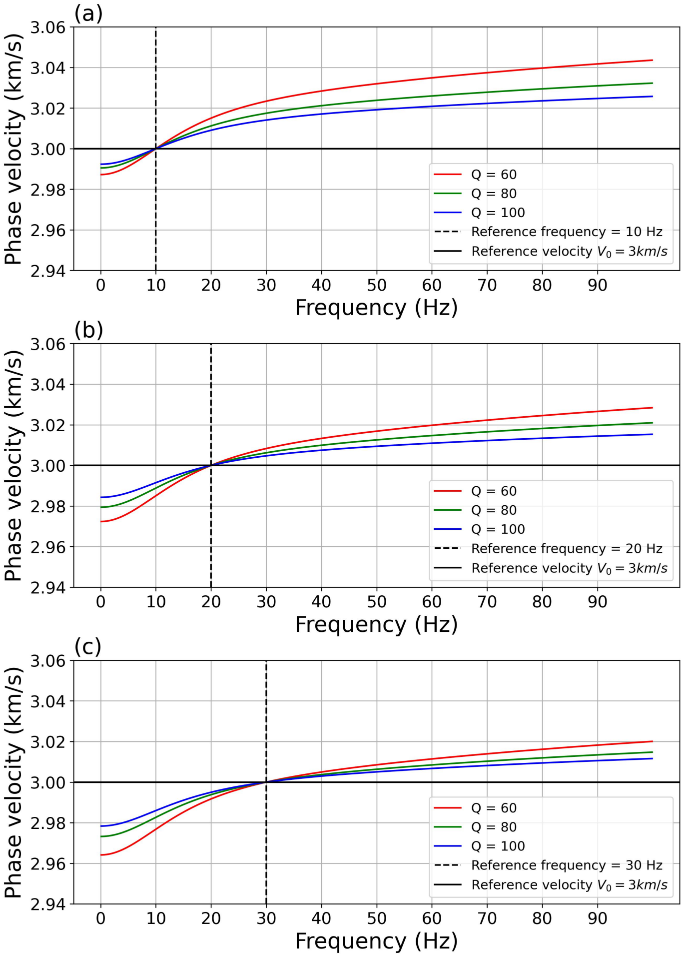

FIG. 5. Phase velocity variation with respect to a frequency band of $1\,\text{Hz} - 100\,\text{Hz}$ with Q values of 60, 80, 100. (a) Reference frequency $10\,\text{Hz}$. (b) Reference frequency $20\,\text{Hz}$. (c) Reference frequency $30\,\text{Hz}$.

In Figure 5, I display how the phase velocity varies with frequencies, $1\,\text{Hz} \sim 100\,\text{Hz}$, for different Q values, i.e., $Q = 60$, 80, and 100. Each Q value has its corresponding relaxation variables generated with the neural network, which are used in equation (10) to model each dispersion curve. In Figure 5(a), I choose a reference frequency of $10\,\text{Hz}$. We can see that the three lines intersect with each other at the 10 Hz frequency on the horizontal axis and a phase velocity of 3 km/s on the vertical axis. This demonstrates the correctness of our calculation since a phase velocity of 3 km/s is the reference velocity I choose at 10 Hz for modeling, and each dispersion curve satisfies this condition. Also, low Q values tend to generate large velocity dispersion, which is consistent with a linear viscoelastic model. Similar observations can be seen in Figure 5(b) and 5(c) when we use different reference frequencies. Again, the curves intersect with each other at the corresponding reference frequencies: 20 km/s for Figure 5(b), and 30 Hz for Figure 5(c) and a phase velocity of 3 km/s. For the same Q value, different reference frequencies cause the dispersion curves to shift up and down, which coincides with the dispersion curves given by Fan et al. (2016) in Figure 4(a).

Generalized standard linear solid viscoelastic wave equation

Following Robertsson et al. (1994), the viscoelastic wave equation for the standard linear solid model can be formulated as:

$$
\begin{aligned}
&\rho\frac{\partial v_x}{\partial t} = \frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xz}}{\partial x} + f_x,\\
&\rho\frac{\partial v_z}{\partial t} = \frac{\partial \sigma_{xz}}{\partial z} + \frac{\partial \sigma_{zz}}{\partial z} + f_z,\\
&\frac{\partial \sigma_{xx}}{\partial t} = \rho V_p^2 \Theta_L^P \left(\frac{\partial v_x}{\partial x} + \frac{\partial v_z}{\partial z}\right) - 2\rho V_s^2 \Theta_L^S \frac{\partial v_z}{\partial z} + R^P + R^S_{xx},\\
&\frac{\partial \sigma_{zz}}{\partial t} = \rho V_p^2 \Theta_L^P \left(\frac{\partial v_x}{\partial x} + \frac{\partial v_z}{\partial z}\right) - 2\rho V_s^2 \Theta_L^S \frac{\partial v_x}{\partial x} + R^P + R^S_{zz},\\
&\frac{\partial \sigma_{xz}}{\partial t} = \rho V_s^2 \Theta_L^S \left(\frac{\partial v_z}{\partial x} + \frac{\partial v_x}{\partial z}\right) + R^S_{xz},\\
&\frac{\partial r_l^P}{\partial t} = -\frac{1}{\tau_{\sigma l}^P}\left[r_l^P + \rho V_p^2 \left(\frac{\tau_{\varepsilon l}^P}{\tau_{\sigma l}^P} - 1\right)\left(\frac{\partial v_x}{\partial x} + \frac{\partial v_z}{\partial z}\right)\right]\\
&\frac{\partial r_{xxl}^S}{\partial t} = -\frac{1}{\tau_{\sigma l}^S}\left[r_{xxl}^S - 2\rho V_s^2 \left(\frac{\tau_{\varepsilon l}^S}{\tau_{\sigma l}^S} - 1\right)\frac{\partial v_z}{\partial z}\right]\\
&\frac{\partial r_{zzl}^S}{\partial t} = -\frac{1}{\tau_{\sigma l}^S}\left[r_{zzl}^S - 2\rho V_s^2 \left(\frac{\tau_{\varepsilon l}^S}{\tau_{\sigma l}^S} - 1\right)\frac{\partial v_x}{\partial x}\right]\\
&\frac{\partial r_{xzl}^S}{\partial t} = -\frac{1}{\tau_{\sigma l}^S}\left[r_{xzl}^S + \rho V_s^2 \left(\frac{\tau_{\varepsilon l}^S}{\tau_{\sigma l}^S} - 1\right)\left(\frac{\partial v_z}{\partial x} + \frac{\partial v_x}{\partial z}\right)\right],
\end{aligned}
\tag{11}
$$

where:

$$
\begin{aligned}
\Theta_L^P &= 1 - \sum_{l=1}^{L}\left(1 - \frac{\tau_{\varepsilon l}^P}{\tau_{\sigma l}^P}\right),\\
\Theta_L^S &= 1 - \sum_{l=1}^{L}\left(1 - \frac{\tau_{\varepsilon l}^S}{\tau_{\sigma l}^S}\right),
\end{aligned}
\tag{12}
$$

and

$$
\begin{aligned}
R^P &= \sum_{l=1}^{L} r_l^P,\\
R_{xx}^S &= \sum_{l=1}^{L} r_{xxl}^S,\\
R_{zz}^S &= \sum_{l=1}^{L} r_{zzl}^S,\\
R_{xz}^S &= \sum_{l=1}^{L} r_{xzl}^S,
\end{aligned}
\tag{13}
$$

The detailed derivation can be seen in Robertsson et al. (1994) and Fan et al. (2016). In equation (13), r_l^P is the relaxation field related to P-wave for the l^{th} SLS relaxation mechanism. r_{xxl}^S is the relaxation field for the stress field σ_{xx} in the l^{th} relaxation mechanism. r_{zzl}^S is the relaxation field for the stress field σ_{zz} in the l^{th} relaxation mechanism. r_{xzl}^S is the relaxation field for the stress field σ_{xz} in the l^{th} relaxation mechanism. This equation is the viscoelastic wave equation I use in this study. I use $L = 2$ in this study for forward modeling and inversion. Although I can see that with $L = 3$ and $L = 4$, I have the relaxation variables generated with the neural network that can better approximate the constant Q model. However, each relaxation mechanism has its own relaxation fields. The more relaxation mechanisms I use, the more relaxation fields I introduce into the forward modeling and inversion, dramatically increasing the computational cost of the inversion. Due to the limited computational resources available, I use $L = 2$ in all our numerical tests.

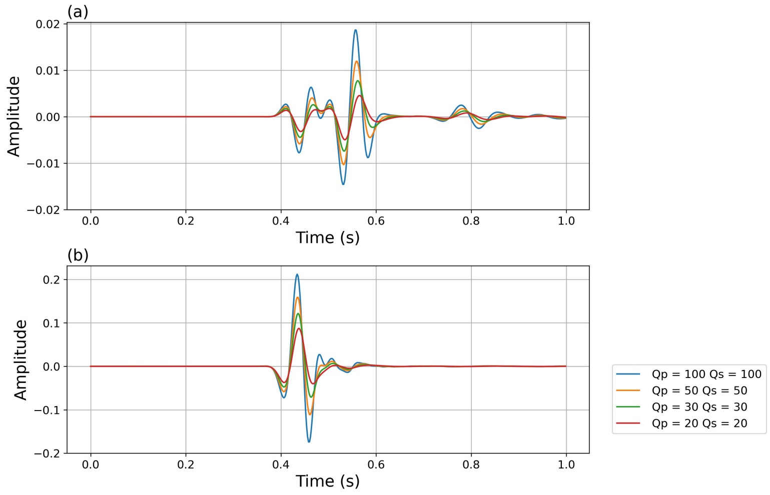

FIG. 6. The influence of different levels attenuation on the phase and amplitude of x and z components of seismic records. (a) Z components of a seismic record. (b) X components of a seismic record. Reference frequency $f_r = 40Hz$.

The choice of the reference frequency for modeling viscoelastic seismic waves should be made carefully. As seen in Figure 5, the choices of reference frequencies influence the phase velocity. If the reference frequency is not chosen properly, it can lead to unrealistic phase effects in wave propagation within viscoelastic media. I illustrate how the choice of reference frequency affects the phase of seismic records in Figures 6, 7, and 8. The seismic

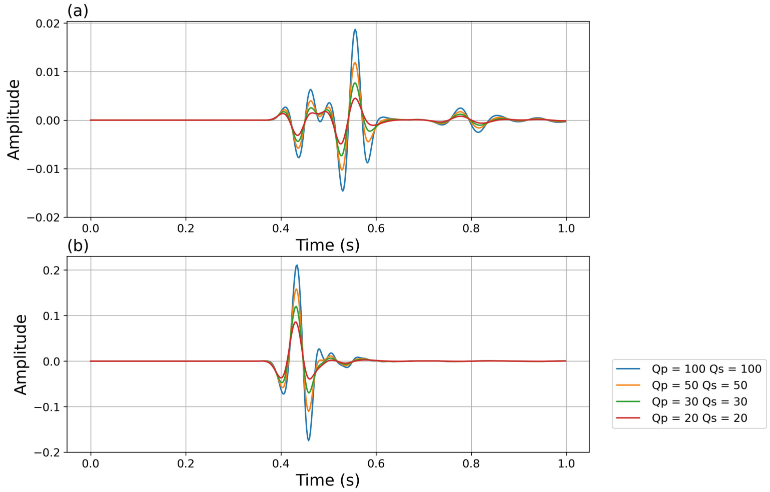

FIG. 7. The influence of different levels attenuation on the phase and amplitude of x and z components of seismic records. (a) Z components of a seismic record. (b) X components of a seismic record. Reference frequency $f_r = 13Hz$.

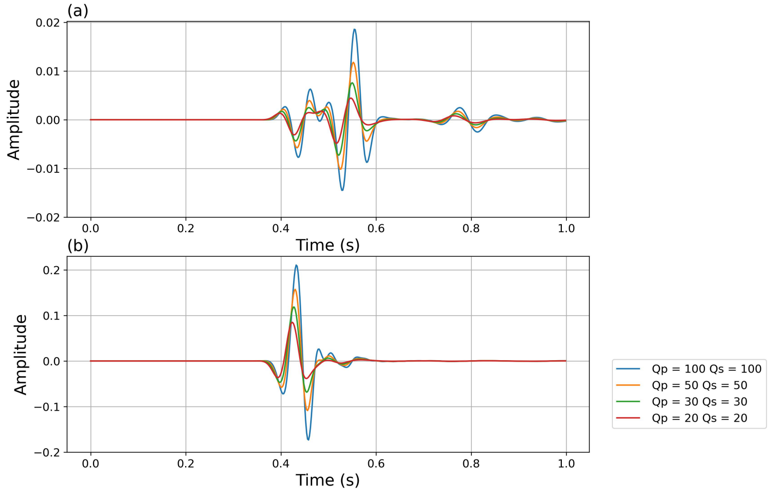

FIG. 8. The influence of different levels attenuation on the phase and amplitude of x and z components of seismic records. (a) Z components of a seismic record. (b) X components of a seismic record. Reference frequency $f_r = 10Hz$.

records are obtained from the viscoelastic media described in Table **??**. The models used are two-layer models with the same V_p, V_s, and ρ models, but different Q values. In each model, Q_p and Q_s have the same attenuation values. The source wavelet is a Ricker wavelet with a main frequency of 13 Hz. In Figure 6, I use a reference frequency of 40 Hz, which is larger than the main frequency of the source wavelet. In Figures 6 (a) and (b), I plot traces of the seismic records for the velocity field recorded in the Z and X components, respectively. As the Q value decreases, the total energy of the records decreases, and the higher frequency components of the records are more attenuated due to higher attenuation levels. The amplitude decaying and the phase delay effect satisfy our expectation of how wavefields propagate in the viscoelastic media Blanch et al. (1995), which is that we expect the wavefield would lose more energy as we increase the attenuation level and the high-frequency components of the records are will be more attenuated compared with the low-frequency componients. In Figure 7, I plot the seismic records recorded using a reference frequency of 13 Hz, which is the same as the main frequency of the source wavelet. I observe similar amplitude decay as the attenuation level increases, as seen in Figure 6, but there are almost no phase delays with this reference frequency. In Figure 8, I plot the records recorded using a reference frequency of 9 Hz. The amplitude decay phenomenon is similar to the aforementioned figures. However, the phase of the records with higher attenuation levels moves forward compared to those with lower attenuation levels, which is not realistic for viscoelastic seismic wavefield propagation. Therefore, I recommend that the reference frequency should be slightly larger than the main frequency of the source wavelet used for modeling.

VISCOELASTIC FWI

Overall inversion diagram

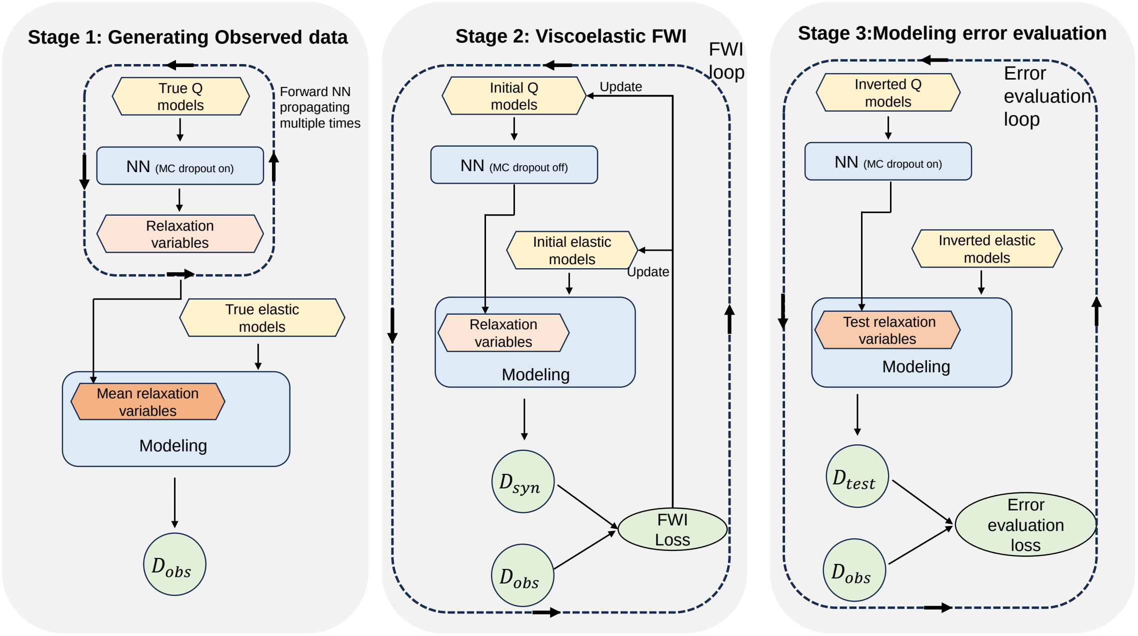

FIG. 9. This study follows an overall numerical testing diagram comprising three stages. In the first stage, I use the mean relaxation variables obtained from forward propagating the neural network multiple times to compute the observed data. In the second stage, I perform viscoelastic Full Waveform Inversion (FWI) using relaxation variables obtained from a single realization of the neural network, resulting in the inverted elastic and Q models. In the third stage, by forward propagating the neural network multiple times, I generate one realization of the test relaxation variables. I collect corresponding synthetic data D_{test} with misfit values close to the final FWI loss. By repeating this process several times, I obtain a data set for D_{test}. The standard deviation of the collected D_{test} dataset is considered as the modeling error. These errors arise from the inability of a single realization of the relaxation variables to represent the constant Q model.

In this section, I perform viscoelastic FWI and attempt to quantify the modeling error arising from the limited capability of the relaxation variables to represent a constant Q model. Our numerical test comprises three stages, and the overall inversion diagram is depicted in Figure 9. In the first stage, I generate the observed data assuming that Q is frequency-independent within the bandwidth of interest. During the observed data generation, I forward propagate the well-trained neural network to produce a set of relaxation variables using the true Q model by enabling the Monte Carlo (MC) dropout in the neural network. As shown in Figures 2, 3, and 4, forward propagation of the well-trained neural network allows the mean relaxation variable to provide a more accurate evaluation of the constant Q model. I use the true elastic model and the mean relaxation variables to compute the observed data D_{obs}.

In stage 2, I calculate the viscoelastic FWI. In this study, I use the recurrent neural network work-based inversion strategy, which uses the automatic differential method to update the elastic models and the Q models directly. During the second stage, I do not enable the Monte Carlo (MC) dropout in the neural network and use a single realization of the relaxation variables from the neural network for inversion. Consequently, the relaxation variables are generated differently from stage 1. If we were to enable the MC dropout, it would increase the nonlinearity between the data and the Q model. Since Full Waveform Inversion (FWI) is already a highly nonlinear inversion method, introducing additional nonlinearity would render the inversion unstable. In contrast to stage 1, where I forward calculated the relaxation variables multiple times, I compute only one realization of the relaxation variables in stage 2 to reduce the computational cost. As a result, the forward modeling used in inversion and data generation differs, ensuring that I do not commit the inverse crime in our numerical test.

In the third stage, I aim to assess the modeling error arising from the limited ability of a single realization of the relaxation field to represent the constant Q model. I achieve this by "turning on" the Monte Carlo (MC) dropout switch of the neural network and letting the neural network generate different relaxation variables and evaluate the corresponding synthetic data. In more detail, I first use the inverted Q model as input to realize the relaxation variables, referred to as the test relaxation variables. Using the inverted elastic models and the test relaxation variables, I obtain the synthetic data D_{test} and compute the misfit between D_{test} and D_{obs}, which I refer to as the error evaluation loss. If the error evaluation loss is equivalent to or within the acceptable range of the final FWI loss, then D_{test} is considered synthetic data that differs from the final FWI synthetic data D_{syn} but does not affect the final FWI loss due to the variation of the relaxation variables. Such D_{test} will be collected into a dataset for quantifying the modeling error. Subsequently, I allow the neural network to generate another realization of the test relaxation variable and evaluate a new D_{test}. By repeating this process several times, I obtain a dataset of D_{test} values that differ solely due to the variation of the relaxation variables but have misfit values close to the final FWI loss. I consider the standard deviation of the collected D_{test} dataset as the modeling errors. These errors arise from the inability of a single realization of the relaxation variable to represent the constant Q model.

Simple layers model viscoelastic FWI

Stage 1

In this section, I calculate viscoelastic FWI on a simple layered model. The true elastic models for V_p, V_s, and ρ are shown in Figures 10 (a), (d), and (g), respectively. These models feature a simple layered structure with a box-shaped anomaly located at the center. The true attenuation models Q_p and Q_s are depicted in Figures 11 (a) and (b). The model dimensions are 40×100 with a grid length of $d_x = d_z = 20m$. In this test, I use a Ricker wavelet as the source wavelet with a central frequency of $13\,\text{Hz}$. The reference frequency f_r is set to $30\,\text{Hz}$. I employ $L = 2$ layers of the Standard Linear Solid (SLS) model for the inversion. As I directly invert both the elastic parameters and the attenuation models, following the suggestions of Song et al. (2021) and Zhang et al. (2021), I incorporate total variation regularization into the objective function to mitigate cross-talk between parameters. The initial models for V_p, V_s, and ρ are shown in Figures 11 (b), (e), and (h), respectively, while the initial models for Q_p and Q_s are depicted in Figures 11 (b) and (e), respectively. The initial models are obtained by smoothing the true models using the Gaussian smoothing method. The maximum recording time is set to $1\,\text{s}$ with a time interval of $0.002\,\text{s}$. We set the maximum number of iterations to 300, and we use the Adam algorithm as the optimization method.

Stage 2

The forward modeling of the viscoelastic wavefields is performed using the theory-guided recurrent neural network, and the gradients are efficiently computed using the automatic differentiation method (Sun et al., 2020; Wang et al., 2021; Sun et al., 2021; Zhang et al., 2020, 2021). This inversion approach eliminates the need to reconstruct the back-propagation operators introduced by Fabien-Ouellet et al. (2017) when the wave equation is not self-adjoint. The inversion results for the elastic models V_p, V_s, and ρ are shown in Figures 11 (c), (f), and (i), respectively. Compared to the true models, I observe that the positions of the layers of the elastic models have been updated at the correct location. The anomaly located at the center of the elastic model has also been updated with the correct shape and value. The box anomalies of the attenuation models have also been updated with acceptable errors at the four edges of the anaomaly. Despite not committing the inverse crime, I still achieve acceptable inversion results compared to the true models.

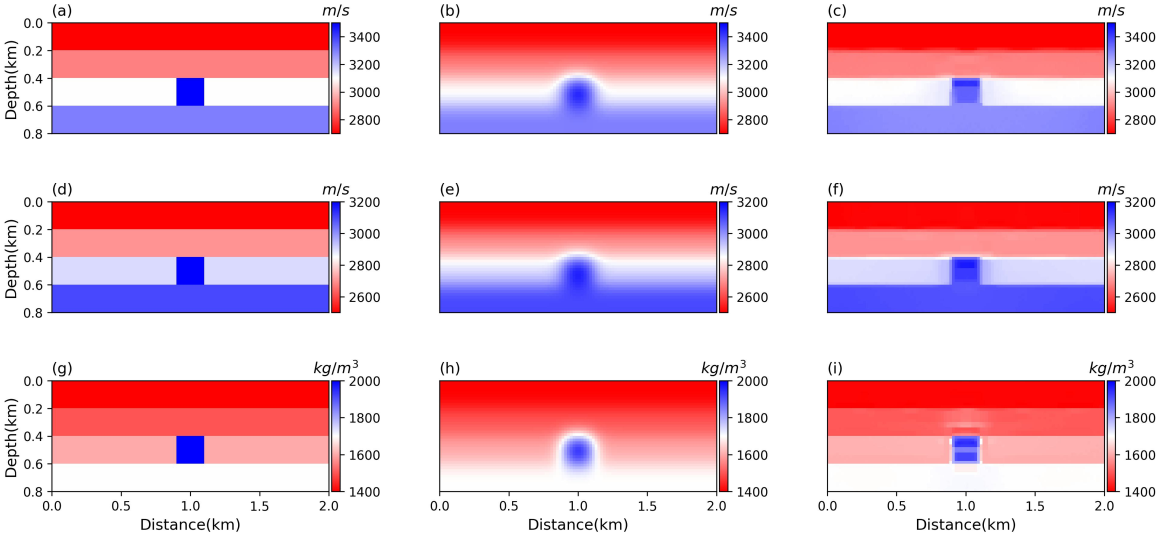

FIG. 10. Layers model V_p, V_s, and ρ viscoelastic FWI. (a), (d) and (g) are the true models for V_p, V_s, and ρ. (b), (e) and (h) are the initial models. (c), (f) and (i) are the inversion results for V_p, V_s, and ρ, respectively.

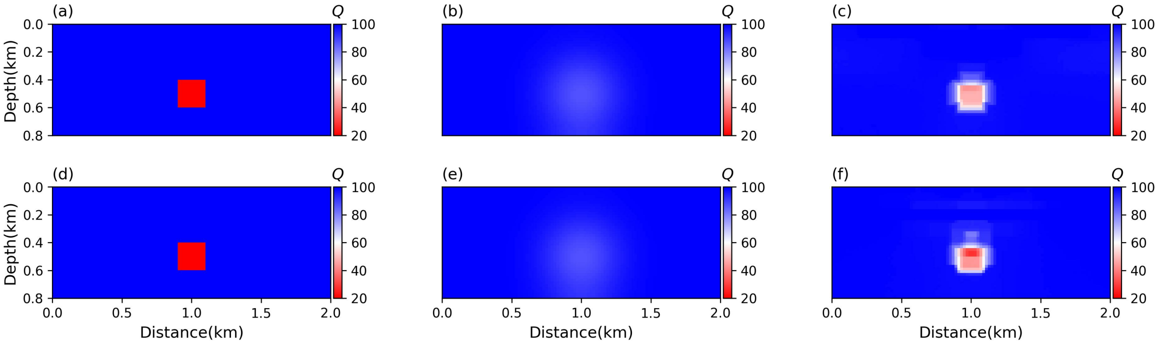

FIG. 11. Layers model Q_p, Q_s viscoelastic FWI. (a) and (d) are the true models for Q_p and Q_s, respectively. (b), (e) are the initial models for Q_p and Q_s, respectively. (c) and (f) are the inversion results for for Q_p and Q_s, respectively.

Stage 3

Next, I will quantify the modeling error, corresponding to the calculation in stage 3 of Figure 9. The MC dropout rate I am using here is 0.5, and I will calculate the loop of the modeling error evaluation in stage 3 for 100. The error evaluation loss values evaluated are plotted as the blue dots in Figure 12. The yellow line represents the FWI loss value at the last iteration. Suppose the error evaluation loss is located within the neighbourhood of FWI loss $\mathring{U}(\mathrm{FWI}_{\mathrm{loss}},\ \pm 2\%\mathrm{FWI}_{\mathrm{loss}})$. In that case, the corresponding D_{test} will be collected to the data set of the modeling error, and the D_{test} with error evaluation loss beyond this range of loss will be discarded. In Figure 12, 5 D_{test} are discarded, and the remaining 95 are taken into consideration.

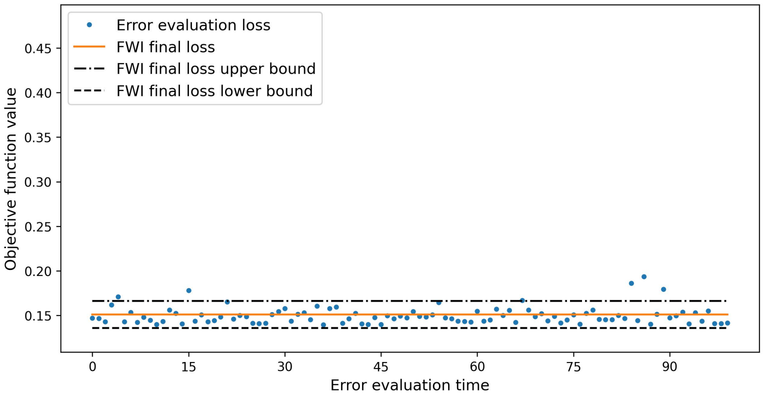

FIG. 12. The loss values of 100 times modeling error evaluation for the layers model, plotted as blue points. The yellow line represents FWI objective function value at the last iteration. If the elastic models evaluated with the modeling error detection calculation falls between the upper and lower bound of the FWI final loss, these elastic models will be considered for further consideration.

I display the observed data, synthetic data from the final iteration of FWI, evaluated modeling error, and the absolute error between the observed data and synthetic data from the final iteration of FWI for a shot located at $1000\,\mathrm{m}$ in the model in Figure 13 (a), (b), (c), and (d), respectively. The modeling error evaluation results, shown in Figure 13 (c), are obtained by calculating the standard deviation of the 95 D_{test} synthetic data set. I do not expect Figure 13 (c) to fully reflect the features of Figure 13 (d), as this is not a fair comparison. Figure 13 (d) contains all the data errors in the inversion calculation. From

Figure 13 (c), I observe that the major evaluated error is located at the direct wave and the reflection wave arriving at around $0.3\,\mathrm{s}$. This reflection error coincides with the error displayed in the absolute error in Figure 13 (d), which, to some extent, aligns with our expectation of quantifying the error caused by the insufficient ability of a single realization of the relaxation variables to model the constant Q model. The modeling error evaluation for another shot located at $1800\,\mathrm{m}$ in the model is shown in Figure 13 (e), (f), (g), and (h). I can also observe that Figure 13 (g) reflects the main features of the absolute errors in Figure 13 (h). To further observce the estimated modeling error, clearly, I plot two traces of the mean and the scaled standard deviation record calculated from the collected D_{test}, in Figure 14 which the location of the traces are marked in Figures 13(b). The standard deviation record are scaled to a factor $\lambda_{scale} = 10$ to let the change of the records caused by the variation of the relaxation variables be better observed. I can see that the variation of the relaxation variables has caused the records to change in amplitude and phase.

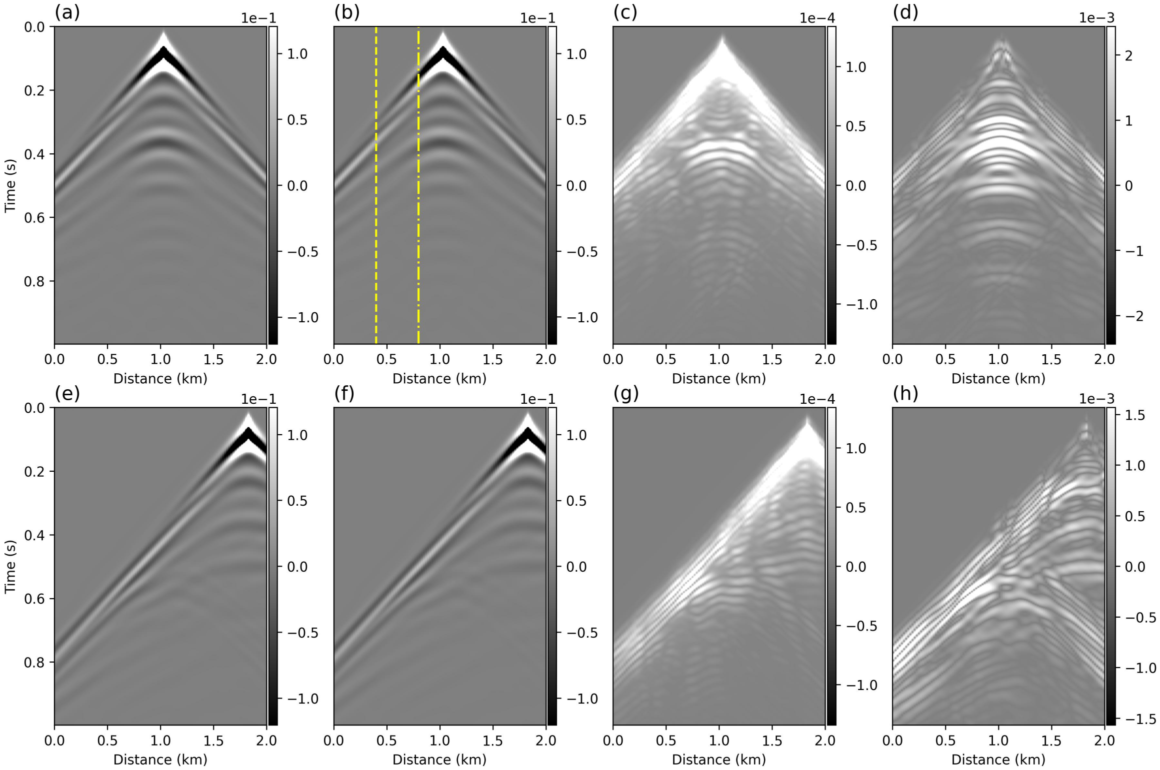

FIG. 13. Modeling error evaluation of the layers model. (a) The observed record at $1000\,\mathrm{m}$ of the model. (b) Synthetic data from the final iteration of FWI. (c) The standard deviation record of the collected D_{test}, which is regarded as the modeling error. (d) Absolute error between (a) and (b). (e) The observed record at $1800\,\mathrm{m}$ of the model. (f) Synthetic data from the final iteration of FWI. (g) The standard deviation record of the collected D_{test}. (h) Absolute error between (e) and (f).

Next, I use the collected 95 D_{test} data sets to perform one iteration of FWI updates for the elastic and attenuation models. As a result, I obtain 95 updated elastic and attenuation models. I plot the standard deviation of these updated models for V_p, V_s, ρ, Q_p, and Q_s in Figure 15, (a), (b), (c), (d), and (e), respectively. I interpret these standard deviation models as the inversion result errors caused by the modeling error. More specifically, these figures represent how the insufficient ability of the relaxation variable to quantify a constant Q model influences the FWI results. I observe no noticeable updates for the V_p, V_s, and ρ

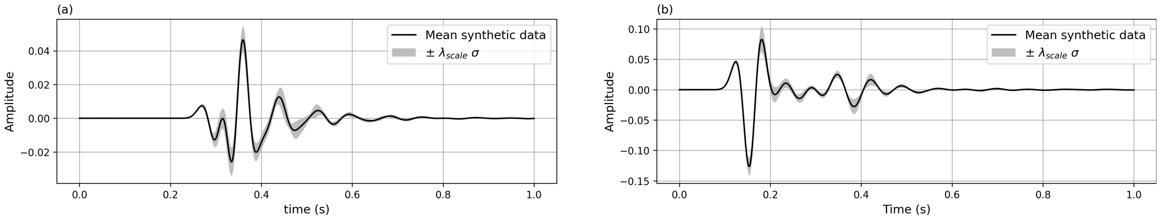

FIG. 14. Examples of the mean and scaled standard deviation of two traces calculated from the collected 95 D_{test} data sets. (a) The 400m record from Figure 13 (b). (b) The 800m record from Figure 13 (b). The scaling factor $\lambda_{\text{scale}} = 10$.

models. However, the updates for Q_p and Q_s are more easily discernible. This suggests that such an inability primarily affects the evaluation of the attenuation models. I also note that the major errors are located at the four edges of the box anomaly, which aligns well with the observations of the inversion error seen in Figure 11.

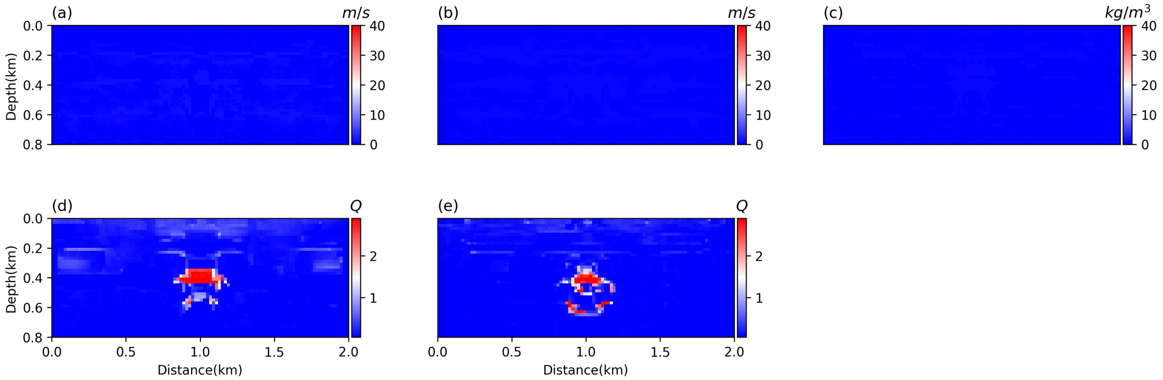

FIG. 15. Standard deviation models of 95 updated models for V_p (a), V_s (b), ρ (c), Q_p (d), and Q_s (e). These Standard deviation models represent how the insufficient ability of relaxation variables to quantify a constant Q model will influence the FWI results.

Marmousi model viscoelastic FWI

Stage 1

In this section, I calcualte viscoelastic FWI on a portion of the Marmousi model, where a reservoir is situated in the upper center of the model. The true models for V_p, V_s, and ρ are shown in Figure 16. The true attenuation models for Q_p and Q_s are depicted in Figure 17 (a) and (d), respectively. The model dimensions are 80×160, with a grid spacing of $d_x = d_z = 20\,\text{m}$. I use $L = 2$ layers of the SLS mechanism and forward propagate the relaxation variables 1000 times to obtain the mean relaxation variables for generating the observed data. The source wavelets are Ricker wavelets with a central frequency of $13\,\text{Hz}$, and the reference frequency used here is $40\,\text{Hz}$. The maximum recording time is $1.5\,\text{s}$ with a time interval of $0.002\,\text{s}$. The initial models for V_p, V_s, and ρ are plotted in Figures 16 (b), (e), and (h). The initial models for the attenuation models are plotted in figures 17 (b), and (e), respectively. The initial models are obtained by smoothing the true models with the Gaussian smoothing method.

Stage 2

As in the simple layers model inversion case, I use total variation regularization in the objective function to mitigate artifacts caused by cross-talk between parameters. I employ the Adam algorithm for optimization and set the maximum iteration number to 300. The relaxation variables during the inversion are obtained with a single forward realization of the neural network. The inversion results for the elastic models V_p, V_s, and ρ are shown in Figures 16 (c), (f), and (i), respectively. The inversion results for the attenuation models Q_p and Q_s are displayed in Figures 17 (c) and (f). Compared to the true models, I obtain promising inversion results for both the elastic and attenuation models. Acceptable model errors are located in the lower right corner of the elastic models. The inversion results for Q_s also exhibit some difficulty in accurately describing the shape of the central reservoir.

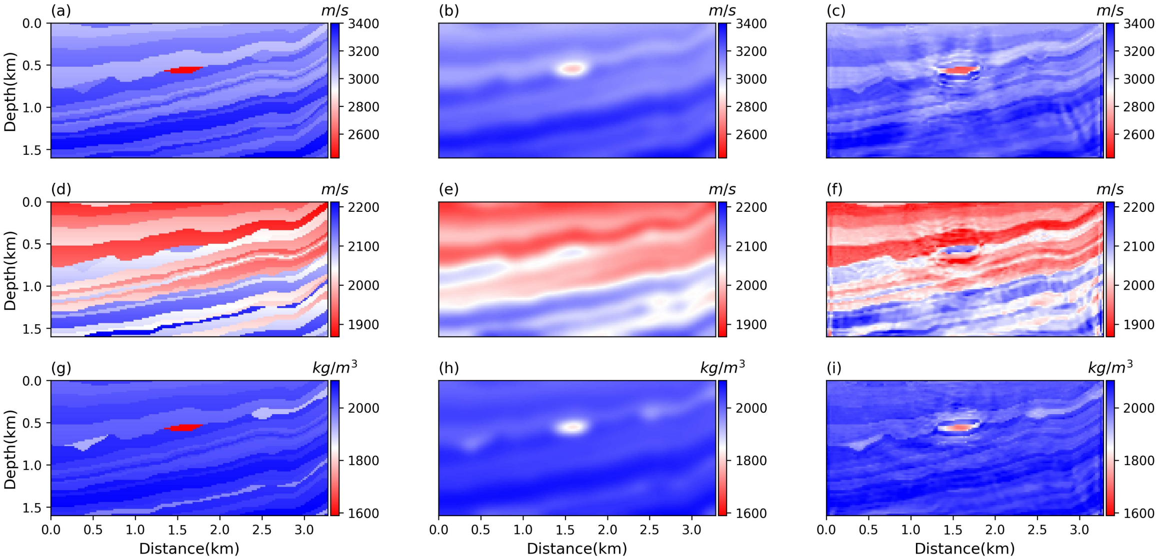

FIG. 16. Part of the Marmousi model V_p, V_s, and ρ viscoelastic FWI. (a), (d) and (g) are the true models for V_p, V_s, and ρ. (b), (e) and (h) are the initial models. (c), (f) and (i) are the inversion results for V_p, V_s, and ρ, respectively.

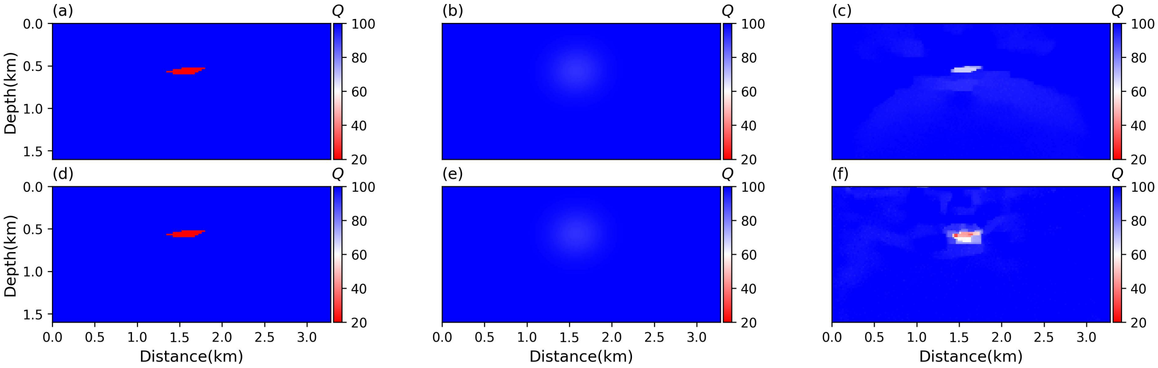

FIG. 17. Part of the Marmousi model Q_p, Q_s viscoelastic FWI. (a) and (d) are the true models for Q_p and Q_s, respectively. (b), (e) are the initial models for Q_p and Q_s, respectively. (c) and (f) are the inversion results for for Q_p and Q_s, respectively.

Stage 3

Next, I will quantify the modeling error, corresponding to the calculation in stage 3 of Figure 9. The MC dropout rate I set here is 0.43, and I will calculate the loop of the modeling error evaluation in stage 3 for 100. The error evaluation loss is plotted as the blue dots in Figure 18, and the final FWI loss is plotted as the horizontal yellow line is Figure 18. As in the simple layers model case, I set the upper and lower loss bound as the $\pm 2\%$ in the neighbourhood of the final FWI loss. I evaluate 100 relaxation variables, and 87 of them have data misfit values which are within the acceptable range. Thus, 87 D_{test} will be collected to evaluate the modeling error. The standard deviation of the model collected are regarded as the modeling error evaluated.

In Figure the first row of 19, I display the modeling error evaluated for the shot located at 1000m of the model. I plot the observed data, synthetic data at the final iteration of FWI, the standard deviation of the model error, and the absolute error between the observed data and the final synthetic data in Figure 19 (a), (b), (c), and (d), respectively. I can see that the modeling error evaluated in Figure 19(c) has reflected the main feature of the data error in Figure 19, especially for the reflection waves arived at around $0.4s$ of the records. Again, I should not expect the modeling error evaluated has the ability to capture all the features of the absolute data error, as Figure 19 (d) includes all the errors in inversion. An example of the modeling error evaluation for another shot, which is located in the 1800m in distance of the model, has been plotted in the second row of Figure 19. I can also see that the modeling error plot can also capture the main feature of the data error. To further observe how the modeling error influence the D_{test}, I plot two traces of the records in Figure 20, where the trace location is marked in Figure 19(b). In Figure 20, the black lines are the synthetic records for the last iteration of FWI, and the gray area represents the change of the range of the signal due to the change of the relaxation variables, quantified by the scaled standard deviation.

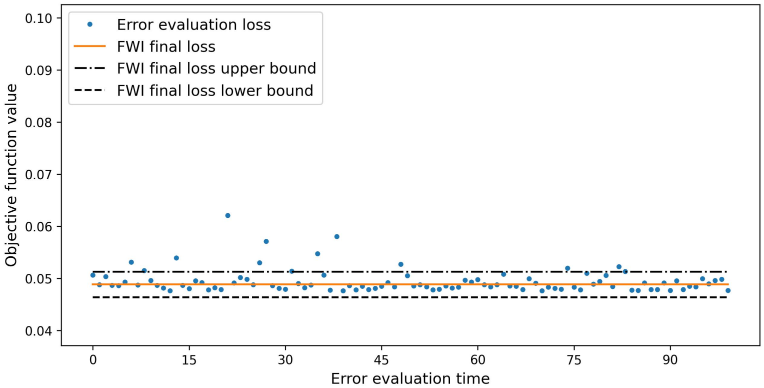

FIG. 18. The loss values of 100 times modeling error evaluation, plotted as blue points. The yellow line represents FWI objective function value at the last iteration. If the elastic models evaluated with the modeling error detection calculation falls between the upper and lower bound of the FWI final loss, these elastic models will be considered.

Next, I use each of the collected 87 D_{test} datasets to calculate a one-iteration FWI update. As a result, I obtain 87 updated models for V_p, V_s, ρ, Q_p, and Q_s. I plot the standard deviation of these updated models for V_p, V_s, ρ, Q_p, and Q_s in Figure 21 (a), (b), (c), (d), and (e), respectively. These figures represent the impact of the relaxation variable's limited

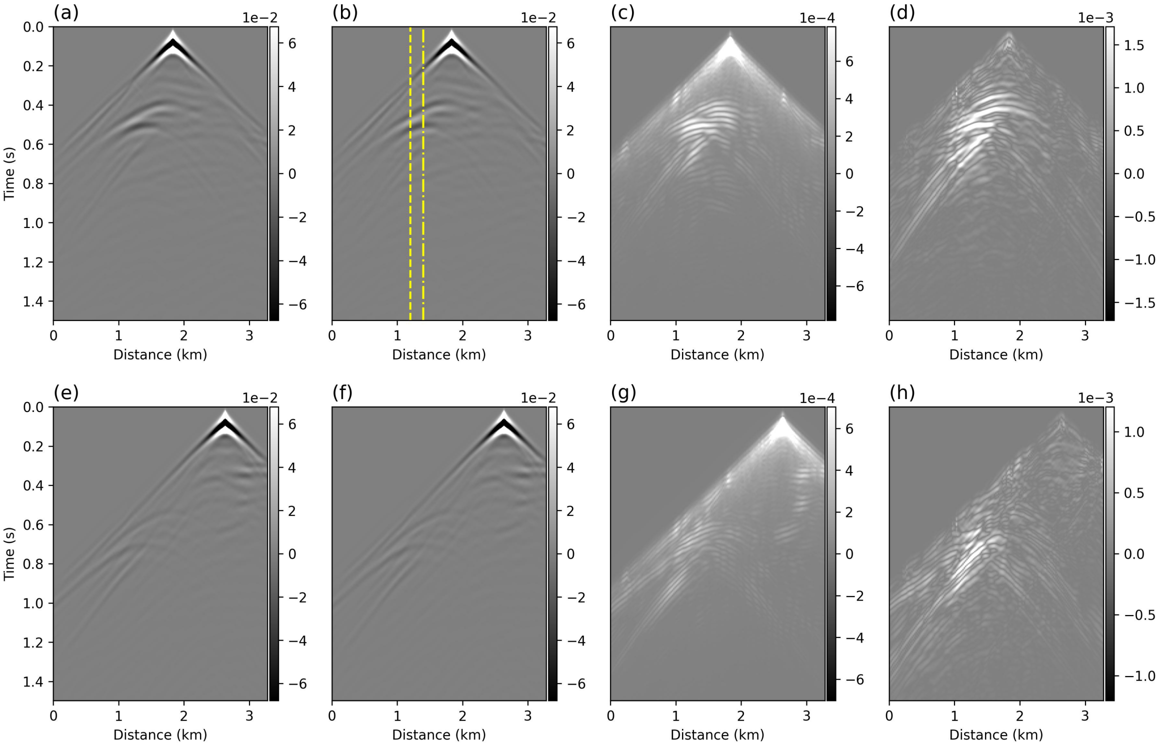

FIG. 19. Modeling error evaluation of the layers model. (a) The observed record at $1000\,$m of the model. (b) Synthetic data from the final iteration of FWI. (c) The standard deviation record of the collected D_{test}, which is regarded as the modeling error. (d) Absolute error between (a) and (b). (e) The observed record at $1800\,$m of the model. (f) Synthetic data from the final iteration of FWI. (g) The standard deviation record of the collected D_{test}. (h) Absolute error between (e) and (f).

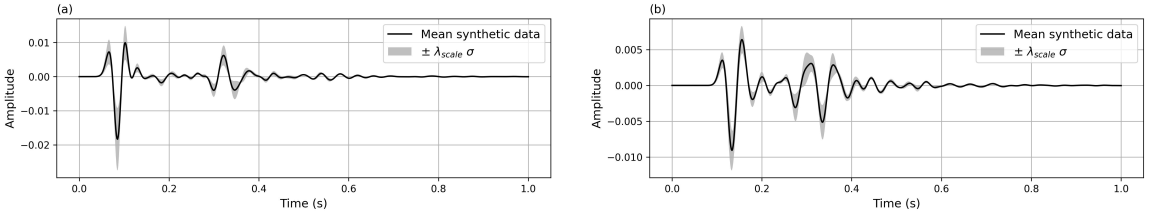

FIG. 20. Synthetic records at the last iteration of FWI and its variation caused by the modeling error evaluation. (a) The 400m record of Figure 19 (b). (b) The 800m record of Figure 19 (b). The scaling factor $\lambda_{\text{scale}} = 10$.

ability to quantify a constant Q model on the FWI results for the elastic Marmousi model section. Similar to the simple layers model case, I do not observe significant variation in the updates of the elastic model at the observable scale. In contrast, the variations in the updates for the attenuation models are readily apparent. This again suggests that such an inability primarily affects the evaluation of the attenuation models, consistent with our observations in the previous numerical test.

CONCLUSION

In this report, we train a neural network to map the attenuation factor Q to the relaxation times for different numbers of the SLS mechanism. I observe that the mean Q spectrum,

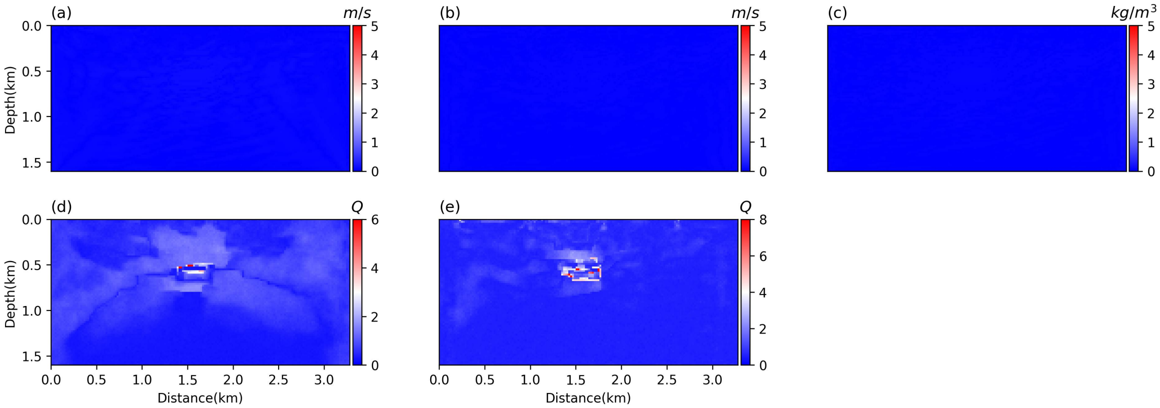

FIG. 21. Standard deviation models of 87 updated models for V_p (a), V_s (b), ρ (c), Q_p (d), and Q_s (e). These Standard deviation models represent how the inability of relaxation variables to quantify a constant Q model will influence the FWI results.

calculated using the relaxation variables generated by the well-trained neural network, can more accurately represent the constant Q model within the seismic frequency bandwidth. The well-trained network can map any Q value to its corresponding relaxation variable, enabling us to directly update the Q model when incorporating the network into the RNN-based inversion framework.

The observed data are generated using the mean relaxation variables obtained from multiple forward propagations of the well-trained neural network, providing a better approximation of the constant Q model. During the inversion, I calculate only one realization of the relaxation variables to maintain the same computational cost. This means that the forward modeling used to obtain the observed data and during the inversion calculation are different, releasing the inverse crime problem that I am facing in the synthetic inversion tests.

After the inversion, by enabling the Monte Carlo dropout method in the neural network, I can evaluate the modeling error caused by the relaxation variables' insufficient ability to model the constant Q model. The overall scale of the modeling error for the seismic records is small, but it can influence the amplitude and phase of the records. The evaluated modeling error can reflect the main features of the absolute error between the observed data and the synthetic data. This modeling error has a greater impact on the evaluation of the attenuation models than on the elastic models.

ACKNOWLEDGMENTS

The sponsors of CREWES are gratefully thanked for continued support. This work was funded by CREWES industrial sponsors, NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 543578-19.

REFERENCES

Blanch, J. O., Robertsson, J. O., and Symes, W. W., 1995, Modeling of a constant q; methodology and algorithm for an efficient and optimally inexpensive viscoelastic technique: Geophysics, **60**, No. 1, 176–184.

Day, S. M., and Minster, J. B., 1984, Numerical simulation of attenuated wavefields using a padé approximant method: Geophysical Journal International, **78**, No. 1, 105–118.

Emmerich, H., and Korn, M., 1987, Incorporation of attenuation into time-domain computations of seismic wave fields: Geophysics, **52**, No. 9, 1252–1264.

Fabien-Ouellet, G., Gloaguen, E., and Giroux, B., 2017, Time domain viscoelastic full waveform inversion: Geophysical Journal International, **209**, No. 3, 1718–1734.

Fan, N., Zhao, L.-F., Xie, X.-B., Ge, Z., and Yao, Z.-X., 2016, Two-dimensional time-domain finite-difference modeling for viscoelastic seismic wave propagation: Geophysical Journal International, **206**, No. 3, 1539–1551.

Fichtner, A., and van Driel, M., 2014, Models and fréchet kernels for frequency-(in) dependent q: Geophysical Journal International, **198**, No. 3, 1878–1889.

Liu, P., and Archuleta, R. J., 2006, Efficient modeling of q for 3d numerical simulation of wave propagation: Bulletin of the Seismological Society of America, **96**, No. 4A, 1352–1358.

O'connell, R., and Budiansky, B., 1978, Measures of dissipation in viscoelastic media: Geophysical Research Letters, **5**, No. 1, 5–8.

Robertsson, J. O., Blanch, J. O., and Symes, W. W., 1994, Viscoelastic finite-difference modeling: Geophysics, **59**, No. 9, 1444–1456.

Song, C., Alkhalifah, T., and Waheed, U. B., 2021, Solving the frequency-domain acoustic vti wave equation using physics-informed neural networks: Geophysical Journal International, **225**, No. 2, 846–859.

Sun, J., Innanen, K. A., and Huang, C., 2021, Physics-guided deep learning for seismic inversion with hybrid training and uncertainty analysis: Geophysics, **86**, No. 3, R303–R317.

Sun, J., Niu, Z., Innanen, K. A., Li, J., and Trad, D. O., 2020, A theory-guided deep-learning formulation and optimization of seismic waveform inversiontheory-guided dl and seismic inversion: Geophysics, **85**, No. 2, R87–R99.

Wang, W., McMechan, G. A., and Ma, J., 2021, Elastic isotropic and anisotropic full-waveform inversions using automatic differentiation for gradient calculations in a framework of recurrent neural networks: Geophysics, **86**, No. 6, R795–R810.

Zhang, T., Innanen, K. A., Sun, J., and Trad, D. O., 2020, Numerical analysis of a deep learning formulation of multi-parameter elastic full waveform inversion, *in* SEG International Exposition and Annual Meeting, OnePetro.

Zhang, T., Sun, J., Innanen, K. A., and Trad, D. O., 2021, A recurrent neural network for l1 anisotropic viscoelastic full-waveform inversion with high-order total variation regularization, *in* First International Meeting for Applied Geoscience & Energy, Society of Exploration Geophysicists, 1374–1378.

Zhou, H., Liu, Y., and Wang, J., 2023, K-space dispersion error compensators for the fractional spatial derivatives based constant-q viscoelastic wave equation modeling: Journal of Computational Physics, **487**, 112,161.

Zhu, T., and Carcione, J. M., 2014, Theory and modelling of constant-q p-and s-waves using fractional spatial derivatives: Geophysical Journal International, **196**, No. 3, 1787–1795.