Reducing the influence of coherent noise on FWI with misfit modification

Luping Qu, Xin Fu, Kris Innanen

ABSTRACT

As field data applications of FWI increase in number and ambition, dealing with the presence of both random and coherent noise in seismic data, and the artifacts they create in FWI models, becomes increasingly important. Though noise suppression methods, including various filters and decompositions, applied before the inversion, can mitigate these to some extent, remnant noise still always exist in the processed data. In this study, we carried out a systematic study of the impact of noise on scalar acoustic FWI models, and sought mitigation strategies. We found that while random noise with low SNR (<=20%) does not exert a strong influence, correlated noise of all types tends to produce a strong negative affect. Mitigation effort here is likely to pay dividends in model accuracy and reliability. We examine the effect of including the data covariance matrix into the misfit function. Through iterative processing, random and correlated noise, and their combination, can be estimated during FWI. As limited frequency bands contribute most of the signal, high frequency bands (>=35 Hz) were cut off. Considering accuracy and computational efficiency, the data can be resampled every two or three points in frequency domain. The intermediate results of misfit and gradient were calculated by individual frequency and then transformed back to time domain after interpolation. The misfits are then calculated by individual frequency and transformed back to the time domain after interpolation. Numerical testing is suggestive that this approach improves on more conventional de-noising methods.

INTRODUCTION

Seismic data are typically contain coherent, incoherent, and nonstationary noise, and combinations of these. To obtain high-quality seismic images, seismic data must be carefully denoised. In real cases, however, noise cannot be completely suppressed, and indeed meaningful distinction between signal and noise may not be available. The objective of noise suppression is to improve signal-to-noise ratios, and remove data variations which add false structure to inverted model results. Random noise is that which follows a Gaussian distribution and is not uncorrelated. Coherent noise is correlated temporally and spatially. Generally, random noise is easier to suppress. Sources of noise include (1) ambient sources (e.g., wind, swell, power line); (2) unwanted waves during wave propagation (e.g., multiples and ground roll); (3) noise due to poor source and receiver coupling to the earth or acquisition footprint; (4) Data processing related noises. Strategies to remove noise include predictive filtering, transformation methods, decompositions, and deep and learning-based approaches (Yu et al., 2019; Zhu et al., 2019). Coherent noise is more resistant to removal via conventional techniques, as it tends to share features of desired signal (Canales, 1984; Neelamani et al., 2008; Bekara and Van der Baan, 2009; Chen and Sacchi, 2015). Targeting and reducing the influence of coherent noise in real data cases is a critical part of inversion and image-formation. After noise removal, furthermore, though reduced to some extent, remnant noise will still tend to be present. This remnant (usually coherent) noise can smear amplitudes, distort waveforms, and in full waveform inversion, aggravate local minima,

leading to inaccurate results.

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). When the observed data contains coherent remnant noise, that noise will tend to produce model artifacts, or false structures added to the model in order to explain the noise. The least-squares misfit does have the capacity to accommodate remnant noise, however, by incorporating the data covariance matrix into the misfit function(Cai and Zelt, 2019). Indeed, remnant noise can in principle be estimated through the data covariance matrix and updated during the inversion. In this way, model parameters and structures with fewer artifacts can be obtained, while the noise is simultaneously estimated.

In this study, we modify the misfit function of acoustic FWI, incorporating the data covariacne matrix. Through this approach, various types of errors can be estimated and separated from the signal. The main challenge in this effort is that the high dimensionality of the data covariance matrix presents a very significant computational expense. The size of the data covariance matrix in an FWI problem is the square of the product of the number of time samples and the number of traces for each shot. This can be partially addressed with matrix sparsification, wherein the effective matrix size is reduced by squeezing out zero elements, but with highly correlated data the matrix may not be sufficiently sparse for this to change the cost significantly. This can be mitigated with Fourier methods, wherein we calculate the misfit in frequency domain. As the misfit is calculated by individual frequency, the size of the covariance matrix can be reduced by the number of time samples. Beside the cut-off of high frequency bands, in frequency domain, the spectra can be resampled each two or three points to save computation and the storage for data covariance matrices. As part of our study, we apply several dimension-reduction methods to reduce the memory requirements.

MODIFYING THE MISFIT FUNCTION TO INCLUDE DATA COVARIANCE

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:

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

for the jth datum, where $\bar{d}$ 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

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

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.

We observe that this is a large matrix; if each trace has Nt samples, and there are Nr traces, its dimension $(Nt \times Nr)^2$. To reduce its effective size, we transform the data residuals and the misfit function as a whole into the frequency domain, i.e., $\mathbf{d}_{\text{obs}} \rightarrow \mathbf{d}_{\text{obs}}^{Fi}$, and $\mathbf{d}_{\text{pre}} \rightarrow \mathbf{d}_{\text{pre}}$, obtaining

$$E = \frac{1}{2N}\left[\left(\mathbf{d}_{\text{pre}}^{Fi} - \mathbf{d}_{\text{obs}}^{Fi}\right)^T \mathbf{C}_{\mathrm{D}_{Fi}}^{-1}\left(\mathbf{d}_{\text{pre}}^{Fi} - \mathbf{d}_{\text{obs}}^{Fi}\right)\right]. \tag{3}$$

Here, i is now an index labelling an individual frequency; the original misfit can be recovered by calculating (3) for each frequency and summing, but the problem at this stage can be made computationally less expensive with a frequency selection.

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

$$\frac{\partial \mathrm{E}(\mathbf{m})}{\partial m_i} = \frac{1}{N}\left(\frac{\partial \mathbf{d}_{\text{pre}}}{\partial m_i}\right)^T \mathbf{C}_{\mathrm{D}}^{-1}\left(\mathbf{d}_{\text{pre}} - \mathbf{d}_{\text{obs}}\right), \tag{4}$$

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.

SYNTHETIC EXAMPLES

We first applied the algorithm described above to inversion of part of the Marmousi model. The model is resampled to a 20m grid. 10 shots and 200 receivers were evenly placed near the surface of the model (Figure 2); two wells were selected, represented by the vertical lines, for monitoring the inversion results. A 10 Hz minimum phase wavelet was used to generate noise-free acoustic data, with a computing time window of 1s and sample interval of 2 ms. We then added different types of errors to the synthetic data to simulate inaccurate observed data. First, we added random noise of 1.7%, 7.6%, 15.9%, respectively, to the noise-free synthetic data in Figure 1. The inverted Vp models after 20 iterations are plotted in the Figure 3. The initial model were obtained by smoothing the true model.

We find that variously-scaled random noise within 20% of the signal does not exert a significant influence on the inversion result after 20 iterations. It is in fact difficult to find an obvious difference among the inverted models, suggestive that the random variations in the data did not distort the phase or amplitude information continuously.

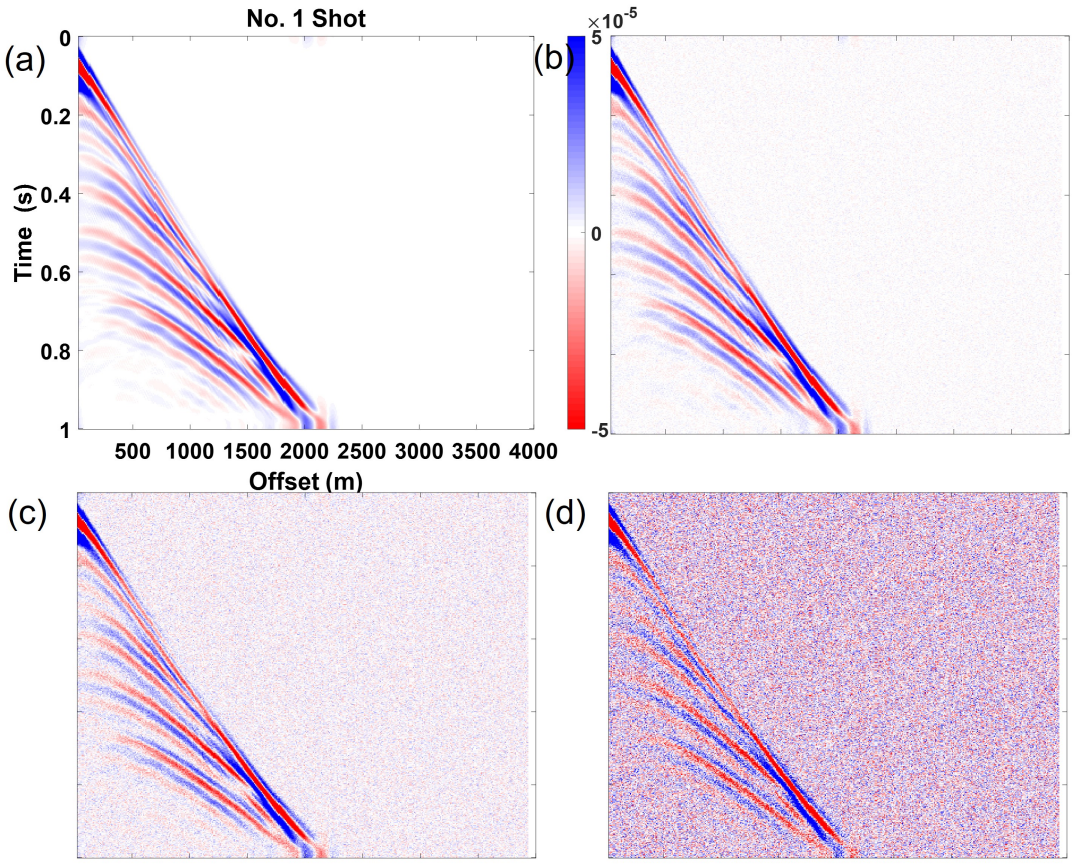

FIG. 1. Noise-free synthetic data and noisy data. (a) Noise-free data, (b) 1.7% noise, (c) 7.6% noise, (d) 15.9% noise.

Next, we added a combination of random and correlated noise. Correlated data error (Figure 13 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, as plotted in Figure 4. The correlated noises and noisy seismogram are shown in Figure 5

A first round of FWI was conducted using the conventional L2 norm misfit function. The results after 20 iterations with conventional FWI are plotted in Figure 6 (b) after 20 iterations. Panels (c) and (d) examine the differences between the inverted model and true model. General structural and velocity information are captured at locations A and B, but in (b) we observe indistinct boundaries and artifacts especially in the shallow parts of the model. The situation has deteriorated, evidently, and we attribute this to the presence of correlated noise in the data.

More detailed information is displayed in Figure 7. In Figure 7(b), the misfit decreased by 8% after 20 iterations, and the model misfit decreased by 10%, both insubstantial. In fact, we surmise that even the structures which can be observed in the model are there because the velocity range of the initial model is similar with that of the true model.

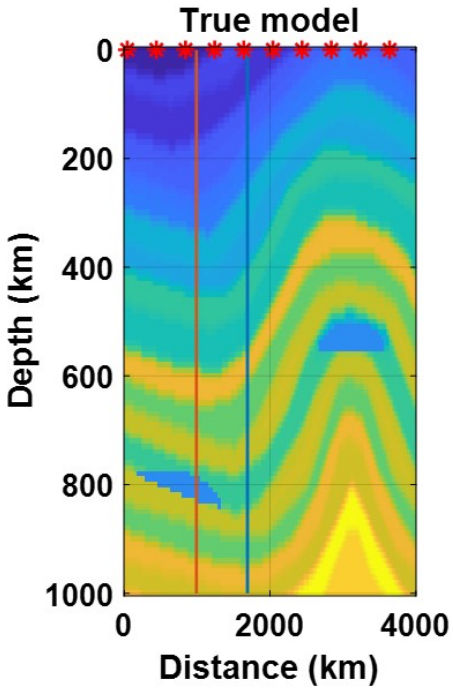

FIG. 2. True model and geometry.

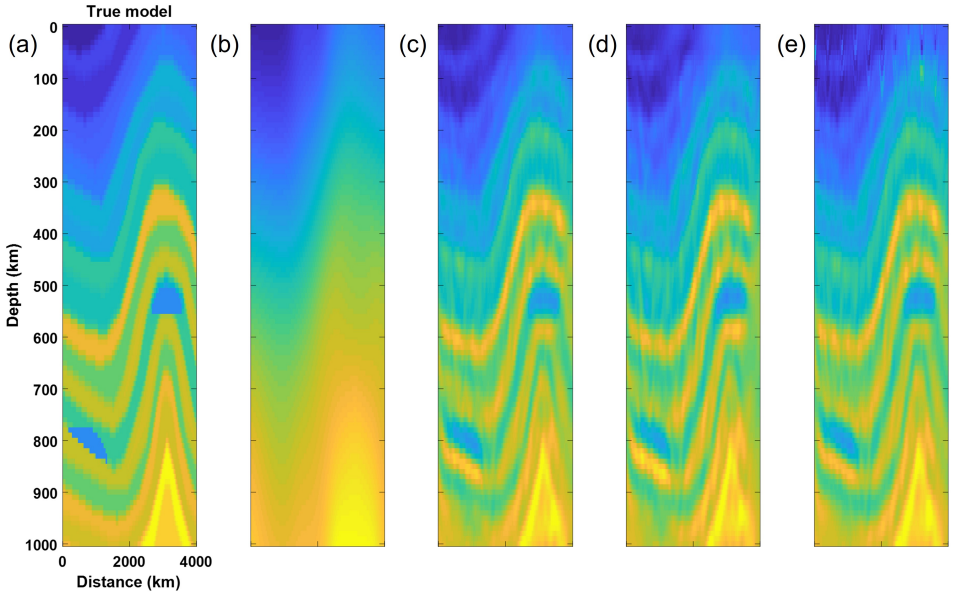

FIG. 3. Models and inversion results. (a) True model, (b) initial model, (c) inverted model with 1.7% noise, (d) 7.6% noise, and (e) 15.9% noise.

Next, we carried out a second pass of FWI with the data covariance matrix in the misfit. To reduce memory requirements, we transformed the data residual to frequency domain to compute the misfit. In frequency domain, the high frequency band (>=35 Hz) was cut off, as this part does not make conspicuous contributions to the signal. Besides, the data residual spectra were resampled by every three points along frequency, so that the storage requirements and computation cost for data covariance matrix have both dropped by at least

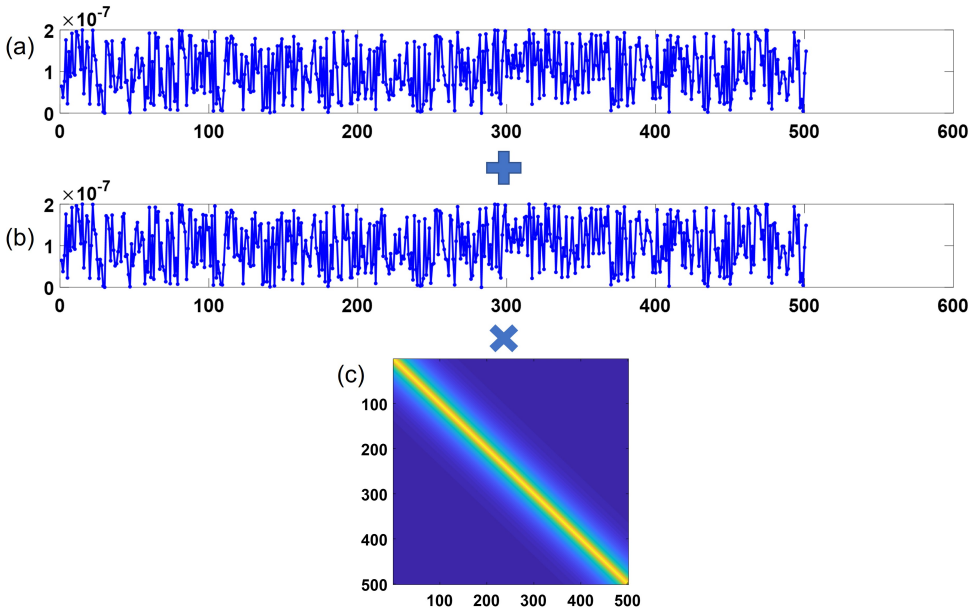

FIG. 4. The creation of complex noise. (a) Random noise sequence one, (b) sequence two, (c) correlation matrix.

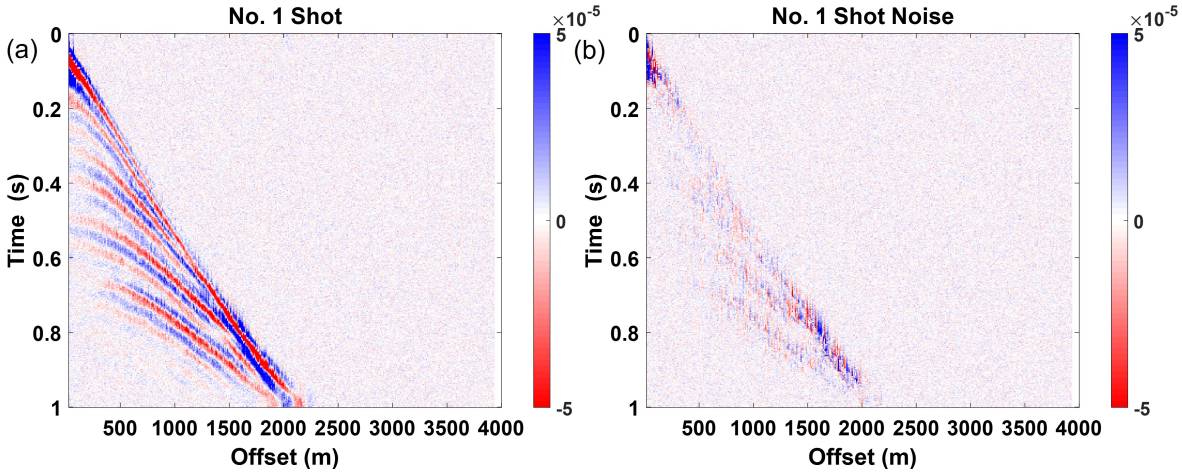

FIG. 5. Seismic data (shot 1) with complex noise. (a) Noisy observed data, (b) noise.

one third (Figure 8). Then before summing over frequency, we passed the intermediate calculation results back to time domain, so that the time-space operator could be adopted for the calculation of gradient.

The data covariance matrix was updated with each iteration. This way, data error from each iteration causes new weights and reduced contribution at the next iteration. In total, 40 iterations, with the first 20 involving the conventional misfit function, and second 20 working from the modified misfit function were carried out. For comparison, we conducted 40 iterations FWI using the conventional misfit function as the reference. The inversion result of the FWI with modified misfit is plotted in Figure 9. Compared with Figure 9(b), we observe an increased clarity in the layers, and boundary outlines, and significantly fewer artifacts. The reference model in Figure 9(d) also improved a lot in imaging, but some small

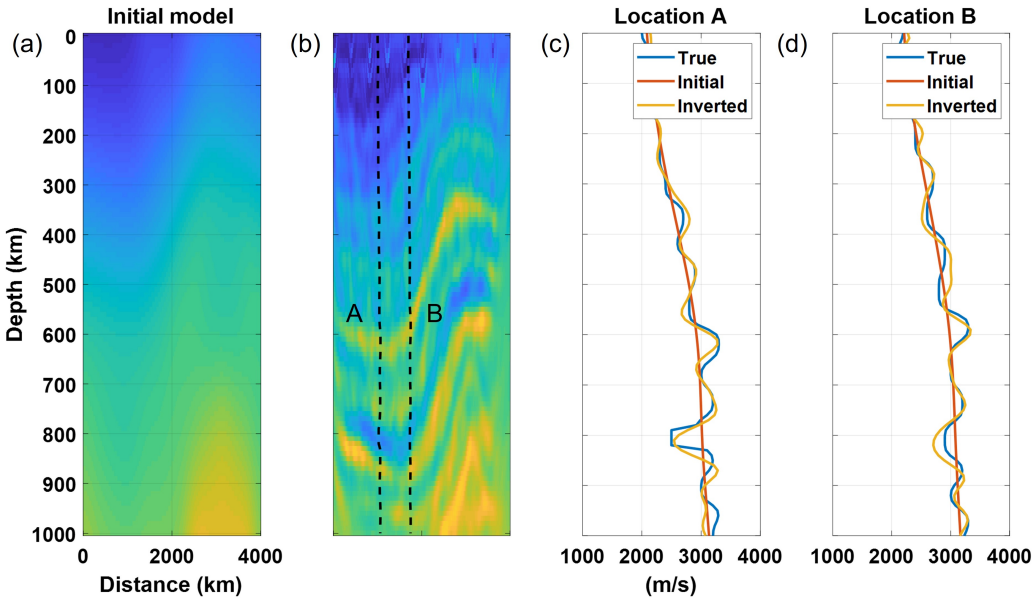

FIG. 6. Initial FWI result. (a) Initial model, (b) inverted model, (c) trace comparison at location A, (d) trace comparison at location B.

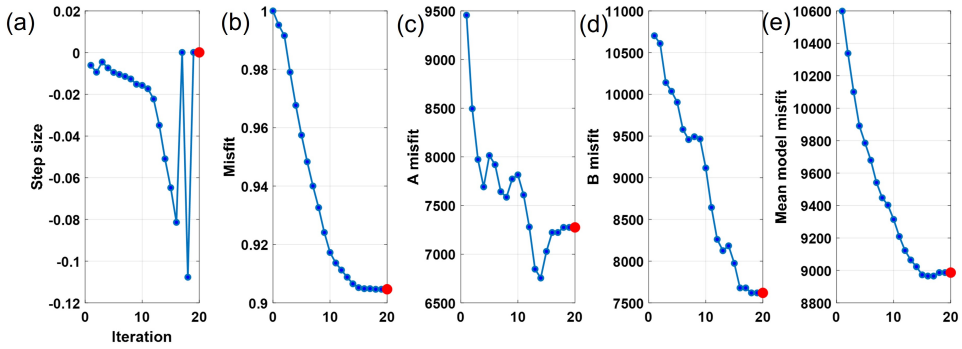

FIG. 7. Evolution of the inversion with iteration: (a) stepsize, (b) misfit, (c) error at location A, (d) error at location B, (e) total model error.

imperfections such as roughness in outlines can be found when compared with (c).

The misfit change of the second time FWI was shown in Figure 10. Though not much, the data misfit dropped to some extent.

The changes in data prediction were also examined. We randomly selected three traces, trace 90, trace 150, and trace 280 of the 5th shot to compare the amplitude and phase information in detail. In Figure 11, we compared the synthetic traces of the first pass of FWI and the second pass, with the noise-free seismic traces and noisy seismic traces included. The blue lines are the noise-free seismic traces, orange lines are the noisy seismic traces, and the yellow land purple lines are the synthetic seismic traces of the first and second passes respectively. Here we observe (arrows show particular examples) very significant differences, and clear evidence that the inversions have made use of the data covariance information to incorporate model structure which (though subtle) has not been introduced to reproduce noise.

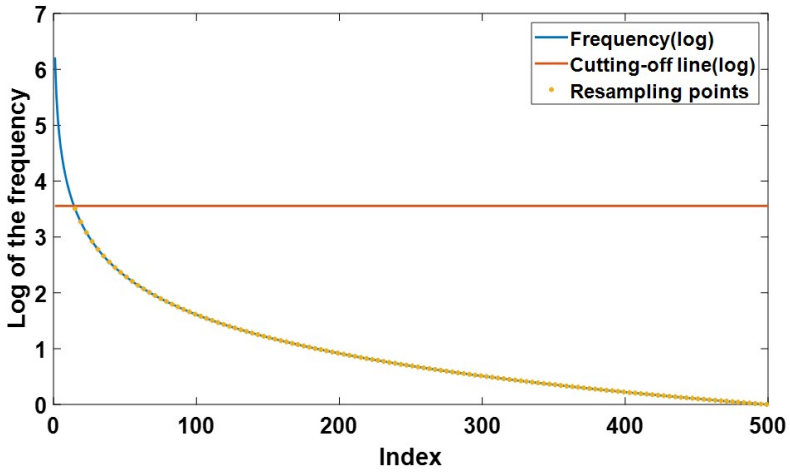

FIG. 8. The resampling and cut-off in frequency domain.

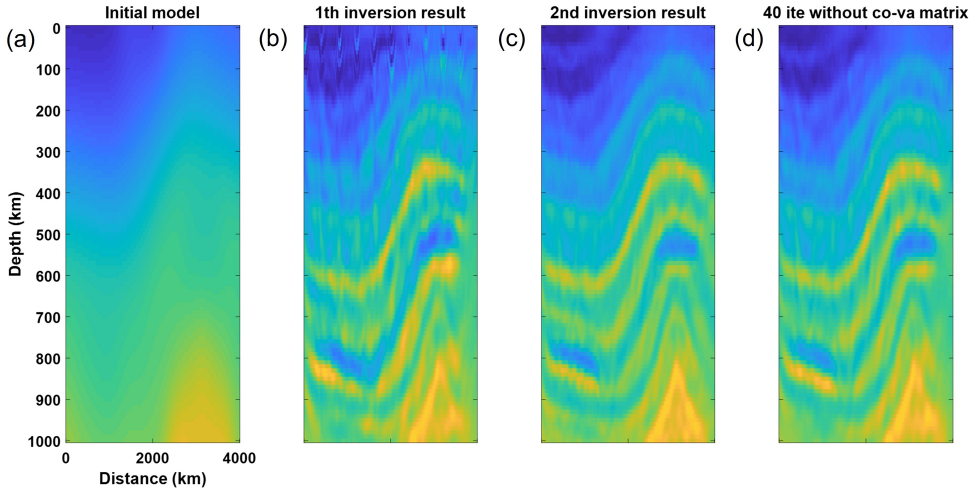

FIG. 9. Inversion results. (a) Initial model, (b) the inversion result of the conventional FWI after 20 iterations, (c) the inversion result of the modified FWI after 20 iterations, (d) the inversion result of the conventional FWI after 40 iterations.

Figure 12 are the comparisons with the conventional FWI after 40 iterations. Here we see few differences between the conventional FWI and modified FWI result. This is mainly because the inversion results of these two FWI are already good enough. The purple and yellow lines almost coincide with the true data (blue lines). In some small parts indicated by the arrows, we can find the modified FWI shows better performances than conventional FWI. The original noise and estimated noise are compared in Figure 13; it appears that the estimated noise captures much of the variability of the input noise.

DISCUSSION AND CONCLUSIONS

In this study, we apply a robust methodology to deal with the remnant noise, especially correlated noise, in seismic data within full waveform inversion. We incorporate the data covariance matrix, and estimate its elements from the data residuals of the conventional FWI result, into the misfit function. We observed that conventional FWI is more resistant

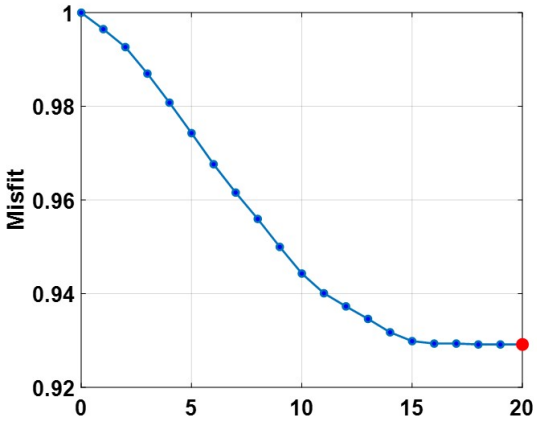

FIG. 10. Evolution of the modified misfit with iteration.

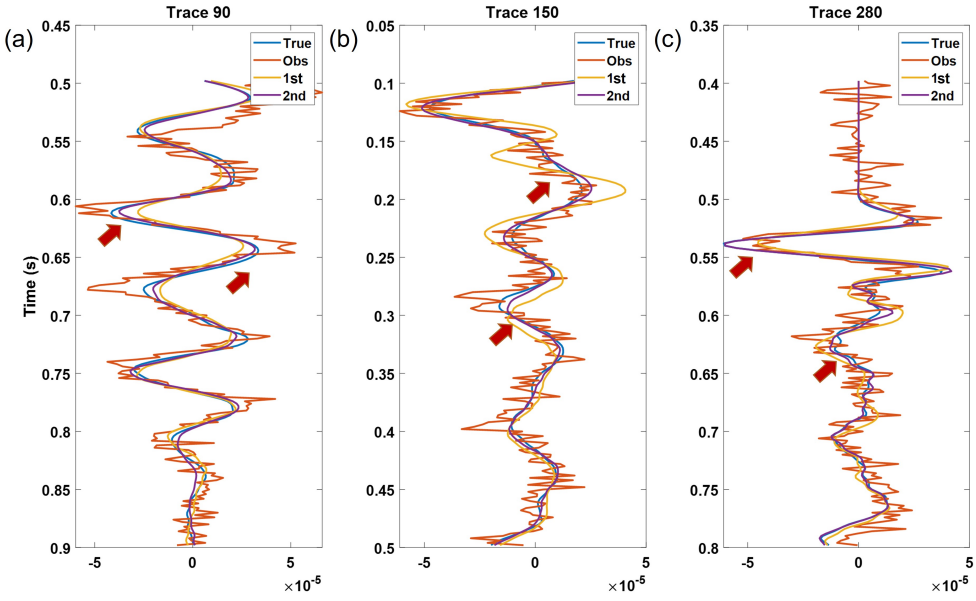

FIG. 11. Seismic trace comparison. (a) Trace 90, (b) trace 150, (c) trace 280.

to the influence of random noises (within 20 %), while the correlated noises impair image quality through distortion of the waveforms. Through incorporation of the data covariance matrix, different types of errors can be well estimated, and less certain regions of the data can be down-weighted in subsequent inversion iterations. Compared with conventional FWI results, the FWI with modified misfit yields more accurate model parameter values with a better estimation of noise. Ultimately, this method is designed to be applied to field FWI applications in monitoring, both onshore and offshore. So, an extension to elastic FWI and real data is required. This makes the size of the data covariance matrix a serious issue. The data covariance matrix is large, so to reduce the dimension of the data covariance matrix, we have set the problem up to involve a transformation of the residuals to the frequency domain, after which the size of data covariance matrix will be reduced by calculating with selected frequencies. Computational efficiency and frequency selection will have to be balanced

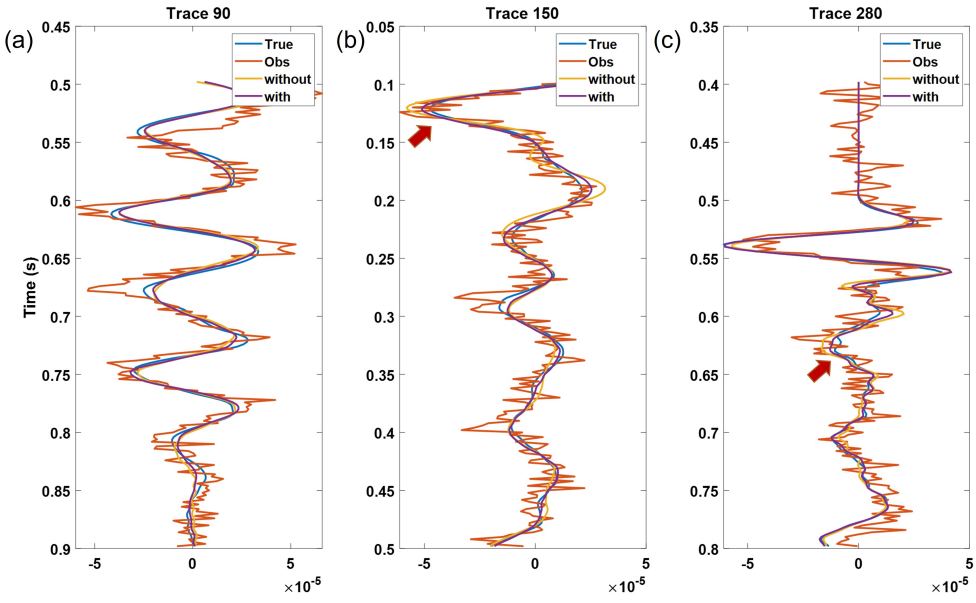

FIG. 12. Seismic trace comparison with conventional FWI. (a) Trace 90, (b) trace 150, (c) trace 280.

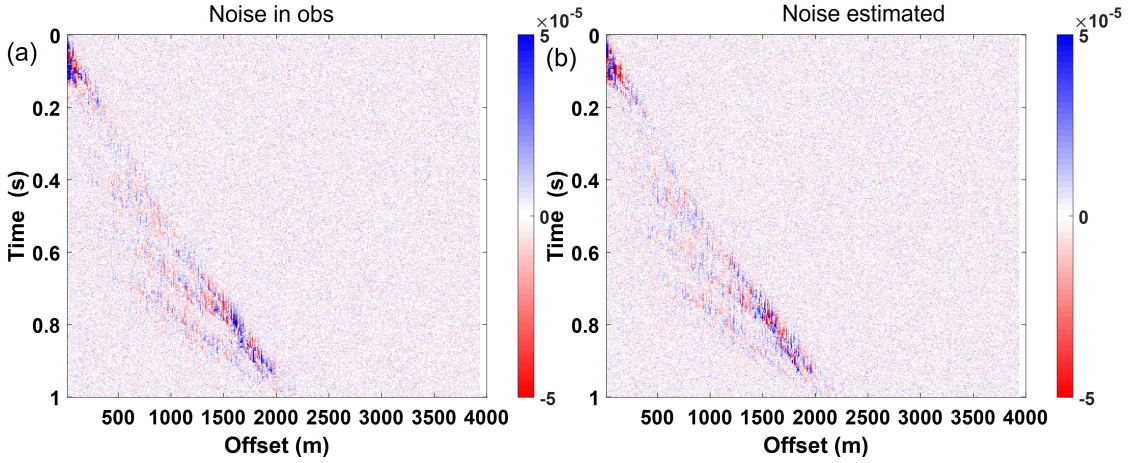

FIG. 13. Real noise and estimated noise. (a) Real noise, (b) estimated noise of the 2nd FWI.

with effectiveness in estimating noise, and the effort to introduce only model structures associated with signal. This will be the focus of the next steps in this research.

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. The first author was partially supported by an SEG scholarship.

REFERENCES

Bekara, M., and Van der Baan, M., 2009, Random and coherent noise attenuation by empirical mode decomposition: Geophysics, **74**, No. 5, V89–V98.

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.

Canales, L. L., 1984, Random noise reduction, *in* SEG Technical Program Expanded Abstracts 1984, Society of Exploration Geophysicists, 525–527.

Chen, K., and Sacchi, M. D., 2015, Robust reduced-rank filtering for erratic seismic noise attenuation: Geophysics, **80**, No. 1, V1–V11.

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.

Neelamani, R., Baumstein, A. I., Gillard, D. G., Hadidi, M. T., and Soroka, W. L., 2008, Coherent and random noise attenuation using the curvelet transform: The Leading Edge, **27**, No. 2, 240–248.

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

Yu, S., Ma, J., and Wang, W., 2019, Deep learning for denoising: Geophysics, **84**, No. 6, V333–V350.

Zhu, W., Mousavi, S. M., and Beroza, G. C., 2019, Seismic signal denoising and decomposition using deep neural networks: IEEE Transactions on Geoscience and Remote Sensing, **57**, No. 11, 9476–9488.