

# Incorporating estimates of data covariance in elastic FWI to combat random and correlated noise

Luping Qu, Scott Keating, Kristopher A. Innanen, and Xin Fu, University of Calgary

## SUMMARY

As field data applications of FWI increase, dealing with both random and coherent noise in seismic data, and the artifacts they create in FWI models, becomes increasingly important; noise suppression or estimation is also increasingly important when we transition to elastic-multiparameter inversion from acoustic approximations. In this study, we analyzed the influence of random and correlated noise on the estimation of model parameter  $Vp$ ,  $Vs$ , and density, and, to mitigate their influence, we adopted a two-stage inversion approach, whose second stage involves a modified FWI misfit. The data covariance matrix is calculated from data residuals obtained from an initial run of FWI, and this is incorporated into the misfit function for a second run. With the elastic FWI conducted in the frequency domain, and the data covariance matrix consequently calculated frequency by frequency, the approach, though not computationally inexpensive, places reasonable demands on memory and storage. Random and correlated noise were examined and estimated, and inversion results were compared with those of otherwise identical conventional FWI runs. The bootstrap approach to inclusion of data covariance estimates in FWI appears to be stable, and to have a strong positive impact especially for correlated data noise.

## INTRODUCTION

Noise in full waveform inversion (FWI) applications refers to variations in the data which are either (1) not meaningfully predictable with the underlying simulation equations or the boundary conditions being used, or (2) predictable, but in a manner too nonlinearly sensitive to medium property changes to be stably used. In many applications, especially those on land, this encompasses environmental noise, ground roll / surface waves, and multiples. Seismic data consequently contain complex combinations of coherent and incoherent, as well as stationary and nonstationary noise. To obtain interpretable seismic images, seismic data must be carefully de-noised. In real cases, however, noise cannot be completely suppressed, and for FWI to proceed we would ideally build into the iterative process weights which accommodate it. Incorporation of data covariance in the definition of the misfit is a clear avenue for this, and we have carried out a systematic analysis of the response of 2D frequency-domain elastic FWI to noise, and adapted it into a two stage procedure, with the second stage involving a misfit with data covariance estimated during the first stage.

Noise impacts elastic parameter updates in complicated ways, differing for each parameter. P- and S-waves have different amplitudes, and therefore the impact of a single noise level on parts of the inversion concerned with each can be quite different. Likewise P-waves typically have larger dominant wavelengths than S-waves, as well as different apparent slowness in data, meaning inversion will tend to respond to coherent noise very differently across the two wave types. Our goal is to arrive at some relatively broad conclusions about these varied effects through simulations / synthetic analysis. To add noise in the frequency domain comparable to noise added in the time domain, we assume the same signal-to-noise ratio in both domains, and construct the real and imaginary parts of the noise separately, with half the variance of the desired total.

The covariance matrix characterizes statistical relationships between each pair of elements in a provided vector. The variances appear on the diagonal and inter-parameter correlations appear in the off-diagonal elements. A misfit function which includes the covariance matrix introduces weights giving more or less relevance to data residuals, based on this information. 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 our study, this misfit function is constructed using simulated data, on top of which additionally simulated *remnant noise* (meaning, noise left over after de-noising procedures are complete), has been added, as the observed data, and clean data simulated with elastic forward modelling in the current model iterate as the predicted data. When the observed data are contaminated by high levels of noise, that noise will tend to produce model artifacts, or false structures added to the model in order to explain the noise. This can in principle be mitigated, if estimates of the noise are available, by incorporating the data covariance matrix into the misfit function (Cai and Zelt, 2019). We here fit noise estimation into FWI by incorporating two stages of FWI, one in which the noise is estimated, and the other in which it is included in the misfit function. We observe that, in frequency domain FWI, this process of inclusion of the data covariance can occur frequency by frequency, such that, provided source and receiver numbers are not too large, memory issues can be straightforwardly avoided.

## METHODS AND SYNTHETIC RESULTS

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 independent, the data covariance matrix can be approximated as diagonal,  $\mathbf{C}_D^{-1} = \sigma^2 \mathbf{I}$ , where  $\mathbf{I}$  is the identity matrix and  $\sigma$  is the standard deviation of the random error. In this case, as  $\sigma$  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 pass 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:

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

for the  $j$ th 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  $\mathbf{C}_D$ . The FWI misfit function incorporating the data covariance matrix  $\mathbf{C}_D$  is

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

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 of the 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  $i$ th model parameter is

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

## Elastic FWI with misfit modification

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.

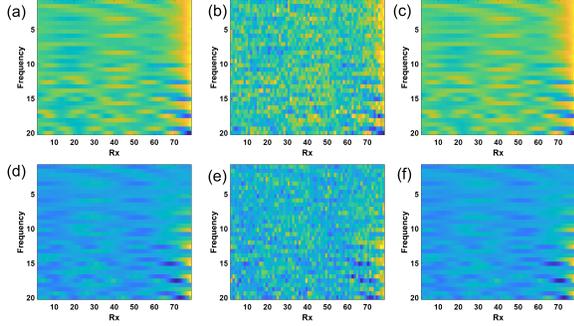


Figure 1: Noise-free synthetic data and noisy data. (a) True data (real part), (b) noisy data with a SNR equals 20 (real part), (c) synthetic data of the inverted model (real part), (d) true data (imaginary part), (e) noisy data with a SNR equals 20 (imaginary part), (f) synthetic data of the inverted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

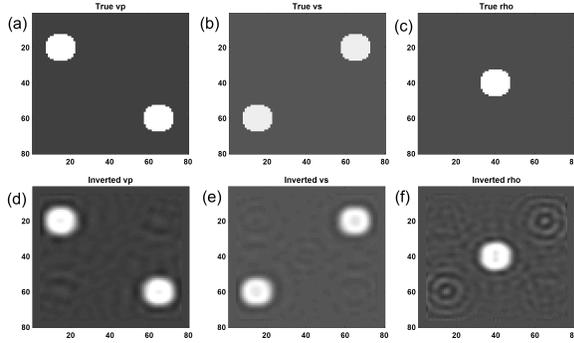


Figure 2: True models and inverted models without noise. (a) true Vp, (b) true Vs, (c) true density, (d) inverted Vp, (e) inverted Vs, (f) inverted density.

We first applied the algorithm described above to a toy model with two circle anomalies. The model spatial interval is 10m. For the first test, to reduce acquisition effects on model estimation, source lines and receiver lines were placed on the top and bottom of the model. We added Gaussian random noise with the SNR of 10 dB, 20 dB, 30 dB, respectively, to the noise-free complex spectra in the frequency domain (see Fig.1). The noise-free inversion results are shown in Fig.2 for reference. The initial models were the homogeneous background models without velocity anomalies. The inverted models from noisy data after 20 iterations with truncated Newton inversion method were shown in the Fig.3.

We observed a strong dependence of the accuracy of inverted model parameters on different scales of random noise. 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.

We then adjusted the experiment to involve a more realistic acquisition, reducing the model size, and acquiring data from surface sources

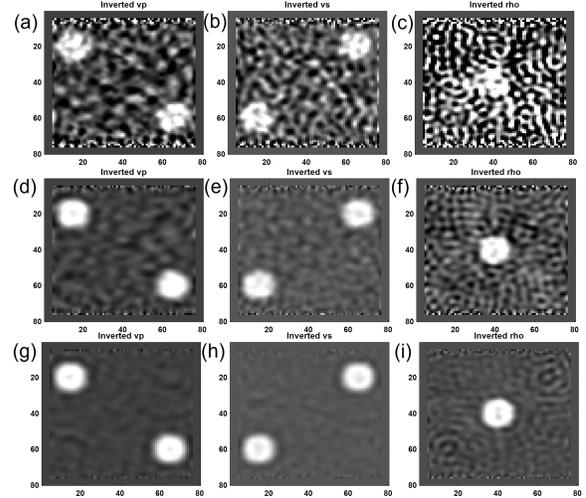


Figure 3: Inversion results of seismic data with SNR equals to 10 dB, 20 dB, 30 dB. (a) inverted Vp with SNR equals 10 dB, (b) inverted Vs with SNR equals 10 dB, (c) inverted density with SNR equals 10 dB, (d) inverted Vp with SNR equals 20 dB, (e) inverted Vs with SNR equals 20 dB, (f) inverted density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

and receivers only. We added random noise with SNR of 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 oscillatory artifacts in the background and mosaic-like artifacts at the top boundary. The density inversion result was poorer than the Vp and Vs both with and without noise. Spectra from the predicted data from the inverted model in Fig.4 g-i are displayed in Fig.5 c and f; simulated data are similar to the true spectra in panels a-d.

Based on the inversion results above, we next organized for a second pass of FWI, to occur with a modified misfit function. The first step was to calculate the data covariance matrix which would be incorporated in the penalty for the first iteration of the second FWI. This matrix was estimated using the data residuals produced during the first pass of FWI. The histogram of the real and imaginary parts of the data residuals are 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 we defined. Then, the data residuals were cross-correlated and given a Toeplitz structure to generate the covariance matrix. From the autocovariance (Fig.6 c), we can find the diagonal contains relatively large values while the values of other points are close to zero. This reflects the fact that the data residuals mainly include independent Gaussian noise.

As the elastic FWI was conducted in the 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 all of the matrices are close to zero.

The second pass of FWI was conducted with the modified misfit function, for 20 iterations; we compared the results thus obtained with the

## Elastic FWI with misfit modification

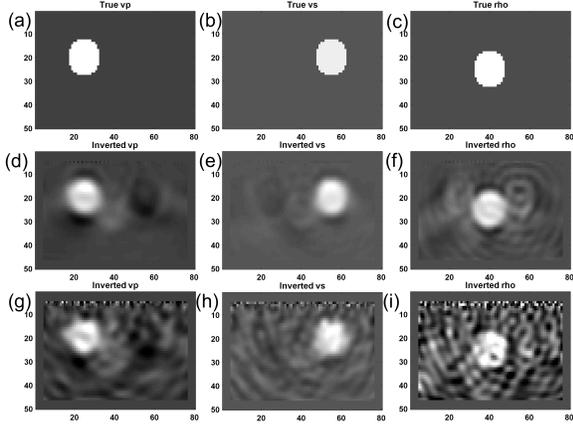


Figure 4: True models and inverted models. (a) True  $V_p$ , (b) true  $V_s$ , (c) true density, (d) inverted  $V_p$  of noise-free data, (e) inverted  $V_s$  of noise-free data, (f) inverted density of noise-free data, (g) inverted  $V_p$  of noisy data with SNR equals 16 dB, (h) inverted  $V_s$  of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

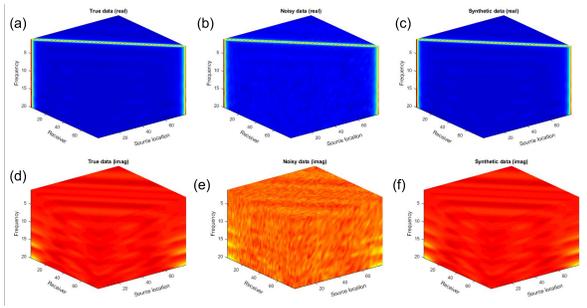


Figure 5: The spectra of true data, noisy data, and synthetic data. (a) True spectra of the real part, (b) noisy spectra of the real part, (c) synthetic spectra of the real part generated by the inverted model, (d) true spectra of the imaginary part, (e) noisy spectra of the imaginary part, (f) synthetic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

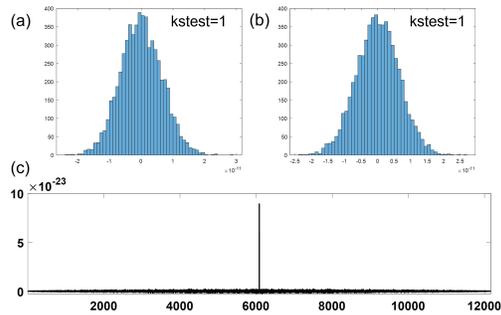


Figure 6: The histogram and autocovariance of data residuals. (a) Histogram of real part data residual, (b) histogram of imaginary part data residual, (c) autocovariance of the data residual in real.

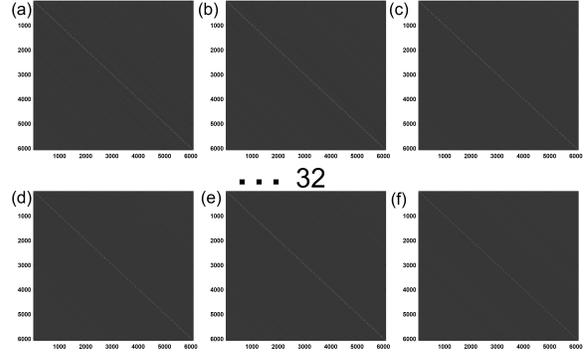


Figure 7: The covariance matrices of 32 frequency bands. (a) to (f) are the matrices of the first 6 bands.

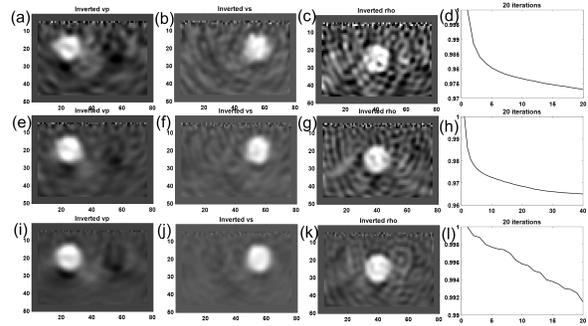


Figure 8: Inversion results. (a) Inverted  $V_p$  model of the first FWI, (b) Inverted  $V_s$  model of the first FWI, (c) Inverted density model of the first model, (d) the misfit change of the first FWI, (e) inverted  $V_p$  model of the conventional FWI using 40 iterations, (f) inverted  $V_s$  model 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  $V_p$  model of the second FWI with the modified misfit, (j) inverted  $V_s$  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 the modified misfit.

conventional method with 40 iterations. The inversion results are displayed in Fig.8. The first row are 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 oscillatory artifacts appear in the background model, especially in the inverted density model. Comparing the third row with the first two rows, we observe that 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 the inversion structures, and fewer 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 com-

## Elastic FWI with misfit modification

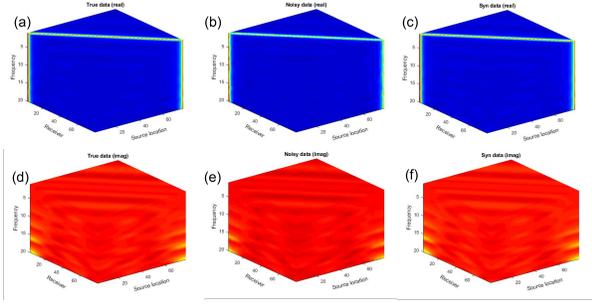


Figure 9: The spectra of true data, noisy data, and synthetic data. (a) True spectra of the real part, (b) noisy spectra of the real part, (c) synthetic spectra of the real part generated by the inverted model, (d) true spectra of the imaginary part, (e) noisy spectra of the imaginary part, (f) synthetic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

bin additional random noise with this correlated noise to obtain the final complex noise to be added. The noisy data spectra are 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.

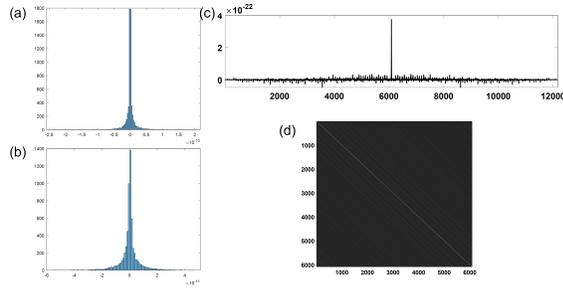


Figure 10: Histogram of data residuals and covariance matrix. (a) histogram of the real part data residual, (b) histogram of the imaginary part data residual, (c) autocovariance of the data residual, (d) the toeplitzed data covariance matrix.

Next, we estimated the data covariance matrices using the same procedure above. From histograms of the real and imaginary parts of the data residuals (Fig.10 a and b), we observe that the data residuals, as expected, do not follow Gaussian distributions. In the autocovariance (Fig.10 c), small spikes are observed on both sides, meaning different regions of the data residuals are correlated to some degree. This is also reflected in the non-zero values of the off-diagonal elements of the covariance matrix (Fig.10 d).

The inversion results are shown in Fig.11. As in the case of Fig.8, the first row are 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 to 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.

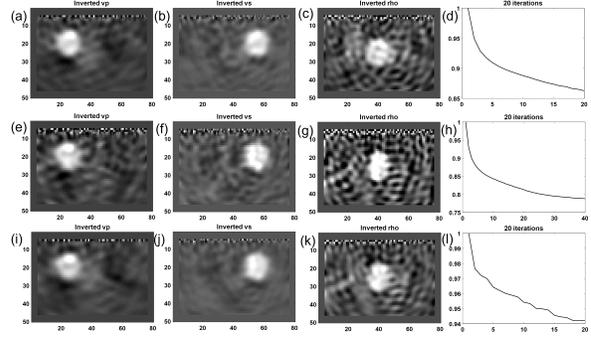


Figure 11: Inversion results. (a) Inverted  $V_p$  model of the first FWI, (b) Inverted  $V_s$  model of the first FWI, (c) Inverted density model of the first model, (d) the misfit change of the first FWI, (e) inverted  $V_p$  model of the conventional FWI using 40 iterations, (f) inverted  $V_s$  model 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  $V_p$  model of the second FWI with the modified misfit, (j) inverted  $V_s$  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 the modified misfit.

## CONCLUSIONS

In this study, we developed a methodology to deal with remnant noise, with particular focus on correlated noise in seismic data. This is implemented by incorporating the data covariance matrix, which is estimated from a conventional FWI result, into the misfit function. We found that random noise in large magnitude have impacts on the accuracy of inversion results. Generally, random noise is less harmful in model estimation than correlated noise. By estimating the noise from analysis of residuals generated through a first FWI pass, and incorporating them within the data covariance matrix, several important types of error can be estimated and suppressed during inversion. Compared with the conventional FWI models, more interpretable images with fewer artifacts result, at moderate increase in computational cost.

## ACKNOWLEDGMENTS

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 the Chevron scholarship from SEG.

## Elastic FWI with misfit modification

### REFERENCES

- Cai, A. and C. A. Zelt, 2019, Data weighted full-waveform inversion with adaptive moment estimation for near-surface seismic refraction data: Presented at the SEG International Exposition and Annual Meeting.
- Dettmer, J., S. E. Dosso, and C. W. Holland, 2007, Uncertainty estimation in seismo-acoustic reflection travel time inversion: The Journal of the Acoustical Society of America, **122**, 161–176.
- Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: SIAM.