Recurrent neural network formulation of viscoelastic VTI full waveform inversion.

Tianze Zhang, Jian Sun, Kristopher Innanen, Daniel Trad

ABSTRACT

In this study, we use the recurrent neural network (RNN) to achieve viscoelastic VTI full waveform inversion. Eight parameters are simultaneously inverted, which are elastic parameters C_{11}, C_{13}, C_{33}, C_{44} and their corresponding attenuation parameters Qc_{11}, Qc_{13}, Qc_{33}, Qc_{44}. The recurrent neural network is built according to the stress velocity VTI viscoelastic wave equation. We also study the acquisition influence on the inversion results. Numerical inversion results show that the combination of cross well data and the surface data can help to better recover the elastic parameters compared with only surface acquisition in which the receivers and shots are all on the surface of the model. To mitigate the cross-talk between the parameters, we also use the high order total variation (TV) to mitigate the cross-talk. The simple structure model and complex part of the overthrust model proves the validation of this method.

INTRODUCTION

Elasticity is usually considered as a good model for seismic forward and inversion problem. However, in real seismic data, the energy of the wave can be converted into different kinds of energy, for instance, heat, due to attenuation (Robertsson et al. (1994)). This means that pure elastic media may not be enough to demonstrate the real underground world in some cases. The quality factor Q is usually used to describe the attenuation level of the subsurface earth. The physical meaning of Q is the number of wavelengths a wave must propagate through the material before its amplitude drops by a factor $e^{-\pi}$. By using Pade approximation, Day and Minster (1984) combined the viscoelastic theory into 2D time-domain modeling methods. Emmerich and Korn (1987) introduced a generalized standard linear solid (GSLS) model to approximate the viscoelastic earth model. Robertsson et al. (1994) developed the staggered grid finite difference method for viscoelastic modeling (Bohlen (2002)). The modeling method we use in this study is the extension of Robertson's method.

Full waveform inversion (FWI) can be considered as a powerful method based on data fitting to invert velocity models. Based on perturbation theory and Born approximation the conventional FWI uses the adjoint state method, which is the zero-lag correlation between forward propagation wavefields and backpropagation wavefields, to calculate the gradients for each parameter. The Hessian matrix which is the second-order derivative of the objective function with respect to the parameters is considered as a way to mitigate cross-talk between the parameters. The attenuation effect can have an important effect in full waveform inversion. Groos et al. (2012) studied the effect of attenuation in shallow seismic surface waves for full waveform inversion. Belahi et al. (2015), by comparing the elastic full waveform inversion and viscoelastic full waveform inversion, stressed the need to add a proper amount of attenuation when inversion long offset seismic data. Belahi et al. (2016) found that inverting for all parameters together is necessary to get access to the short-wavelength features of the subsurface model because the short-wavelength attenuation model is required to properly treat reflections and converted waves close to the critical angle. Fabien-Ouellet et al. (2016) reformulated the wave equations for the viscoelastic wave equation and derive the adjoint state equations for viscoelastic media and then perform the viscoelastic full waveform inversion is actual field data. Fabien-Ouellet et al. (2017) uses parallel High-Performance Computers to calculate viscoelastic full waveform inversion in the time domain. Yang et al. (2016) studied 3-D viscoelastic full waveform inversion in viscoelastic media using generalized Maxwell/Zener body including an arbitrary number of attenuation mechanisms.

In this study, we formulate VTI viscoelastic recurrent neural network for full waveform inversion, which forms the theory-guided neural network. The network is built with several RNN cells and each RNN cell is designed according to the VTI viscoelastic stress velocity wave equation. In this article, we first introduce the based theory about the viscoelastic VTI wave equation. Second, we introduce how the RNN network is formulated based on this equation. Third, we introduced the regulation terms for the misfits to mitigate the cross-talk noise. Fourth, we use two numerical tests to prove the validation of our method.

THEORY

The constitutive relationship for a vicoelastic media can be expressed as:

$$\sigma_{ij} = C_{ijkl} * \dot{\epsilon}_{kl} = \dot{C}_{ijkl} * \epsilon_{kl}, \tag{1}$$

where the $\dot{C}_{ijkl}$ is a tensor called the relaxation stiffness matrix parameters. $*$ means the time convolution. The dot means the time differential. σ_{ij} means the stress and ϵ_{kl} demonstrate the strain. In a GSLS framework, the relaxation function could be described as :

$$C(t) = C\left(1 - \sum_{l=1}^{L}\left(1 - \frac{\tau_{\varepsilon l}^{C}}{\tau_{\sigma l}^{C}}\right)e^{-t/\tau_{\sigma l}^{C}}\right)\theta(t), \tag{2}$$

where C is the elastic modulus, L is the number of layers for viscoelastic model. $\tau_{\sigma l}^{C}$ and $\tau_{\varepsilon l}^{C}$ are the relaxation time for stress and strain. The time differential of C is :

$$\dot{C} = C\left(\frac{1}{\tau_{\sigma l}^{C}}\sum_{\ell=1}^{L}\left(1 - \frac{\tau_{e\ell}^{C}}{\tau_{\sigma\ell}^{C}}\right)\mathrm{e}^{-t/\tau_{\sigma l}^{C}}\right)\theta(t) + C\left(1 - \sum_{\ell=1}^{L}\left(1 - \frac{\tau_{e\ell}^{C}}{\tau_{\sigma\ell}^{C}}\right)\mathrm{e}^{-t/\tau_{\sigma l}^{C}}\right)\delta(t), \tag{3}$$

Thus for each stiffness parameters we have:

$$\dot{C_{11}} = C_{11}\left(\frac{1}{\tau_{\sigma l}^{C_{11}}}\sum_{\ell=1}^{L}\left(1 - \frac{\tau_{e\ell}^{C_{11}}}{\tau_{\sigma\ell}^{C_{11}}}\right)\mathrm{e}^{-t/\tau_{\sigma l}^{C_{11}}}\right)\theta(t) + C_{11}\left(1 - \sum_{\ell=1}^{L}\left(1 - \frac{\tau_{e\ell}^{C_{11}}}{\tau_{\sigma\ell}^{C_{11}}}\right)\mathrm{e}^{-t/\tau_{\sigma l}^{C_{11}}}\right)\delta(t), \tag{4}$$

$$\dot{C_{13}} = C_{13}\left(\frac{1}{\tau_{\sigma l}^{C_{13}}}\sum_{\ell=1}^{L}\left(1 - \frac{\tau_{e\ell}^{C_{13}}}{\tau_{\sigma\ell}^{C_{13}}}\right)\mathrm{e}^{-t/\tau_{\sigma l}^{C_{13}}}\right)\theta(t) + C_{13}\left(1 - \sum_{\ell=1}^{L}\left(1 - \frac{\tau_{e\ell}^{C_{13}}}{\tau_{\sigma\ell}^{C_{13}}}\right)\mathrm{e}^{-t/\tau_{\sigma l}^{C_{13}}}\right)\delta(t), \tag{5}$$

$$\dot{C}_{33} = C_{33}\left(\frac{1}{\tau_{\sigma l}^{C_{33}}}\sum_{\ell=1}^{L}\left(1-\frac{\tau_{e\ell}^{C_{33}}}{\tau_{\sigma\ell}^{C_{33}}}\right)\mathrm{e}^{-t/\tau_{\sigma l}^{C_{33}}}\right)\theta(t) + C_{33}\left(1-\sum_{\ell=1}^{L}\left(1-\frac{\tau_{e\ell}^{C_{33}}}{\tau_{\sigma\ell}^{C_{33}}}\right)\mathrm{e}^{-t/\tau_{\sigma l}^{C_{33}}}\right)\delta(t), \tag{6}$$

$$\dot{C}_{44} = C_{44}\left(\frac{1}{\tau_{\sigma l}^{C_{44}}}\sum_{\ell=1}^{L}\left(1-\frac{\tau_{e\ell}^{C_{44}}}{\tau_{\sigma\ell}^{C_{44}}}\right)\mathrm{e}^{-t/\tau_{\sigma l}^{C_{44}}}\right)\theta(t) + C_{44}\left(1-\sum_{\ell=1}^{L}\left(1-\frac{\tau_{e\ell}^{C_{44}}}{\tau_{\sigma\ell}^{C_{44}}}\right)\mathrm{e}^{-t/\tau_{\sigma l}^{C_{44}}}\right)\delta(t), \tag{7}$$

For viscoelastic VTI media the stiffness matrix is:

$$\mathbf{C_{ANVTI}} = \begin{bmatrix} \dot{C}_{11} & \dot{C}_{13} & 0 \\ \dot{C}_{13} & \dot{C}_{33} & 0 \\ 0 & 0 & \dot{C}_{44} \end{bmatrix}, \tag{8}$$

Thus the stress velocity relationship between the stress and strain in the viscoelastic media can be expressed as:

$$\begin{bmatrix} \partial_t\sigma_{xx} \\ \partial_t\sigma_{zz} \\ \partial_t\sigma_{xz} \end{bmatrix} = \begin{bmatrix} \dot{C}_{11} & \dot{C}_{13} & 0 \\ \dot{C}_{13} & \dot{C}_{33} & 0 \\ 0 & 0 & \dot{C}_{44} \end{bmatrix} * \begin{bmatrix} \partial_x v_x \\ \partial_z v_z \\ \partial_x v_z + \partial_z v_x \end{bmatrix}, \tag{9}$$

where * represents the time convolution. Now, we substitute equation (3) to (7) into equation (9). Thus we have the following formula:

$$\begin{aligned} \partial_t\sigma_{xx} =& \dot{C}_{11}\partial_x v_x + \dot{C}_{13}\partial_z v_z \\ =& \left[C_{11}\left(\frac{1}{\tau_{\sigma l}^{C_{11}}}\sum_{\ell=1}^{L}\left(1-\frac{\tau_{e\ell}^{C_{11}}}{\tau_{\sigma\ell}^{C_{11}}}\right)\mathrm{e}^{-t/\tau_{\sigma l}^{C_{11}}}\right)\theta(t)\right] * \partial_x v_x + \\ & \left[C_{11}\left(1-\sum_{\ell=1}^{L}\left(1-\frac{\tau_{e\ell}^{C_{11}}}{\tau_{\sigma\ell}^{C_{11}}}\right)\mathrm{e}^{-t/\tau_{\sigma l}^{C_{11}}}\right)\delta(t)\right] * \partial_x v_x + \\ & \left[C_{13}\left(\frac{1}{\tau_{\sigma l}^{C_{13}}}\sum_{\ell=1}^{L}\left(1-\frac{\tau_{e\ell}^{C_{13}}}{\tau_{\sigma\ell}^{C_{13}}}\right)\mathrm{e}^{-t/\tau_{\sigma l}^{C_{13}}}\right)\theta(t)\right] * \partial_z v_z + \\ & \left[C_{13}\left(1-\sum_{\ell=1}^{L}\left(1-\frac{\tau_{e\ell}^{C_{13}}}{\tau_{\sigma\ell}^{C_{13}}}\right)\mathrm{e}^{-t/\tau_{\sigma l}^{C_{13}}}\right)\delta(t)\right] * \partial_z v_z, \end{aligned} \tag{10}$$

$$\begin{aligned}\partial_t\sigma_{xx} =&\dot{C}_{11}\partial_x v_x + \dot{C}_{13}\partial_z v_z\\ =&\left[C_{11}\left(\frac{1}{\tau_{\sigma l}^{C_{11}}}\sum_{\ell=1}^{L}\left(1-\frac{\tau_{e\ell}^{C_{11}}}{\tau_{\sigma\ell}^{C_{11}}}\right)\mathrm{e}^{-t/\tau_{\sigma l}^{C_{11}}}\right)\theta(t)\right]*\partial_x v_x+\\ &\left[C_{11}\left(1-\sum_{\ell=1}^{L}\left(1-\frac{\tau_{e\ell}^{C_{11}}}{\tau_{\sigma\ell}^{C_{11}}}\right)\right)\right]\times\partial_x v_x+\\ &\left[C_{13}\left(\frac{1}{\tau_{\sigma l}^{C_{13}}}\sum_{\ell=1}^{L}\left(1-\frac{\tau_{e\ell}^{C_{13}}}{\tau_{\sigma\ell}^{C_{13}}}\right)\mathrm{e}^{-t/\tau_{\sigma l}^{C_{13}}}\right)\theta(t)\right]*\partial_z v_z+\\ &\left[C_{13}\left(1-\sum_{\ell=1}^{L}\left(1-\frac{\tau_{e\ell}^{C_{13}}}{\tau_{\sigma\ell}^{C_{13}}}\right)\right)\right]\times\partial_z v_z,\end{aligned}\tag{11}$$

Here we note that $R_{xx}^{C_{11}}$ and $R_{zz}^{C_{13}}$ are the relaxation fields, and the detail expression for these wave fields are:

$$R_{xx}^{C_{11}} = \left[C_{11}\left(\frac{1}{\tau_{\sigma l}^{C_{11}}}\sum_{\ell=1}^{L}\left(1-\frac{\tau_{e\ell}^{C_{11}}}{\tau_{\sigma\ell}^{C_{11}}}\right)\mathrm{e}^{-t/\tau_{\sigma l}^{C_{11}}}\right)\theta(t)\right]*\partial_x v_x,\tag{12}$$

$$R_{zz}^{C_{13}} = \left[C_{13}\left(\frac{1}{\tau_{\sigma l}^{C_{13}}}\sum_{\ell=1}^{L}\left(1-\frac{\tau_{e\ell}^{C_{13}}}{\tau_{\sigma\ell}^{C_{13}}}\right)\mathrm{e}^{-t/\tau_{\sigma l}^{C_{13}}}\right)\theta(t)\right]*\partial_z v_z,\tag{13}$$

With equation (12) and (13) σ_{xx} can be written as:

$$\begin{aligned}\partial_t\sigma_{xx} = \dot{C}_{11}\partial_x v_x + \dot{C}_{13}\partial_z v_z =&\\ R_{xx}^{C_{11}} + \left[C_{11}\left(1-\sum_{\ell=1}^{L}\left(1-\frac{\tau_{e\ell}^{C_{11}}}{\tau_{\sigma\ell}^{C_{11}}}\right)\right)\right]&\times\partial_x v_x+\\ R_{zz}^{C_{13}} + \left[C_{13}\left(1-\sum_{\ell=1}^{L}\left(1-\frac{\tau_{e\ell}^{C_{13}}}{\tau_{\sigma\ell}^{C_{13}}}\right)\right)\right]&\times\partial_z v_z,\end{aligned}\tag{14}$$

When we choose $l = 1$ and take the partial derivative of the relaxation fields with respect to time, we have :

$$\begin{aligned}\partial_t R_{xx}^{C_{11}} =&(\partial_x v_x) * \left[\pi\frac{-1}{{\tau_{\sigma l}^{C_{11}}}^2}\left(1-\frac{\tau_{\varepsilon l}^{C_{11}}}{\tau_{\sigma l}^{C_{11}}}\right)e^{-t/\tau_{\sigma l}^{C_{11}}}\theta(t)+\pi\frac{1}{\tau_{\sigma l}^{C_{11}}}\left(1-\frac{\tau_{\varepsilon l}^{C_{11}}}{\tau_{\sigma l}^{P}}\right)e^{-t/\tau_{\sigma l}^{C_{11}}}\delta(t)\right]\\ =&-\frac{1}{\tau_{\sigma l}^{C_{11}}}R_{xx}^{C_{11}}-\frac{1}{\tau_{\sigma l}^{C_{11}}}C_{11}\left(\frac{\tau_{\varepsilon l}^{C_{11}}}{\tau_{\sigma l}^{C_{11}}}-1\right)\partial_x v_x,\end{aligned}\tag{15}$$

$$\begin{aligned}\partial_t R_{zz}^{C_{13}} =& (\partial_z v_z) * \left[\pi \frac{-1}{{\tau_{\sigma l}^{C_{13}}}^2}\left(1-\frac{\tau_{\varepsilon l}^{C_{13}}}{\tau_{\sigma l}^{C_{13}}}\right) e^{-t/\tau_{\sigma l}^{C_{13}}}\theta(t) + \pi\frac{1}{\tau_{\sigma l}^{C_{13}}}\left(1-\frac{\tau_{\varepsilon l}^{C_{13}}}{\tau_{\sigma l}^{P}}\right) e^{-t/\tau_{\sigma l}^{C_{13}}}\delta(t)\right] \\ =& -\frac{1}{\tau_{\sigma l}^{C_{13}}} R_{xx}^{C_{13}} - \frac{1}{\tau_{\sigma l}^{C_{13}}} C_{13}\left(\frac{\tau_{\varepsilon l}^{C_{13}}}{\tau_{\sigma l}^{C_{13}}}-1\right)\partial_z v_z,\end{aligned} \tag{16}$$

Thus when we are using one relaxation scheme. The stress in viscoelastic VTI media can be expressed as:

$$\begin{aligned}\partial_t \sigma_{xx} =& \dot{C}_{11}\partial_x v_x + \dot{C}_{13}\partial_z v_z = \\ & C_{11}R_{xx}^{C_{11}} + \left[C_{11}\left(\frac{\tau_{e\ell}^{C_{11}}}{\tau_{\sigma\ell}^{C_{11}}}\right)\right]\partial_x v_x + \quad C_{13}R_{zz}^{C_{13}} + \left[C_{13}\left(\frac{\tau_{e\ell}^{C_{13}}}{\tau_{\sigma\ell}^{C_{13}}}\right)\right]\partial_z v_z,\end{aligned} \tag{17}$$

where $R_{xx}^{C_{11}}$ and $R_{zz}^{C_{13}}$ can be expressed as:

$$\partial_t R_{xx}^{C_{11}} = -\frac{1}{\tau_{\sigma l}^{C_{11}}} R_{xx}^{C_{11}} - \frac{1}{\tau_{\sigma l}^{C_{11}}} C_{11}\left(\frac{\tau_{\varepsilon l}^{C_{11}}}{\tau_{\sigma l}^{C_{11}}}-1\right)\partial_x v_x \tag{18}$$

$$\partial_t R_{zz}^{C_{13}} = -\frac{1}{\tau_{\sigma l}^{C_{13}}} R_{zz}^{C_{13}} - \frac{1}{\tau_{\sigma l}^{C_{13}}} C_{13}\left(\frac{\tau_{\varepsilon l}^{C_{13}}}{\tau_{\sigma l}^{C_{13}}}-1\right)\partial_z v_z \tag{19}$$

$$\begin{aligned}\partial_t \sigma_{zz} =& \dot{C}_{13}\partial_x v_x + \dot{C}_{33}\partial_z v_z = \\ & C_{13}R_{xx}^{C_{13}} + \left[C_{13}\left(\frac{\tau_{e\ell}^{C_{13}}}{\tau_{\sigma\ell}^{C_{13}}}\right)\right]\partial_x v_x + \quad C_{33}R_{zz}^{C_{33}} + \left[C_{33}\left(\frac{\tau_{e\ell}^{C_{33}}}{\tau_{\sigma\ell}^{C_{33}}}\right)\right]\partial_z v_z,\end{aligned} \tag{20}$$

where $R_{xx}^{C_{13}}$ and $R_{zz}^{C_{33}}$ can be expressed as :

$$\partial_t R_{xx}^{C_{13}} = -\frac{1}{\tau_{\sigma l}^{C_{13}}} R_{xx}^{C_{13}} - \frac{1}{\tau_{\sigma l}^{C_{13}}} C_{13}\left(\frac{\tau_{\varepsilon l}^{C_{13}}}{\tau_{\sigma l}^{C_{13}}}-1\right)\partial_x v_x \tag{21}$$

$$\partial_t R_{zz}^{C_{33}} = -\frac{1}{\tau_{\sigma l}^{C_{33}}} R_{zz}^{C_{33}} - \frac{1}{\tau_{\sigma l}^{C_{33}}} C_{33}\left(\frac{\tau_{\varepsilon l}^{C_{33}}}{\tau_{\sigma l}^{C_{33}}}-1\right)\partial_z v_z \tag{22}$$

$$\begin{aligned}\partial_t \sigma_{xz} =& \dot{C}_{44}\left(\partial_z v_x + \partial_x v_z\right) = \\ & C_{44}R_{zz}^{C_{44}} + \left[C_{44}\left(\frac{\tau_{e\ell}^{C_{44}}}{\tau_{\sigma\ell}^{C_{44}}}\right)\right]\left(\partial_z v_x + \partial_x v_z\right)\end{aligned} \tag{23}$$

where $R_{xx}^{C_{44}}$ can be expressed as :

$$\partial_t R_{xx}^{C_{44}} = -\frac{1}{\tau_{\sigma l}^{C_{44}}} R_{xx}^{C_{44}} - \frac{1}{\tau_{\sigma l}^{C_{44}}} C_{44}\left(\frac{\tau_{\varepsilon l}^{C_{44}}}{\tau_{\sigma l}^{C_{44}}}-1\right)\left(\partial_z v_x + \partial_x v_z\right) \tag{24}$$

Algorithm 1 Sequence of calculations in the viscoelastic VTI RNN cell

```
Input: Source: $s_x$, $s_z$; Space partial derivative convolution kernel. $\mathbf{k}_{x_2}$,$\mathbf{k}_{z_1}$,$\mathbf{k}_{x_1}$,$\mathbf{k}_{z_2}$ time step: $dt$. Stiffness parameters: $C_{11}$,$C_{13}$,$C_{33}$,$C_{44}$,
Output: Update velocity field at $t+\frac{1}{2}$ and stress fields at $t+1$
$\sigma_{xx}^t \leftarrow \sigma_{xx}^t + s_x$
$\sigma_{zz}^t \leftarrow \sigma_{zz}^t + s_z$
$\partial_x\sigma_{xx}^t \leftarrow (\sigma_{xx}^t * \mathbf{k}_{x_1})/\rho$
$\partial_z\sigma_{xz}^t \leftarrow (\sigma_{xz}^t * \mathbf{k}_{z_2})/\rho$
$\partial_x\sigma_{xz}^t \leftarrow (\sigma_{xz}^t * \mathbf{k}_{x_2})/\rho$
$\partial_z\sigma_{zz}^t \leftarrow (\sigma_{zz}^t * \mathbf{k}_{z_1})/\rho$
$v_x^{t+\frac{1}{2}} \leftarrow v_x^{t-\frac{1}{2}} + dt(\partial_x\sigma_{xx}^t) + dt(\partial_z\sigma_{xz}^t)$
$v_z^{t+\frac{1}{2}} \leftarrow v_z^{t-\frac{1}{2}} + dt(\partial_z\sigma_{zz}^t) + dt(\partial_x\sigma_{xz}^t)$
$\partial_x v_x^{t+\frac{1}{2}} \leftarrow v_x^{t+\frac{1}{2}} * \mathbf{k}_{x_2}$
$\partial_z v_x^{t+\frac{1}{2}} \leftarrow v_x^{t+\frac{1}{2}} * \mathbf{k}_{z_1}$
$\partial_x v_z^{t+\frac{1}{2}} \leftarrow v_z^{t+\frac{1}{2}} * \mathbf{k}_{x_1}$
$\partial_z v_z^{t+\frac{1}{2}} \leftarrow v_z^{t+\frac{1}{2}} * \mathbf{k}_{z_2}$
$\sigma_{zz}^{t+1} = \sigma_{zz}^t + dt\left\{C_{13}R_{xx}^{C_{13}} + \left[C_{13}\left(\frac{\tau_{e\ell}^{C_{13}}}{\tau_{\sigma\ell}^{C_{13}}}\right)\right]\partial_x v_x + C_{33}R_{zz}^{C_{33}} + \left[C_{33}\left(\frac{\tau_{e\ell}^{C_{33}}}{\tau_{\sigma\ell}^{C_{33}}}\right)\right]\partial_z v_z\right\}$
$\sigma_{xx}^{t+1} = \sigma_{xx}^t + dt\left\{C_{11}R_{xx}^{C_{11}} + \left[C_{11}\left(\frac{\tau_{e\ell}^{C_{11}}}{\tau_{\sigma\ell}^{C_{11}}}\right)\right]\partial_x v_x + C_{13}R_{zz}^{C_{13}} + \left[C_{13}\left(\frac{\tau_{e\ell}^{C_{13}}}{\tau_{\sigma\ell}^{C_{13}}}\right)\right]\partial_z v_z\right\}$

$\sigma_{xz}^{t+1} = \sigma_{xz}^t + dt\left\{C_{44}R_{zz}^{C_{44}} + \left[C_{44}\left(\frac{\tau_{e\ell}^{C_{44}}}{\tau_{\sigma\ell}^{C_{44}}}\right)\right](\partial_z v_x + \partial_x v_z)\right\}$
$\partial_t R_{xx}^{C_{11}} = -\frac{1}{\tau_{\sigma l}^{C_{11}}}R_{xx}^{C_{11}} - \frac{1}{\tau_{\sigma l}^{C_{11}}}C_{11}\left(\frac{\tau_{\varepsilon l}^{C_{11}}}{\tau_{\sigma l}^{C_{11}}} - 1\right)\partial_x v_x$
$\partial_t R_{xx}^{C_{13}} = -\frac{1}{\tau_{\sigma l}^{C_{13}}}R_{xx}^{C_{13}} - \frac{1}{\tau_{\sigma l}^{C_{13}}}C_{13}\left(\frac{\tau_{\varepsilon l}^{C_{13}}}{\tau_{\sigma l}^{C_{13}}} - 1\right)\partial_x v_x$
$\partial_t R_{zz}^{C_{13}} = -\frac{1}{\tau_{\sigma l}^{C_{13}}}R_{xx}^{C_{13}} - \frac{1}{\tau_{\sigma l}^{C_{13}}}C_{13}\left(\frac{\tau_{\varepsilon l}^{C_{13}}}{\tau_{\sigma l}^{C_{13}}} - 1\right)\partial_z v_z$
$\partial_t R_{zz}^{C_{33}} = -\frac{1}{\tau_{\sigma l}^{C_{33}}}R_{xx}^{C_{33}} - \frac{1}{\tau_{\sigma l}^{C_{33}}}C_{33}\left(\frac{\tau_{\varepsilon l}^{C_{33}}}{\tau_{\sigma l}^{C_{33}}} - 1\right)\partial_z v_z$
$\partial_t R_{xz}^{C_{44}} = -\frac{1}{\tau_{\sigma l}^{C_{44}}}R_{xx}^{C_{44}} - \frac{1}{\tau_{\sigma l}^{C_{44}}}C_{44}\left(\frac{\tau_{\varepsilon l}^{C_{44}}}{\tau_{\sigma l}^{C_{44}}} - 1\right)(\partial_z v_x + \partial_x v_z)$
```

Algorithm 1 shows the basic structure of the recurrent neural network. At each time step the discrete sources s_x and s_z act as inputs; the velocity and stress information, $v_x^{t-\frac{1}{2}}$, $v_z^{t-\frac{1}{2}}$, σ_{xx}^t, σ_{zz}^t, and σ_{xz}^t, is communicated between the RNN cells ; the partial derivative fields, $\partial_x\sigma_{xx}^t$, $\partial_z\sigma_{zz}^t$,$\partial_x\sigma_{xz}^t$, $\partial_z\sigma_{xz}^t$, $\partial_x v_x^{t+\frac{1}{2}}$, $\partial_z v_x^{t+\frac{1}{2}}$,$\partial_x v_z^{t+\frac{1}{2}}$, $\partial_z v_z^{t+\frac{1}{2}}$ are the internal variables in each RNN cell. Stiffness parameters: C_{11},C_{13},C_{33},C_{44}, are the stiffness matrix parameters we are going to invert for the inversion and $\tau_{\sigma\ell}^{C_{11}}$,$\tau_{\epsilon\ell}^{C_{11}}$ are the stress and strain relaxation time for parameter C_{11}. $\tau_{\sigma\ell}^{C_{13}}$,$\tau_{\epsilon\ell}^{C_{13}}$ are the stress and strain relaxation time for parameter C_{13}.

$\tau_{\sigma\ell}^{C_{44}}$,$\tau_{\epsilon\ell}^{C_{33}}$ are the stress and strain relaxation time for parameter C_{33}. $\tau_{\sigma\ell}^{C_{44}}$,$\tau_{\epsilon\ell}^{C_{44}}$ are the stress and strain relaxation time for parameter C_{44}. We find that this image convolution operator is also capable of calculating space partial derivatives if the convolution kernel is designed according to the finite difference coefficients. dx, dz are the grid intervals, and the image convolution kernels are: $\mathbf{k}_{x_1} = \mathbf{a}/dx$, $\mathbf{k}_{x_2} = \mathbf{b}/dx$, $\mathbf{k}_{z_1} = \mathbf{a}^T/dz$, and $\mathbf{k}_{z_2} = \mathbf{b}^T/dz$, where $\mathbf{a} = [0, 1/24, -9/8, 9/8, -1/24]$ and $\mathbf{b} = [1/24, -9/8, 9/8, -1/24, 0]$. $\mathbf{a}$ and $\mathbf{b}$ are $1{\times}5$ dimension arrays. $\mathbf{k}_{x_1}$ and $\mathbf{k}_{x_2}$ are kernels, for the image convolution process, responsible for calculating the staggered grid space partial derivative in x direction. $\mathbf{k}_{z_1}$ and $\mathbf{k}_{z_2}$ are kernels, for the image convolution process, responsible for calculating the staggered grid space partial derivative in z direction, and that is also why the arrays, $\mathbf{a}$ and $\mathbf{b}$, are transposed in $\mathbf{k}_{z_1}$ and $\mathbf{k}_{z_2}$. Space partial derivative calculated in this way is, mathematically, the same with conventional staggered grid method.

INVERSION WITH TOTAL VARIATION REGULARIZATION

Here we fist introduce the elastic RNN misfits based on l_2 norm with high order TV regularization:

$$\begin{aligned}
\Phi_{l2}^{TV}(\mathbf{C}_{11}, \mathbf{C}_{13}, \mathbf{C}_{33}, \mathbf{C}_{44}, \mathbf{Qc}_{11}, \mathbf{Qc}_{13}, \mathbf{Qc}_{33}, \mathbf{Qc}_{44},) = \\
\frac{1}{2}\|\mathbf{D}_{syn}(\mathbf{C}_{11}, \mathbf{C}_{13}, \mathbf{C}_{33}, \mathbf{C}_{44}, \mathbf{Qc}_{11}, \mathbf{Qc}_{13}, \mathbf{Qc}_{33}, \mathbf{Qc}_{44}) - \mathbf{D}_{obs}\|_2^2 + \\
\alpha_1^{c11}\Theta_{TV}(\mathbf{C}_{11}) + \alpha_1^{c13}\Theta_{TV}(\mathbf{C}_{13}) + \alpha_1^{c33}\Theta_{TV}(\mathbf{C}_{33}) + \alpha_1^{c44}\Theta_{TV}(\mathbf{C}_{44}) + \\
\alpha_1^{Qc11}\Theta_{TV}(\mathbf{Qc}_{11}) + \alpha_1^{Qc13}\Theta_{TV}(\mathbf{Qc}_{13}) + \alpha_1^{Qc33}\Theta_{TV}(\mathbf{Qc}_{33}) + \alpha_1^{Qc44}\Theta_{TV}(\mathbf{Qc}_{44}) \\
\alpha_2^{c11}\Upsilon_{TV}(\mathbf{C}_{11}) + \alpha_2^{c13}\Upsilon_{TV}(\mathbf{C}_{13}) + \alpha_2^{c33}\Upsilon_{TV}(\mathbf{C}_{33}) + \alpha_2^{c44}\Upsilon_{TV}(\mathbf{C}_{44}) + \\
\alpha_2^{Qc11}\Upsilon_{TV}(\mathbf{Qc}_{11}) + \alpha_2^{Qc13}\Upsilon_{TV}(\mathbf{Qc}_{13}) + \alpha_2^{Qc33}\Upsilon_{TV}(\mathbf{Qc}_{33}) + \alpha_2^{Qc44}\Upsilon_{TV}(\mathbf{QC}_{44}),
\end{aligned} \tag{25}$$

where α_1^{c11}, α_1^{c13}, α_1^{c33}, α_1^{c44}, are values of Lagrange multipliers. Θ_{TV} and Υ_{TV} represent first and second order TV regularization functions respectively. $\mathbf{D}_{syn}$represents the synthetic data, which is the function of the model parameters. $\mathbf{D}_{obs}$ is the observed data. Θ_{TV} and Υ_{TV} represent functions for calculating the first and second order TV relularizations for the models.

The first order TV regularization term can be expressed as:

$$\begin{aligned}
TV_1((\mathbf{m})) = \sum_{i=1}^{n-1}\sum_{j=1}^{m-1}|M_{i+1,j} - M_{i,j}| + \sum_{i=1}^{n-1}\sum_{j=1}^{m-1}|M_{i,j+1} - M_{i,j}| = \\
\left(\nabla_x, \nabla_z\right)\begin{pmatrix}\mathbf{m}\\ \mathbf{m}\end{pmatrix} = \left(\mathcal{L}_{\mathbf{x}}, \mathcal{L}_{\mathbf{z}}\right)\begin{pmatrix}\mathbf{m}\\ \mathbf{m}\end{pmatrix} = \Theta_{TV}(\mathbf{m}),
\end{aligned} \tag{26}$$

where $M_{i,j}$ represents the element in model parameter vector. n and m are the grid numbers in x and z directions respectively. The second order TV regularization term can be expressed as:

$$
\begin{aligned}
TV_2((\mathbf{m})) &= \sum_{i=1}^{n-1}\sum_{j=1}^{m-1} |M_{i+1,j} - 2M_{i,j} + M_{i-1,j}| + \sum_{i=1}^{n-1}\sum_{j=1}^{m-1} |M_{i,j+1} - 2M_{i,j} + M_{i,j-1}| \\
&= \big(\nabla_{xx}, \nabla_{zz}\big)\begin{pmatrix}\mathbf{m}\\ \mathbf{m}\end{pmatrix} = \big(\mathcal{K}_{\mathbf{xx}}, \mathcal{K}_{\mathbf{zz}}\big)\begin{pmatrix}\mathbf{m}\\ \mathbf{m}\end{pmatrix} = \Upsilon_{TV}(\mathbf{m}).
\end{aligned}
\tag{27}
$$

$\mathcal{L}_{\mathbf{x}}$ and $\mathcal{L}_{\mathbf{z}}$ are the first order differential vectors to calculate the first order total variations in x and z directions respectively. $\mathcal{K}_{\mathbf{xx}}$ and $\mathcal{K}_{\mathbf{zz}}$ are the second order differential vectors to calculate the second order total variations in x and z directions respectively.

The derivative of Φ_{l2}^{TV} for each parameter, which is the gradient for $\mathbf{C}_{11}$, $\mathbf{C}_{13}$ and $\mathbf{C}_{33}$, $\mathbf{C}_{44}$ based on the l_2^{TV} norm, can be expressed as:

$$
\begin{pmatrix}
\frac{\partial \Phi_{l2}^{TV}}{\partial \mathbf{C_{11}}}\\
\frac{\partial \Phi_{l2}^{TV}}{\partial \mathbf{C_{13}}}\\
\frac{\partial \Phi_{l2}^{TV}}{\partial \mathbf{C_{33}}}\\
\frac{\partial \Phi_{l2}^{TV}}{\partial \mathbf{C_{44}}}\\
\frac{\partial \Phi_{l2}^{TV}}{\partial \mathbf{Qc_{11}}}\\
\frac{\partial \Phi_{l2}^{TV}}{\partial \mathbf{Qc_{13}}}\\
\frac{\partial \Phi_{l2}^{TV}}{\partial \mathbf{Qc_{33}}}\\
\frac{\partial \Phi_{l2}^{TV}}{\partial \mathbf{Qc_{44}}}
\end{pmatrix}
=
\begin{pmatrix}
\mathbf{G_{l2_{c11}}}\\
\mathbf{G_{l2_{c13}}}\\
\mathbf{G_{l2_{c33}}}\\
\mathbf{G_{l2_{c44}}}\\
\mathbf{G_{l2_{Qc11}}}\\
\mathbf{G_{l2_{Qc13}}}\\
\mathbf{G_{l2_{Qc33}}}\\
\mathbf{G_{l2_{Qc44}}}
\end{pmatrix}
+
\begin{pmatrix}
\mathbf{R_{l2_{c11}}}\\
\mathbf{R_{l2_{c13}}}\\
\mathbf{R_{l2_{c33}}}\\
\mathbf{R_{l2_{c44}}}\\
\mathbf{R_{l2_{Qc11}}}\\
\mathbf{R_{l2_{Qc13}}}\\
\mathbf{R_{l2_{Qc33}}}\\
\mathbf{R_{l2_{Qc44}}}
\end{pmatrix},
\tag{28}
$$

where $\mathbf{G_{l2_{c11}}}$,$\mathbf{G_{l2_{c13}}}$,$\mathbf{G_{l2_{c33}}}$,$\mathbf{G_{l2_{c44}}}$, $\mathbf{G_{l2_{Qc11}}}$,$\mathbf{G_{l2_{Qc13}}}$,$\mathbf{G_{l2_{Qc33}}}$,$\mathbf{G_{l2_{Qc44}}}$,are the gradient vectors for parameters C_{11},C_{13}, C_{33}, C_{44}, Qc_{11},Qc_{13}, Qc_{33}, Qc_{44} respectively. In this study we use the Automatic Differential method, which is to generate the gradients for the model parameters.

Now we rewrite the misfit function as:

$$
\Phi^{TV} = J_D + J_{r1} + J_{r2},
\tag{29}
$$

where J_D represents the any kind of norm misfit between observed data and synthetic data. $J_{r1} = \alpha_1^{c11}\Theta_{TV}(\mathbf{C}_{11}) + \alpha_1^{c13}\Theta_{TV}(\mathbf{C}_{13}) + \alpha_1^{c33}\Theta_{TV}(\mathbf{C}_{33}) + \alpha_1^{c44}\Theta_{TV}(\mathbf{C}_{44}) + \alpha_1^{Qc11}\Theta_{TV}(\mathbf{Qc}_{11}) + \alpha_1^{Qc13}\Theta_{TV}(\mathbf{Qc}_{13}) + \alpha_1^{Qc33}\Theta_{TV}(\mathbf{Qc}_{33}) + \alpha_1^{Qc44}\Theta_{TV}(\mathbf{Qc}_{44})$. $J_{r2} = \alpha_2^{c11}\Theta_{TV}(\mathbf{C}_{11}) + \alpha_2^{c13}\Theta_{TV}(\mathbf{C}_{13}) + \alpha_2^{c33}\Theta_{TV}(\mathbf{C}_{33}) + \alpha_2^{c44}\Theta_{TV}(\mathbf{C}_{44}) + \alpha_2^{Qc11}\Theta_{TV}(\mathbf{Qc}_{11}) + \alpha_2^{Qc13}\Theta_{TV}(\mathbf{Qc}_{13}) + \alpha_2^{Qc33}\Theta_{TV}(\mathbf{Qc}_{33}) + \alpha_2^{Qc44}\Theta_{TV}(\mathbf{Qc}_{44})$. The values for Lagrange multiples are chosen according to the following formula:

$$
K = \frac{J_D}{J_{r1} + J_{r2}}.
\tag{30}
$$

We should keep the balance between the influence of the regulation terms and data misfit term. If K is too large, the data misfit J_D would dominate the final misfit value and the regularization term does not have enough influence on the inversion results. If K is too small, the regularization term would have too much influence on the inversion results. In this study, K value is chosen to between 0.1 and 1, to make sure the regularization terms have reasonable influence in the inversion. K should be relatively large when noise occurs in the data.

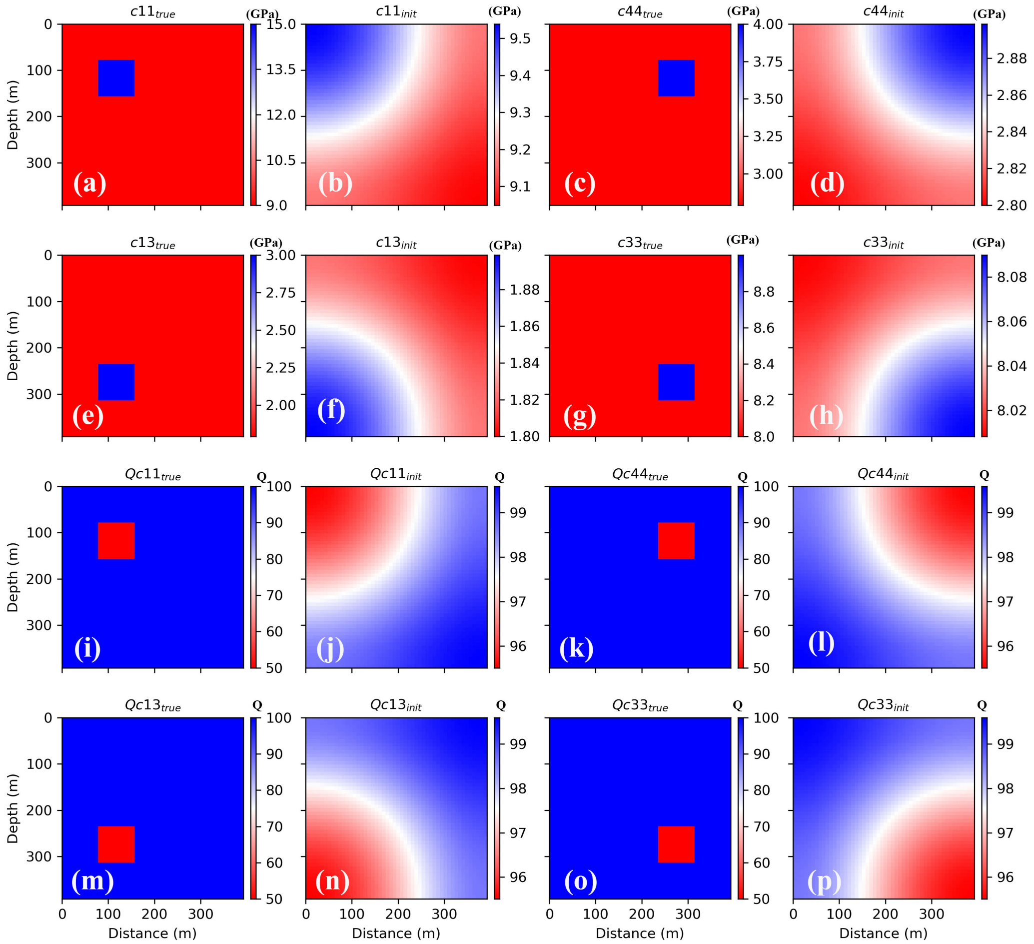

FIG. 1. True and initial viscoelastic VTI models. (a) True C_{11} model (b) Initial C_{11} model. (c) True C_{44} model (d) Initial C_{44} model. (e) True C_{13} model (f) Initial C_{13} model. (g) True C_{33} model (h) Initial C_{33} model. (i) True Qc_{11} model (j) Initial Qc_{11} model. (k) True Qc_{44} model (l) Initial Qc_{44} model. (m) True Qc_{13} model (n) Initial Qc_{13} model. (o) True Qc_{33} model (p) Initial Qc_{33} model.

NUMERICAL TEST

In this numerical test, the size of the model is 50×50. $dx = dz = 7m$. The source is the Rickre's wavelet with main frequency 25Hz. Figure 1 (a) and (b) are the true and initial C_{11}. Figure 1 (c) and (d) are the true and initial C_{44}. Figure 1 (e) and (f) are the true and initial C_{13}. Figure 1 (g) and (h) are the true and initial C_{33}. Figure 1 (i) and (j) are the true and initial Qc_{11}. Figure 1 (k) and (l) are the true and initial Qc_{44}. Figure 1 (m) and (n) are the true and initial Qc_{13}. Figure 1 (o) and (p) are the true and initial Qc_{33}.

Figure 2 shows the inversion results by using the surface acquisition. In this study the sources and the receivers are all located at the surface of the models. Figure 2 (a) is the inversion results for C_{11}. Figure 2 (b) is the inversion results for C_{44}. Figure 2 (c) is the

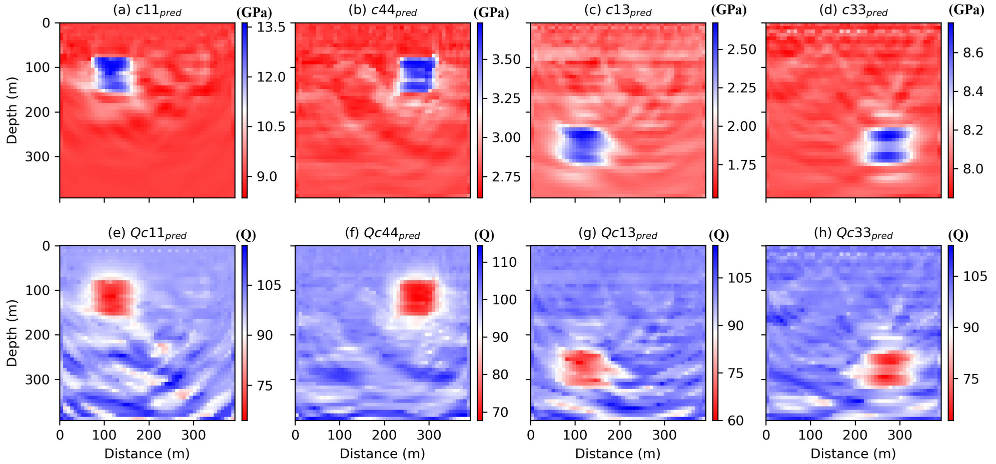

FIG. 2. Surface acquisition RNN viscoelastic VTI models the inversion. (a) C_{11} inversion result. (b) C_{44} inversion result. (c) C_{13} inversion result. (d) C_{33} inversion result. (e) Qc_{11} inversion result. (f) Qc_{44} inversion result. (g) Qc_{13} inversion result. (h) Qc_{33} inversion result.

inversion results for C_{13}. Figure 2 (d) is the inversion results for C_{33}. Figure 2 (e) is the inversion results for Qc_{11}. Figure 2 (f) is the inversion results for Qc_{44}. Figure 2 (g) is the inversion results for Qc_{13}. Figure 2 (h) is the inversion results for Qc_{33}. From the comparison between the inversion results and the true models, we can see that the general structure of the box model has been recovered, however, the inversion shows a certain amount of noise be which does not appear in the true models. This may due to the inversion is based on gradient-based method and the inversion is influenced by the cross-talk between the parameters.

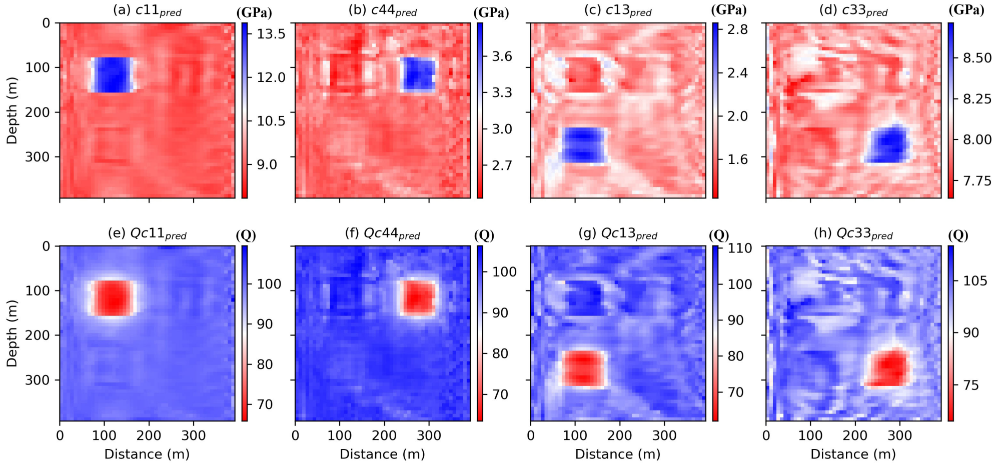

FIG. 3. RNN viscoelastic VTI models the inversion based on cross well data. (a) C_{11} inversion result. (b) C_{44} inversion result. (c) C_{13} inversion result. (d) C_{33} inversion result. (e) Qc_{11} inversion result. (f) Qc_{44} inversion result. (g) Qc_{13} inversion result. (h) Qc_{33} inversion result.

Figure 3 shows the inversion results by using the cross well data. In this numerical test, the shots are generated by the vertical well on the right-hand side of the model and the receivers are on both the surface and another well on the left side of the model. 10 shots are distributed in the well on the right side of the model. 50 receivers are distributed on the surface and the well on the left side of the model.Figure 3 (a) is the inversion results for C_{11}. Figure 3 (b) is the inversion results for C_{44}. Figure 3 (c) is the inversion results for C_{13}. Figure 3 (d) is the inversion results for C_{33}. Figure 3 (e) is the inversion results for Qc_{11}. Figure 3 (f) is the inversion results for Qc_{44}. Figure 3 (g) is the inversion results for Qc_{13}. Figure 3 (h) is the inversion results for Qc_{33}. From the comparison between Figure 2 and Figure 3. We can see that in Figure 3 the box structure of the models is more clearly recovered, especially for C_{11}, Qc_{11}, C_{44} and Qc_{44}. The cross-talk noise in the inversion still appears in the inversion results, which means that the inversion still suffers from the cross-talk issue. Cross well data has a better ability to recover the true structure of the velocity models.

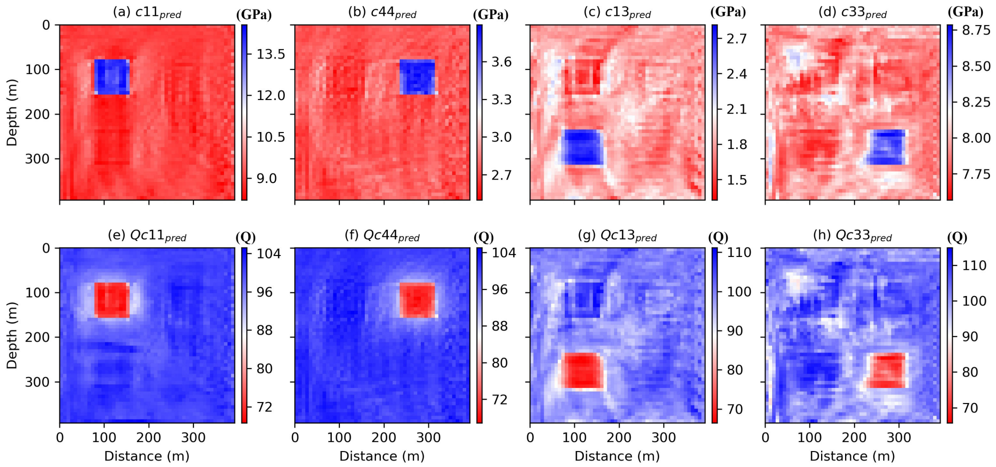

FIG. 4. Cross well data and surface data acquisition RNN viscoelastic VTI FWI with high order total variation. (a) C_{11} inversion result. (b) C_{44} inversion result. (c) C_{13} inversion result. (d) C_{33} inversion result. (e) Qc_{11} inversion result. (f) Qc_{44} inversion result. (g) Qc_{13} inversion result. (h) Qc_{33} inversion result.

Figure 4 shows the inversion results by using the cross well data and the surface data acquisition with total variation regularization. Figure 4 (a) is the inversion results for C_{11}. Figure 4 (b) is the inversion results for C_{44}. Figure 4 (c) is the inversion results for C_{13}. Figure 4 (d) is the inversion results for C_{33}. Figure 4 (e) is the inversion results for Qc_{11}. Figure 4 (f) is the inversion results for Qc_{44}. Figure 4 (g) is the inversion results for Qc_{13}. Figure 4 (h) is the inversion results for Qc_{33}. With this high order TV regulation the inversion of the box structure is more clearly reflected, i.e. Figure 4 (a) , Figure 4 (b), Figure 4 (e), Figure 4 (f). The cross-talk noise between the parameters is also mitigated in this test. Thus we believe the inversion with the combination of surface and cross well data with high order TV regulation has the best inversion results.

Figure 5 to Figure 12 shows the inversion for part of the Marmusi model. The size of the model is 125×125. $dx = dz = 10m$. The source of the wavelet is the Ricker's wavelet. The main frequency of the wavelet is 30Hz. We also use both the combination of the cross

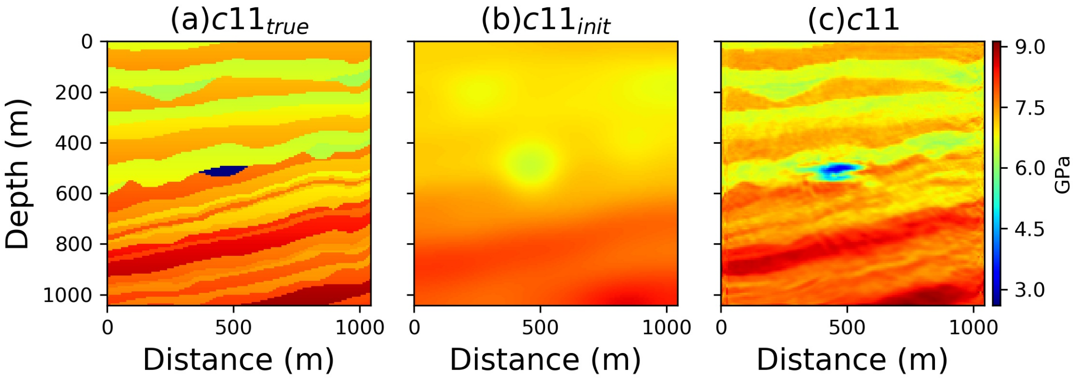

FIG. 5. C_{11} inversion. (a) True C_{11}. (b) Initial C_{11}. (c) Inversion result for C_{11}.

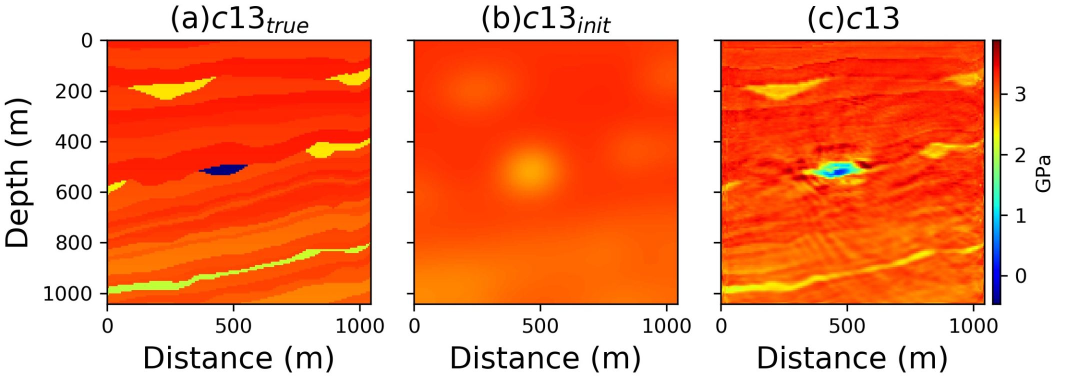

FIG. 6. C_{13} inversion. (a) True C_{13}. (b) Initial C_{13}. (c) Inversion result for C_{13}.

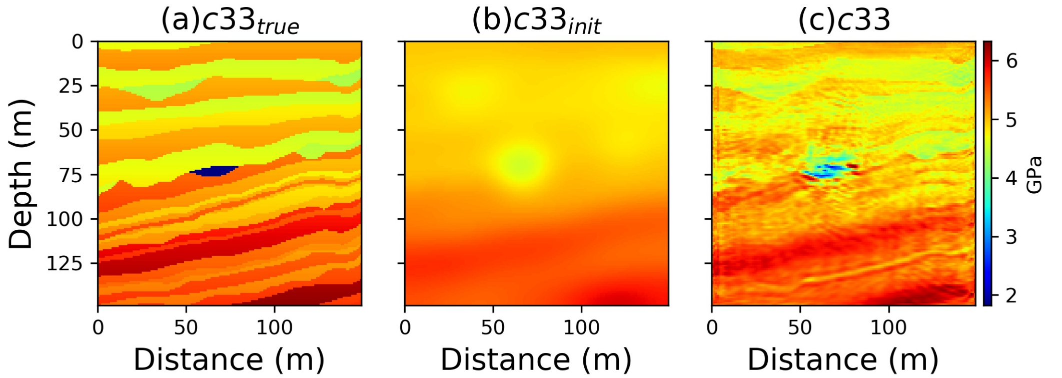

FIG. 7. C_{33} inversion. (a) True C_{33}. (b) Initial C_{33}. (c) Inversion result for C_{33}.

well data and the surface data acquisition to invert the models. We use the Adam algorithm as the optimization method. The maximum iteration is 100. From Figure 5 to Figure 8 we can see that the inversion gives accurate inversion results for the elastic modulus, especially for parameter C_{11} and C_{13}, Though noise appears in the inversion results for C_{13} and C_{44}, the inversion still has the tendency to converge to the right inversion results. Figure 9 to 12 shows the inversion results for the attenuation models.The inversion suffers from the dispersion noise occurs between the fined layers. This effect may due to the dispersion effect of the forward modeling and the cross-talk between the parameters. How to mitigate

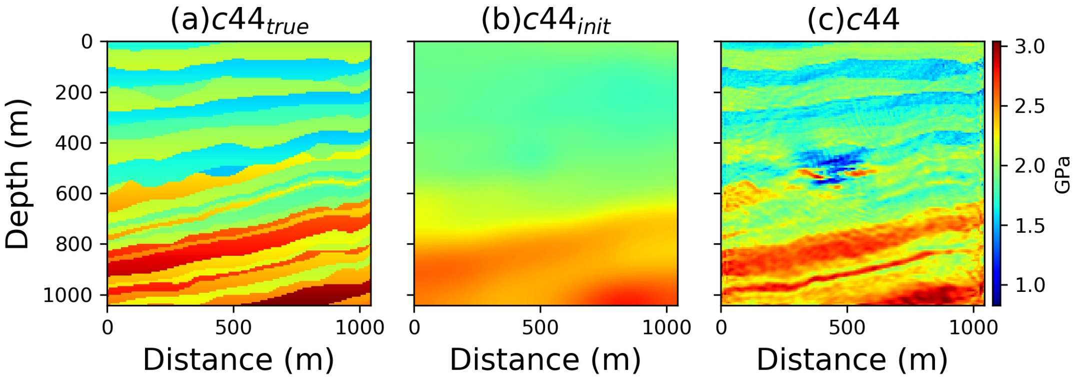

FIG. 8. C_{44} inversion. (a) True C_{44}. (b) Initial C_{44}. (c) Inversion result for C_{44}.

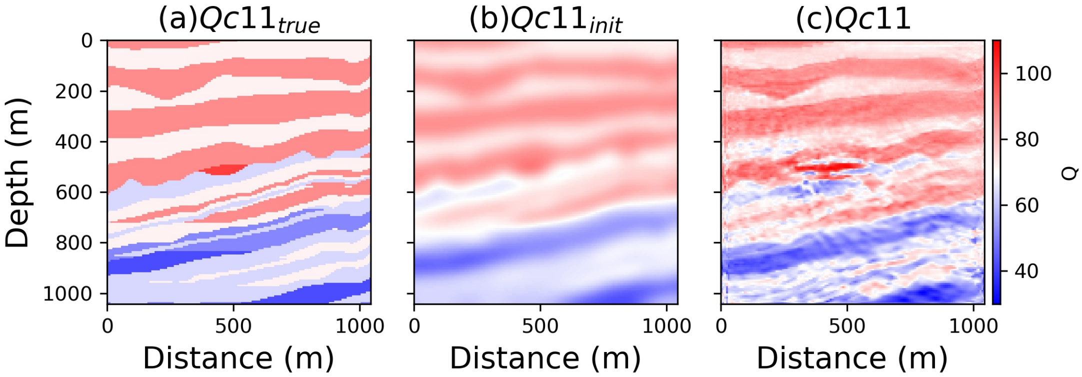

FIG. 9. Qc_{11} inversion. (a) True Qc_{11}. (b) Initial Qc_{11}. (c) Inversion result for Qc_{11}.

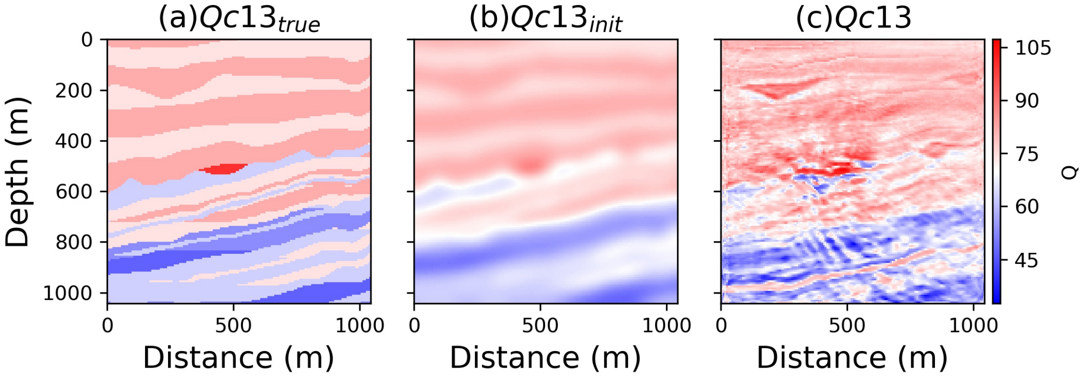

FIG. 10. Qc_{13} inversion. (a) True Qc_{13}. (b) Initial Qc_{13}. (c) Inversion result for Qc_{13}.

such kind of effect still need to be further studied.

CONCLUSIONS

In this study, based on the viscoelastic VTI staggered grid stress velocity wave equation, we build the VTI anelastic RNN cell and performed the viscoelastic full waveform inversion, which forms the theory based machine learning full waveform inversion. This inversion framework would generate the gradient automatically, and the gradients are the

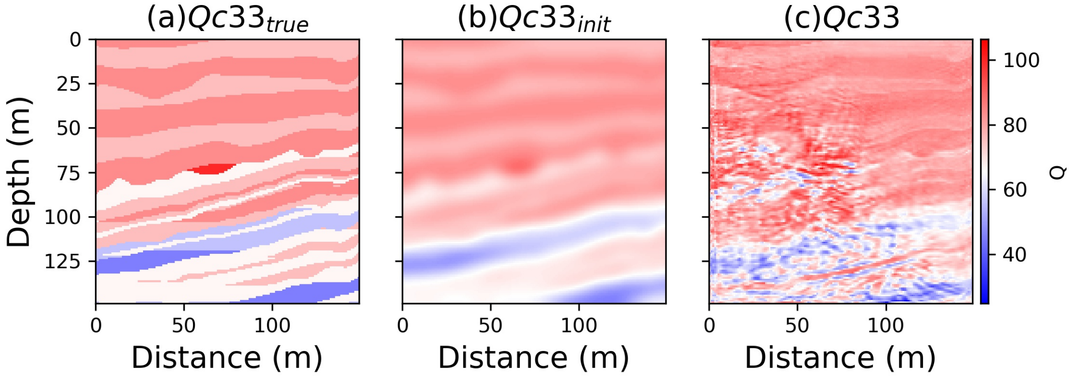

FIG. 11. Qc_{11} inversion. (a) True Qc_{11}. (b) Initial Qc_{11}. (c) Inversion result for Qc_{11}.

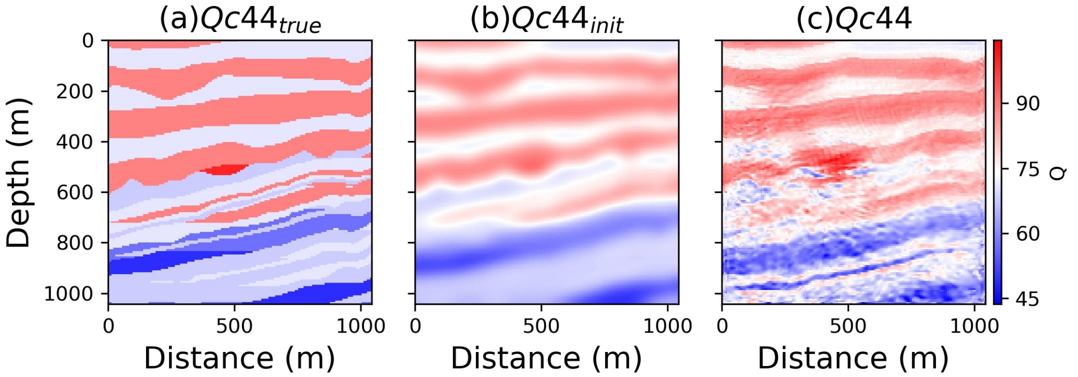

FIG. 12. Qc_{44} inversion. (a) True Qc_{44}. (b) Initial Qc_{44}. (c) Inversion result for Qc_{44}.

exact gradients based on the forward computational graph.

Numerical testes show that the inversion framework we introduced would generate the correct inversion results for C_{11}, C_{13}, C_{33}, C_{44} and Qc_{11}, Qc_{13}, Qc_{33}, Qc_{44}. The numerical tests shows that the inversion of θ the title angle is the hardest to recover. We also use high order total variation (TV) regularization to mitigate the cross-talks between each parameters. With Automatic differential method the gradient of the complex media parameters based on complex misfits can be calculated. The numerical tests also proved the validation of this method.

ACKNOWLEDGMENT

We thank the sponsors of CREWES for continued support. This work was funded by CREWES industrial sponsors, NSERC (Natural Science and Engineering Research Council of Canada) through the grants CRDPJ 461179-13 and CRDPJ 543578-19. Partial funding also came from the Canada First Research Excellence Fund. The author is supported by China Scholarship Counsel.

REFERENCES

Belahi, T., Fuji, N., and Singh, S., 2015, Elastic versus viscoelastic full waveform inversion of near-offset and wide-angle data in the presence of attenuation, *in* 77th EAGE Conference and Exhibition 2015.

Belahi, T., Singh, S., and Fuji, N., 2016, Viscoelastic full waveform inversion of sea bottom long offset seismic data in presence of attenuation, *in* 78th EAGE Conference and Exhibition 2016.

Bohlen, T., 2002, Parallel 3-d viscoelastic finite difference seismic modelling: Computers & Geosciences, **28**, No. 8, 887–899.

Day, S., and Minster, J., 1984, Numerical simulation of wavefields using a pade approximation method: Geophys. JR astr. Soc, 78–105.

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., 2016, The adjoint state method for the viscoelastic wave equation in the velocity-stress formulation, *in* 78th EAGE Conference and Exhibition 2016.

Fabien-Ouellet, G., Gloaguen, E., and Giroux, B., 2017, Time-domain seismic modeling in viscoelastic media for full waveform inversion on heterogeneous computing platforms with opencl: Computers & Geosciences, **100**, 142–155.

Groos, L., Schäfer, M., Forbriger, T., and Bohlen, T., 2012, On the significance of viscoelasticity in a 2d full waveform inversion of shallow seismic surface waves, *in* 74th EAGE Conference and Exhibition incorporating EUROPEC 2012.

Lin, Y., and Wu, Y., 2018, Inversionnet: A real-time and accurate full waveform inversion with convolutional neural network: The Journal of the Acoustical Society of America, **144**, No. 3, 1683–1683.

Mosser, L., Dubrule, O., and Blunt, M., 2018, Stochastic seismic waveform inversion using generative adversarial networks as a geological prior, *in* First EAGE/PESGB Workshop Machine Learning.

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

Sun, H., and Demanet, L., 2018, Low frequency extrapolation with deep learning, *in* SEG Technical Program Expanded Abstracts 2018, Society of Exploration Geophysicists, 2011–2015.

Trinh, P.-T., Brossier, R., Métivier, L., and Virieux, J., 2018, Data-oriented strategy and v p/v s model-constraint for simultaneous v p and v s reconstruction in 3d visco-elastic fwi: Application to the seam ii foothill dataset, *in* SEG Technical Program Expanded Abstracts 2018, Society of Exploration Geophysicists, 1213–1217.

Yang, F., and Ma, J., 2019, Deep-learning inversion: a next generation seismic velocity-model building method: Geophysics, **84**, No. 4, 1–133.

Yang, P., Brossier, R., Métivier, L., and Virieux, J., 2016, A review on the systematic formulation of 3-d multiparameter full waveform inversion in viscoelastic medium: Geophysical Journal International, **207**, No. 1, 129–149.