Non-Stationary Low Frequency De-noising using Prediction Error Filters

Milad Bader and Robert Clapp[1]

ABSTRACT

Following the low-frequency de-noising method introduced previously, where a stationary prediction-error filter is estimated from the high-frequency data, we extend the process to the non-stationary case. We account for the non-stationarity by estimating a bank of prediction-error filters defined on a regular grid. Two sets of filters are built, one for the high-frequency signal and one for the low-frequency data. The first is expanded in time and space, then used with the second in a regularized inverse problem to estimate the low-frequency signal. The method is tested on a synthetic shot gather.

INTRODUCTION

Low-frequency seismic data ($< 4\,Hz$) has become essential in recent years for building high-fidelity models using full waveform inversion (FWI) (Sirgue and Pratt, 2004). It allows the latter to start from a crude model by preventing cycle-skipping and filling the low wavenumbers when a multi-scale strategy is followed (Bunks et al., 1995). It also extends the maximum depth of the inversion, and yields reliable results in more complex settings such as in the presence of salt bodies (Shen et al., 2018). This has lead to large expenditures in the industry in order to acquire more and better low-frequency data (Pool et al., 2018; Brenders et al., 2018). Nevertheless, these low frequencies often exhibits a low signal-to-noise ratio (SNR) and are difficult to de-noise.

In Bader et al. (2019), we have introduced a new method of de-noising low-frequency data using prediction-error filters (PEF). It consisted of estimating a PEF from high-frequency data adequately chosen with sufficiently high SNR. Then the PEF is expanded in each dimension and applied to the low-frequency data in an iterative process in order to remove any energy which dips are inconsistent with the high-frequency data. This concept of expanding/shrinking the filter axis is tightly linked to the pyramid transform (Hung et al., 2004). It has been used to interpolate missing traces in the F-X domain (Spitz, 1991), and to interpolate high-frequency aliased data in the T-X domain (Crawley, 1998).

[1]Email: nmbader@sep.stanford.edu, bob@sep.stanford.edu

Here we extend the method to cover the 2D non-stationary case by estimating a bank of PEFs defined on a sparse regular grid (Ruan et al., 2015; Lapilli et al., 2019). The de-noising problem is transformed into a regularized problem with two sets of PEFs, one for the signal and one for the noise (Abma, 1995). We apply our method to a synthetic shot gather from the Marmousi model.

In the first part of this paper, we review the general de-noising problem using PEFs and how it is adapted to the low frequency de-noising case. Then, we discuss the possible approaches to account for the non-stationarity and justify our choice. Finally, we show our synthetic results where three types of noise are added to the low-frequency signal.

METHOD

De-noising using PEFs

Assuming that the data vector $\mathbf{d}$ can be written as the sum of signal and noise components $\mathbf{d} = \mathbf{s} + \mathbf{n}$, Abma (1995) formulated the signal-noise separation as the following regularized least-squares minimization problem:

$$\begin{aligned} \mathbf{0} &\approx \mathbf{r_n} = \mathbf{N}\mathbf{n} \\ \mathbf{0} &\approx \mathbf{r_s} = \epsilon\mathbf{S}\mathbf{s} \\ s.t. \quad \mathbf{d} &= \mathbf{s} + \mathbf{n} \end{aligned} \tag{1}$$

where $\mathbf{N}$ and $\mathbf{S}$ are the noise and signal PEFs respectively, and $\approx$ designate the approximate equality in a least squares sense.

This formulation was previously used to remove different types of noise from the data, such as multiples and ground-roll (Clapp and Brown, 2000; Guitton et al., 2001; Guitton, 2005) where a noise model could be constructed a priori. It proved its superiority over conventional adaptive least squares subtraction thanks to its ability to exploit the signal and noise spatio-temporal patterns.

In this report, we start from the same formulation in order to de-noise the low-frequency data without having any a priori knowledge of the noise. Spitz (1999) pointed out that, when signal and noise are uncorrelated, the signal PEF $\mathbf{S}$ can be approximated from the data PEF $\mathbf{D}$ and the noise PEF $\mathbf{N}$ by:

$$\mathbf{S} \sim \mathbf{D}\mathbf{N}^{-\mathbf{1}} \tag{2}$$

where $\sim$ designates a spectral similarity where the PEFs are "weak", meaning that their corresponding predicted components are strong (Claerbout and Fomel, 2008).

It follows that the noise PEF can be approximated by $\mathbf{N} \sim \mathbf{S^{-1}D}$.

As for the low-frequency signal PEF, it is obtained from the high-frequency data PEF after expansion in every dimension **?**. The high-frequency data is assumed to have a SNR sufficiently high compared to the low-frequency data to de-noise.

Replacing in (1) and solving for $\mathbf{s}$, we obtain the regularized problem:

$$\tilde{\mathbf{s}} = \min \left(\|\mathbf{S^{-1}D}(\mathbf{d}-\mathbf{s})\|_{\mathbf{2}}^{\mathbf{2}} + \epsilon^{\mathbf{2}} \|\mathbf{Ss}\|_{\mathbf{2}}^{\mathbf{2}} \right) \tag{3}$$

This new problem requires inverting the non-stationary signal PEF $\mathbf{S}$, supposed invertible. We can use the matrix induced norm (Trefethen and Bau III, 1997) to bound the objective function to be minimized in (3):

$$\begin{aligned} \|\mathbf{S^{-1}D}(\mathbf{d}-\mathbf{s})\|_{\mathbf{2}}^{\mathbf{2}} + \epsilon^{\mathbf{2}} \|\mathbf{Ss}\|_{\mathbf{2}}^{\mathbf{2}} &\leq \|\mathbf{S^{-1}}\|_{\mathbf{2}}^{\mathbf{2}} . \|\mathbf{D}(\mathbf{d}-\mathbf{s})\|_{\mathbf{2}}^{\mathbf{2}} + \epsilon^{\mathbf{2}} \|\mathbf{Ss}\|_{\mathbf{2}}^{\mathbf{2}} \\ &= \|\mathbf{S^{-1}}\|_{\mathbf{2}}^{\mathbf{2}} . \left(\|\mathbf{D}(\mathbf{d}-\mathbf{s})\|_{\mathbf{2}}^{\mathbf{2}} + \beta^{\mathbf{2}} \|\mathbf{Ss}\|_{\mathbf{2}}^{\mathbf{2}} \right) \end{aligned} \tag{4}$$

where $\|\mathbf{S^{-1}}\|_{\mathbf{2}}$ is the matrix norm induced by the vector L2 norm, and $\beta^2 = \frac{\epsilon^2}{\|\mathbf{S^{-1}}\|_2^2}$. Dividing both sides of the inequality by $\|\mathbf{S^{-1}}\|_{\mathbf{2}}^{\mathbf{2}}$ and setting $\hat{\mathbf{S}}^{\mathbf{-1}} = \frac{\mathbf{S^{-1}}}{\|\mathbf{S^{-1}}\|_{\mathbf{2}}}$, the inequality becomes:

$$\|\hat{\mathbf{S}}^{\mathbf{-1}}\mathbf{D}(\mathbf{d}-\mathbf{s})\|_{\mathbf{2}}^{\mathbf{2}} + \beta^{\mathbf{2}} \|\mathbf{Ss}\|_{\mathbf{2}}^{\mathbf{2}} \leq \|\mathbf{D}(\mathbf{d}-\mathbf{s})\|_{\mathbf{2}}^{\mathbf{2}} + \beta^{\mathbf{2}} \|\mathbf{Ss}\|_{\mathbf{2}}^{\mathbf{2}} \tag{5}$$

Now we can solve the simplified problem on the RHS of (5):

$$\tilde{\mathbf{s}} = \min \left(\|\mathbf{D}(\mathbf{d}-\mathbf{s})\|_{\mathbf{2}}^{\mathbf{2}} + \beta^{\mathbf{2}} \|\mathbf{Ss}\|_{\mathbf{2}}^{\mathbf{2}} \right) \tag{6}$$

however, a solution of (6) is not necessarily a solution of (3). Nevertheless, if we ensure that the RHS in (5) is small, it automatically follows that the LHS will also be small. Note that a minimizer of the LHS in (5) is a solution of (3).

If the minimization in (6) is carried out iteratively and $\mathbf{s}$ is initialized to $\mathbf{s} = \mathbf{d}$, then both the LHS and RHS in (5) will have initially the same value equal to $\beta^{\mathbf{2}} \|\mathbf{Sd}\|_{\mathbf{2}}^{\mathbf{2}}$. Therefore, reducing the RHS will also reduce the LHS by at least the same proportion. Figure 1 illustrates the two objective functions figuring in the inequality (5). They are both quadratic, one is always higher than the other and they meet at one single point corresponding to $\mathbf{s} = \mathbf{d}$. The solution $\tilde{\mathbf{s}}_{\mathbf{modified}}$ of the modified problem (6) approaches the solution $\tilde{\mathbf{s}}_{\mathbf{original}}$ of the original problem (3).

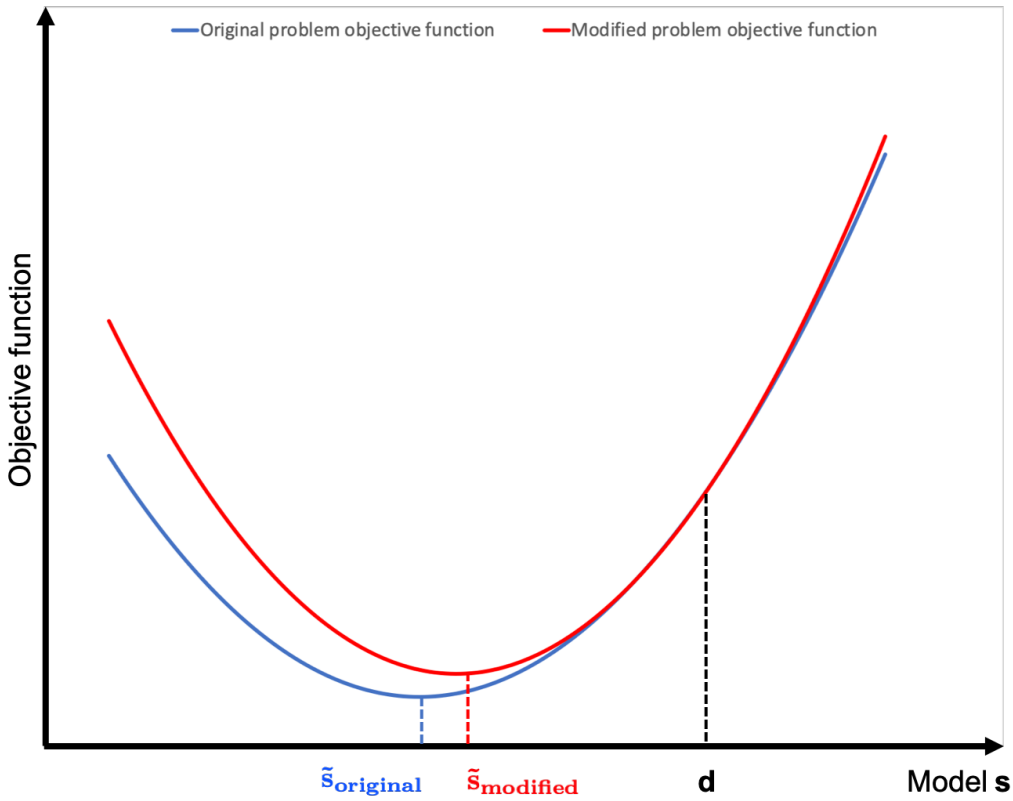


Figure 1: Sketch showing the objective functions figuring on either sides of the inequality (5). [**NR**]

Non-stationary PEF estimation

All the PEF convolution operators form the previous section are assumed to be non-stationary in every dimension. In order to estimate such filters, several methods are available. The most straightforward (and least desirable) one is to treat the data from which the PEFs are estimated as an assembly of stationary patches, possibly overlapping (Schwab, 1998), but this does not account effectively for non-stationarity and tends to create edge effects. A similar method consists of keeping the PEFs constant in micro-patches and adding a smoothing regularizer in their estimation (Crawley, 2000; Guitton, 2005), but this requires a larger memory and the convergence may be slowed down by the regularization. An alternative method is to define a sparse regular grid of PEFs and linearly interpolate between them during the forward operation (**?**). Hence, at every data point, a linear combination of four PEFs is used, corresponding to the four closest corners from the sparse grid. This is the approach adopted in this report. Typically, the distance between the sparse grid points is few times the size of the individual PEFs, which makes the problem of estimating the non-stationary PEF over-determined and saves us from using a regularization (the interpolation itself acts as a smoothing operator). Moreover, the size of the resulting PEFs is smaller than the data and can be easily stored and used in any iterative process.

Note that another potential method for estimating the non-stationary PEF is the streaming method (Fomel and Claerbout, 2016) which is tightly related to the least mean squares algorithm in 1D (Widrow and Stearns, 1985) and allows for PEF estimation and application "on the fly" as the data samples stream in. This method is not considered in this report.

NUMERICAL EXAMPLES

We applied the method described above to a synthetic shot gather modeled with the scalar wave equation and the Marmousi velocity model. The shot was modeled with a band limited wavelet [2 - 10 Hz] and is shown on Figure 2.

The shot is split into two frequency bands: [2 - 4 Hz] and [4 - 10 Hz] (Figures 3a and 3b). A noise model is built and added to the low frequency component to create a low frequency data to be de-noised. It consists of random noise, two monochromatic waves with opposite dips and frequencies of 2.5 Hz and 3.5 Hz, and band-limited coherent events with opposite dips. Figures 4a and 4b show the noisy low frequency data as well as the generated noise model. The SNR of the data in this case is estimated from the ratio between the signal and the noise RMS: $\text{SNR} = \frac{RMS_{signal}}{RMS_{noise}} \approx 1$.

The high frequency component is used to estimate a bank of PEFs on a sparse grid of 40 traces x 50 time samples, using a CGLS algorithm (Aster et al., 2012). Each individual PEF has 11 traces and 11 time samples with the leading 1 at the middle of the first column. The set of PEFs is stored in a single dataset shown on Figures 5a

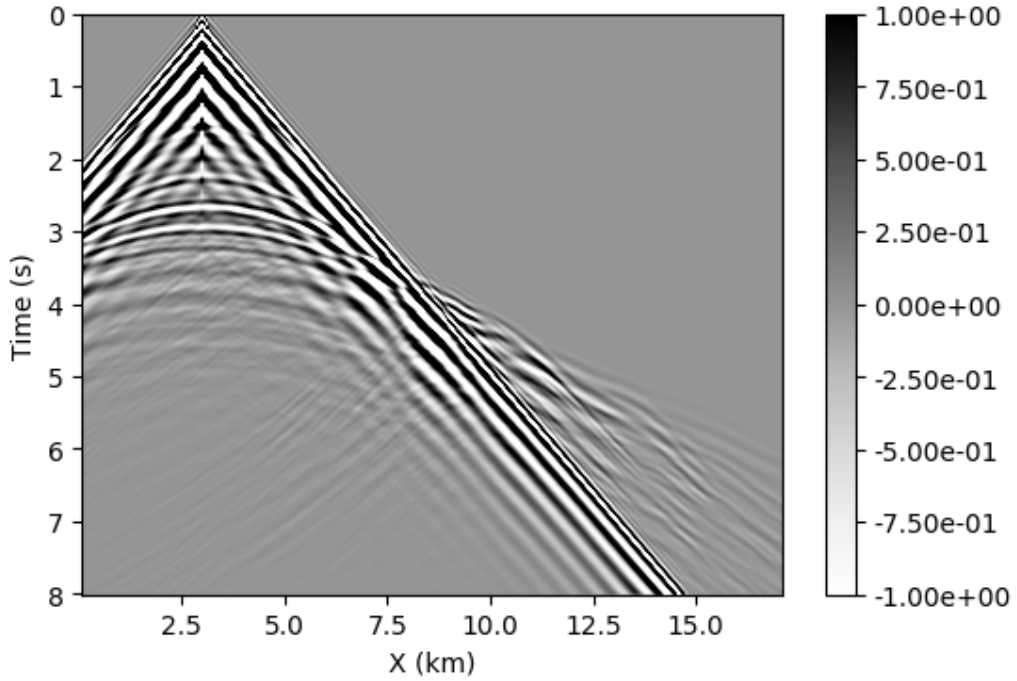
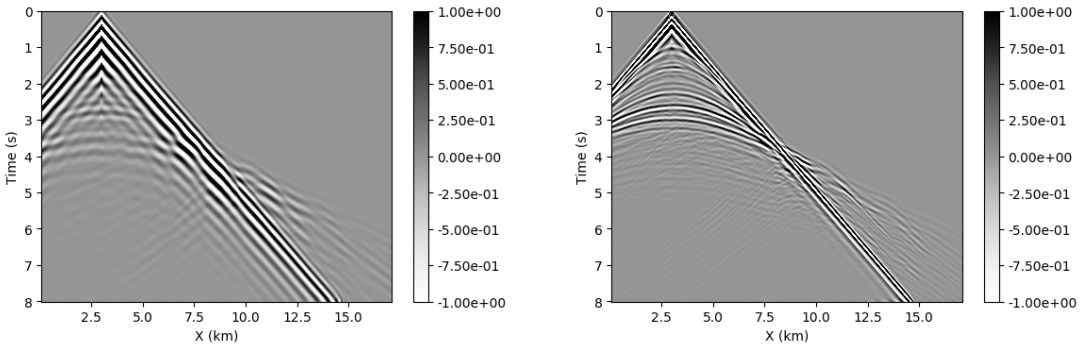
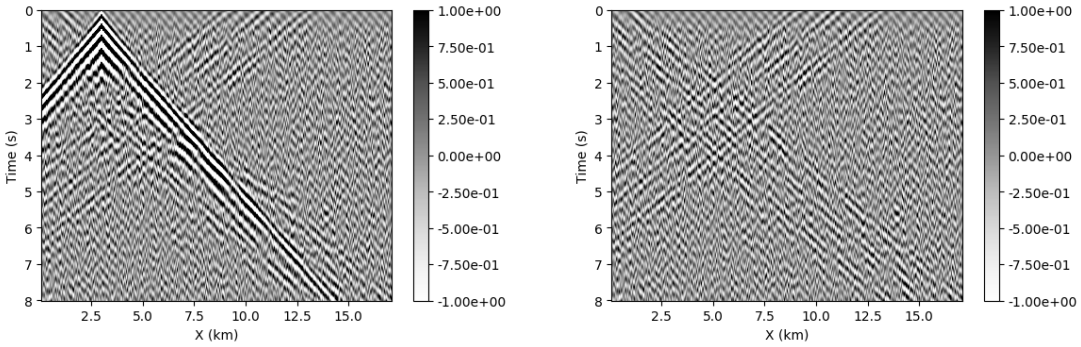


Figure 2: Band-limited ([2 - 10 Hz]) synthetic shot gather from Marmousi model. [**ER**]



(a) (b)

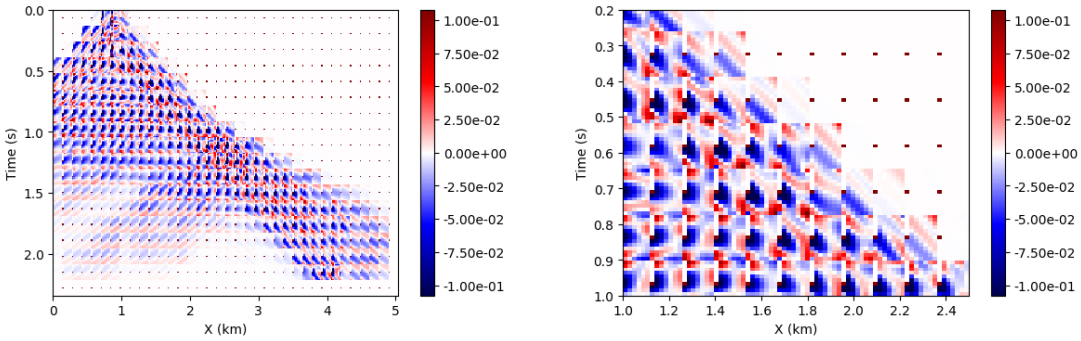
Figure 3: Marmousi synthetic shot gather split into two frequency bands: [2 - 4 Hz] (a) - [4 - 10,Hz] (b). [**ER**]



(a) (b)

Figure 4: Noisy low frequency data (a) - Noise model (b). [**ER**]

and 5b. The axis expansion is performed implicitly when solving (6). Similar PEFs are estimated from the low frequency data to build the operator $\mathbf{D}$ and from the noise model to build the operator $\mathbf{N}$. The latter is used to solve the original problem (3) for verification. Neither $\mathbf{D}$ nor $\mathbf{N}$ need to be expanded.



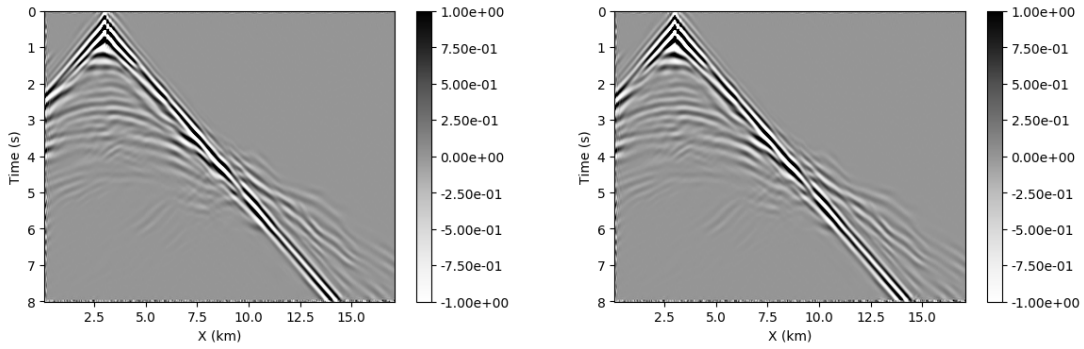
(a) (b)

Figure 5: Bank of PEFs estimated on a sparse grid (a) - zoom-in (b). [**ER**]

Solving (6) iteratively with $\beta = 1$ and $\mathbf{s_0} = \mathbf{d}$ yields after 20 iterations the result shown on Figure 6a. Problem (3) is also solved using the "exact" noise PEF $\mathbf{N}$ and with $\epsilon = \beta = 1$. The result is shown on Figure 6b.

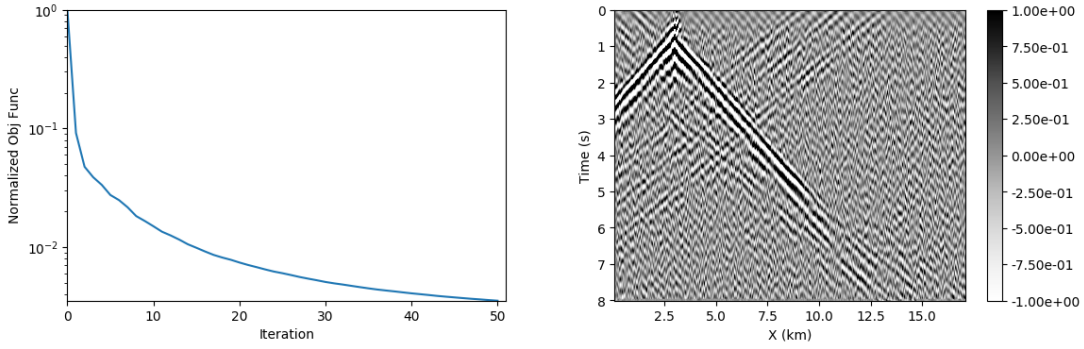
The normalized objective function for the modified problem (6) is shown on Figure 7a. The noise that has been removed by solving that problem is shown on Figure 7b.

In order to conduct a more rigorous comparison, we should set the Lagrange multiplier for the modified problem to $\beta^2 = \frac{\epsilon^2}{\|\mathbf{S}^{-1}\|_2^2}$, but this will require computing an operator norm. This task is unnecessary in practice since the choice of the Lagrange multiplier is user-defined and is based on the quality of the output.



(a) (b)

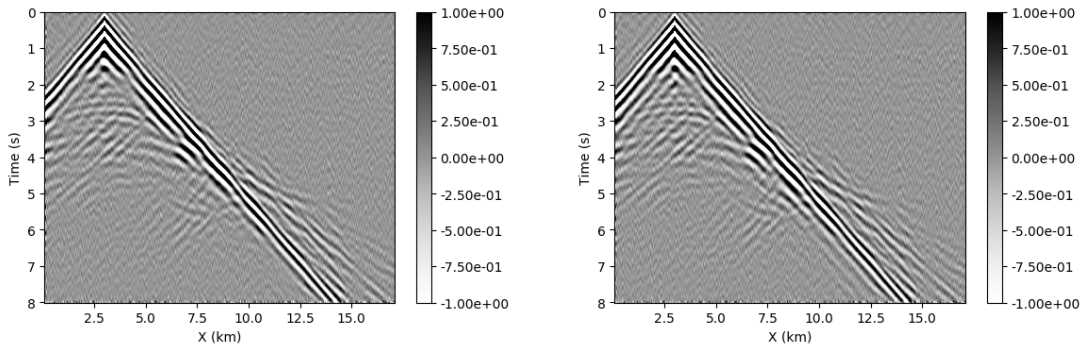
Figure 6: De-noised low frequency data without prior knowledge of the noise model (a) - and with perfect knowledge of the noise model (b) where $\epsilon = \beta = 1$. [**ER**]



(a) (b)

Figure 7: Normalized objective function (a) - and noise being removed (b) by solving (6). [**ER**]

Notice that the results shown on Figures 6a and 6b are very similar. This could be due to the dominance of the term $\beta^2\|\mathbf{Ss}\|_2^2$ in both problems. We performed another test with $\epsilon = \beta = 0.1$ and the results are shown on Figures 8a and 8b. Even though the two solutions started off quite differently, they converged to similar results although the direct arrival is better preserved when using the exact noise PEF. We believe that in this test, the bulk of the de-noising work is done by the operator $\mathbf{S}$, whereas the noise PEF has an auxiliary role and mainly prevents the solution from converging towards the trivial one $\tilde{\mathbf{s}} = \mathbf{0}$. We expect the noise PEF to play a more important role when the noise model is more complicated.



(a) (b)

Figure 8: De-noised low frequency data without prior knowledge of the noise model (a) - and with perfect knowledge of the noise model (b) where $\epsilon = \beta = 0.1$. [**ER**]

One may argue that the same "trick" used to obtain the modified problem (6) can also be used to get rid of the operator $\mathbf{D}$. The minimization problem becomes:

$$\tilde{\mathbf{s}} = \min\left(\|\mathbf{d} - \mathbf{s}\|_2^2 + \lambda^2\|\mathbf{Ss}\|_2^2\right) \tag{7}$$

where $\lambda^2 = \frac{\beta^2}{\|\mathbf{D}\|_2^2}$. Numerical tests (not shown in this report) have shown that this gives inferior results compared to the problem (6). The reason is that the noise PEF $\mathbf{N}$ is somehow embedded in the data PEF $\mathbf{D}$ through Spitz approximation $\mathbf{D} \sim \mathbf{SN}$, therefore it is more effective in separating noise from signal than the identity operator implicit in (7). Nevertheless, it does not dispense from looking into inverting the operator $\mathbf{S}$ in order to ensure a better signal and noise separation.

CONCLUSION

We have introduced a new method for de-noising low-frequency data in a 2D non-stationary setting using high-frequency prediction-error filters and without any prior knowledge of the noise model. The method yields promising results verified on a

synthetic shot gather from Marmousi model. Its impact is to be assessed by running FWI with the de-noised data, and it is yet to be verified on a field dataset.

REFERENCES

Abma, R. L., 1995, Least-squares separation of signal and noise using multidimensional filters: Stanford University.

Aster, R. C., C. H. Thurber, and B. Borchers, 2012, Parameter estimation and inverse problems.

Bader, M., R. G. Clapp, and B. Biondi, 2019, Low frequency denoising using high frequency prediction error filters: SEP-Report, **176**, 75–84.

Brenders, A., J. Dellinger, C. Kanu, Q. Li, and S. Michell, 2018, The wolfspar® field trial: Results from a low-frequency seismic survey designed for fwi: 88th annual international meeting, seg, expanded abstracts, 1083–1087.

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Claerbout, J. F. and S. Fomel, 2008, Image estimation by example: geophysical soundings image construction: multidimensional autoregression: Citeseer.

Clapp, R. G. and M. Brown, 2000, $(t-x)$ domain, pattern-based multiple separation: SEP-Report, **103**, 201–210.

Crawley, S., 1998, Shot interpolation for radon multiple suppression, *in* SEG Technical Program Expanded Abstracts 1998, 1238–1241, Society of Exploration Geophysicists.

———, 2000, Seismic trace interpolation with nonstationary prediction-error filters: Stanford University.

Fomel, S. and J. Claerbout, 2016, Streaming prediction-error filters, *in* SEG Technical Program Expanded Abstracts 2016, 4787–4791, Society of Exploration Geophysicists.

Guitton, A., 2005, Multiple attenuation in complex geology with a pattern-based approach: Geophysics, **70**, V97–V107.

Guitton, A., M. Brown, J. Rickett, and R. Clapp, 2001, A pattern-based technique for ground-roll and multiple attenuation: SEP-Report, **108**, 249–274.

Hung, B., C. Notfors, and S. Ronen, 2004, Seismic trace interpolation using the pyramid transform, *in* SEG Technical Program Expanded Abstracts 2004, 2017–2020, Society of Exploration Geophysicists.

Lapilli, C., C. Kostov, A. Rushdy, D. Nichols, F. Xavier de Melo, and R. Clapp, 2019, Pattern-matching adaptive subtraction with nonstationary prediction-error filters: Requirements for applications to high-dimensional data sets, *in* SEG Technical Program Expanded Abstracts 2019, 4530–4534, Society of Exploration Geophysicists.

Pool, R., C. Kanu, A. Brenders, and J. Dellinger, 2018, The wolfspar® field trial: Design and execution of a low-frequency seismic survey in the gulf of mexico, *in* SEG Technical Program Expanded Abstracts 2018, 91–96, Society of Exploration Geophysicists.

Ruan, K., J. Jennings, E. Biondi, R. G. Clapp, S. A. Levin, and J. Claerbout, 2015,

Industrial scale high-performance adaptive filtering with pef applications: SEP-Report, **160**, 177–188.

Schwab, M., 1998, Enhancement of discontinuities in seismic 3-d images using a java estimation library: PhD thesis, Stanford University.

Shen, X., I. Ahmed, A. Brenders, J. Dellinger, J. Etgen, and S. Michell, 2018, Full-waveform inversion: The next leap forward in subsalt imaging: The Leading Edge, **37**, 67b1–67b6.

Sirgue, L. and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, 231–248.

Spitz, S., 1991, Seismic trace interpolation in the fx domain: Geophysics, **56**, 785–794.

———, 1999, Pattern recognition, spatial predictability, and subtraction of multiple events: The Leading Edge, **18**, 55–58.

Trefethen, L. N. and D. Bau III, 1997, Numerical linear algebra, volume **50**: Siam.

Widrow, B. and S. D. Stearns, 1985, Adaptive signal processing prentice-hall: Englewood Cliffs, NJ.