Making f-x projection filters robust to erratic noise

Ke Chen and Mauricio D. Sacchi, University of Alberta*

SUMMARY

Linear prediction filters are efficient for reducing random seismic noise but not for removing erratic noise. We propose a robust f-x projection filtering scheme for simultaneous erratic noise and Gaussian random noise attenuation. Instead of adopting the ℓ_2 norm, as commonly used in the conventional design of f-x filters, we utilize the hybrid ℓ_1/ℓ_2 norm to penalize the energy of the additive noise. The estimation of the prediction error filter and the additive noise sequence are performed in an alternating fashion. First, the additive noise sequence is fixed and the prediction error filter is estimated via the least-squares solution of a system of linear equations. Then, the prediction error filter is fixed and the additive noise sequence is estimated through a cost function containing hybrid ℓ_1/ℓ_2 norm that prevents erratic noise to influence the final solution. Synthetic and field data examples are used to evaluate the performance of the proposed algorithm.

INTRODUCTION

The f-x prediction filtering methods for random seismic noise reduction have been widely used in industry. Canales (1984) proposed the f-x prediction technique for seismic random noise reduction. This method implicitly utilizes the autoregressive (AR) model to represent data in the f-x domain. The method is often named f-x deconvolution (Gulunay, 1986).

f-x deconvolution is known to damage the signal if the signal-to-noise ratio (SNR) is low. A large order AR model can be used to better represent the data (Ulrych and Sacchi, 2005). However, long AR filters will also model the noise and therefore, one will not be able to attenuate random noise. Harris and White (1997) suggests to "clean up" the correlation matrix that is required to estimate the prediction error filter via the truncated SVD, a methodology first described in Tufts and Kumaresan (1982). Soubaras (1994, 1995) proposed the f-x projection filtering technique. The latter utilizes the additive noise model and the concept of quasi-predictability to estimate additive random noise. The additive noise is estimated via the application of an autodeconvolved prediction error filter (called the projection filter) to the data. Sacchi and Kuehl (2001) pointed out that the model for seismic data in f-x is actually a special autoregressive moving-average (ARMA) model (Ulrych and Clayton, 1976) in the sense that the parameters of the AR portion are identical to the parameters of the MA portion of the model. The prediction error filter in Sacchi and Kuehl (2001) is the solution of an eigen-decomposition problem. The additive noise is estimated by a least-squares procedure equivalent to the method outlined by Soubaras (1994).

The aforementioned methods are based on the least-squares approach. They are efficient for Gaussian noise elimination. However, it is well known that the least-squares estimation is very sensitive to erratic noise (non-Gaussian errors). Unfortunately, seismic data often contain erratic noise such as noise bursts, power-line noise, traffic noise, swell noise, etc. Several methods based on outlier detection have been proposed to denoise seismic data contaminated by erratic noise. For instance, in each frequency slice or frequency band, the traces containing impulsive noise are first detected, invalidated and then interpolated by f-x projection filters (Soubaras, 1995) or f-x prediction filters (Schonewille et al., 2008). Instead of outlier detection techniques followed by least-squares estimation, we propose to apply direct robust estimation. The proposed robust f-x projection method can simultaneously remove random Gaussian noise and erratic noise. The misfit between the observed data and the modeled signal is measured by the hybrid ℓ_1/ℓ_2 norm (Bube and Langan, 1997) instead of the classical ℓ_2 norm. The estimation of the prediction error filter and the clean signal is a nonlinear problem because these two are coupled together as convolution. In this article, the aforementioned problem is tackled by an alternating minimization scheme where the noise sequence and the prediction error filter are alternately updated.

THEORY

Additive noise model

The seismic signal is usually corrupted with seismic noise resulting from various sources. We will consider the situation where a signal in the f-x domain is corrupted by not only Gaussian noise but also erratic (impulsive) noise. In each frequency slice, the observed seismic data can be represented by an additive noise model (Ulrych and Clayton, 1976; Soubaras, 1995; Sacchi and Kuehl, 2001)

$$y_n = x_n + n_n + i_n, \tag{1}$$

where y_n is a wide-sense stationary random process, process x_n represents the clean signal, n_n is the complex white Gaussian noise and i_n is a stationary process representing erratic noise. The signal x_n, the Gaussian noise n_n and the impulsive noise i_n are assumed to be mutually independent. We use $e_n = n_n + i_n$ to represent the mixture of additive noises. Equation 1 can be rewritten as

$$y_n = x_n + e_n. \tag{2}$$

We will use lower case bold fonts to indicate the realizations of random processes in vector form. For instance, $\mathbf{y}$ is a realization of process y_n expressed in a column vector form.

Signal model: quasi-predictability

A seismic signal that contains p linear events with distinct dips manifests itself as a superposition of p complex sinusoids in the f-x domain. The signal in the channel n at angular fre-

DOI http://dx.doi.org/10.1190/segam2014-0893.1

quency ω can be represented by

$$x_n(\omega) = \sum_{k=1}^{p} a_k(\omega) e^{-i\omega\eta_k(n-1)\Delta s}, \tag{3}$$

where $a_k(\omega)$ is the Fourier transform of the source wavelet corresponding to the kth event, η_k is the kth dip and Δs is the spatial interval between two channels, and $i = \sqrt{-1}$. For convenience, we will omit the symbol ω and understand that the analysis is carried out for all frequencies. The exponential signal is represented by line spectra consisting of p impulses in the wavenumber domain. It can be shown that the exponential signal satisfies the pth-order homogeneous difference equation

$$x_n + f_1 x_{n-1} + f_2 x_{n-2} + \ldots + f_p x_{n-p} = 0. \tag{4}$$

In other words, x_n is perfectly predictable based on it preceding values. Clearly, the series x_n is completely deterministic. The elements $f_0 = 1, f_1, f_2, \ldots, f_p$ are the coefficients of the so-called prediction error filter. In this particular case, the prediction error filter is also the annihilator of x_n. This complex sinusoidal process is equivalent to a "special AR process" with innovation equal to zero (Kay and Marple, 1981).

In realistic cases, the noise-free f-x seismic signal cannot be perfectly modeled as a sum of a finite number of exponentials. The concept of quasi-predictability (Soubaras, 1995) will allow us to cope with situation where the innovation of the special AR process is not equal to zero. The deterministic process x_n in equation 4 is approximated by an AR process

$$x_n = -\sum_{k=1}^{p} f_k x_{n-k} + u_n, \tag{5}$$

where $f_k, k = 1, 2, \ldots, p$ are the AR coefficients and u_n indicates white noise sequence (innovation). The random process x_n is quasi-predictable from its preceding samples. Substituting $x_n = y_n - e_n$ into equation 5 leads to

$$\sum_{k=0}^{p} f_k y_{n-k} = \sum_{k=0}^{p} f_k e_{n-k} + u_n. \tag{6}$$

The above equation is an ARMA process similar to the model studies by Ulrych and Clayton (1976) and Sacchi and Kuehl (2001), however, the process now contains innovation term. The ARMA parameter estimation problem is nonlinear (Kay and Marple, 1981). We tackle it via an alternating minimization scheme. First, the additive noise sequence is fixed and the prediction error filter is estimated. Then, the prediction error filter is fixed and the additive noise sequence is estimated. The two stages are iterated until reaching convergence.

Additive noise sequence and PEF estimation

Random process y_n is observed over an interval of N in space that results in a N-point data sequence $\mathbf{y}$. In this paper, we adopt the forward and backward prediction method (modified covariance method) (Ulrych and Clayton, 1976). In matrix vector form, the forward and backward prediction can be expressed as

$$\mathbf{F}(\mathbf{y} - \mathbf{e}) = \mathbf{u} \tag{7}$$

where $\mathbf{F}$ is a convolutional matrix containing the elements of the unknown filter coefficients. Our task is to estimate the prediction error filter $\mathbf{f}$ and the noise sequence $\mathbf{e}$. It is a nonlinear problem. We will first simplify the problem by assuming that the prediction error filter is known. The estimation of $\mathbf{e}$ from equation 7 is an ill-posed problem. Soubaras (1994) solved the problem via the damped least-squares method resulting in the well-known Tikhonov regularized least-squares solution. In this paper, we propose to adopt a constraint that minimizes the hybrid ℓ_1/ℓ_2 norm of the noise sequence $\mathbf{e}$. The estimation of the noise sequence reduces as minimizing the cost function

$$\mathscr{J} = \frac{1}{2}||\mathbf{F}(\mathbf{y} - \mathbf{e})||_2^2 + \lambda \mathscr{H}(\mathbf{e}), \tag{8}$$

where $\lambda = \xi^2/\sigma$ is a trade-off parameter, ξ is the standard deviation of the innovation and σ is the scale parameter for the noise sequence $\mathbf{e}$. The functional $\mathscr{H}(\mathbf{e}) = \sum_{i=1}^{N} h(e_i)$ is the hybrid ℓ_1/ℓ_2 norm of the complex vector $\mathbf{e}$ with the hybrid function given by

$$h(e) = \sqrt{\sigma^2 + |e|^2} - \sigma. \tag{9}$$

Bube and Langan (1997) recommended to choose σ approximately equal to 0.6 times the standard deviation of the random variable e. Setting $\frac{\partial \mathscr{J}(\mathbf{e})}{\partial \mathbf{e}^*} = \mathbf{0}$, leads to the "nonlinear normal equations"

$$\left(\mathbf{F}^H\mathbf{F} + \lambda \mathbf{W}\right)\mathbf{e} = \mathbf{F}^H\mathbf{F}\mathbf{y}, \tag{10}$$

where $\mathbf{W}$ is a $N \times N$ diagonal weight matrix with diagonal elements given by $W_{jj} = 1/\sqrt{\sigma^2 + |e_j|^2}$. The nonlinear equations can be solved by the iteratively reweighed least-squares (IRLS) algorithm (Bube and Langan, 1997). The kth iteration is solved with weights computed from the iteration $k-1$, $\mathbf{W}^{k-1}$

$$W_{jj}^{(k-1)} = 1/\sqrt{\sigma^2 + |e_j^{(k-1)}|^2}, \; j = 1, 2, \ldots, N. \tag{11}$$

The iterative solution is given by

$$\mathbf{e}^{(k)} = \left(\mathbf{F}^H\mathbf{F} + \lambda \mathbf{W}^{(k-1)}\right)^{-1} \mathbf{F}^H\mathbf{F}\mathbf{y}. \tag{12}$$

Now we turn our attention to the estimation of the filter $\mathbf{f}$. Clearly, once we have estimated the noise sequence $\mathbf{e}$, we can compute an estimation of the clear signal $\mathbf{x} = \mathbf{y} - \mathbf{e}$. Moreover, given that the regularization term does not depend on $\mathbf{f}$, the problem of estimating $\mathbf{f}$ reduces as minimizing

$$\mathscr{J} = \frac{1}{2}\|\mathbf{F}\mathbf{x}\|_2^2. \tag{13}$$

Given the commutative property of the convolution operator, minimizing $\|\mathbf{F}\mathbf{x}\|_2^2$ is equivalent to minimize $\|\mathbf{X}\mathbf{f}\|_2^2$ where $\mathbf{X}$ is the matrix containing the elements of the $\mathbf{x}$ and representing the convolution of $\mathbf{f}$ with $\mathbf{x}$. Given that $f_0 = 1$, we estimate the prediction filter $\mathbf{g}$

$$\hat{\mathbf{g}} = (\bar{\mathbf{X}}^H\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}^H\bar{\mathbf{x}}, \tag{14}$$

$\bar{\mathbf{X}}$, $\bar{\mathbf{x}}$ are the partitioned matrix and vector of $\mathbf{X}$ such that $\mathbf{X} = \left(\bar{\mathbf{x}}|\bar{\mathbf{X}}\right)$. Finally, the estimated prediction error filter is given by the vector

$$\hat{\mathbf{f}} = (1, -\hat{\mathbf{g}}^T)^T. \tag{15}$$

DOI http://dx.doi.org/10.1190/segam2014-0893.1

Iterative algorithm, hyperparameter selection and stopping criteria

The algorithm is applied to each temporal frequency with special attention paid to Fourier domain symmetries to save computational cost. The algorithm can be summarized as follows

1. Initialize the signal $\mathbf{x}$ and compute an initial noise term $\mathbf{e} = \mathbf{y} - \mathbf{x}$.
2. Estimate of the prediction error filter $\mathbf{f}$ via equation 14 and 15.
3. Estimate of the noise sequence $\mathbf{e}$ by minimizing cost function 8.
4. Iterate steps 2 to 3 until convergence.

For a given temporal frequency slice, the signal component $\mathbf{x}$ is initialized by the estimated signal $\mathbf{x}$ from the preceding frequency slice. The solution at the first frequency slice, $\mathbf{x}$, is generated by the traditional least-squares f-x projection (Sacchi and Kuehl, 2001). The parameter σ is fixed and the trade-off parameter λ is tuned by examining the residuals. This is similar to the strategy often used in f-x deconvolution for parameter selection. The algorithm has two groups of iterations: an internal iteration (IRLS) to estimate $\mathbf{e}$ and an external iteration for alternating minimization. We have two convergence criteria to stop to reduce the number of iterations. We monitor the cost function $\mathscr{J}$ and terminate the external loop when the relative change of the cost function between two consecutive iterations is less that a tolerance tol_1. A second tolerance tol_2 is used to control the number of IRLS iterations that are required to estimate $\mathbf{e}$.

EXAMPLES

Synthetic example

Our algorithm is first tested with a synthetic example. We compare the results of robust f-x projection, f-x deconvolution and the conventional f-x projection. The f-x deconvolution used here averages the forward and backward predicted values and uses prediction matrix corresponds to transient-free formulation. The conventional f-x projection filter used here is a modification of Sacchi and Kuehl (2001)'s method that uses the modified covariance method. Figure 1a shows a 2-D synthetic data with noise. The central frequency of the Ricker wavelet is 30 Hz. Figure 1b shows the Gaussian noise with signal to noise ratio (SNR) equal to 1.2 (SNR is defined as the ratio of the maximum amplitudes of signal and noise). Figure 1c shows the high amplitude erratic noise. The maximum amplitude of the erratic noise is approximately 5 times the maximum amplitude of the signal in Figure 1a. The processing frequency band ranges from 0 to 100 Hz. The length of the prediction error filter for the robust f-x projection filtering is set to 4. The scale parameter σ and the trade-off parameter λ are 6 and 0.1, respectively. The two stopping criteria are $tol_1 = 10^{-8}$ and $tol_2 = 10^{-8}$. The length of prediction error filter of the f-x deconvolution is 9. The length of prediction error filter in the conventional f-x projection filtering method

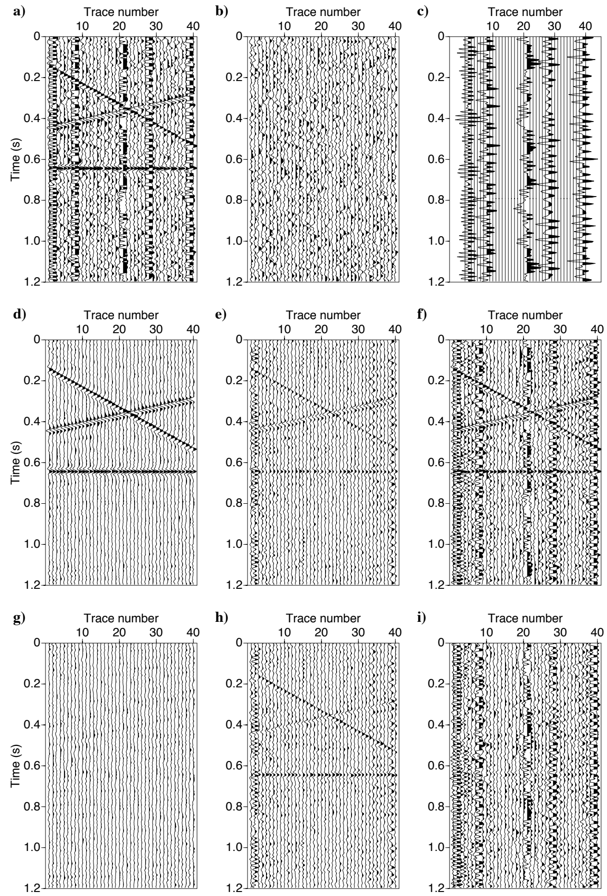

Figure 1: (a) Noisy synthetic data (after clipping). (b) Gaussian noise with SNR = 1.2. (c) Erratic noise. (d) Denoising via robust f-x projection. (e) Denoising via f-x deconvolution. (f) Denoising via least-squares f-x projection. (g) Difference section for robust f-x projection. (h) Difference section for f-x deconvolution. (i) Difference section for f-x projection filter.

is 4 and the trade-off parameter is 3. The filtered data by robust f-x projection, f-x deconvolution and least-squares f-x projection are shown in Figure 1d, Figure 1e and Figure 1f, respectively. Only the robust f-x projection filter was able to suppress the erratic noise and Gaussian noise. Difference sections (noise free data minus filtered data) in Figure 1g, Figure 1h and Figure 1i show that the robust f-x projection preserves the original signal. On the other hand, f-x deconvolution damages the signal. We evaluate the performance of the algorithms in decibels via the expression $Q = 10\,\log\frac{\|\mathbf{D}_0\|_F^2}{\|\mathbf{D}_0 - \hat{\mathbf{D}}\|_F^2}$, where $\mathbf{D}_0$ denotes noise-free data, $\hat{\mathbf{D}}$ denotes filtered data and $\|\cdot\|_F$ is the Frobenius norm of a matrix. Larger value of Q means better denoising performance. The Q value for the robust f-x projection filter is 7.6. The Q value for the f-x deconvolution is -3.3. The Q value for the f-x projection is -19.9.

Field data example

We tested our proposed algorithm on a post-stack field dataset from the Western Canadian Sedimentary Basin (WCSB). The

DOI http://dx.doi.org/10.1190/segam2014-0893.1

performance of robust f-x projection, f-x deconvolution and conventional f-x projection are compared. Figure 2a is a post-stack data section with erratic noise and random Gaussian noise. The complete data are divided into overlapped windows. All windows are processed and then added back. In the spatial direction, each window has 50 traces and the overlap between two adjacent windows is 25 traces. In the temporal direction, each window has 300 samples (0.6 s) and the overlap between two adjacent windows is 100 samples (0.2 s). All the three filtering methods are applied for frequencies in the band of 1-80 Hz. We choose the length of the prediction error filter for robust f-x projection filtering as 4. The scale parameter σ and the trade-off parameter λ are 2 and 0.4, respectively. The two stopping criterion values are $tol_1 = 10^{-6}$ and $tol_2 = 10^{-5}$. The length of prediction error filter for the f-x deconvolution is 7. The length of perdition error filter for the conventional f-x projection filter method is 4, and the trade-off parameter is 0.1. We show the results for data in two windows that are highlighted in Figure 2a. The results for the window to the left in Figure 2a are shown in Figure 3. Figure 4 shows the results of the window on the right. It is clear that the robust f-x projection has performed better than the classical projection filter and the f-x deconvolution.

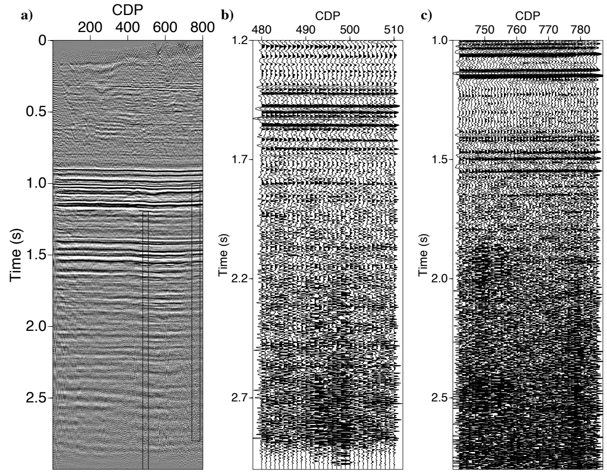

Figure 2: (a) Poststack data from WCSB with erratic noise and random Gaussian noise. (b) The data in the left rectangular window. (c) The data in the right rectangular window.

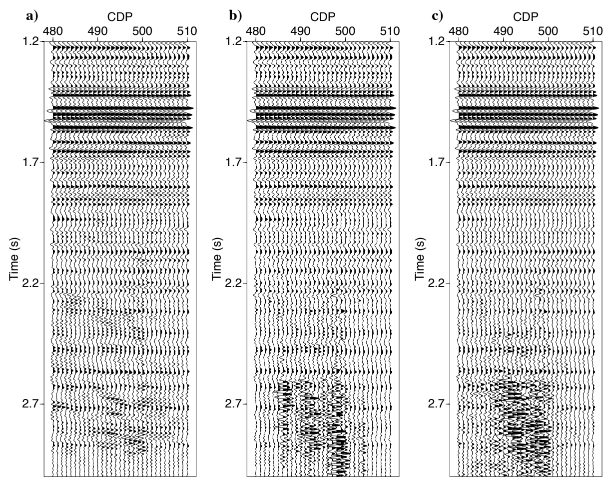

Figure 3: The comparison of filtered results of the data in the left window. (a) The result of robust f-x projection. (b) The result of f-x deconvolution. (c) The result of f-x projection.

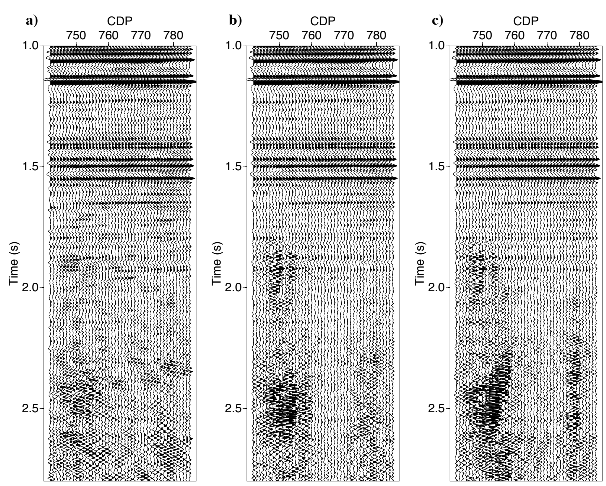

Figure 4: The comparison of filtered results of the data in the right window. (a) The result of robust f-x projection. (b) The result of f-x deconvolution. (c) The result of f-x projection.

CONCLUSIONS

In this paper, we propose a robust f-x projection denoising method that is robust to erratic noise. The method is also efficient for Gaussian noise attenuation. Instead of using the ℓ_2 norm of the additive noise, we adopted the hybrid ℓ_1/ℓ_2 norm to penalize the energy of the additive noise in order to promote robustness to erratic noise. The estimation of the noise sequence and the estimation of the prediction error filter are conducted via an alternating minimization algorithm. Synthetic data examples and a field data example show that the proposed robust algorithm can remove erratic noise with a minimal degradation of the signal.

ACKNOWLEDGMENTS

We wish to thank the sponsors of the Signal Analysis and Imaging Group (SAIG) at the University of Alberta for supporting this research.

http://dx.doi.org/10.1190/segam2014-0893.1

EDITED REFERENCES

Note: This reference list is a copy-edited version of the reference list submitted by the author. Reference lists for the 2014 SEG Technical Program Expanded Abstracts have been copy edited so that references provided with the online metadata for each paper will achieve a high degree of linking to cited sources that appear on the Web.

REFERENCES

Bube, K., and R. Langan, 1997, Hybrid minimization with applications to tomography: Geophysics, **62**, 1183–1195, http://dx.doi.org/10.1190/1.1444219.

Canales, L. L., 1984, Random noise reduction: 54th Annual International Meeting, SEG, Expanded Abstracts, 525–527.

Gulunay, N., 1986, FXDECON and complex Wiener prediction filter: 56th Annual International Meeting, SEG, Expanded Abstracts, 279–281.

Harris, P., and R. White, 1997, Improving the performance of *f-x* prediction filtering at low signal-to-noise ratios: Geophysical Prospecting, **45**, no. 2, 269–302, http://dx.doi.org/10.1046/j.1365-2478.1997.00347.x.

Kay, S., and S. Marple, 1981, Spectrum analysis — A modern perspective: Proceedings of the IEEE, **69**, no. 11, 1380–1419, http://dx.doi.org/10.1109/PROC.1981.12184.

Sacchi, M. D., and H. Kuehl, 2001, ARMA formulation of FX prediction error filters and projection filters: Journal of Seismic Exploration, **9**, 185–197.

Schonewille, M., A. Vigner, and A. Ryder, 2008, Swell-noise attenuation using an iterative FX prediction filtering approach: 78th Annual International Meeting, SEG, Expanded Abstracts, 2647–2651.

Soubaras, R., 1994, Signal-preserving random noise attenuation by the *f-x* projection: 64th Annual International Meeting, SEG, Expanded Abstracts, 1576–1579.

Soubaras, R., 1995, Prestack random and impulsive noise attenuation by *f-x* projection filtering: 65th Annual International Meeting, SEG, Expanded Abstracts, 711–714.

Tufts, D., and R. Kumaresan, 1982, Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood: Proceedings of the IEEE, **70**, no. 9, 975–989, http://dx.doi.org/10.1109/PROC.1982.12428.

Ulrych, T. J., and R. W. Clayton, 1976, Time series modelling and maximum entropy: Physics of the Earth and Planetary Interiors, **12**, no. 2-3, 188–200, http://dx.doi.org/10.1016/0031-9201(76)90047-9.

Ulrych, T. J., and M. D. Sacchi, 2005, Information-based inversion and processing with applications: Elsevier.