geoconvention
Calgary • Canada • June 20-22 2022

Reducing the influence of remnant noises on elastic FWI with misfit modification

Luping Qu, Scott Keating, and Kris Innanen, Xin Fu
University of Calgary (CREWES)

Summary

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 elastic 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. The data covariance matrices calculated from the data residuals, which were obtained from a first round of FWI, were generated and applied in an following iterative inversion. As the elastic FWI was conducted in frequency domain, the dimension size of the data covariance matrix is calculated by each frequency, and not likely to cause out-of-memory problem. Two types of noises including random noises and correlated noises were added to the true spectra and estimated in the modified FWI. The inversion results were compared with the results of the conventional FWI with the same iteration number.

Theory

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:

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

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

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

where N is the number of data, and $\mathbf{d}_{\mathrm{pre}}$ and $\mathbf{d}_{\mathrm{obs}}$ 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.

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

$$\frac{\partial \mathrm{E}(\mathbf{m})}{\partial m_i} = \frac{1}{N}\left(\frac{\partial \mathbf{d}_{\mathrm{pre}}}{\partial m_i}\right)^T \mathbf{C}_{\mathrm{D}}^{-1}\left(\mathbf{d}_{\mathrm{pre}} - \mathbf{d}_{\mathrm{obs}}\right), \quad (3)$$

where $\partial \mathbf{d}_{\mathrm{pre}}/\partial m_i$ 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 tests and Results

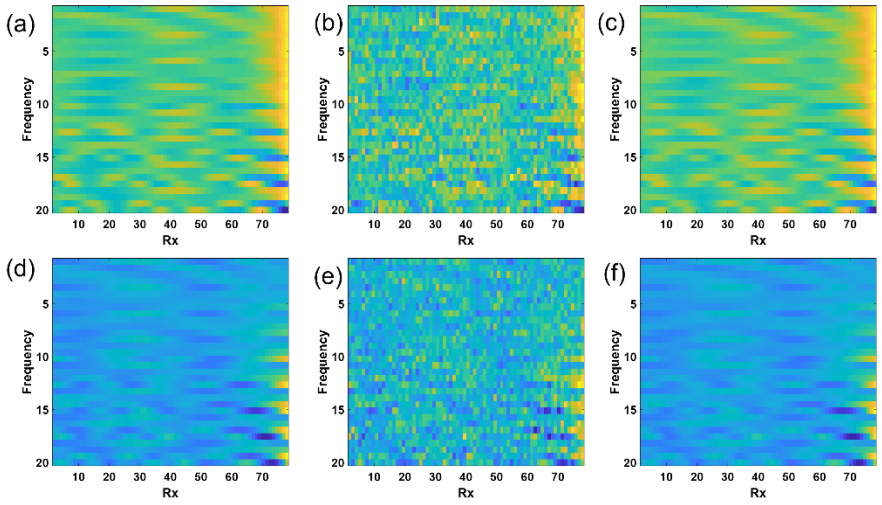

Figure 1: Noise-free synthetic data and noisy data. (a) True data (real part), (b) noisy data with a SNR equals 20 (real part), (c) synthetic data of the inverted model (real part), (d) true data (imaginary part), (e) noisy data with a SNR equals 20 (imaginary part), (f) synthetic data of the inverted model (imaginary part).

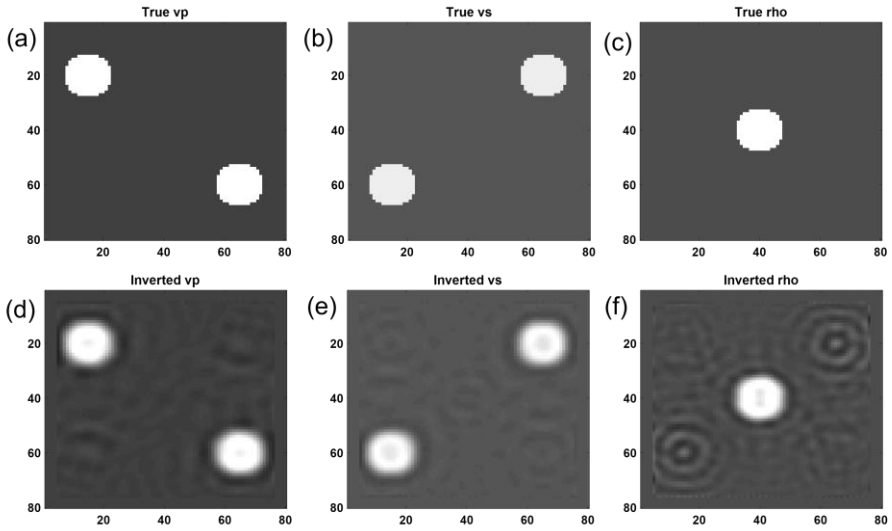

Figure 2: True models and inverted model without noise. (a) true Vp , (b) true Vs, (c) true density, (d) inverted Vp, (e) inverted Vs, (f) inverted density.

We first applied the algorithm described above to a toy model with two circle anomalies. The model spatial interval is 10 m. For the first test, two source lines and two receiver lines were evenly placed on the top and bottom of the model. In order to check how the noises can affect the inversion results, we added Gaussian distributed random noises with the SNR equals to 10 dB, 20 dB, 30 dB, respectively, to the noise-free complex spectra in frequency domain shown in Fig.1. The noise-free inversion results were shown in Fig.2 for reference. The initial models were the background model without velocity anomalies. The inverted models of noisy data after 20 iterations with truncated newton inversion method were shown in the Fig.3.

From our inversion results, we found with different scaled random noises, the imaging quality of accuracy of inverted model parameters were influenced. When the SNR is equal to or smaller than 20 dB (Fig.2d-f), there are obvious artifacts in the inverted models. When the SNR is equal to 10 dB, the outline of the circle anomalies is not clear. The imaging quality is seriously influenced by the noises in data. From the comparison with the noise-free inversion results (Fig.2), these figures (Fig.3) illustrate the random noises in large magnitude have impacts on the accuracy of inversion results.

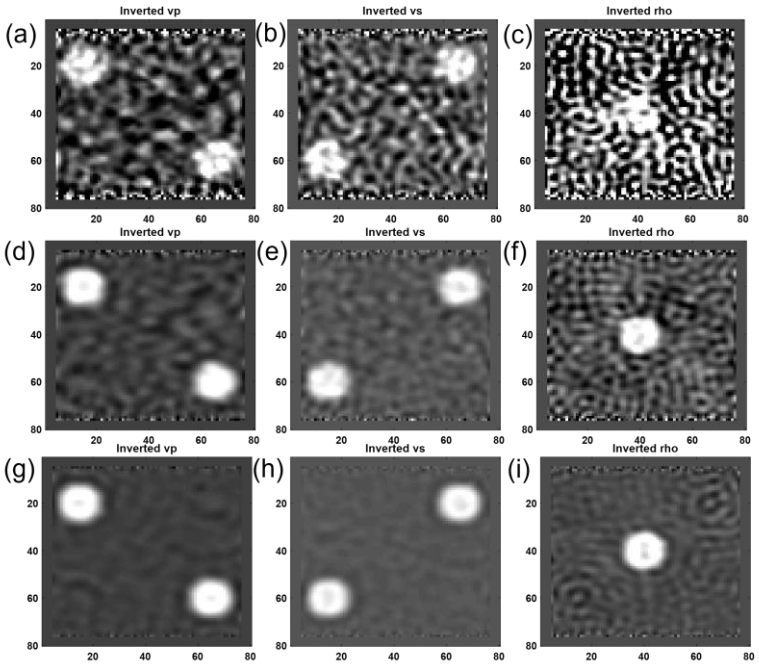

Figure 3: Inversion results of seismic data with SNR equals to 10 dB, 20 dB, 30 dB. (a) inverted Vp with SNR equals 10 dB, (b) inverted Vs with SNR equals 10 dB, (c) inverted density with SNR equals 10 dB, (d) inverted Vp with SNR equals 20 dB, (e) inverted Vs with SNR equals 20 dB, (f) inverted density with SNR equals 20 dB, (g) inverted Vp with SNR equals 30 dB, (h) inverted Vs with SNR equals 30 dB, (i) inverted density with SNR equals 30 dB.

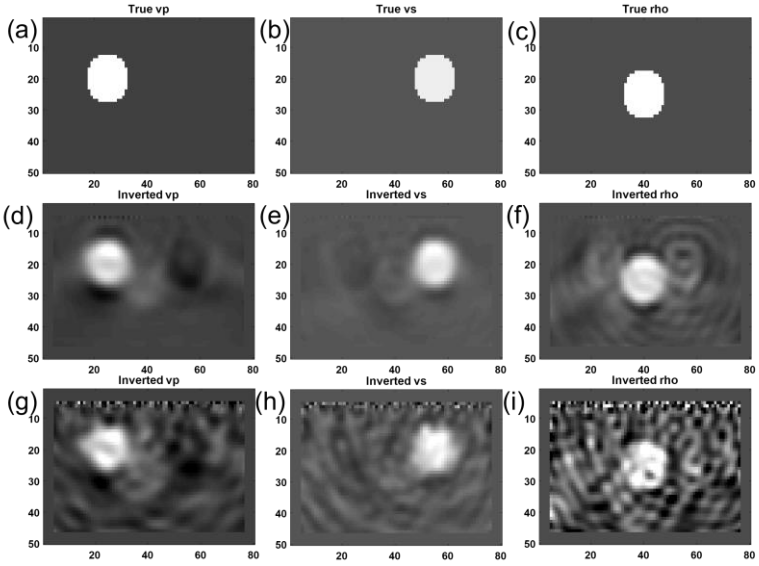

Figure 4: True models and inverted models. (a) True Vp, (b) true Vs, (c) true density, (d) inverted Vp of noise-free data, (e) inverted Vs of noise-free data, (f) inverted density of noise-free data, (g) inverted Vp of noisy data with SNR equals 16 dB, (h) inverted Vs of noisy data with SNR equals 16 dB, (i)inverted density of noisy data with SNR equals 16 dB.

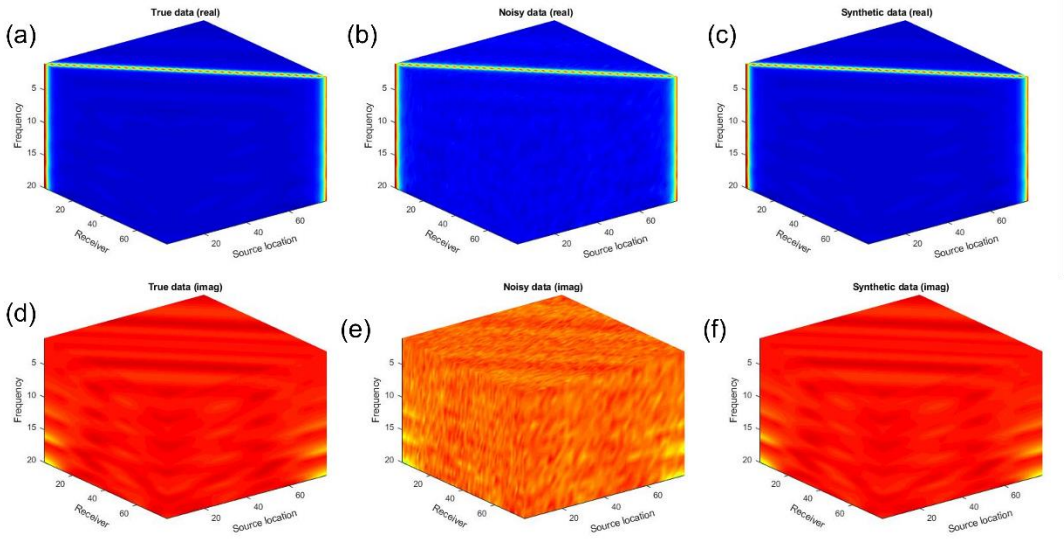

Figure 5: The spectra of true data, noisy data, and synthetic data. (a) True spectra of the real part, (b) noisy spectra of the real part, (c) synthetic spectra of the real part generated by the inverted model, (d) true spectra of the imaginary part, (e) noisy spectra of the imaginary part, (f) synthetic spectra of the imaginary part generated by the inverted model.

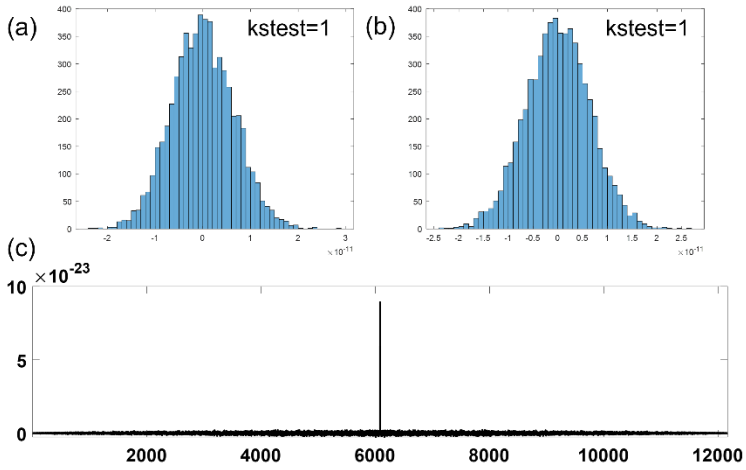

Figure 6: The histogram and autocovariance of data residuals. (a) Histogram of real part data residual, (b) histogram of imaginary part data residual, (c) autocovariance of the data residual in real.

In the test above, we found the geometry of the acquisition is a bit ideal. Therefore, we reduced the model size, and only adopted the surface excitation and reception geometry in the following tests. We added the random noises with SNR equals to 16 dB to the data simulated from the new model shown in Fig.4 a-c. The true synthetic data and noisy data are shown in Fig.5. In Fig.4, d-f are inverted models of the noise-free data after 20 iterations using the truncated newton method, and g-i figures are inverted models of the noisy data after 20 iterations. In d-f figures, we found the inverted models are close to the true models, though slight cross-talk artifacts exist. By

contrast, in g-i figures, there are more ripple-shaped artifacts in the background and Mosaic-like artifacts at the top boundary. It is conceivable to obtain these results with the mottled noisy data in Fig.5 b and e. Besides, we can find the density inversion result is not as good as the Vp and Vs both with or without noises. We generated the synthetic data using the inverted model in Fig.4 g-i. The obtained spectra are displayed in Fig.5 c and f. The synthetic spectra are similar with the true spectra in figures a, and d.

Based on the inversion results above, we implemented the second time FWI with the modified misfit function. The first step is to calculate the data covariance matrix which would be incorporated in the penalty for the first iteration of the second FWI. This matrix is estimated using the data residual of the first FWI. The histogram distribution of the real and imaginary part of the data residual were plotted in Fig.6 a, and b. The Kstest results for both real and imaginary part residuals are equal to one, which means the data residuals follow Gaussian distribution as what we defined. Then, the data residuals were auto-covarianced and toeplitzed to generate the covariance matrix. From the autocovariance (Fig.6 c), we can find the diagonal line has a large value while the values of other points are close to zero. This also further proves the data residuals are mainly independent Gaussian noises.

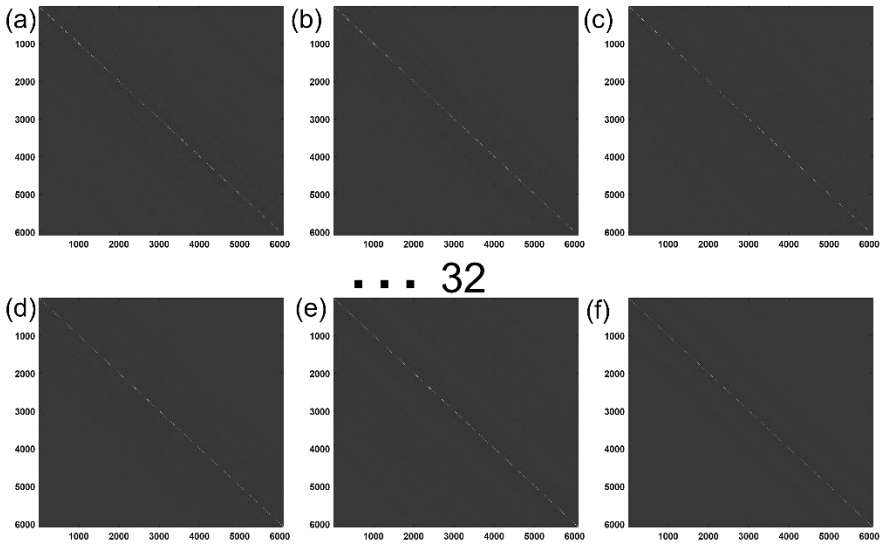

Figure 7: The covariance matrices of 32 frequency bands. (a) to (f) are the matrices of the first 6 bands.

As the elastic FWI were conducted in frequency domain, the misfit was calculated each frequency by each frequency with a total number of 32. Thus, 32 data covariance matrices were generated, as shown in Fig.7. The non-diagonal values in the matrices are close to zero.

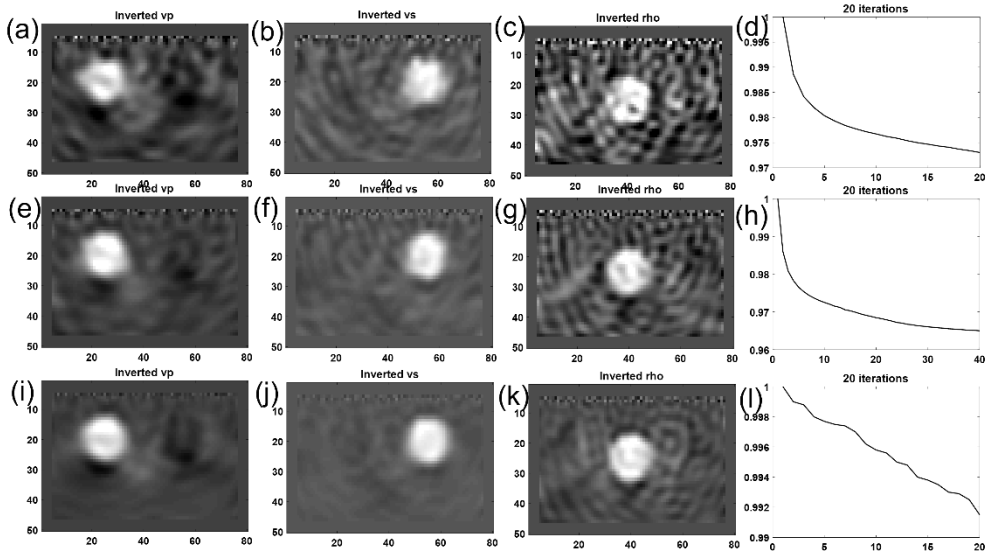

Figure 8: Inversion results. (a) Inverted Vp model of the first FWI, (b) Inverted Vs model of the first FWI, (c) Inverted density model of the first model, (d) the misfit change of the first FWI, (e) inverted Vp model of the conventional FWI using 40 iterations, (f) inverted Vs model of the conventional FWI using 40 iterations, (g) inverted density model of the conventional FWI using 40 iterations, (h) the misfit change of the conventional FWI with 40 iterations, (i) inverted Vp model of the second FWI with the modified misfit, (j) inverted Vs model of the second FWI with the modified misfit, (k) inverted density model of the second FWI with the modified misfit, (l) the misfit change of the second FWI with modified misfit.

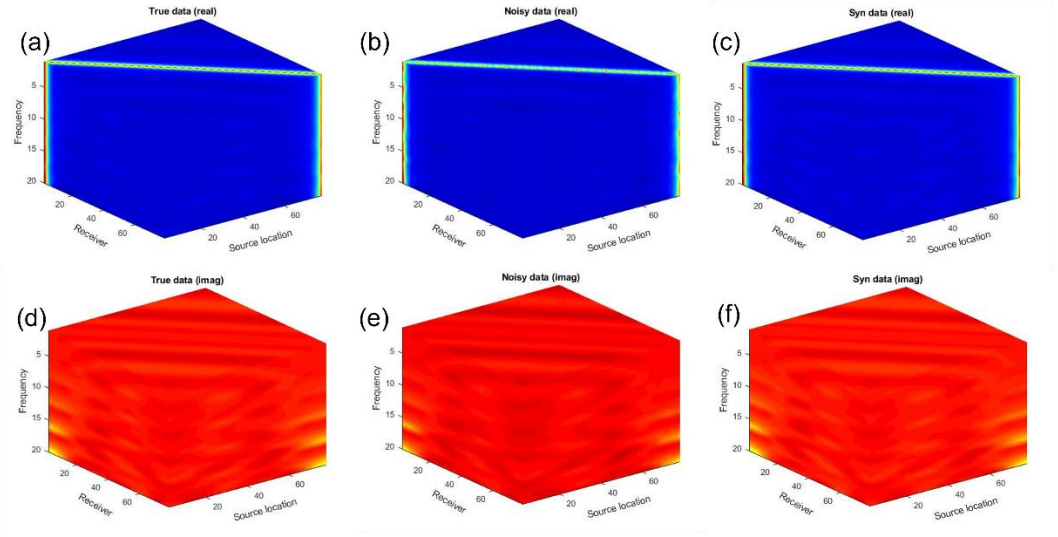

Figure 9: The spectra of true data, noisy data, and synthetic data. (a) True spectra of the real part, (b) noisy spectra of the real part, (c) synthetic spectra of the real part generated by the inverted model, (d) true spectra of the imaginary part, (e) noisy spectra of the imaginary part, (f) synthetic spectra of the imaginary part generated by the inverted model.

Then, we conducted the second FWI with the modified misfit function for 20 iterations and compared the results with conventional method using totally 40 iterations. The inversion results are displayed in Fig.8. The first row is the first-time conventional FWI results, the second row are the conventional FWI results using totally 40 iterations, and the third row are the second FWI results using modified misfit. From the comparison of the first two rows, we found though with more iterations, the results of 40 iterations seem no better than the results of 20 iterations. More rippled artifacts appear in the background model, especially in the inverted density model. When compare the third row with the first two rows, the results of the new method are better than the first FWI results or conventional method results after the same iteration number. The improvements are mainly reflected by the clear outline of anomalies of less artifacts in the background model.

Next, we added a more complex type of noise onto the synthetic data, which is a combination of random noise with correlated noise. Correlated data error was generated by multiplying a Gaussian random array with the Cholesky decomposition of a constructed data error covariance matrix with non-zero decaying off-diagonal terms. We combined another random noise with this correlated noise to obtain the complex noise. The generated noisy data spectra were plotted in Fig.9(b) and (e). Their differences compared with the true spectra can be observed from the edges and top faces of the cubes.

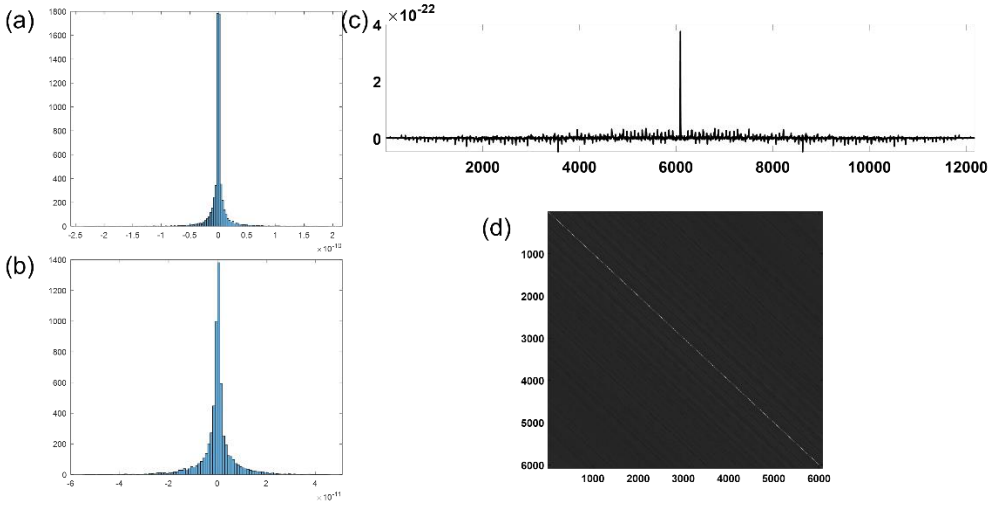

Figure 10: Histogram of data residuals and covariance matrix. (a) histogram of the real part data residual, (b) histogram of the imaginary part data residual, (c) autocovariance of the data residual, (d) the toeplitzed data covariance matrix.

Next, we estimated the data covariance matrices using the same procedure above. From the histogram of data residuals in real and imaginary (Fig.10 a and b), we can find the data residuals do not follow a Gaussian random distribution. In the autocovariance figure (Fig.10 c), there are some small spikes in both sides, which are non-zero. This means the values at different points in the data residuals are correlated to some degree, which can also be reflected by the non-zero values of the non-diagonal points in the covariance matrix (Fig.10 d).

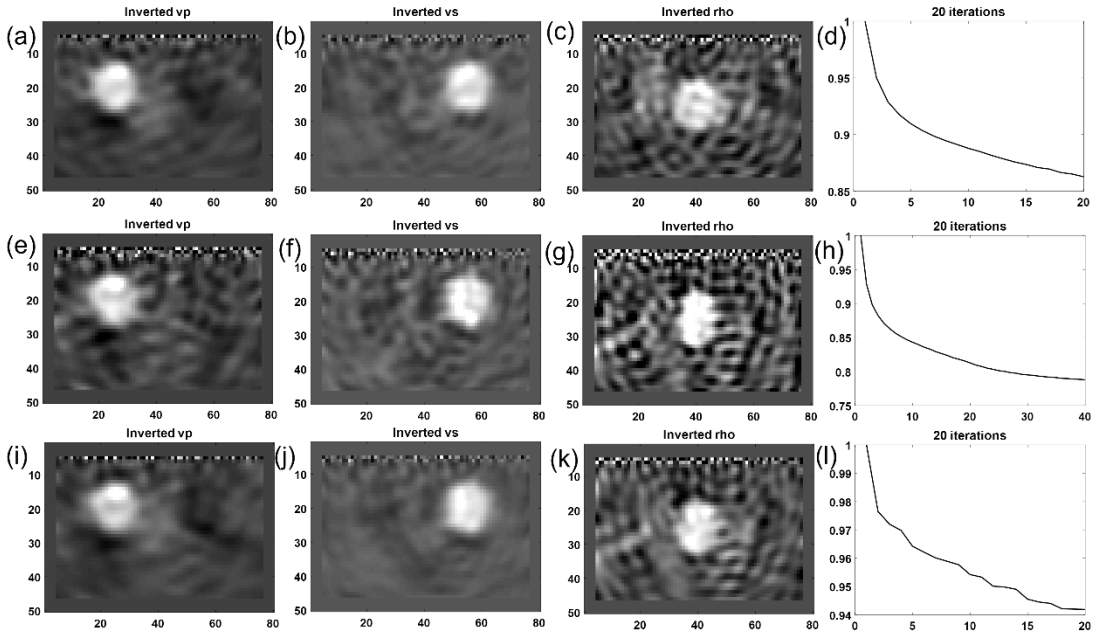

Figure 11: Inversion results. (a) Inverted Vp model of the first FWI, (b) Inverted Vs model of the first FWI, (c) Inverted density model of the first model, (d) the misfit change of the first FWI, (e) inverted Vp model of the conventional FWI using 40 iterations, (f) inverted Vs model of the conventional FWI using 40 iterations, (g) inverted density model of the conventional FWI using 40 iterations, (h) the misfit change of the conventional FWI with 40 iterations, (i) inverted Vp model of the second FWI with the modified misfit, (j) inverted Vs model of the second FWI with the modified misfit, (k) inverted density model of the second FWI with the modified misfit, (l) the misfit change of the second FWI with modified misfit.

The inversion results are shown in Fig.11. As explained for Fig.8, the first row is the first-time conventional FWI results, the second row are the conventional FWI results using totally 40 iterations, and the third row are the second FWI results using modified misfit. Same as above, the results after 40 iterations seem have more artifacts than results after 20 iterations. The results of the modified FWI results have less artifacts than the results of conventional method.

Comparing the spectra cubes of the synthetic data generated from the first FWI and modified FWI, it is hard to observe the difference with naked eyes. Therefore, these figures are not shown here. From the inversion results comparison above, improvements in imaging have been found already.

Conclusions

In this study, we develop a new methodology to deal with the remnant noises especially the correlated noise in seismic data. This is implemented through incorporating the data covariance matrix, which is estimated from a conventional FWI result, into the misfit function. Based on the experiments, we found the random noises in large scale have impacts on the accuracy of inversion results. Generally, seismic data with random noises are more resistant to the influence of noises than the seismic data with correlated noises. Through estimating the noises using the data covariance matrix, difference types of errors can be well estimated and suppressed in the

signal during inversion. Compared with the conventional method, the new methods yield better imaging with less artifacts.

Discussion

As this method is targeted for solving problems in real condition, an extension to real data applications are required. The covariance matrix is large in size. Some sparsification methods can be adopted to reduce the storage demand. Another issue we have been concerned is the local minimum problem. With the incorporation of the data covariance matrix, the points with large data residual values will be down-weighted in the following iterations. The differences between the seismic data generated by true model and inverted model will be diminished. This means the inversion is more likely to converge to the previous model. When trapped in a region, it has smaller possibility to jump out of the local minimum. A global optimization method may mitigate this problem to some extent, but for a gradient based inversion method like the truncated newton, this can be a problem. This can also be discussed from another respective. For the first time FWI, the inverted models are not exactly true models, and the data residuals are not exactly noises. The second time FWI with modified misfit functions in essence diminished the contribution of the data residuals rather than true noises, this may lead to erroneous inverted models due to solution nonuniqueness. Therefore, it is necessary to figure out solutions to deal with the local minimum problem of this method.

Acknowledgements

We thank Jan Dettmer for the inspiration of this research and valuable conversations. 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

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.

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.

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