Incorporating estimates of data covariance in FWI to combat coherent noise based on auto-regressive process

Luping Qu, Scott Keating, Xin Fu, He Liu, Kris Innanen

ABSTRACT

An important step in seismic data processing is noise attenuation, which typically improves the subsurface seismic image and signal-to-noise ratio (SNR). In seismic records, coherent noise, which correlates spatially or temporally, are more difficult to attenuate or process as it can interfere with signals and be mistakenly recognized as signals. Through incorporating data covariance matrix into the misfit function, both the model parameters and noise can be estimated. When implementing this, the data covariance matrix with random noise can be simplified to be a vector. However, the data covariance matrix with coherent noise still need to be completely computed and stored. We find the serial data-error correlations can be characterized by adding the forward model with a autoregressive error model. As autoregressive error models do not estimate error with point estimates, the inverse of data-error covariance matrix does not need to be computed. The order of the autoregressive process required to fit the data is determined by the residual data-fitting examination. To avoid overfitting, estimates with several different orders were conducted and adopted in the following rounds of FWI.

INTRODUCTION

The potential advantage and significant impact of seismic data in exploration, production, and development are increased with the aid of appropriate and acceptable noise attenuation strategies. The interpretation of data for an interested area's geology has been significantly impacted by improvements in signal processing methodologies. The term "noise" in seismic exploration refers specifically to an undesired or unintelligible portion of recorded seismic signals for a variety of reasons. These undesirable events may be regarded as signals, but they typically provide insufficient or conflicting information about the subsurface and are referred to as random noise and coherent noise. Without regard to the noise source, the physical characteristics of noise typically fall into two categories: the first is coherent noise (Ground roll, Guided Waves, Multiples, and Power Line), and the second is random noise.

The main components of seismic data are signals and various types of noise. Any recorded waves that obstruct the desired signal are considered noise, according to the general definition. Separating signal from noise can be difficult and problematic at times due to the variety of noise types. But for high resolution imaging, effective noise suppression is crucial. Removing the noise from seismic data is a crucial step toward interpretations with high confidence. To obtain reliable processing results, usually various types of noise are handled differently and may call for the application of additional strategies. It is in fact very difficult to reduce high-amplitude noises like ground roll when processing seismic data. Though the coherent noise gets attenuated to some degree, it is difficult to get removed completely.

The covariance matrix characterizes the difference of each pair of elements of a provided vector. It not only includes the variances on the diagonal but also contains the correlation relationship between elements on the off-diagonal. As covariance matrix extends the variance with an additional dimension, the calculation of misfit with covariance matrix introduces more information of data residuals.

The least squares (L2) norm of the misfit between the observed and predicted data is the most commonly used penalty function for FWI (Tarantola, 2005). In this misfit function, the observing data, containing remnant noises after denoising, are treated as the true data. During the iterations, penalty function is minimized to obtain subsurface model parameters which contribute to the observing data. When the real data are seriously contaminated by noises, that noise will tend to produce model artifacts, or false structures added to the model in order to explain the noise. This has motivated us to include the noise estimation in FWI by incorporating the data covariance matrix into the misfit function(Cai and Zelt, 2019). The various types of remnant noises can be estimated through the data covariance matrix and updated during the inversion. As implemented in frequency domain, the data covariance can be calculated frequency by frequency. If source number and receiver number are not too large, the size of the data covariance matrix will not be very large. However, when the model size is large, it is challenging to incorporating such a large covariance matrix into the misfit computation.

Here, autoregressive (AR) models (Andel, 1971; Lütkepohl, 2013; Shibata, 1976) are brought in to characterize the serial data-error correlation. In such way, the inverse or determinant of the data covariance matrix does not need to be computed. AR model can describe the correlation relationship with few parameters. However, the data information is traded off between the forward-modeling model and the AR error model (Dettmer et al., 2012). Thus, several orders of AR models were tested.

GENERALIZED MISFIT FUNCTION WITH AUTOREGRESSIVE DATA-ERROR MODELS

In Bayesian inversion(e.g., Dettmer et al., 2007), the likelihood formulation includes the data uncertainty distribution, which embodies both modeling errors and measurement errors. In theory, the likelihood can be formulated and applied with arbitrary uncertainty distributions. However, in practice, the error distribution is unknown in advance. Therefore, a mathematically simple distribution (e.g., Gaussian) is usually assumed initially.

Different approaches can be adopted to estimate the covariance matrices. If the error is assumed to be random, the data covariance matrix can be approximated as diagonal, $\mathbf{C}_{\mathrm{D}}^{-1} = \sigma^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix and σ is the standard deviation of the random error. In this case, as σ is a scalar, the negative log likelihood is similar to the conventional L2 norm misfit function. A more sophisticated approach, beyond assuming that the statistics are simple, or known, is to analyze the data residuals to incorporate error correlations into the inversion. The data covariance matrix is estimated from the data residuals in a first past through FWI, assuming uncorrelated errors. The data covariance matrix can be estimated from the autocovariance of the data residual after some fixed number of iterations:

$$c_j = \frac{1}{N} \sum_{k=0}^{N-j-1} \left(\mathbf{d}_{\text{obs}}^{j+k} - \bar{\mathbf{d}} \right) \left(\mathbf{d}_{\text{obs}}^{k} - \bar{\mathbf{d}} \right), \tag{1}$$

for the jth datum, where $\bar{\mathbf{d}}$ is the mean of the samples. As we do not have many observing data samples, the synthetic data generated using the conventional FWI result were utilized to approximate the sample mean. These values are arranged in the covariance matrix $\mathbf{C}_{\mathrm{D}}$. The FWI misfit function incorporating the data covariance matrix $\mathbf{C}_{\mathrm{D}}$ is

$$E = \frac{1}{2N}\left[(\mathbf{d}_{\text{pre}} - \mathbf{d}_{\text{obs}})^T \mathbf{C}_{\mathrm{D}}^{-1} (\mathbf{d}_{\text{pre}} - \mathbf{d}_{\text{obs}})\right], \tag{2}$$

where N is the number of data, and $\mathbf{d}_{\text{pre}}$ and $\mathbf{d}_{\text{obs}}$ are the predicted and observing seismic data, respectively.

For each frequency, the dimension pf covariance matrix is $(Ns \times Nr)^2$. In most cases, this size is within the tolerance of the computer memory capacity.

The AR data error model can replace the covariance matrix through characterizing the correlation relationship with few parameters. AR models are usually applied in time-series analysis and its general expression is

$$d_t = c + \sum_{i=1}^{p} a_i d_{t-i} + \epsilon_t \tag{3}$$

where t represents data index, p is the AR model order, a_i are the parameters for AR model, ϵ_t is the uncorrelated Gaussian noise term. The model and its order are commonly abbreviated as AR(p).

To apply the AR models in FWI inversion, the misfit function can be written

$$E = \frac{1}{2N}\left[(\mathbf{d}_{\text{pre}} - \mathbf{d}_{\text{obs}} - \mathbf{d}(\mathbf{a}))^T \left(\sigma^2 \mathbf{I}\right)^{-1} (\mathbf{d}_{\text{pre}} - \mathbf{d}_{\text{obs}} - \mathbf{d}(\mathbf{a}))\right], \tag{4}$$

where a is the vector of AR parameters, $\mathbf{I}$ is the identity matrix, σ is the standard deviation for uncorrelated Gaussian noise. In such way, the correlation part is removed from the data covariance matrix.

Assuming the errors to be independent of model parameters, the gradient of the misfit function incorporating data covariance matrix with respect to the ith model parameter is

$$\frac{\partial \mathrm{E}(\mathbf{m})}{\partial m_i} = \frac{1}{N}\left(\frac{\partial \mathbf{d}_{\text{pre}}}{\partial m_i}\right)^T \mathbf{C}_{\mathrm{D}}^{-1} (\mathbf{d}_{\text{pre}} - \mathbf{d}_{\text{obs}}), \tag{5}$$

where $\partial \mathbf{d}_{\text{pre}} / \partial m_i$ is the Fréchet derivative, and we observe that the wavefield residuals have been weighted by the data covariances before back propagation. Data residual regions which, through the iterative estimation, appear to contain large errors, are in this calculation down-weighted and contribute less to the inversion results.

When the AR mdoel is applied, the gradient is

$$\frac{\partial \mathrm{E}(\mathbf{m})}{\partial m_i} = \frac{1}{N}\left(\frac{\partial \mathbf{d}_{\text{pre}}}{\partial m_i}\right)^T (\mathbf{d}_{\text{pre}} - \mathbf{d}_{\text{obs}} - \mathbf{d}(\mathbf{a})). \tag{6}$$

The complete AR model is given by $\sigma = (\sigma, a)$. Usually, the order of AR mdoel should be chosen to be small enough to avoid unnecessary complexity or overfitting.

SYNTHETIC EXAMPLES

Figure 1 shows simulated examples of several orders' AR processes to display that the correlations can be modelled with AR models.

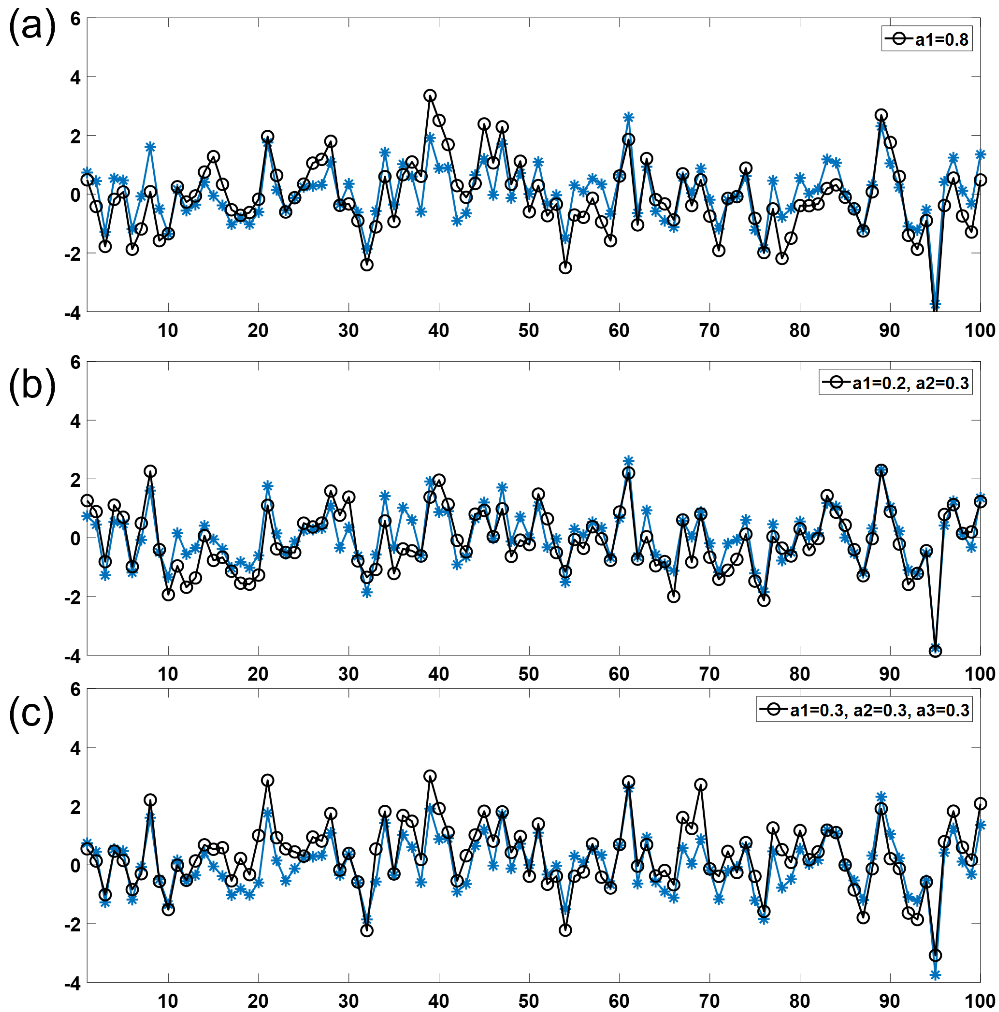

FIG. 1. The histogram and autocovariance of data residuals. (a) Histogram of real part data residual , (b) histogram of imaginary part data residual, (c) autocovariance of the data residual in real.

In the 1st round FWI, inversion with the conventional misfit function without considering noise estimation was conducted for both noise-free data and noisy data. The true models and initial models are shown in Figure 3.

We firstly applied this method in acoustic FWI in frequency domain. we added a combination of random and correlated noise. Correlated data error (Figure 2 a) was generated by multiplying an Gaussian random array with the Cholesky decomposition of a constructed data error covariance matrix with non-zero decaying off-diagonal terms. We combined purely random noise with this correlated noise to obtain the input complex noise.

After the 1st round FWI, we checked the data residuals. Figure 6 shows the data residual

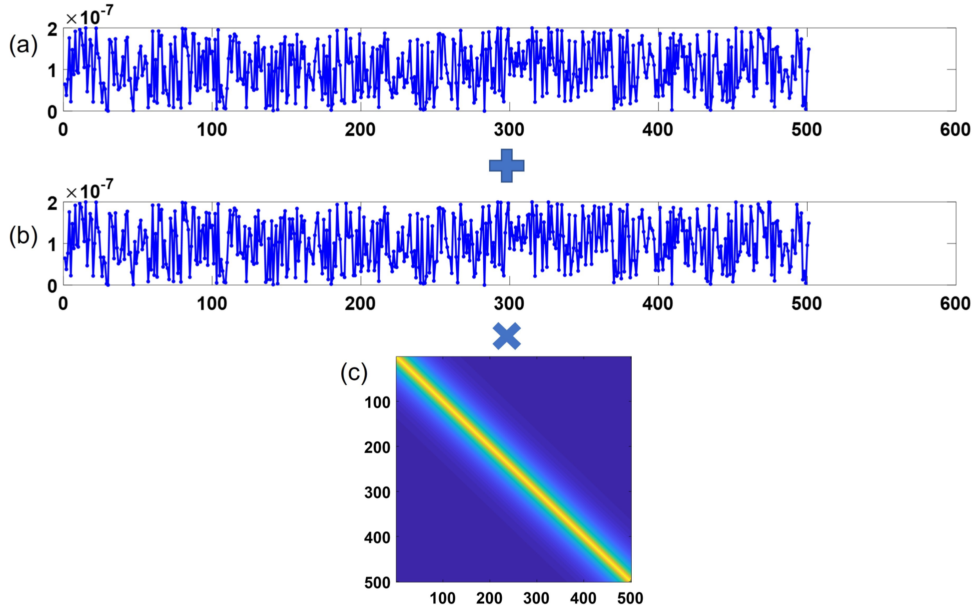

FIG. 2. The creation of complex noise. (a) Random noise sequence one, (b) sequence two, (c) correlation matrix.

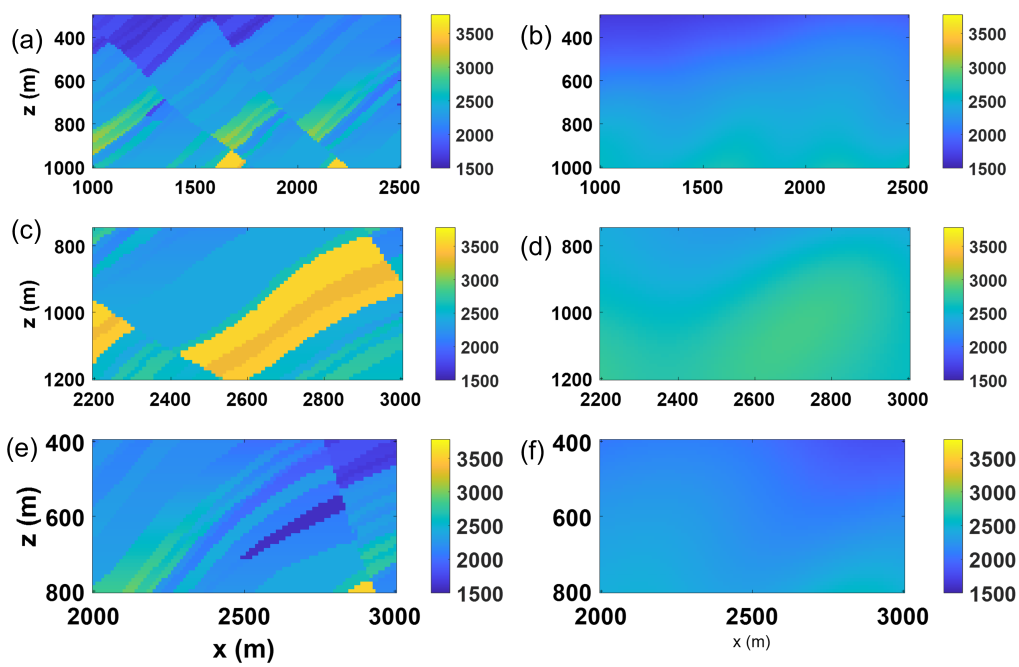

FIG. 3. True models and initial models.(a) (c) (e) are three true models for inversion, (b) (d) (f) are the corresponding initial models for inversion.

sample. In our processing, the data residuals vector length is much longer.

Based on the inversion results above, we implemented the second round FWI with the generalized misfit function. We generated autoregressive models with orders equal to 2 and 3. This time, the data covariance matrix does not need to be calculated, as the autoregressive

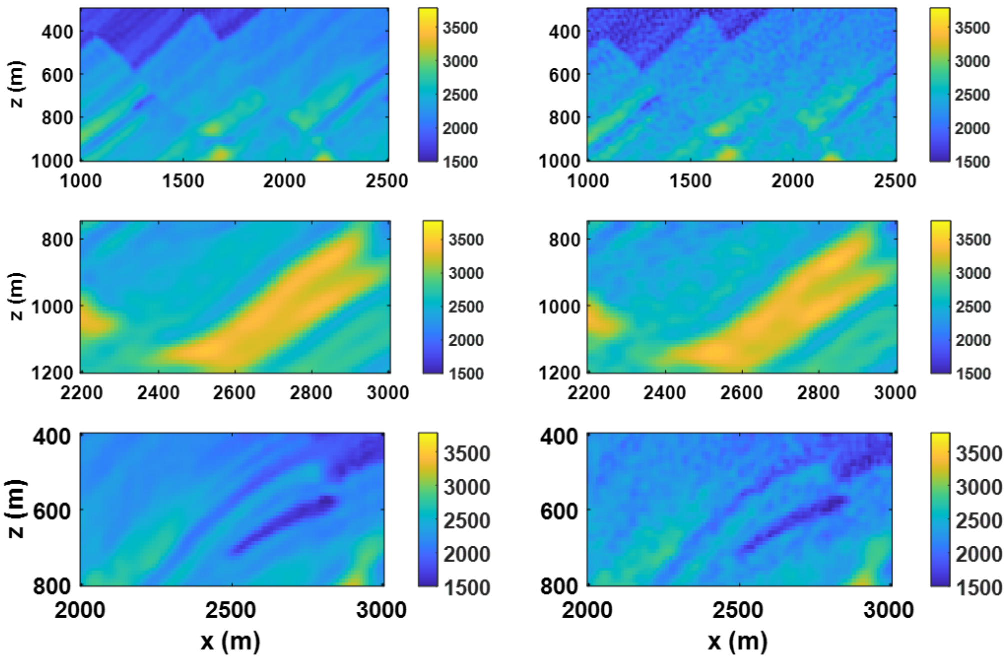

FIG. 4. Noise-free inversion results and noisy data inversion results using conventional misfit after 50 iterations of the 1st round FWI. (a) (c) (e) are inversion results for noise-free data, (b) (d) (f) are inversion results for noisy data.

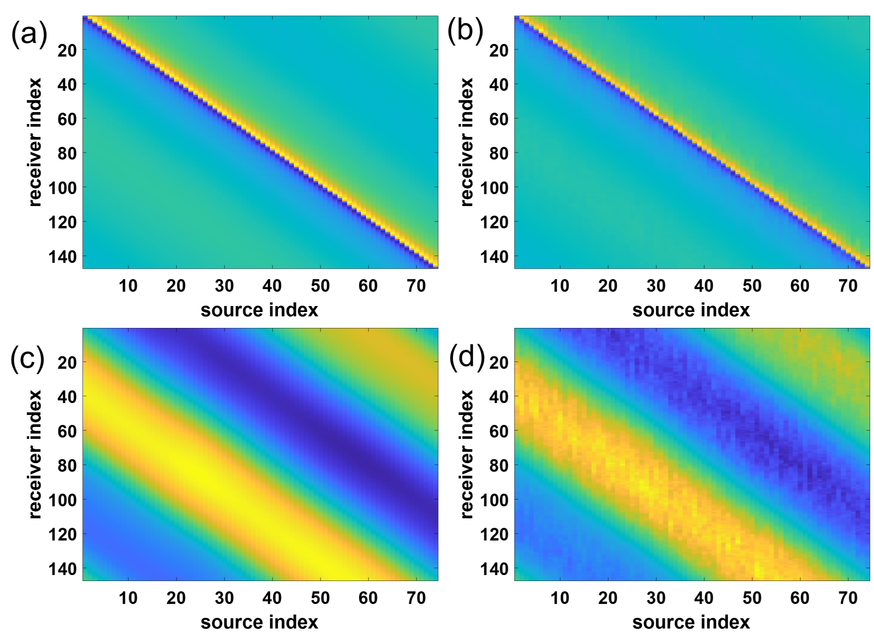

FIG. 5. The real and imaginary part of the noise-free and noisy data of one model. (a) Noisy data (real), (b) noise-free data (real), (c) noisy data (imaginary), (d) noise-free data (imaginary).

model replaced the covariance matrix in the misfit.

The inversion results are shown in Fig.9. By comparison, when AR model order equals 2, the imaging resolution does not improve much. When AR model order equals 3, the

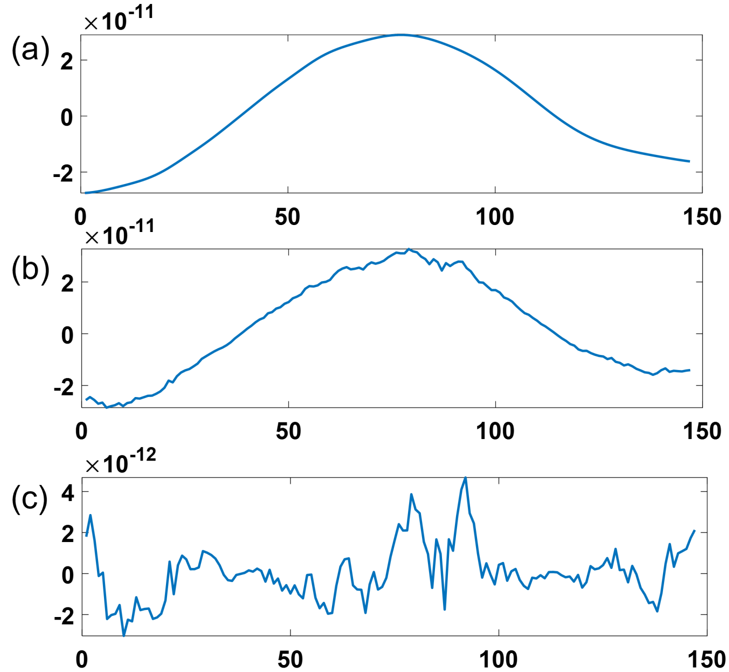

FIG. 6. A data residuals sample of the imaginary-part data. (a) noise-free data, (b) noisy data, (c) data residual.

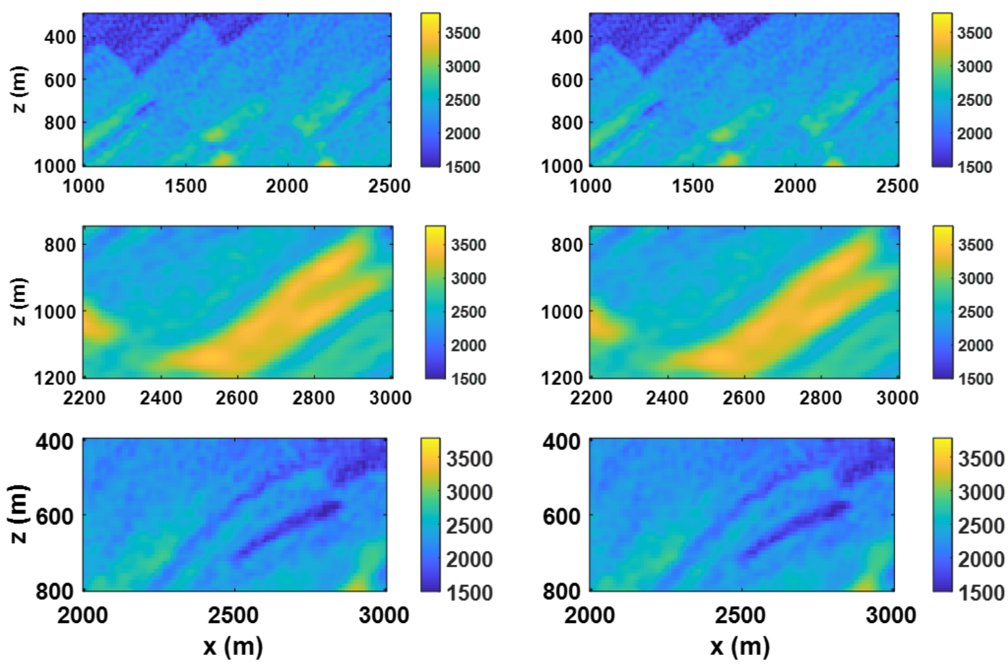

FIG. 7. Inversion results of the FWI with conventional misfit function and generalized misfit function. The left column is the inversion result using conventional misfit after 80 iterations, the right column is the inversion result using generalized misfit function with the same iteration number and AR model order equals 2.

imaging resolution shows some improvements.

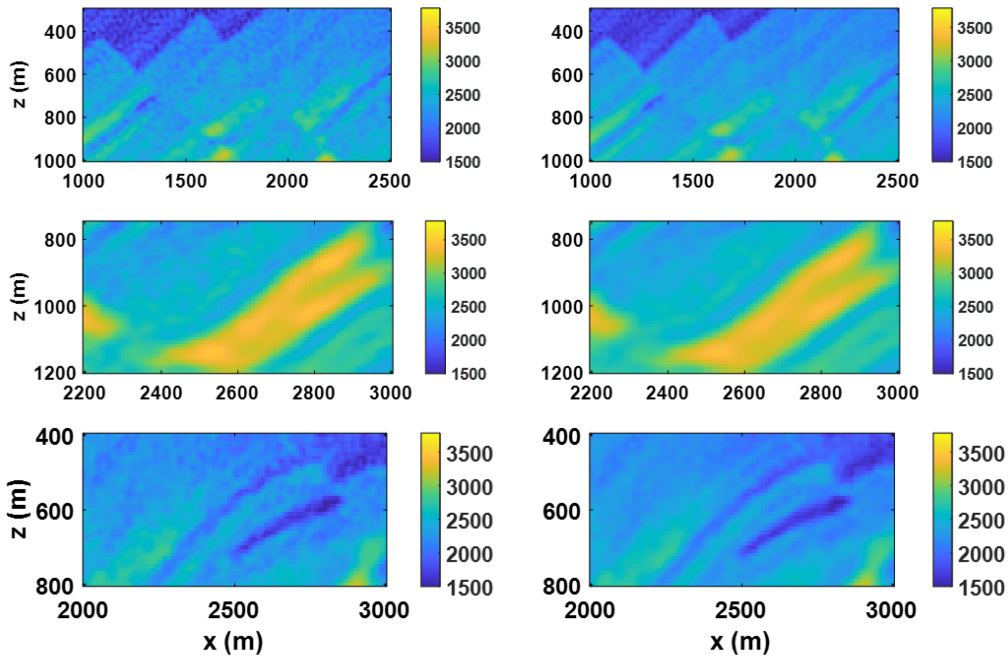

FIG. 8. Inversion results of the FWI with conventional misfit function and generalized misfit function. The left column is the inversion result using conventional misfit after 80 iterations, the right column is the inversion result using generalized misfit function with the same iteration number and AR model order equals 3.

CONCLUSIONS

In this study, we develop a strategy based on the proposed generalized misfit function to avoid data covariance matrix computation and storage while estimating the general noise during FWI. Through adopting autoregressive models, the 2D matrix can be simplified into a vector, which greatly reduced the computation and memory requirement. From the numerical tests above, we found the selection of AR model order can influence the characterization of correlation, which in turns influence the inversion results. However, an intrinsic trade-off exist between correlated error estimation and data features which are related with velocity models. Thus, we conducted several rounds of FWI, with the first one ignoring the correlations to judge the effects of AR model on the results. We hope to apply this method to some small-sized real data tests.

ACKNOWLEDGEMENTS

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 grant CRDPJ 543578-19.

REFERENCES

Andel, J., 1971, Local properties of the autoregressive series: The Annals of Mathematical Statistics, 67–74.

Cai, A., and Zelt, C. A., 2019, Data weighted full-waveform inversion with adaptive moment estimation for near-surface seismic refraction data, *in* SEG International Exposition and Annual Meeting, OnePetro.

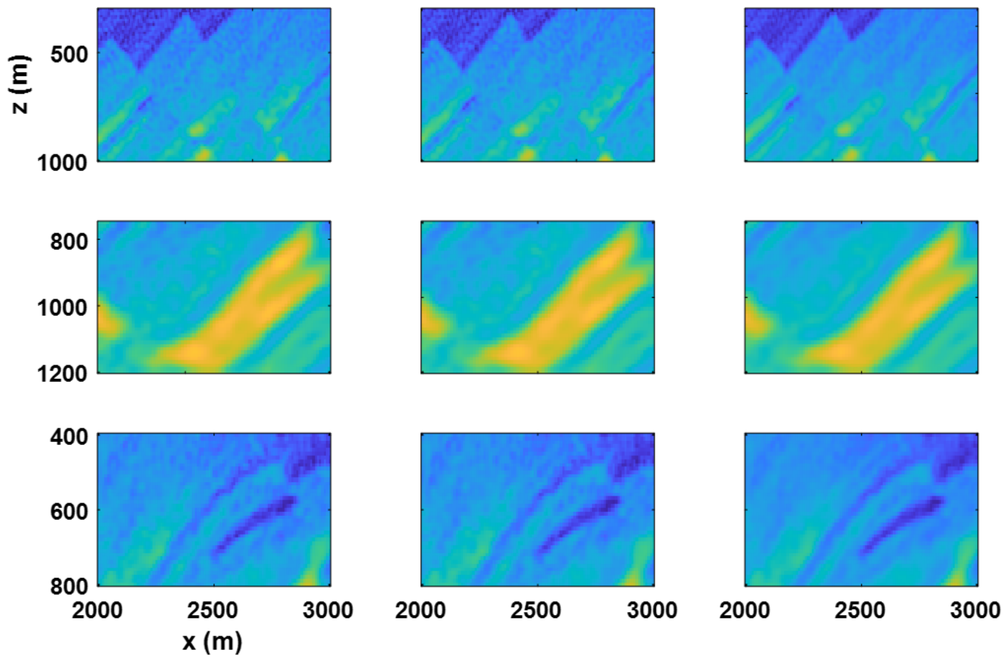

FIG. 9. For a better comparison of the above two figure. The left column are the inversion results using conventional misfit function, the middle column are the inversion results using generalized misfit function with AR model order equals 2, the right column are the inversion results using generalized misfit function with AR model order equals 3.

Dettmer, J., Dosso, S. E., and Holland, C. W., 2007, Uncertainty estimation in seismo-acoustic reflection travel time inversion: The Journal of the Acoustical Society of America, **122**, No. 1, 161–176.

Dettmer, J., Molnar, S., Steininger, G., Dosso, S. E., and Cassidy, J. F., 2012, Trans-dimensional inversion of microtremor array dispersion data with hierarchical autoregressive error models: Geophysical Journal International, **188**, No. 2, 719–734.

Lütkepohl, H., 2013, Vector autoregressive models, *in* Handbook of Research Methods and Applications in Empirical Macroeconomics, Edward Elgar Publishing, 139–164.

Shibata, R., 1976, Selection of the order of an autoregressive model by akaike's information criterion: Biometrika, **63**, No. 1, 117–126.

Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: SIAM.