Title
Robust f-x projection filtering for simultaneous random and erratic seismic noise attenuation

Journal
Geophysical Prospecting, 65(3)

ISSN
0016-8025

Authors
Chen, Ke
Sacchi, Mauricio D

Publication Date
2017-05-01

DOI
10.1111/1365-2478.12429

Peer reviewed

Robust *f-x* projection filtering for simultaneous random and erratic seismic noise attenuation

Ke Chen

Mauricio D. Sacchi
First published: 25 August 2016
https://doi.org/10.1111/1365-2478.12429
UC-eLinks

ABSTRACT

Linear prediction filters are an effective tool for reducing random noise from seismic records. Unfortunately, the ability of prediction filters to enhance seismic records deteriorates when the data are contaminated by erratic noise. Erratic noise in this article designates non-Gaussian noise that consists of large isolated events with known or unknown distribution. 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 a hybrid ℓ_1/ℓ_2-norm that prevents erratic noise to influence the final solution. In other words, we proposed and designed a robust M-estimate of a special autoregressive moving-average model in the *f-x*domain. 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 adopted by the industry. Canales (1984) proposed the *f-x* prediction technique for seismic random noise reduction. This method implicitly utilizes the autoregressive (AR) model (Yule 1927) 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 because the AR model is only an approximation to the true process. 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) suggest to “clean up” the linear prediction data matrix that is required to estimate the prediction error filter via truncated singular value decomposition and use relatively large filter length, 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 moving-average (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).

Recently, the singular spectrum analysis method (Sacchi 2009), also known as Cadzow filtering (Trickett 2008), was introduced to attenuate random seismic noise and for seismic data reconstruction (Oropeza and Sacchi 2011). It is also based on the spatial predictability of seismic signal in the *f-x* domain.

The aforementioned methods are based on the least-squares approach. They are efficient for Gaussian noise elimination. However, it is well known that least-squares estimation is very sensitive to erratic noise (non-Gaussian errors). Unfortunately, seismic data often contain erratic noise such as noise bursts, polarity reversals, power-line noise, traffic noise, swell noise, and harmonic noise. The contaminated data samples are named outliers in robust statistics (Huber 1981; Maronna, Martin, and Yohai 2006). The large size of the modern dataset makes the conventional manual trace editing impractical. Several automatic methods based on outlier detection have been proposed to denoise seismic data contaminated by erratic noise (Elboth, Presterud, and Hermansen 2010; Bekara and van der Baan 2010). 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 (Cambois and Frelet 1995; Soubaras 1995) or detected, clipped, and iteratively interpolated by *f-x* prediction filters (Schonewille, Vigner, and Ryder 2008). Empirical studies show that the outlier detection-based methods may be not able to deal with multiple outliers because of the so-called masking effect (one outlier may hide the presence of others) (Hampel 1985). The breakdown point of least-squares estimation after outlier rejection is lower than that of robust estimation (M-estimators) (Hampel 1985). Therefore, instead of outlier detection techniques followed by least-squares estimation, we propose to apply direct robust estimation (Chen and Sacchi 2014). The proposed robust *f-x* projection method can simultaneously remove random Gaussian noise and erratic noise and preserve the signal amplitude. The misfit between the observed data and the modelled 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 via 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.

Robust estimation (inversion) has been used in geophysics for seismic deconvolution (Claerbout and Muir 1973; Taylor, Banks, and McCoy 1979; Gholami and Sacchi 2012), travel-time tomography (Scales and Gersztenkorn 1988; Bube and Langan 1997), full-waveform inversion (Crase *et al*. 1990; Ha, Chung, and Shin 2009; Brossier, Operto, and Virieux 2010; Aravkin, van Leeuwen, and Herrmann 2011), velocity analysis (Guitton and Symes 2003; Li, Zhang, and Claerbout 2012), simultaneous source separation (Ibrahim and Sacchi 2014), and matrix rank reduction-based erratic noise removal and interpolation (Chen and Sacchi 2013; Chen 2013; Chen and Sacchi 2015). The main contribution of this paper is the introduction of a robust inversion methodology to the problem of estimating projection filters for seismic noise suppression.

THEORY

Sinusoids in additive noise

The seismic signal is usually corrupted with seismic noise resulting from various sources. We will consider the typical situation where a signal in the *f-x* domain is corrupted by not only Gaussian noise but also erratic (impulsive) noise. The observed seismic data at one frequency $(y_1(\omega), y_2(\omega), \ldots, y_N(\omega))$ can be regarded as a discrete time series. We will omit the symbol ω in the following text and understand that the analysis is carried out for different frequencies in a predetermined frequency band. The time series is a sample realization from a wide-sense stochastic process. The discrete stochastic process modelling the time series can be expressed as

$$\tilde{y}_n = \tilde{x}_n + \tilde{n}_n + \tilde{i}_n, \quad (1)$$

where we use tilded letters to represent stochastic processes and normal letters to represent time series ($y_n = x_n + n_n + i_n$). For instance, $\tilde{x}_n$ is a deterministic process representing the signal, $\tilde{n}_n$ is a stationary stochastic process representing the additive complex white Gaussian noise, and $\tilde{i}_n$ is a stationary stochastic process representing additive impulsive noise (Fox 1972). The signal $\tilde{x}_n$, the Gaussian noise $\tilde{n}_n$, and the impulsive noise $\tilde{i}_n$ are assumed to be mutually independent. The process $\tilde{e}_n = \tilde{n}_n + \tilde{i}_n$ represents the mixture of additive noises. Equation 1 can be rewritten as

$$\tilde{y}_n = \tilde{x}_n + \tilde{e}_n, \quad (2)$$

where the additive noise $\tilde{e}_n$ only affects the current observation and will not affect the subsequent observations. A noise-free 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 discrete noise-free signal process $\tilde{x}_n$ can be represented by

$$\tilde{x}_n = \sum_{k=1}^{p} A_k(\omega) e^{-i\omega\eta_k(n-1)\Delta s}, \quad (3)$$

where $A_k(\omega)$ is the Fourier transform of the source wavelet corresponding to the *k*th event, η_k is the *k*th dip, Δs is the spatial interval between two channels, and $i = \sqrt{-1}$. Equation 2 and 3 are the *sinusoidal model* representing the sinusoids embedded in white noise (Stoica and Moses 2005). It can be shown that the exponential signal satisfies the *p*th-order homogeneous difference equation

$$\tilde{x}_n + f_1\tilde{x}_{n-1} + f_2\tilde{x}_{n-2} + \cdots + f_p\tilde{x}_{n-p} = 0, \quad (4)$$

or in *z*-domain notation

$$\tilde{X}(z)F(z) = 0, \quad (5)$$

with $F(z) = \prod_{k=1}^{p}(1 - z_k^{-1}z)$ and $z_k = e^{i\omega\eta_k\Delta s}$. The coefficients $f_0 = 1, f_1, f_2, \ldots, f_p$ are the so-called prediction error filter (Canales 1984; Gulunay 1986). In this particular case, the prediction error filter is also the *annihilating filter* of the discrete signal $\tilde{x}_n$ (Blu *et al*. 2008). The exponential signal $\tilde{x}_n$ is represented by line spectra consisting of *p* impulses located at sinusoidal frequencies $\omega\eta_k$ ($k = 1, \ldots, p$). The latter is also a "degenerate AR process" with innovations equal to zero and poles that lie on the unit circle. Substituting $\tilde{x}_n = \tilde{y}_n - \tilde{e}_n$ into equation 4, yields

$$\sum_{k=0}^{p} f_k\tilde{y}_{n-k} = \sum_{k=0}^{p} f_k\tilde{e}_{n-k}, \quad (6)$$

or in *z*-domain notation

$$\tilde{Y}(z)F(z) = \tilde{E}(z)F(z). \quad (7)$$

This ARMA model is special in the sense that the AR and MA coefficients ($f_0, f_1, \ldots, f_p$) are identical (Ulrych and Clayton 1976). Its poles and zeros are located on the unit circle, and they overlap each other.

A special ARMA model for observed time series

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

$$\tilde{x}_n = -\sum_{k=1}^{p} f_k\tilde{x}_{n-k} + \tilde{\epsilon}_n, \quad (8)$$

where $f_k, k = 1, 2, \ldots, p$, are the AR coefficients and $\tilde{\epsilon}_n$ indicates white noise sequence (innovation). Substituting $\tilde{x}_n = \tilde{y}_n - \tilde{e}_n$ into equation 8 leads to

$$\sum_{k=0}^{p} f_k\tilde{y}_{n-k} = \sum_{k=0}^{p} f_k\tilde{e}_{n-k} + \tilde{\epsilon}_n. \quad (9)$$

This equation is an ARMA process (Kay 1978) similar to equation 6; however, the process now contains an innovation term.

Estimation of ARMA parameters and additive noise

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 (ARMA parameters) is estimated. Then, the prediction error filter is fixed, and the additive noise sequence is estimated. The two stages are iterated until reaching convergence. The random process $\tilde{y}_n$ is observed over a spatial interval of *N* points leading to the observation vector $\mathbf{y} = (y_1, y_2, \ldots, y_N)^T$. The realization of signal plus additive noise (equation 2) is given by

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{pmatrix}. \quad (10)$$

The forward and backward linear prediction method (modified covariance method) (Ulrych and Clayton 1976) is used to represent the AR process in equation 8

$$\begin{pmatrix} x_{p+1} & x_p & \cdots & x_1 \\ x_{p+2} & x_{p+1} & \cdots & x_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_N & x_{N-1} & \cdots & x_{N-p} \\ x_1^* & x_2^* & \cdots & x_{p+1}^* \\ x_2^* & x_3^* & \cdots & x_{p+2}^* \\ \vdots & \vdots & \ddots & \vdots \\ x_{N-p}^* & x_{N-p+1}^* & \cdots & x_N^* \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \\ \vdots \\ f_p \end{pmatrix} = \begin{pmatrix} \epsilon_{p+1} \\ \epsilon_{p+2} \\ \vdots \\ \epsilon_N \\ \varphi_{p+1}^* \\ \varphi_{p+2}^* \\ \vdots \\ \varphi_N^* \end{pmatrix}, \quad (11)$$

where $\epsilon_i, i = p+1, \ldots, N$, are the forward prediction errors and $\varphi_i, i = p+1, \ldots, N$, are the backward prediction errors. Equation 11 can be notated as

$$\mathbf{Xf} = \begin{pmatrix} \boldsymbol{\epsilon} \\ \boldsymbol{\varphi}^* \end{pmatrix}, \quad (12)$$

where $\boldsymbol{\epsilon}$ is the forward prediction error vector and $\boldsymbol{\varphi}$ is the backward prediction error vector. Fig. 1 provides a diagram highlighting forward and backward prediction.

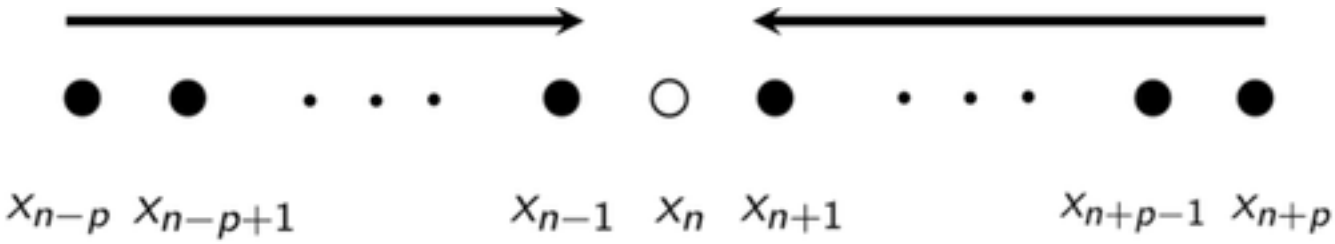

Figure 1

Open in figure viewerPowerPoint

Forward and backward prediction scheme. Black dot represents already predicted data sample, and white dot indicates the data sample to be predicted. Arrows indicate the direction of prediction.

Caption

Given the commutative property of convolution, equation 11 can be written as follows:

$$\begin{pmatrix} f_p & f_{p-1} & \cdots & f_0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & f_p & \cdots & f_1 & f_0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & f_p & \cdots & f_0 \\ f_0^* & f_1^* & \cdots & f_p^* & 0 & \cdots & 0 & \cdots & 0 \\ 0 & f_0^* & \cdots & f_{p-1}^* & f_p^* & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & f_0^* & \cdots & f_p^* \end{pmatrix} \times \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} = \begin{pmatrix} \epsilon_{p+1} \\ \epsilon_{p+2} \\ \vdots \\ \epsilon_N \\ \varphi_{p+1} \\ \varphi_{p+2} \\ \vdots \\ \varphi_N \end{pmatrix}, \quad (13)$$

or in matrix formulation

$$\mathbf{Fx} = \begin{pmatrix} \boldsymbol{\epsilon} \\ \boldsymbol{\varphi} \end{pmatrix}. \quad (14)$$

With $\mathbf{x} = \mathbf{y} - \mathbf{e}$, this equation changes to

$$\mathbf{F}(\mathbf{y} - \mathbf{e}) = \begin{pmatrix} \boldsymbol{\epsilon} \\ \boldsymbol{\varphi} \end{pmatrix}, \quad (15)$$

where **F** is a convolutional matrix containing the elements of the unknown filter coefficients. Our task is to estimate the prediction error filter **f** and the noise sequence **e**. Soubaras (1994) and Sacchi and Kuehl (2001) constrain the noise term **e** by ℓ_2-norm. In this paper, we propose adopting a constraint that minimizes the hybrid ℓ_1/ℓ_2-norm (Bube and Langan 1997; Li *et al*. 2012) of the noise sequence **e**. We estimate the prediction error filter **f** and the noise sequence **e** via minimizing the cost function

$$\mathcal{J}(\mathbf{e},\mathbf{f}) = \frac{1}{2}||\mathbf{F}(\mathbf{y}-\mathbf{e})||_2^2 + \lambda\mathcal{H}(\mathbf{e}), \tag{16}$$

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

$$h(e) = \sqrt{\sigma^2 + |e|^2} - \sigma. \tag{17}$$

In Fig. 2, we compare the normalized hybrid function $f(x) = \sqrt{1+|x|^2} - 1$ and the normalized quadratic function $f(x) = \frac{1}{2}|x|^2$. In the current situation, *x* is a normalized additive noise, i.e., $x = e/\sigma$, where σ is the scale parameter.

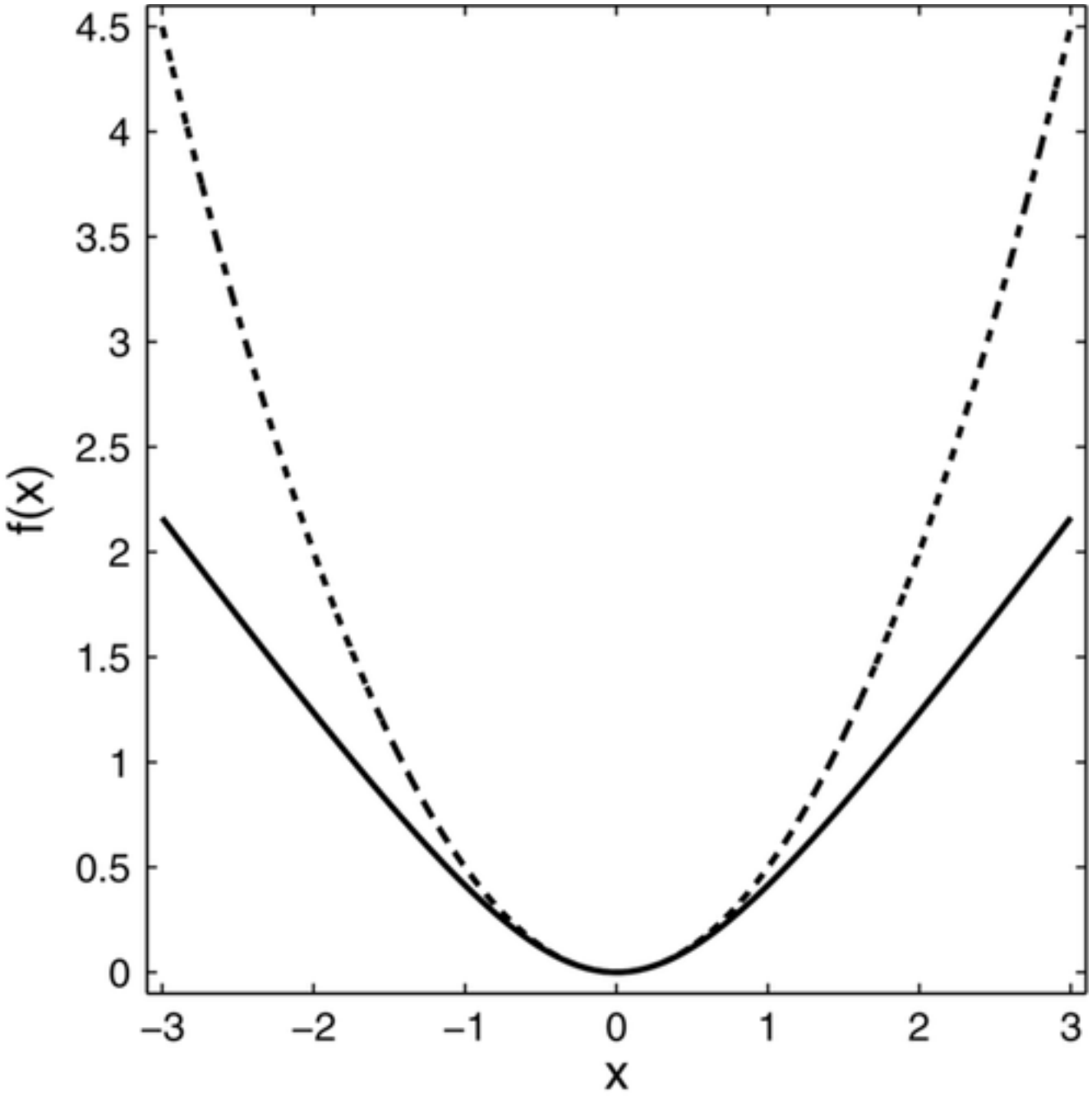

Figure 2

Open in figure viewerPowerPoint

The dashed line represents the quadratic function $\frac{1}{2}|x|^2$. The solid line represents the hybrid function $\sqrt{1+|x|^2}-1$.

Caption

The optimization problem (equation 16) is nonlinear because of the coupling of the two unknowns. It can be tackled by an alternating minimization technique (Golub and Pereyra 1973; Kaufman 1975). We will first simplify the problem by assuming that the prediction error filter is

already estimated from the previous iteration. The estimation of the noise sequence reduces as minimizing the cost function

$$\mathcal{J}(\mathbf{e}) = \frac{1}{2}\|\mathbf{F}(\mathbf{y}-\mathbf{e})\|_2^2 + \lambda\mathcal{H}(\mathbf{e}), \quad (18)$$

Setting $\frac{\partial \mathcal{J}}{\partial \mathbf{e}^*} = 0$ leads to the "nonlinear normal equations" (details are given in Appendix)

$$\left(\mathbf{F}^H\mathbf{F} + \lambda\mathbf{W}\right)\mathbf{e} = \mathbf{F}^H\mathbf{F}\mathbf{y}, \quad (19)$$

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

$$W_{jj}^{(k-1)} = 1/\sqrt{\sigma^2 + |e_j^{(k-1)}|^2}, \quad j = 1, 2, \ldots, N, \quad (20)$$

The iterative solution is given by

$$\mathbf{e}^{(k)} = (\mathbf{F}^H\mathbf{F} + \lambda\mathbf{W}^{(k-1)})^{-1}\mathbf{F}^H\mathbf{F}\mathbf{y}. \quad (21)$$

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

$$\mathcal{J}(\mathbf{f}) = \frac{1}{2}\|\mathbf{F}\mathbf{x}\|_2^2. \quad (22)$$

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

$$\hat{\mathbf{g}} = (\bar{\mathbf{X}}^H\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}^H\bar{\mathbf{x}}, \quad (23)$$

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

$$\hat{\mathbf{f}} = (1, -\hat{\mathbf{g}}^T)^T. \quad (24)$$

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 **x** by least-squares estimation (AR modelling or ARMA modelling) and compute an initial noise term $\mathbf{e} = \mathbf{y} - \mathbf{x}$.
- 2. Estimate the prediction error filter **f** via equations 23 and 24.

- 3. Estimate the noise sequence **e** by minimizing the cost function (equation 18).
- 4. Iterate steps 2–3 until convergence.

Parameter σ is fixed, and the tradeoff parameter λ is tuned by examining the residuals. This is similar to the strategy often used in *f-x* deconvolution for parameter selection. For a very wide range of parameters, the erratic noise is well removed. For fixed scale parameter σ, a smaller value of λ will result in a cleaner section. Meanwhile, the signal will be damaged if λ is chosen too small.

The algorithm has two groups of iterations, namely, an internal iteration (IRLS) to estimate **e**and an external iteration for alternating minimization. We have two convergence criteria to reduce the number of iterations. We monitor the cost function $\mathcal{J}(\mathbf{e},\mathbf{f})$ and terminate the external loop when the relative change of the cost function between two consecutive iterations is less than a tolerance tol_1. A second tolerance tol_2 is used to control the number of IRLS iterations that are required to estimate **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 in this paper averages the forward and backward predicted values and uses prediction matrix corresponds to transient-free formulation (Sacchi 2008). The conventional *f-x* projection filter used here is a modification of the method of Sacchi and Kuehl (2001) that uses the modified covariance method. Figure 3a shows 2-D synthetic data with noise. The central frequency of the Ricker wavelet is 20 Hz. Figure 3b shows band-limited 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 3c shows the high-amplitude erratic noise. The maximum amplitude of the erratic noise is approximately five times the maximum amplitude of the signal in Fig. 3a.

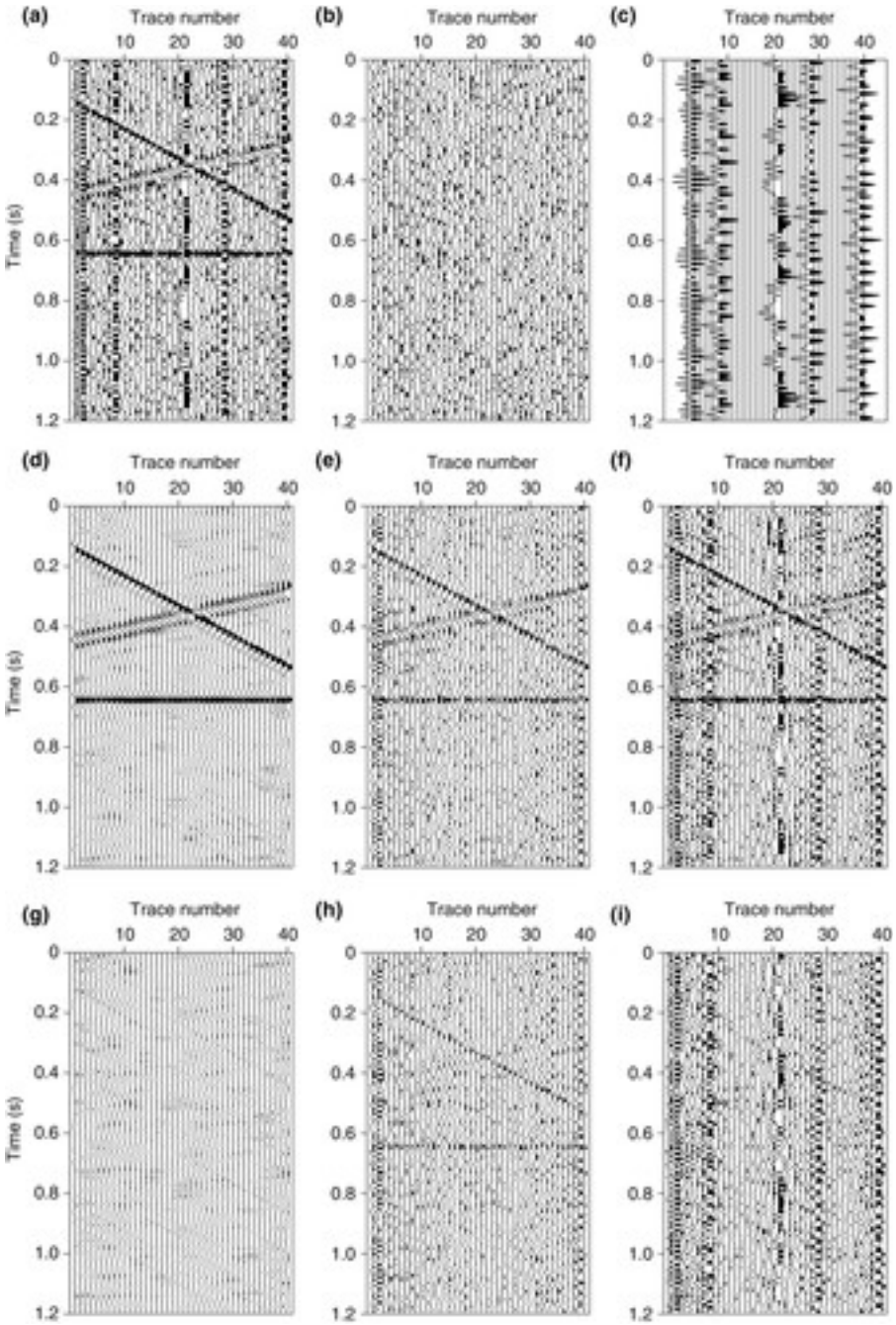

Figure 3

Open in figure viewerPowerPoint

(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 least-squares *f-x* projection filter.

Caption

The processing frequency band ranges from 1 Hz to 60 Hz. The length of the prediction error filter for the robust *f-x* projection filtering is set to 4. Scale parameter σ and trade-off parameter λ

are 6 and 0.1, respectively. The length of the prediction error filter of the *f-x*deconvolution is 11. The length of the prediction error filter in the conventional *f-x* projection filtering method is 4, and the pre-whitening parameter is 3. The filtered data by robust *f-x*projection, *f-x* deconvolution, and least-squares *f-x* projection are shown in Figs. 3d, 3e, and 3f, 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 Figs. 3g, 3h, and 3i show that the robust *f-x* projection preserves the original signal. On the other hand, *f-x*deconvolution damages the signal. We tested *f-x* deconvolution and least-squares *f-x*projection with a variety of parameters, but we never managed to produce fully satisfying results when the data are contaminated by high-amplitude erratic noise. 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 the noise-free data section, $\hat{\mathbf{D}}$ denotes the filtered data section, 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 13.1. The Q value for the *f-x* deconvolution is -1.2. The Q value for the *f-x*projection is -15.5. The *f-k* spectra in Fig. 4 show that the proposed robust *f-x* projection reasonably recovers the true spectrum but *f-x* deconvolution and least-squares *f-x* projection cannot.

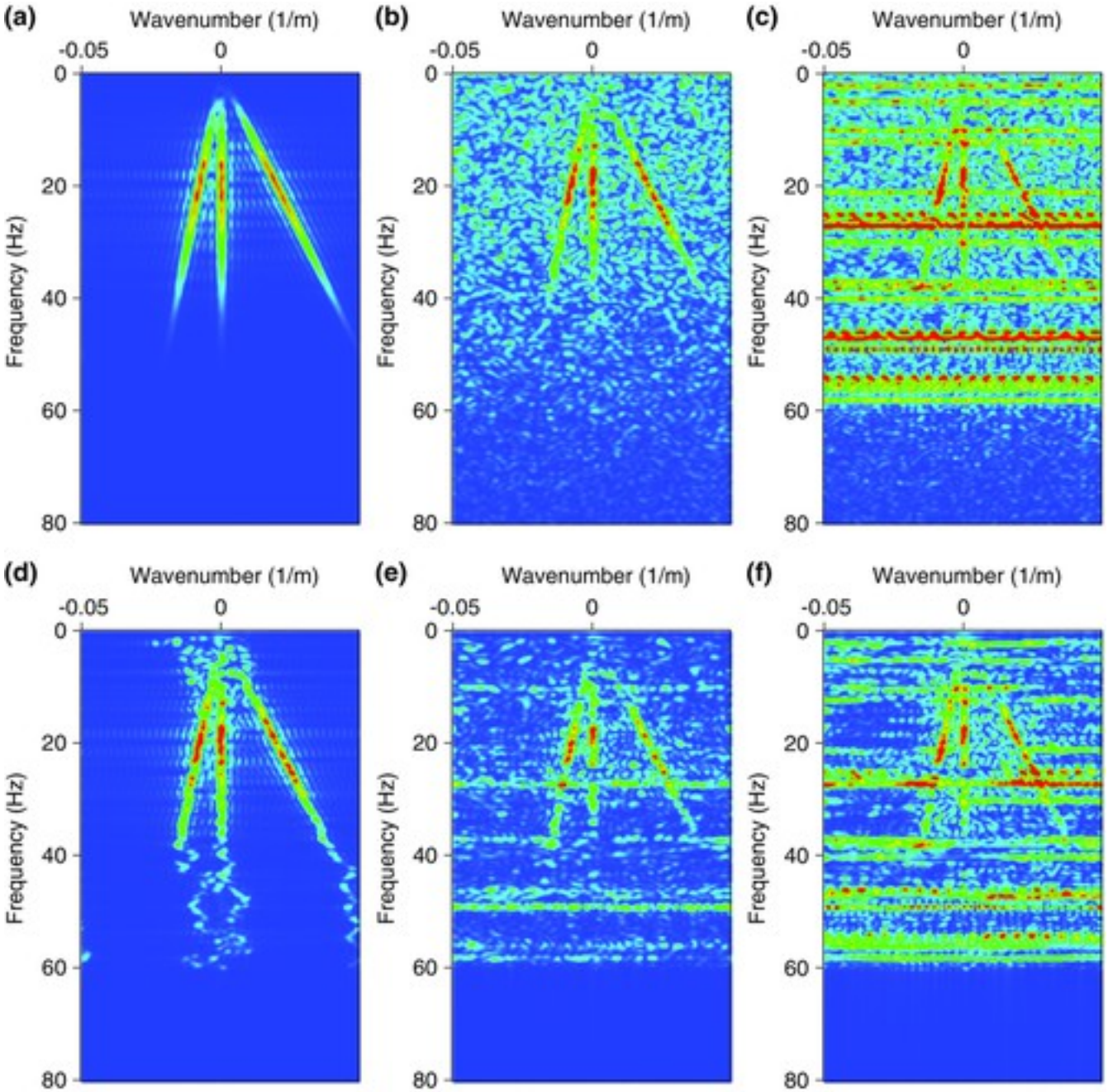

Figure 4

Open in figure viewerPowerPoint

(a) *f-k* spectrum of noise-free data. (b) *f-k* spectrum of data corrupted with Gaussian noise. (c) *f-k*spectrum of the input noisy data (Fig. 3a). (d) *f-k* spectrum of robust *f-x* projection filtered data (Fig. 3d). (e) *f-k* spectrum of *f-x* deconvolution filtered data (Fig. 3e). (f) *f-k* spectrum of least-squares *f-x* projection filtered data (Fig. 3f).

Caption

Poststack field data example

We tested our proposed algorithm on a poststack field dataset from the Western Canadian Sedimentary Basin (WCSB). The performances of robust *f-x* projection, *f-x* deconvolution, and conventional *f-x* projection are compared. Figure 5a is a poststack data section with erratic noise and random Gaussian noise. Magnified portions of data in the left and right rectangular windows highlighted in Fig. 5a are shown in Figs. 5b and 5c, respectively. The complete data in Fig. 5a are divided into overlapping windows. All windows are processed and then added back. Each window has 50 traces with 50% overlap and 300 time samples (0.6 s) with 33% overlap. All the three filtering methods are applied for frequencies in the band of 1 Hz–80 Hz. The length of the prediction error filter for robust *f-x* projection filtering is 4. Scale parameter σ and tradeoff parameter λ are 10^{-3} and 0.1, respectively. The length of the prediction error filter for the *f-x* deconvolution is 5. The length of the prediction error filter for the conventional *f-x*projection filter method is 4, and the pre-whitening parameter is 0.1.

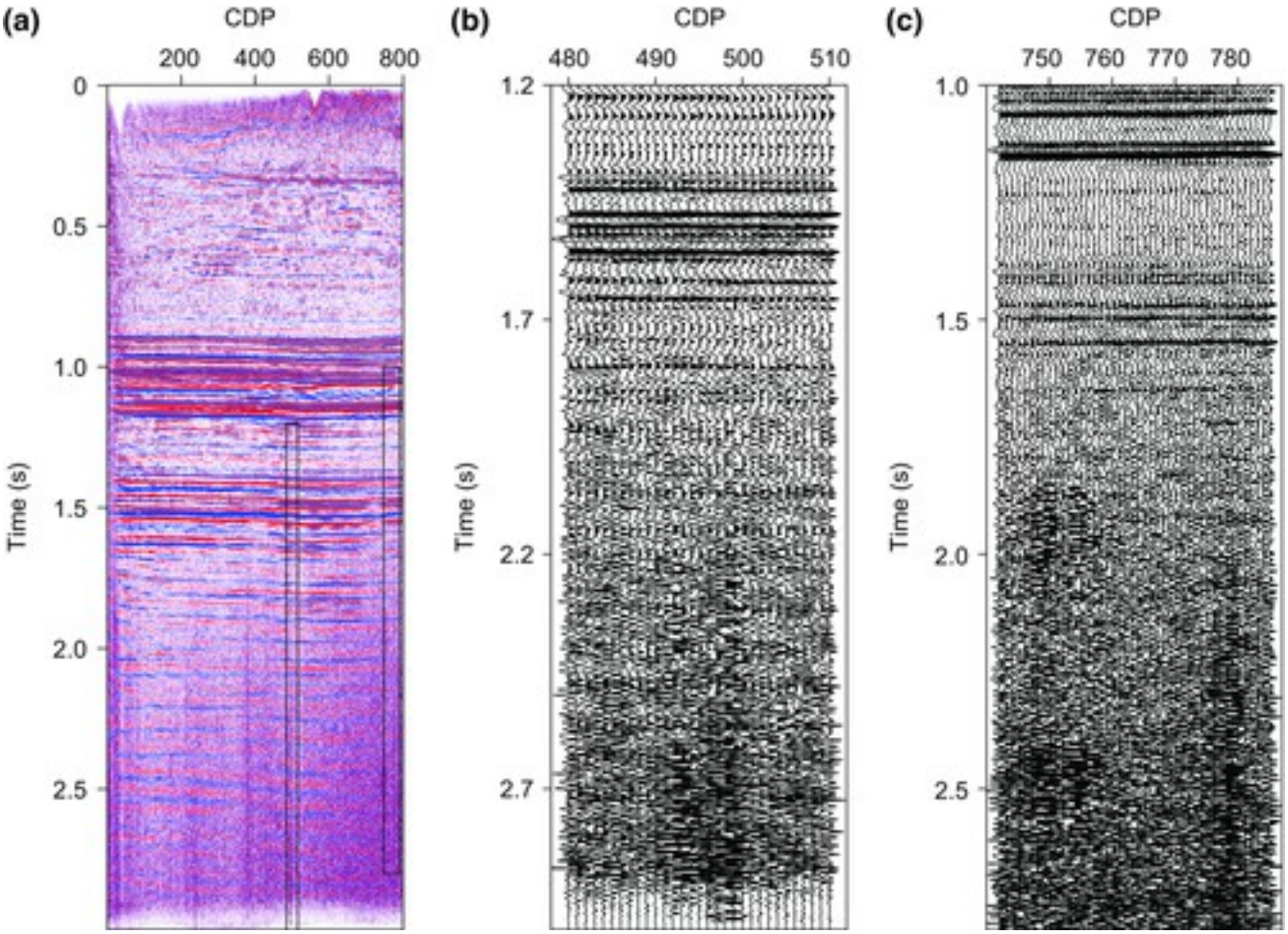

Figure 5

Open in figure viewerPowerPoint

(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.

Caption

The denoising results of the robust *f-x* projection, *f-x* deconvolution, and least-squares *f-x*projection are shown in Figs. 6a, 6b, and 6c, respectively. The result of robust *f-x* projection is cleaner than the other two. The difference sections in Fig. 7 show that the robust *f-x* projection preserves the original seismic signal. To show the details more clearly, we display the zoomed results and difference sections for the left window in Fig. 5a in Figs. 8 and 9, respectively. The zoomed results and difference sections for the right window in Fig. 5a are shown in Figs. 10and 11, respectively. The proposed robust *f-x* projection removes more erratic noise than *f-x*deconvolution and least-squares *f-x* projection.

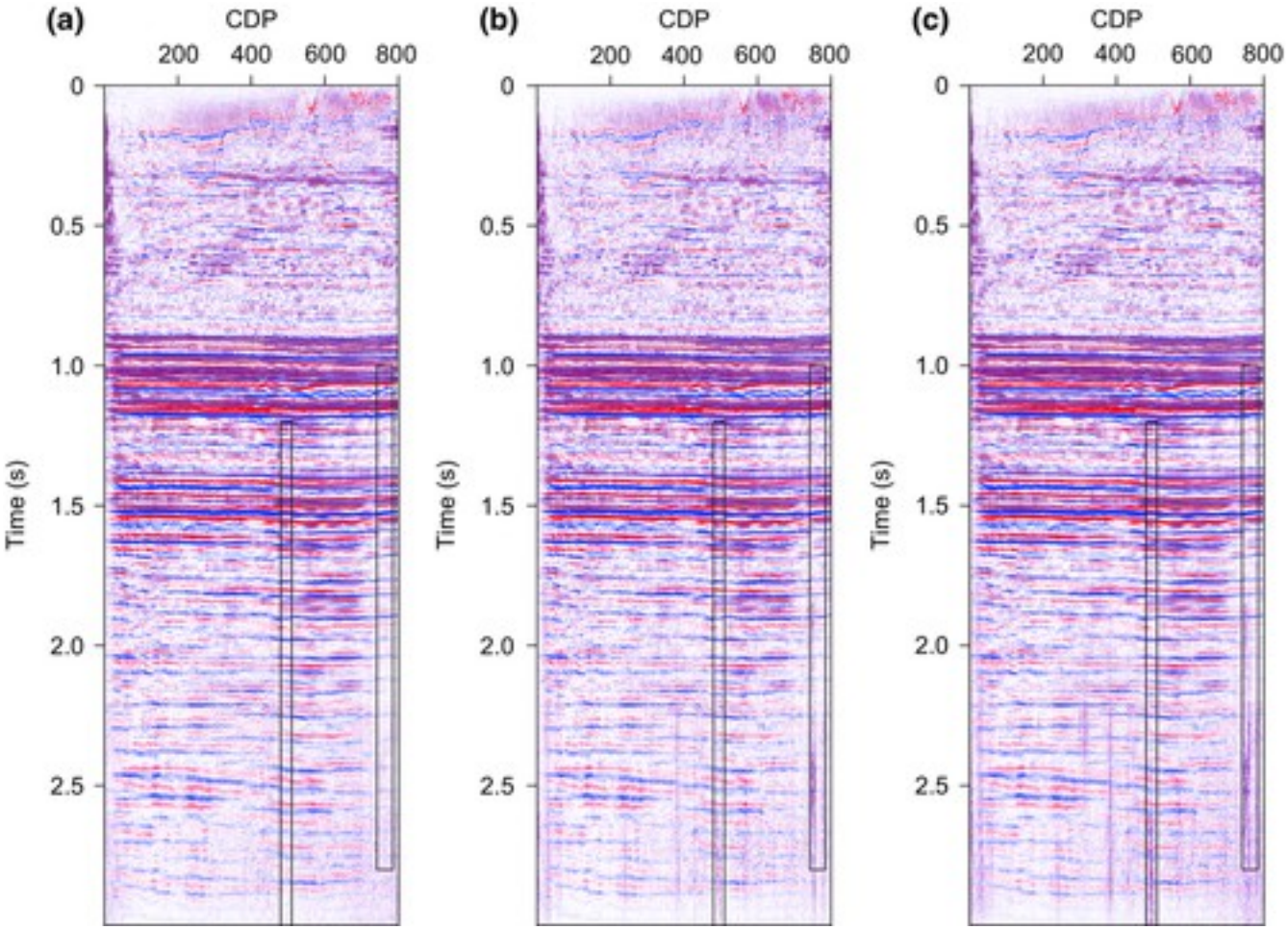

Figure 6

Open in figure viewerPowerPoint

(a) Data after robust *f-x* projection filtering. (b) Data after *f-x* deconvolution filtering. (c) Data after least-squares *f-x* projection filtering.

Caption

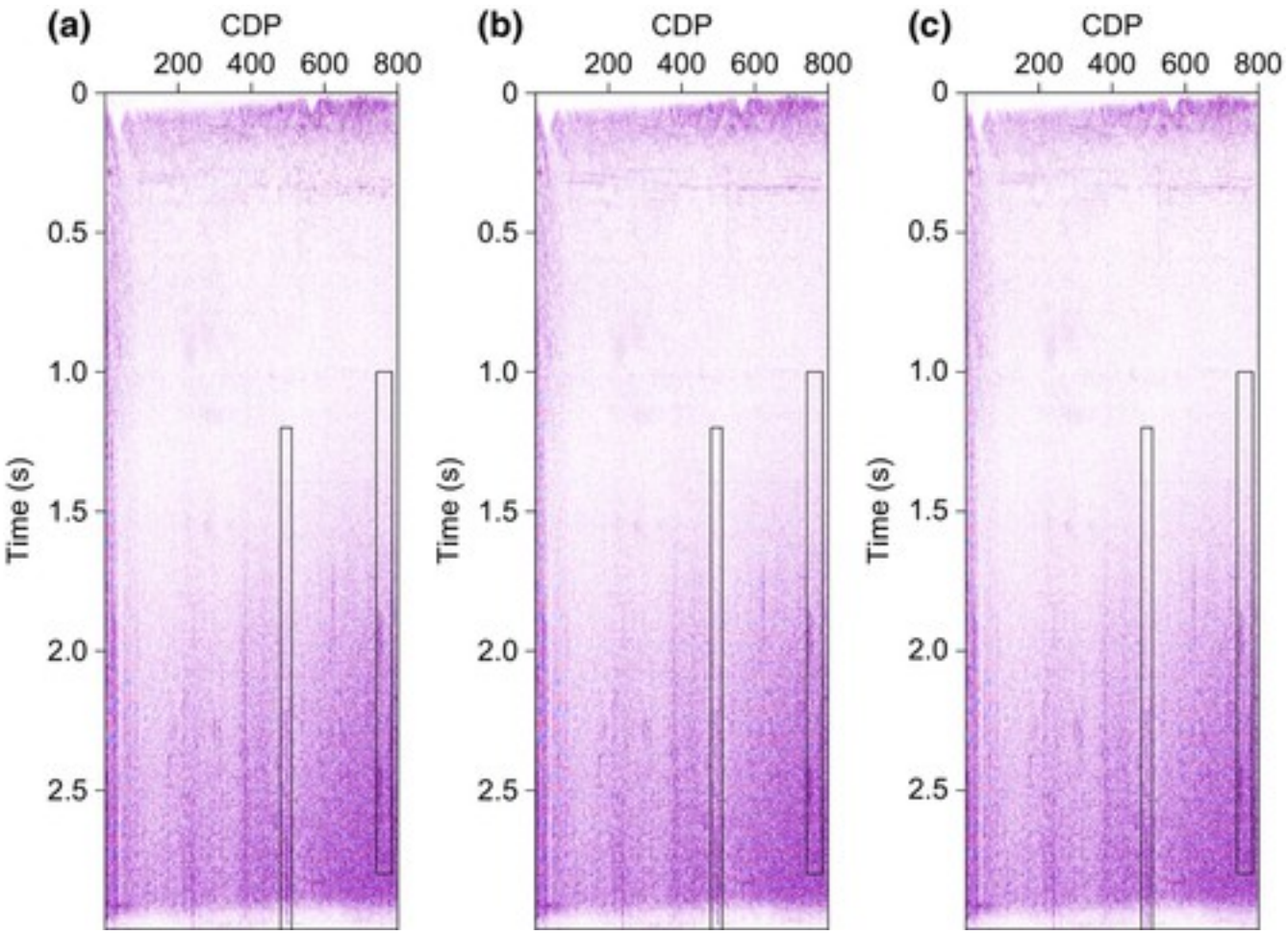

Figure 7

Open in figure viewerPowerPoint

Difference sections (input noisy data minus filtered data) of (a) robust *f-x* projection, (b) *f-x*deconvolution, and (c) least-squares *f-x* projection.

Caption

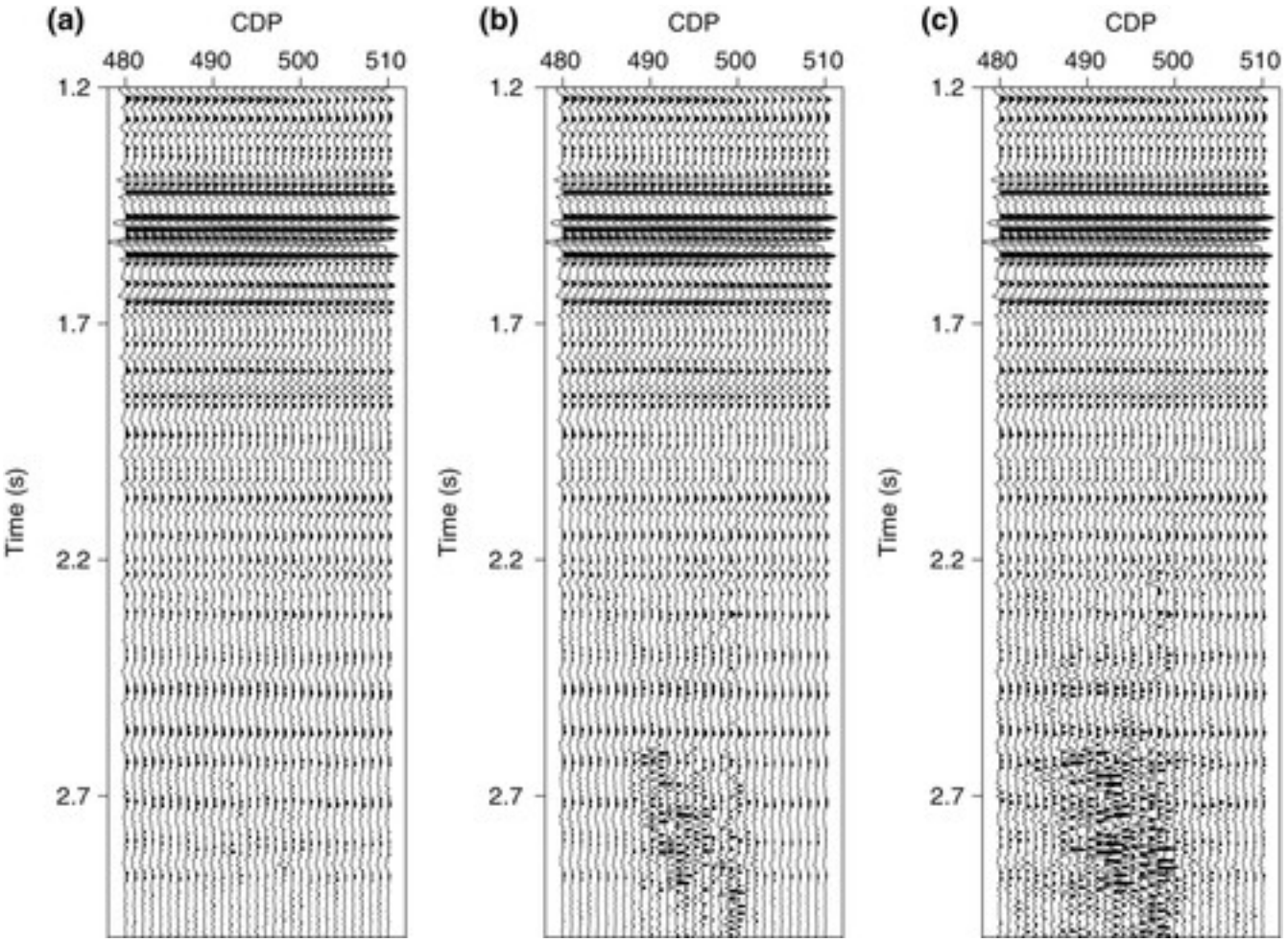

Figure 8

Open in figure viewerPowerPoint

The comparison of the filtered results of the data in the left rectangular window highlighted in Fig. 5a. (a) The result of robust *f-x* projection. (b) The result of *f-x* deconvolution. (c) The result of least-squares *f-x* projection.

Caption

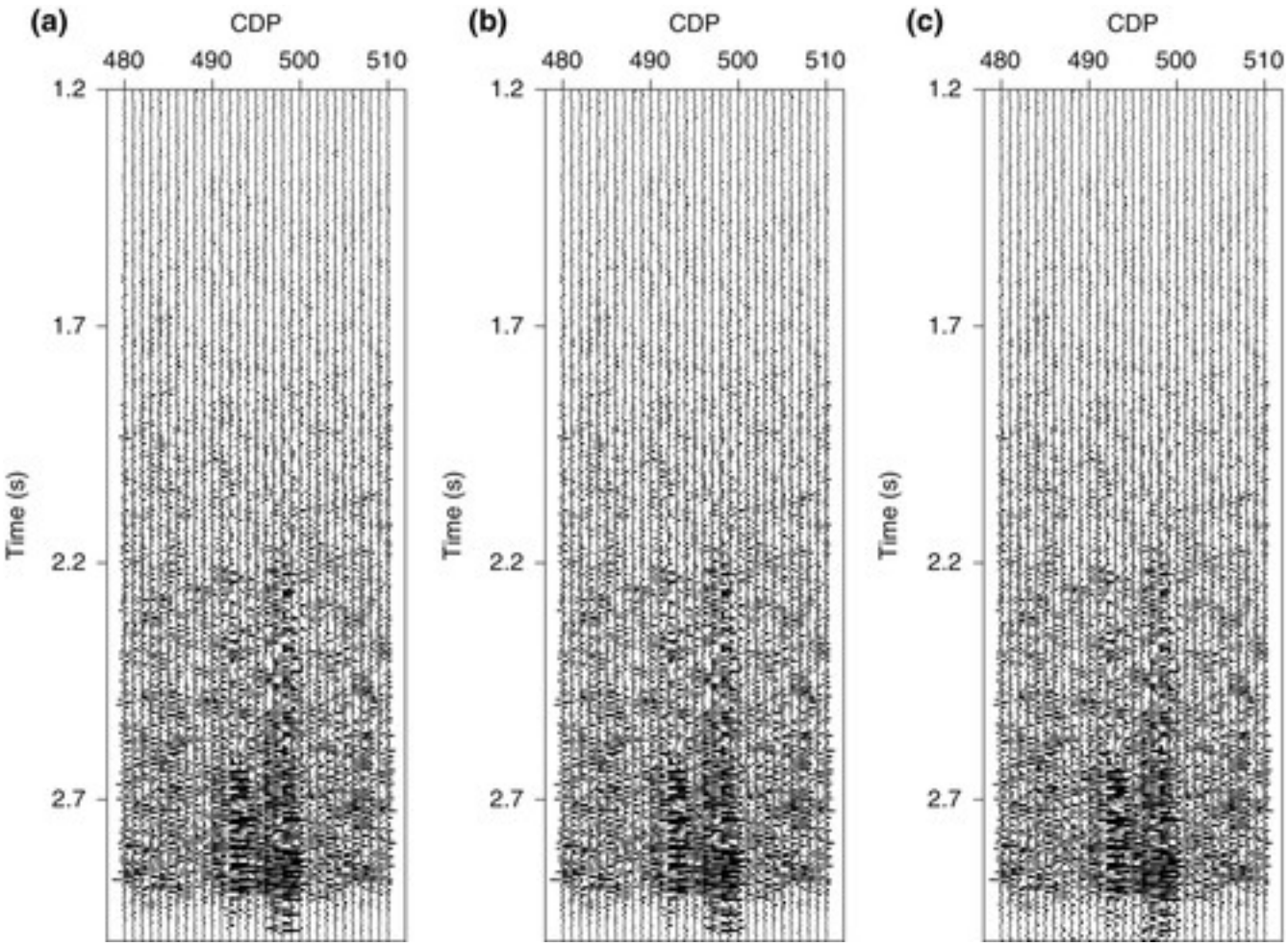

Figure 9

Open in figure viewerPowerPoint

The comparison of the difference sections of three different methods in the left rectangular window highlighted in Fig. 5a. Difference sections of (a) robust *f-x* projection, (b) *f-x*deconvolution, and (c) least-squares *f-x* projection.

Caption

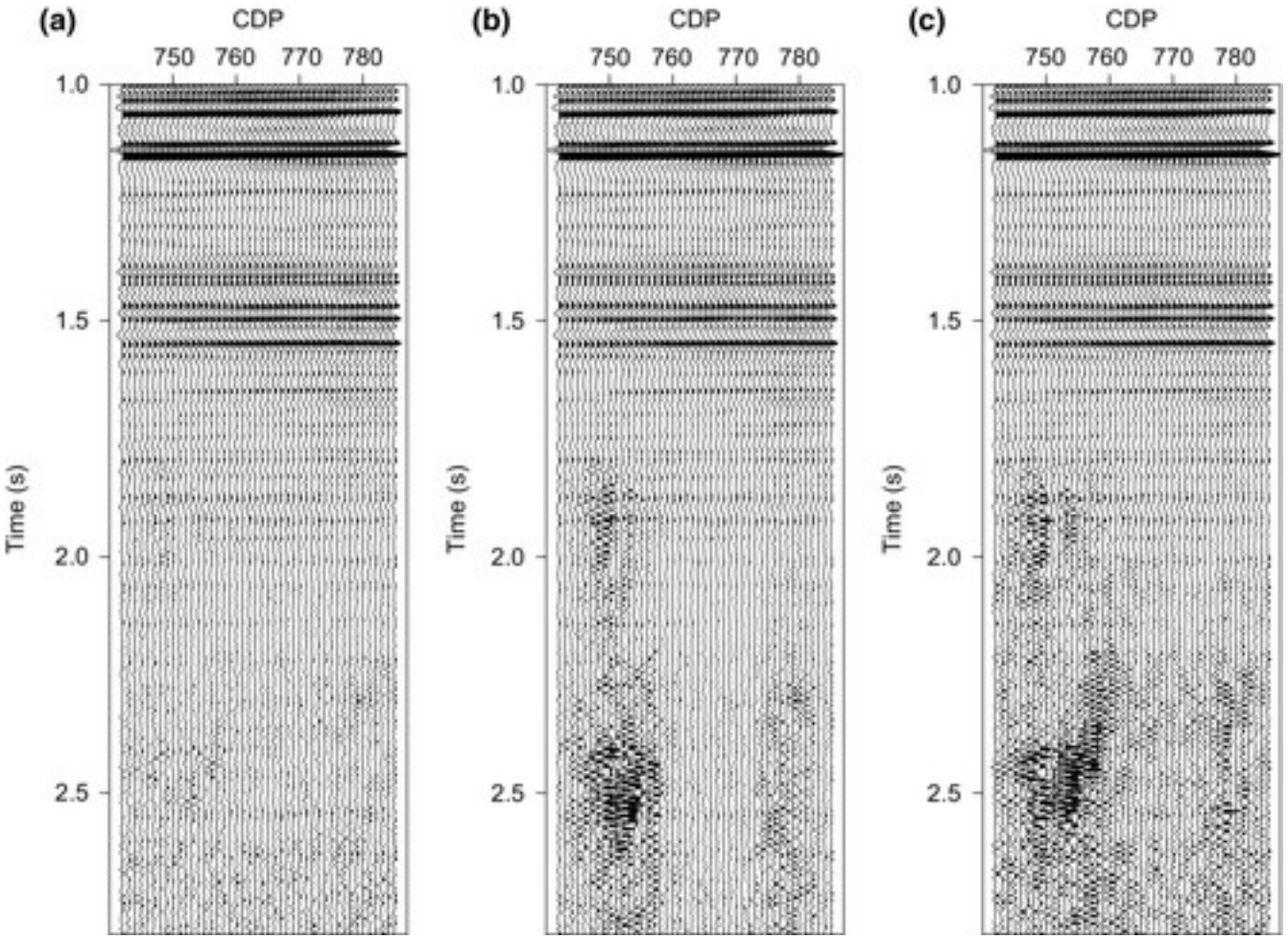

Figure 10

Open in figure viewerPowerPoint

The comparison of the filtered results of the data in the right rectangular window highlighted in Fig. 5a. (a) The result of robust *f-x* projection. (b) The result of *f-x* deconvolution. (c) The result of least-squares *f-x* projection.

Caption

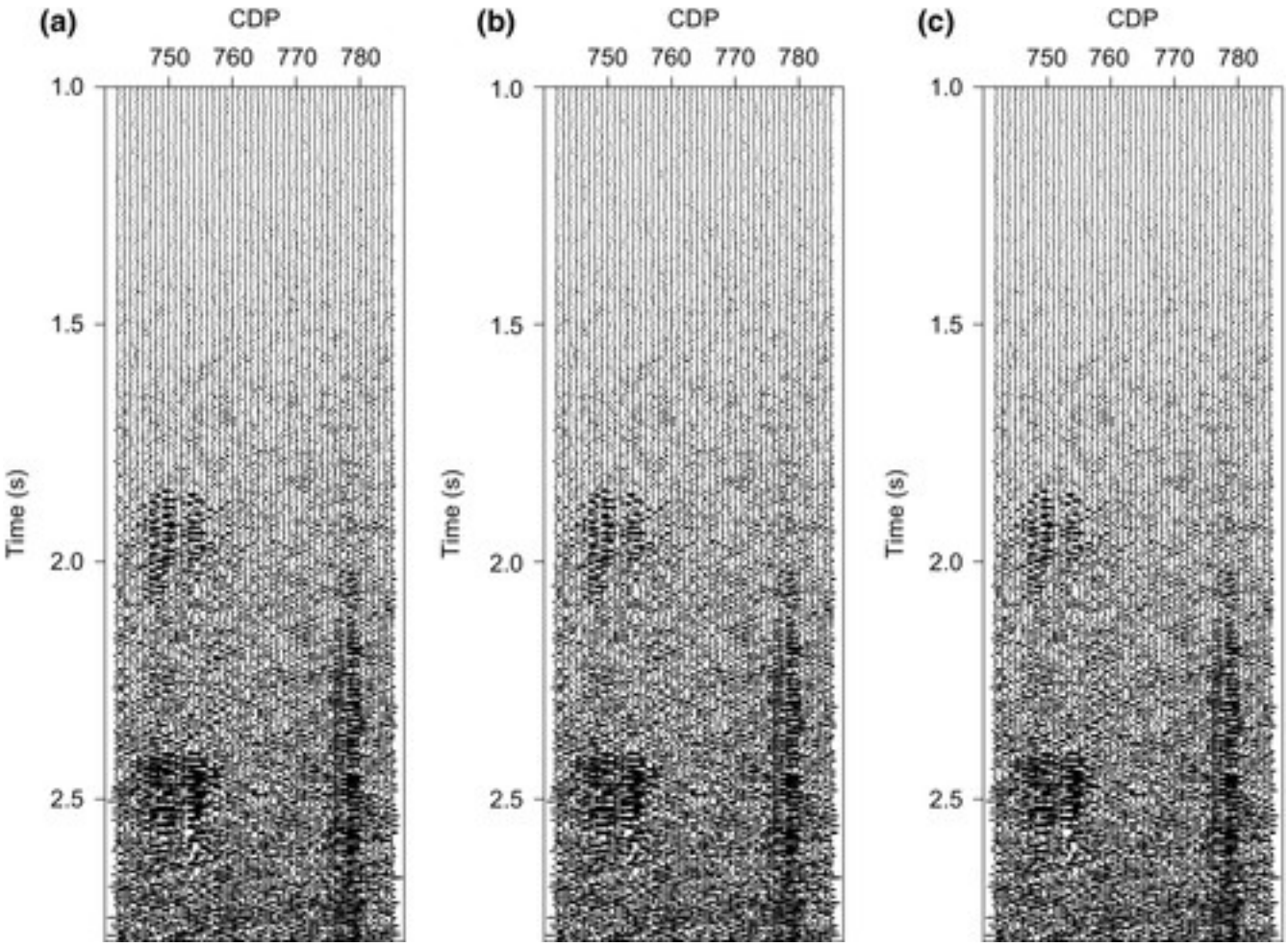

Figure 11

Open in figure viewerPowerPoint

The comparison of the difference sections of three different methods in the right rectangular window highlighted in Fig. 5a. Difference sections of (a) robust *f-x* projection, (b) *f-x*deconvolution, and (c) least-squares *f-x* projection.

Caption

Pre-stack marine data example with swell noise

The proposed algorithm was also tested on a pre-stack marine shot gather (Fig. 12a) that contains swell noise. This is a benchmark dataset used in Elboth *et al*. (2010) and Bekara and van der Baan (2010). Swell noise usually manifests as high-amplitude and low-frequency (2 Hz–10 Hz) vertical stripes (Elboth, Reif, and Andreassen 2009). If not removed, it will pose problems for the following processing steps. Similarly, the shot gather is divided into overlapping windows for processing. The window size is 50 traces by 250 time samples (1 s). The overlapping percentages in time and space are both 50%. The frequency band processed is 1 Hz–120 Hz. The length of the prediction error filter for robust *f-x* projection filtering is 5. Scale parameter σ and tradeoff parameter λ are 10^{-3} and 0.1, respectively. The length of the prediction error filter for the *f-*

x deconvolution is 6. The length of the prediction error filter for the conventional *f*-*x* projection filter method is 5, and the pre-whitening parameter is 0.1. We can see that the robust *f*-*x* projection filter almost completely removed the swell noise (Fig. 12b). While, conventional *f*-*x* deconvolution (Fig. 12c) and *f*-*x* projection (Fig. 12d) do not perform well for swell noise attenuation. The difference sections (Fig. 13) show that the robust *f*-*x* projection preserves the original seismic signal. Figs. 14 and 15 show the filtered results and difference sections of the data in the left rectangular window highlighted in Fig. 12a, respectively. Similarly, Figs. 16 and 17 show the filtered results and difference sections of the data in the right rectangular window highlighted in Fig. 12a, respectively. All of them demonstrate that robust *f*-*x* projection successfully removed the high-amplitude swell noise and preserves the signals.

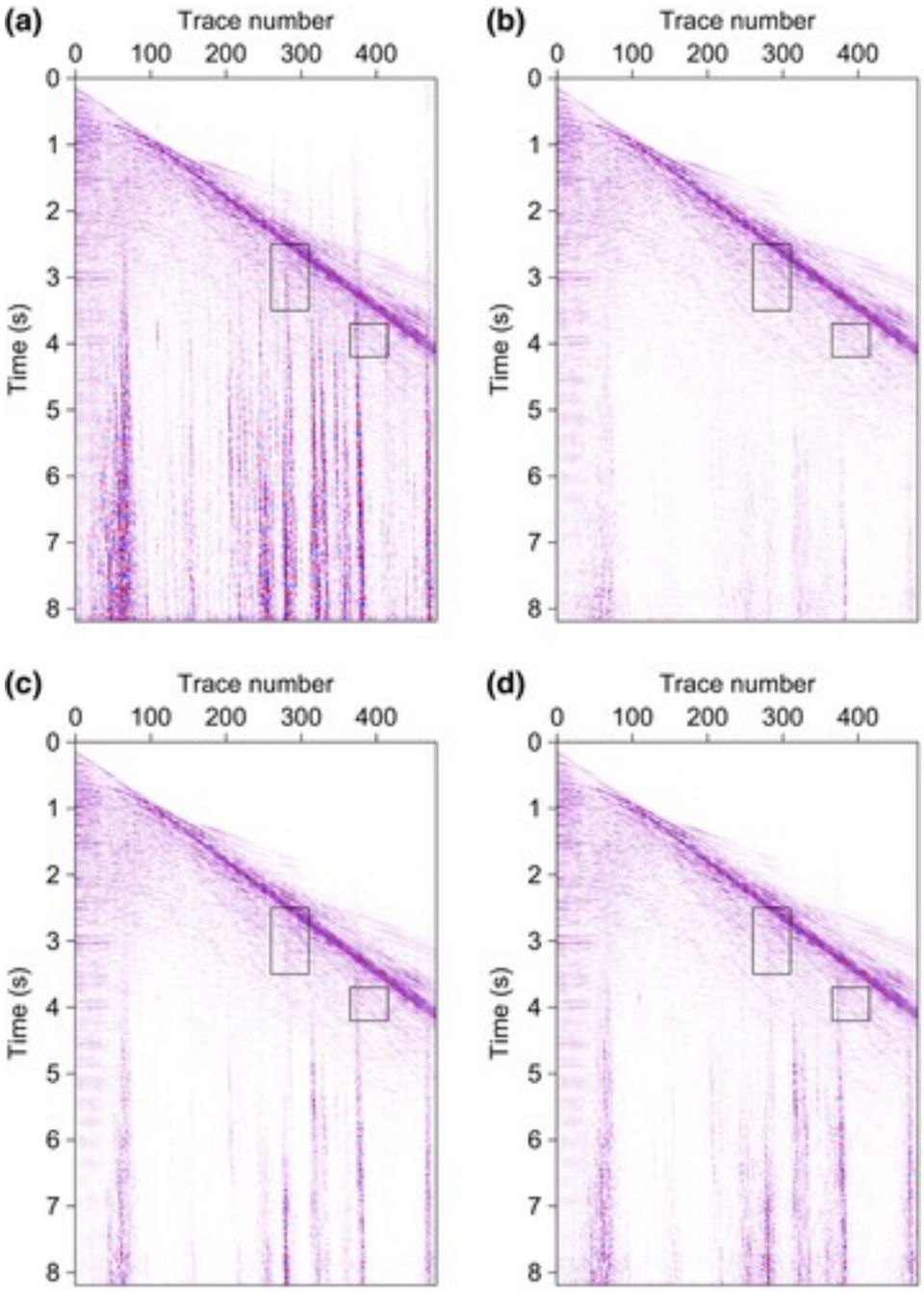

Figure 12

Open in figure viewerPowerPoint

(a) Input marine shot gather with strong swell noise. (b) Data after robust *f-x* projection filtering. (c) Data after *f-x* deconvolution filtering. (d) Data after least-squares *f-x* projection filtering.

Caption

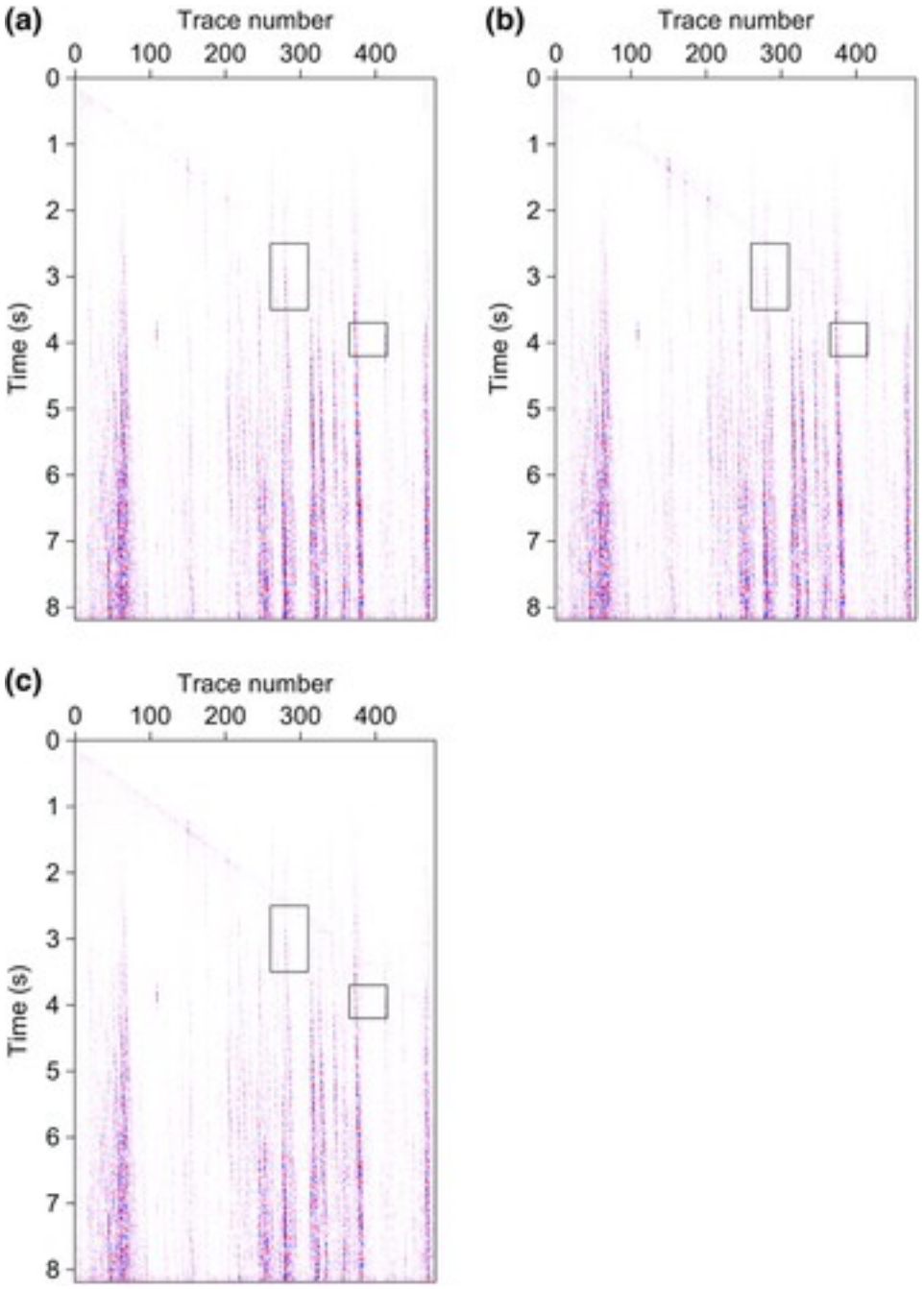

Figure 13

Open in figure viewerPowerPoint

Difference sections of (a) robust *f-x* projection, (b) *f-x* deconvolution, and (c) least-squares *f-x*projection.

Caption

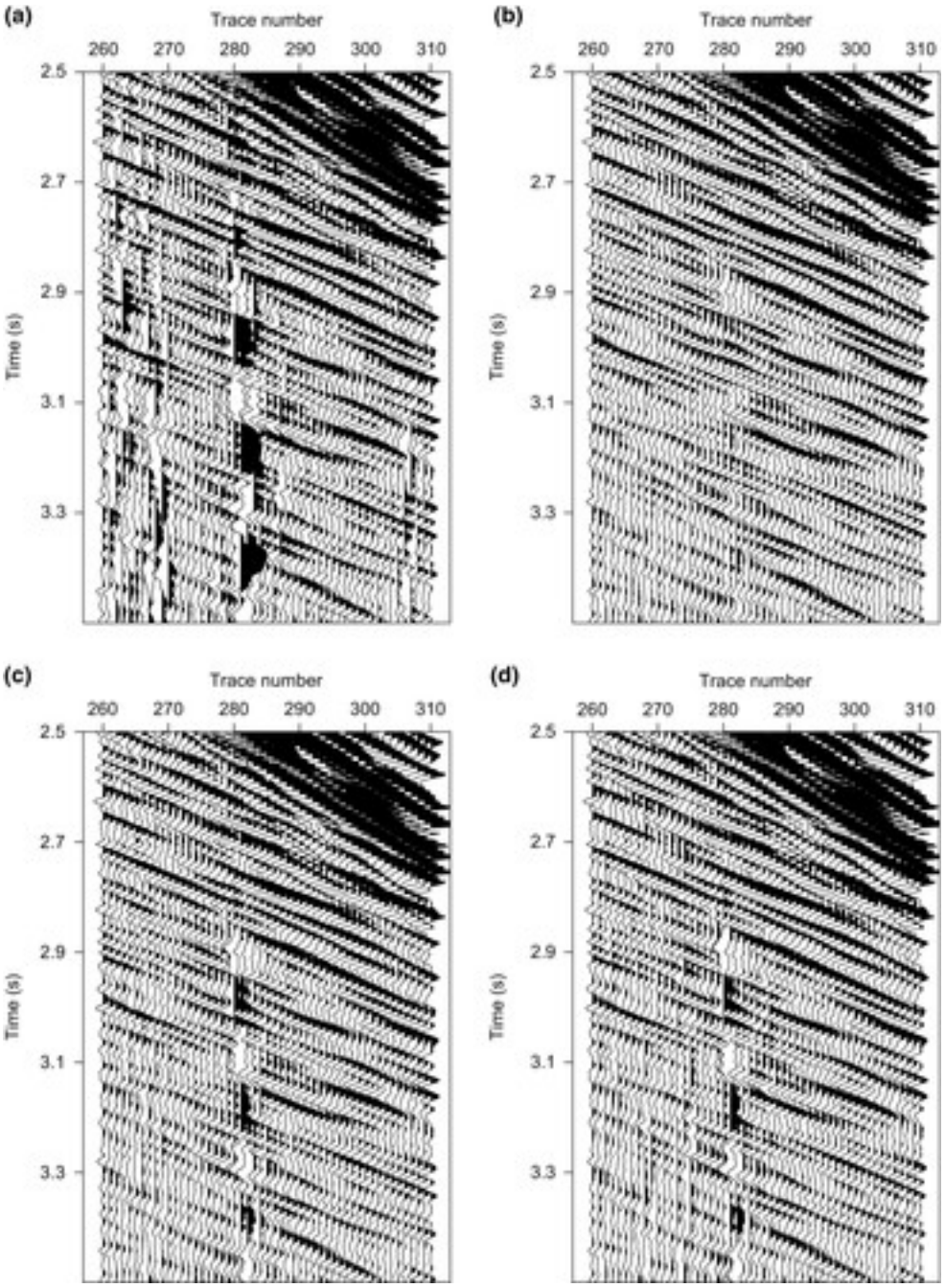

Figure 14

Open in figure viewerPowerPoint

(a) The data in the left rectangular window highlighted in Fig. 12a. (b) The result of robust *f-x*projection. (c) The result of *f-x* deconvolution. (d) The result of least-squares *f-x* projection.

Caption

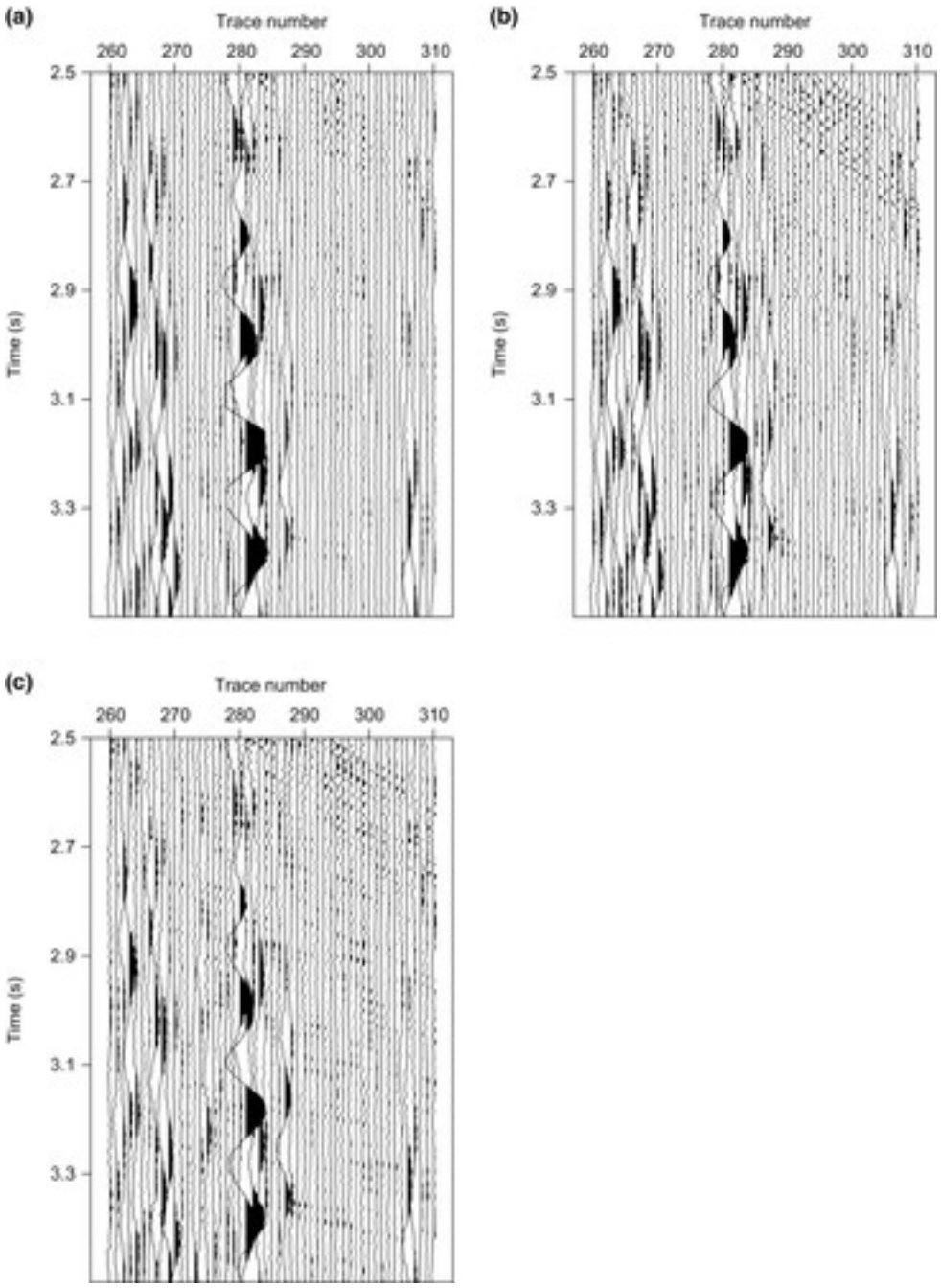

Figure 15

Open in figure viewerPowerPoint

The comparison of the difference sections of three different methods in the left rectangular window highlighted in Fig. 12a. Difference sections of (a) robust *f-x* projection, (b) *f-x*deconvolution, and (c) least-squares *f-x* projection.

Caption

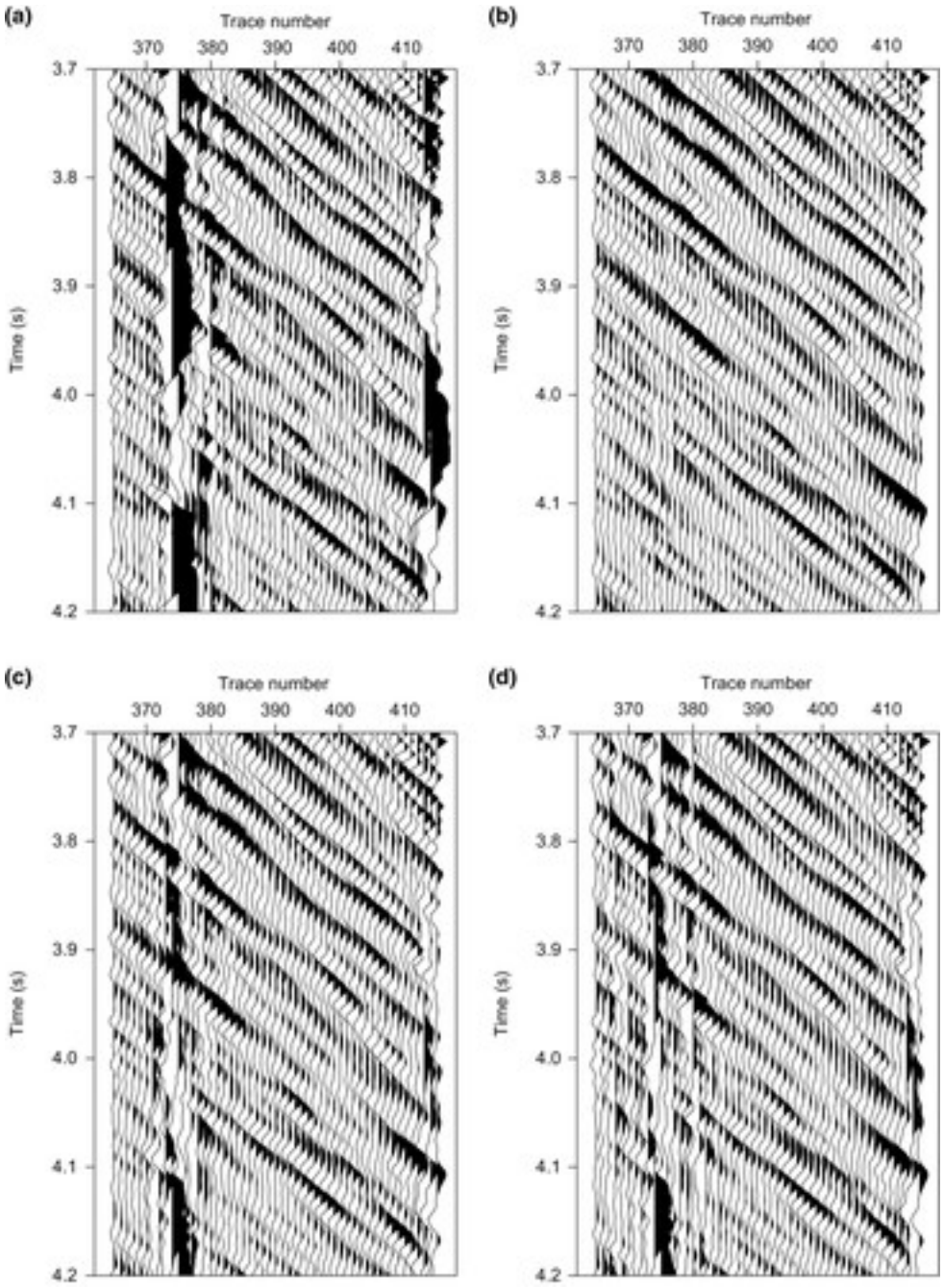

Figure 16

Open in figure viewerPowerPoint

(a) The data in the right rectangular window highlighted in Fig. 12a. (b) The result of robust *f-x*projection. (c) The result of *f-x* deconvolution. (d) The result of least-squares *f-x* projection.

Caption

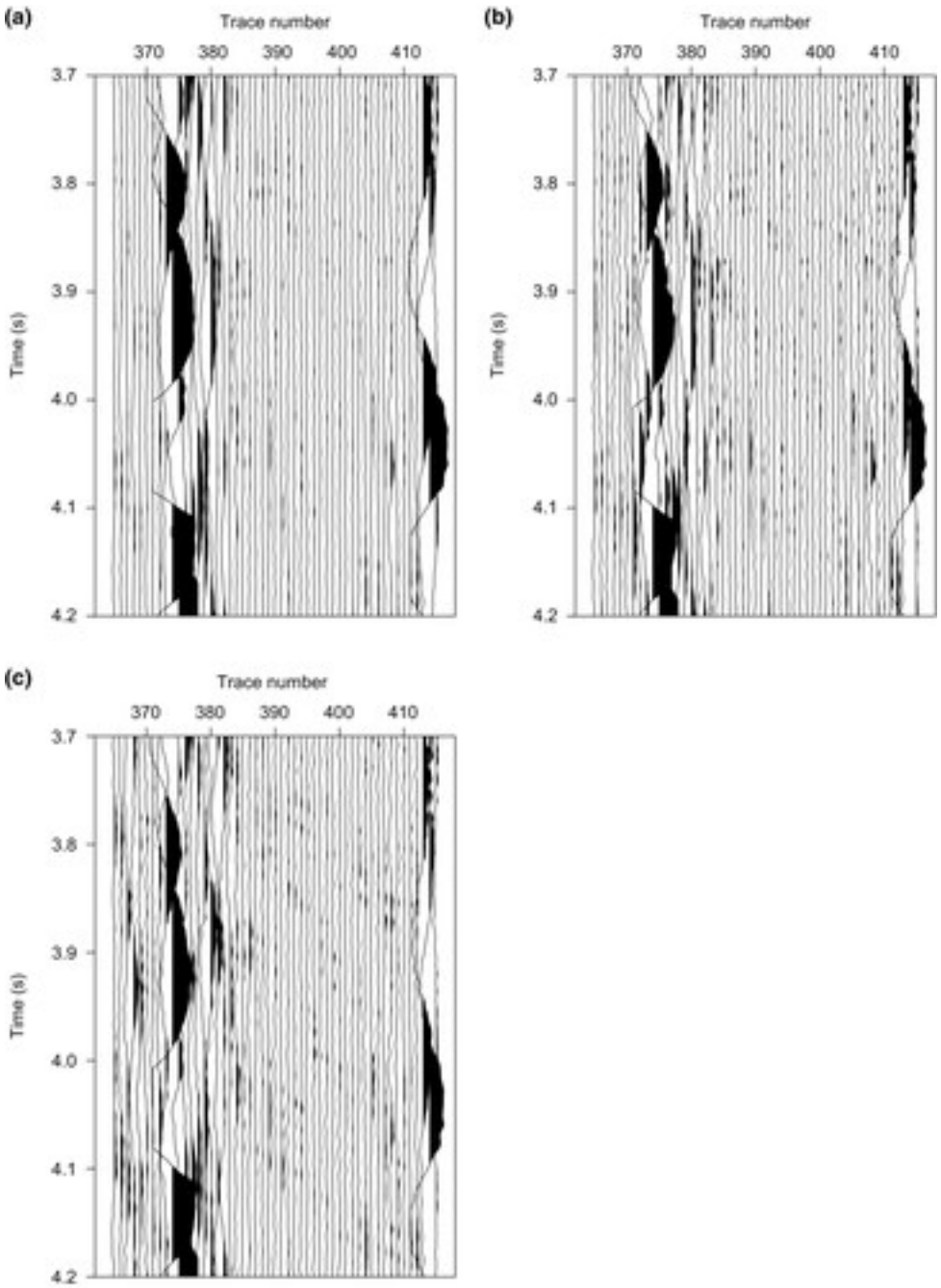

Figure 17

Open in figure viewerPowerPoint

The comparison of the difference sections of three different methods in the right rectangular window highlighted in Fig. 12a. Difference sections of (a) robust *f-x* projection, (b) *f-x*deconvolution, and (c) least-squares *f-x* projection.

Caption

DISCUSSION

The proposed method is based on robust inversion. There are two loops of iterations (one for alternating minimization and another for IRLS). Thus, the computation is more expensive than outlier detection-based methods. We derived the 2-D robust *f-x* projection filtering algorithm. It can be extended to a 3-D case using 2-D convolution (Chase 1992; Soubaras 2000; Naghizadeh and Sacchi 2010). For the current implementation, the algorithm uses the same parameter setting for different windows of the data. One potential future research is developing a data-adaptive parameter determination scheme that will give better tradeoff on signal-preserving and noise attenuation in different windows. It is known that *f-x* prediction and projection filters may produce small artificial events in the filtered results (Abma and Claerbout 1995; Soubaras 1994, 1995; Ozdemir *et al*. 1999; Soubaras 2000). This problem can be alleviated by designing filters in the time domain (Abma and Claerbout 1995; Liu, Liu, and Liu 2015). It is important to point out that the idea of our robust *f-x* projection filter can also be adapted for designing robust *t-x* prediction filters.

CONCLUSIONS

In this paper, we have proposed 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 two field data examples show that the proposed robust algorithm can remove erratic noise with a minimal degradation of the signal.

ACKNOWLEDGEMENTS

The authors would like to thank the sponsors of the Signal Analysis and Imaging Group (SAIG) at the Department of Physics, University of Alberta, for supporting this research. They would also like to thank CGG for providing the real marine dataset and for permission to publish. Special thanks go to Dr. Thomas Elboth and anonymous reviewers for their valuable suggestions and comments that improved the manuscript.

APPENDIX : PROOF OF EQUATION

The system of “nonlinear normal equations” (equation 19) is obtained by setting the derivative of the cost function $\mathcal{J}$ to zero

$$\frac{\partial \mathcal{J}}{\partial \mathbf{e}^*} = \frac{\partial}{\partial \mathbf{e}^*}\left[\frac{1}{2}||\mathbf{F}(\mathbf{y}-\mathbf{e})||_2^2 + \lambda \mathcal{H}(\mathbf{e})\right]$$

$$= \frac{1}{2}\mathbf{F}^H(\mathbf{F}\mathbf{e}-\mathbf{F}\mathbf{y}) + \lambda \frac{\partial \mathcal{H}(\mathbf{e})}{\partial \mathbf{e}^*}. \quad \text{(A-1)}$$

The derivative of $\mathcal{H}(\mathbf{e})$ with respect to $\mathbf{e}^*$ is equal to $(\partial \mathcal{H}(\mathbf{e})/\partial e_1^*, \ldots, \partial \mathcal{H}(\mathbf{e})/\partial e_N^*)^T$ with

$$\frac{\partial \mathcal{H}(\mathbf{e})}{\partial e_j^*} = \frac{\partial}{\partial e_j^*}\sum_{i-1}^{N}\left(\sqrt{\sigma^2 + |e_i|^2} - \sigma\right)$$

$$= \frac{1}{2}\frac{e_j}{\sqrt{\sigma^2 + |e_j|^2}}$$

$$= \frac{1}{2}W_{jj}e_j. \quad \text{(A-2)}$$

Rewrite equation A-2 in matrix form

$$\frac{\partial \mathcal{H}(\mathbf{e})}{\partial \mathbf{e}^*} = \frac{1}{2}\mathbf{W}\mathbf{e}, \quad \text{(A-3)}$$

where **W** is a diagonal weighting matrix. After substituting equation A-3 into equation A-1, we obtain the "nonlinear normal equations"

$$\left(\mathbf{F}^H\mathbf{F} + \lambda\mathbf{W}\right)\mathbf{e} = \mathbf{F}^H\mathbf{F}\mathbf{y}, \quad \text{(A-4)}$$

where the weighting matrix **W** depends on **e**.