Adaptive multiple subtraction using regularized nonstationary regression[a]

[a]Published in Geophysics, 74, no. 1, V25-V33, (2009)

Sergey Fomel

ABSTRACT

Stationary regression is the backbone of different seismic data processing algorithms including match filtering, which is commonly applied for adaptive multiple subtraction. However, the assumption of stationarity is not always adequate for describing seismic signals. I present a general method of nonstationary regression and show its application to nonstationary match filtering. The key idea is the use of shaping regularization for constraining the variability of nonstationary regression coefficients.

As shown by simple computational experiments, shaping regularization has clear advantages over conventional Tikhonov's regularization, incuding a more intuitive selection of parameters and a faster iterative convergence.

Using benchmark synthetic data examples, I demonstrate successful applications of this method to the problem of adaptive subtraction of multiple reflections.

INTRODUCTION

Many natural phenomena, including geological events and geophysical data, are fundamentally nonstationary. They may exhibit stationarity on the short scale but eventually change their behavior in space and time. Nonstationary adaptive filtering is a well-developed field in signal processing (Haykin, 2001). In seismic signal processing, nonstationary filters were analyzed by Margrave (1998) and applied to many important problems, including multiple suppression (Rickett et al., 2001), data interpolation (Crawley et al., 1999; Curry, 2003), migration deconvolution (Guitton, 2004; Valenciano et al., 2006).

In this paper, I present a general approach to designing nonstationary operators, including the case of nonstationary matching and prediction-error filters. The key idea is the application of shaping regularization (Fomel, 2007b) for constraining the continuity and smoothness of the filter coefficients. Regularization makes the estimation problem well-posed and leads to fast numerical algorithms. Advantages of shaping regularization in comparison with the classic Tikhonov regularization include an easier control of the regularization parameters and a faster iterative convergence resulting from the better conditioning of the inverted matrix.

Adaptive subtraction is a method for matching and removing coherent noise, such as multiple reflections, after their prediction by a data-driven technique (Verschuur et al., 1992). Adaptive subtraction involves a matching filter to compensate for the amplitude, phase, and frequency distortions in the predicted noise model. Different techniques for matching filtering and adaptive subtraction have been developed and discussed by a number of authors (Verschuur et al., 1992; Monk, 1993; Spitz, 1999; van Borselen et al., 2003; Wang, 2003; Guitton and Verschuur, 2004; Lu and Mao, 2005; Abma et al., 2005; Denisov et al., 2006). The regularized non-stationary regression technique, proposed in this paper, allows the matching filter to become smoothly nonstationary without the need to break the input data into local windows.

The paper is organized as follows. I start with an overview of stationary and nonstationary regression theory and introduce a method of regularized nonstationary regression. Next, I demonstrate this method using toy examples of line fitting and nonstationary deconvolution. Finally, I apply it to the adaptive multiple suppression problem and test its performance using a number of benchmark synthetic data examples.

STATIONARY AND NONSTATIONARY REGRESSION

Consider a "master" signal $m(\mathbf{x})$, where $\mathbf{x}$ represents the coordinates of a multidimensional space, and a collection of "slave" signals $s_k(\mathbf{x})$, $k = 1, 2, \ldots, N$. The goal of stationary regression is to estimate coefficients a_k, $k = 1, 2, \ldots, N$ such that the prediction error

$$e(\mathbf{x}) = m(\mathbf{x}) - \sum_{k=1}^{N} a_k \, s_k(\mathbf{x}) \tag{1}$$

is minimized in the least-squares sense. Particular examples include:

Line fitting

Let $\mathbf{x}$ be one-dimensional (denoted by x), $N = 2$, $s_1(x) = 1$, and $s_2(x) = x$. The problem of minimizing $e(x)$ amounts to fitting a straight line $a_1 + a_2\, x$ to the master signal.

Match filtering

If the slave signals are translates of the same signal $s(\mathbf{x})$, the regression problem corresponds to the problem of match filtering between $s(\mathbf{x})$ and $m(\mathbf{x})$. In the 1-D case, one can take, for example,

$$s_k(x) = s(x - k + N/2) \ ,$$

which turns the sum in equation 1 into a convolution with an unknown matching filter.

Prediction-error filtering

If the slave signals are causal translates of the master signal, the regression problem corresponds to the problem of autoregressive prediction-error filtering. In the 1-D case, one can take, for example,

$$s_k(x) = m(x-k) \ .$$

Nonstationary regression

Non-stationary regression uses a definition similar to equation 1 but allows the coefficients a_k to change with $\mathbf{x}$. The error turns into

$$e(\mathbf{x}) = m(\mathbf{x}) - \sum_{k=1}^{N} b_k(\mathbf{x})\, s_k(\mathbf{x}) \ , \tag{2}$$

and the problem of its minimization becomes ill-posed, because one can get more unknown variables than constraints. The remedy is to include additional constraints that limit the allowed variability of the estimated coefficients.

The classic regularization method is Tikhonov's regularization (Tikhonov, 1963; Engl et al., 1996), which amounts to minimization of the following functional:

$$F[\mathbf{b}] = \|e(\mathbf{x})\|^2 + \epsilon^2 \sum_{k=1}^{N} \|\mathbf{D}\left[b_k(\mathbf{x})\right]\|^2 \ , \tag{3}$$

where $\mathbf{D}$ is the regularization operator (such as the gradient or Laplacian filter) and ϵ is a scalar regularization parameter. If $\mathbf{D}$ is a linear operator, least-squares estimation reduces to linear inversion

$$\mathbf{b} = \mathbf{A}^{-1}\,\mathbf{d} \ , \tag{4}$$

where $\mathbf{b} = \left[\begin{array}{cccc} b_1(\mathbf{x}) & b_2(\mathbf{x}) & \cdots & b_N(\mathbf{x}) \end{array}\right]^T$,
$\mathbf{d} = \left[\begin{array}{cccc} s_1(\mathbf{x})\, m(\mathbf{x}) & s_2(\mathbf{x})\, m(\mathbf{x}) & \cdots & s_N(\mathbf{x})\, m(\mathbf{x}) \end{array}\right]^T$, and the elements of matrix $\mathbf{A}$ are

$$A_{ij}(\mathbf{x}) = s_i(\mathbf{x})\, s_j(\mathbf{x}) + \epsilon^2\, \delta_{ij}\, \mathbf{D}^T\, \mathbf{D} \ . \tag{5}$$

Shaping regularization (Fomel, 2007b) formulates the problem differently. Instead of specifying a penalty (roughening) operator $\mathbf{D}$, one specifies a shaping (smoothing) operator $\mathbf{S}$. The regularized inversion takes the form

$$\mathbf{b} = \widehat{\mathbf{A}}^{-1}\,\widehat{\mathbf{d}} \ , \tag{6}$$

where

$$\widehat{\mathbf{d}} = \left[\begin{array}{cccc} \mathbf{S}\left[s_1(\mathbf{x})\, m(\mathbf{x})\right] & \mathbf{S}\left[s_2(\mathbf{x})\, m(\mathbf{x})\right] & \cdots & \mathbf{S}\left[s_N(\mathbf{x})\, m(\mathbf{x})\right] \end{array}\right]^T \ ,$$

the elements of matrix $\widehat{\mathbf{A}}$ are

$$\widehat{A}_{ij}(\mathbf{x}) = \lambda^2 \, \delta_{ij} + \mathbf{S} \left[s_i(\mathbf{x}) \, s_j(\mathbf{x}) - \lambda^2 \, \delta_{ij} \right] \ , \tag{7}$$

and λ is a scaling coefficient.The main advantage of this approach is the relative ease of controlling the selection of λ and $\mathbf{S}$ in comparison with ϵ and $\mathbf{D}$. In all examples of this paper, I define $\mathbf{S}$ as Gaussian smoothing with an adjustable radius and choose λ to be the median value of $s_i(\mathbf{x})$. As demonstrated in the next section, matrix $\widehat{\mathbf{A}}$ is typically much better conditioned than matrix $\mathbf{A}$, which leads to fast inversion with iterative algorithms.

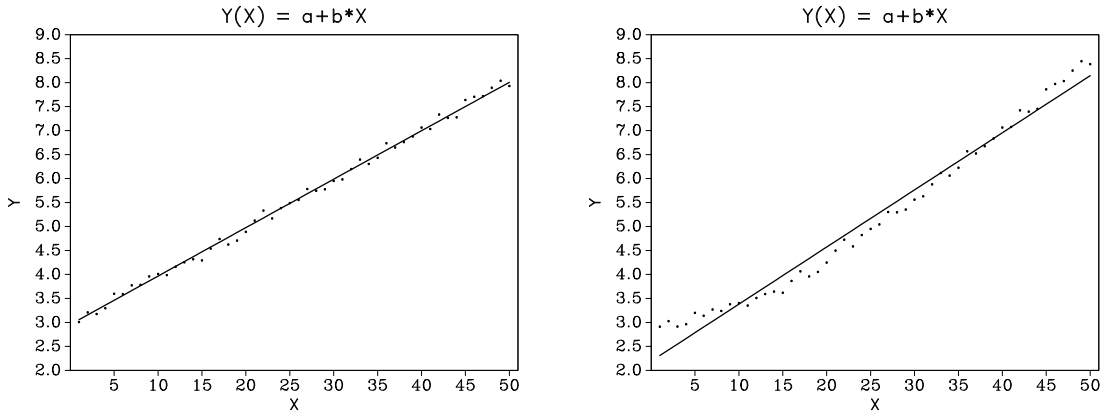
In the case of $N = 1$ (regularized division of two signals), a similar construction was applied before to define local seismic attributes (Fomel, 2007a).

TOY EXAMPLES

In this section, I illustrate the general method of regularized nonstationary regression using simple examples of nonstationary line fitting and nonstationary deconvolution.

Nonstationary line fitting

Figure 1a shows a classic example of linear regression applied as a line fitting problem. When the same technique is applied to data with a non-stationary behavior (Figure 1b), stationary regression fails to produce an accurate fit and creates regions of consistent overprediction and underprediction.



a b

Figure 1: Line fitting with stationary regression works well for stationary data (a) but poorly for non-stationary data (b).

One remedy is to extend the model by including nonlinear terms (Figure 2a), another is to break the data into local windows (Figure 2b). Both solutions work

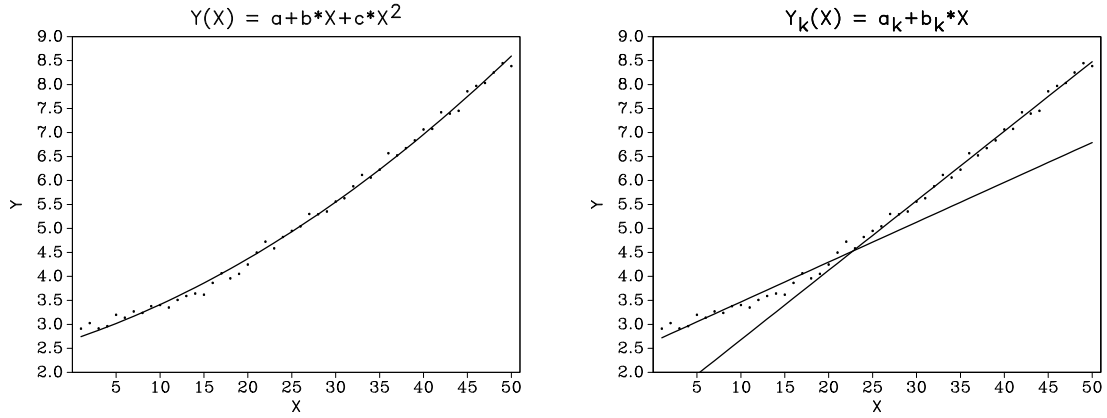


Figure 2: Nonstationary line fitting using nonlinear terms (a) and local windows (b).

to a certain extent but are not completely satisfactory, because they decreases the estimation stability and introduce additional non-intuitive parameters.

The regularized nonstationary solution, defined in the previous section, is shown in Figure 3. When using shaping regularization with smoothing as the shaping operator, the only additional parameter is the radius of the smoothing operator.

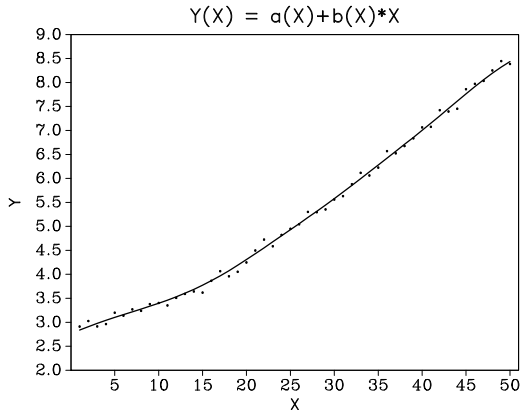


Figure 3: Nonstationary line fitting by regularized nonstationary regression.

This toy example makes it easy to compare shaping regularization with the more traditional Tikhonov's regularization. Figures 4 and 5 show inverted matrix $\mathbf{A}$ from equation 5 and the distribution of its eigenvalues for two different values of Tikhonov's regularization parameter ϵ, which correspond to mild and strong smoothing constraints. The operator $\mathbf{D}$ in this case is the first-order difference. Correspondingly, Figures 6 and 7 show matrix $\widehat{\mathbf{A}}$ from equation 7 and the distribution of its eigenvalues for mild and moderate smoothing implemented with shaping. The operator $\mathbf{S}$ is Gaussian smoothing controlled by the smoothing radius.

When a matrix operator is inverted by an iterative method such as conjugate

gradients, two characteristics control the number of iterations and therefore the cost of inversion (Golub and Van Loan, 1996; van der Vorst, 2003):

1. the condition number κ (the ratio between the largest and the smallest eigenvalue)
2. the clustering of eigenvalues.

Large condition numbers and poor clustering lead to slow convergence. Figures 4–7 demonstrate that both the condition number and the clustering of eigenvalues are significantly better in the case of shaping regularization than in the case of Tikhonov's regularization. In toy problems, this difference in behavior is not critical, because one can easily invert both matrices exactly. However, this difference becomes important in large-scale applications, where inversion is iterative and saving the number of iterations is crucial for performance.

As the smoothing radius increases, matrix $\widehat{\mathbf{A}}$ approaches the identity matrix, and the result of non-stationary regression regularized by shaping approaches the result of stationary regression. This intuitively pleasing behavior is difficult to emulate with Tikhonov's regularization.

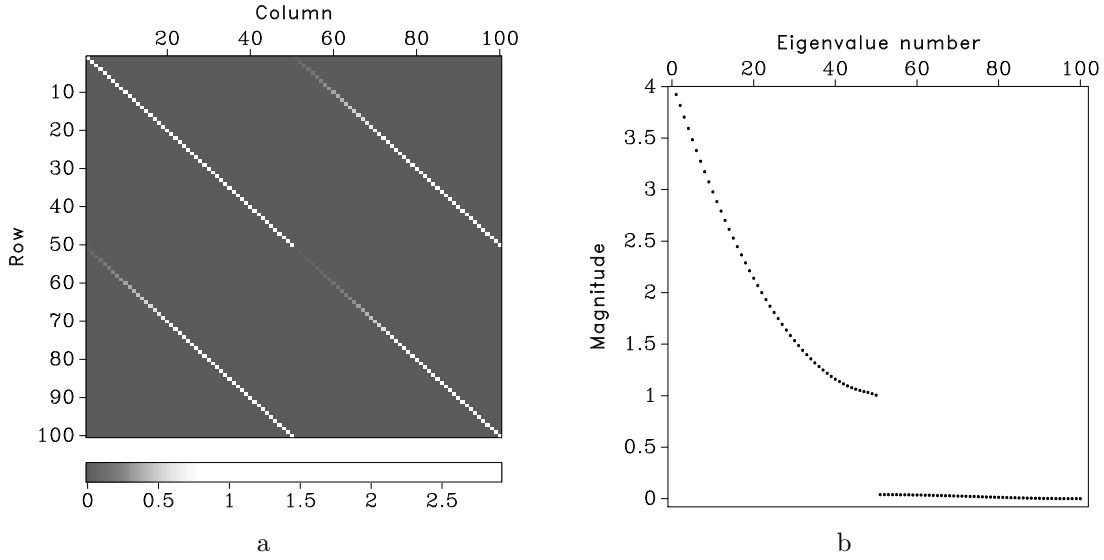


Figure 4: Matrix inverted in Tikhonov's regularization applied to nonstationary line fitting (a) and the distribution of its eigenvalues (b). The regularization parameter $\epsilon = 0.1$ corresponds to mild smoothing. The condition number is $\kappa \approx 888888$.

Nonstationary deconvolution

Figure 8 shows an application of regularized nonstationary regression to a benchmark deconvolution test from Claerbout (2008). The input signal is a synthetic trace that

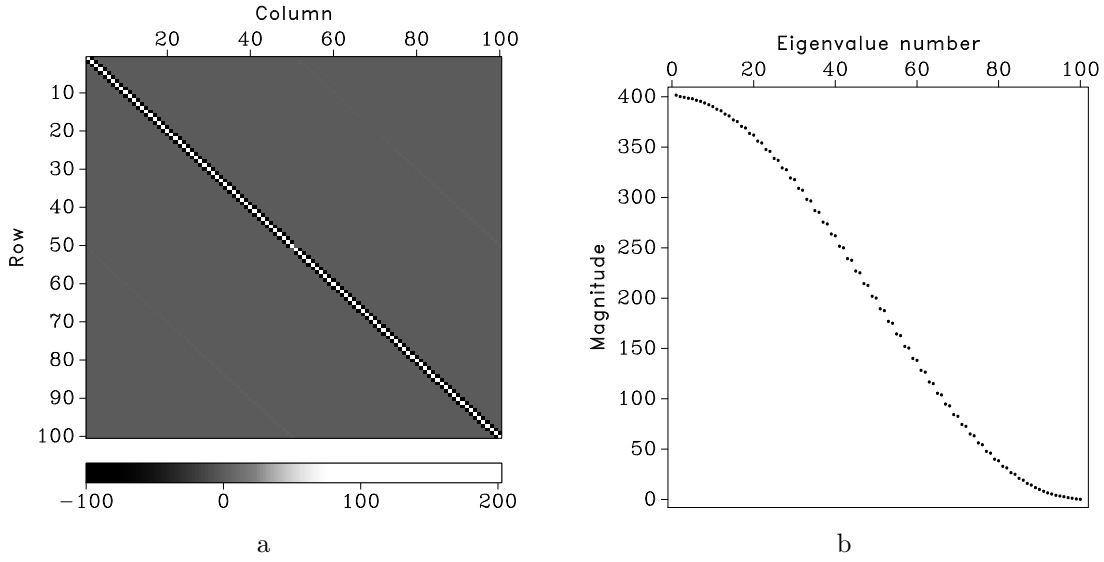


Figure 5: Matrix inverted in Tikhonov's regularization applied to nonstationary line fitting (a) and the distribution of its eigenvalues (b). The regularization parameter $\epsilon = 10$ corresponds to strong smoothing. The condition number is $\kappa \approx 14073$. Eigenvalues are poorly clustered.

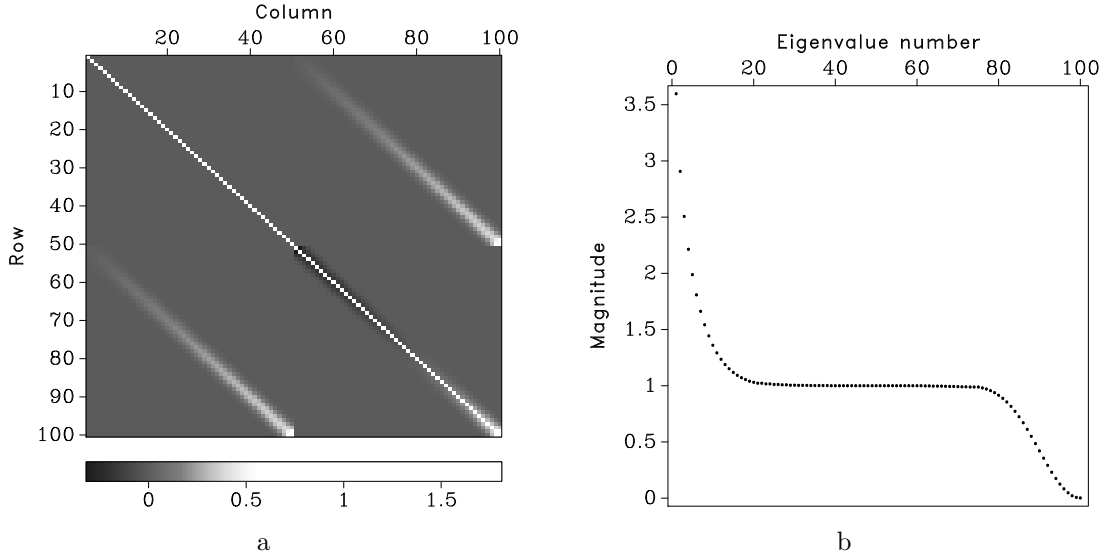


Figure 6: Matrix inverted in shaping regularization applied to nonstationary line fitting (a) and the distribution of its eigenvalues (b). The smoothing radius is 3 samples (mild smoothing). The condition number is $\kappa \approx 6055$.

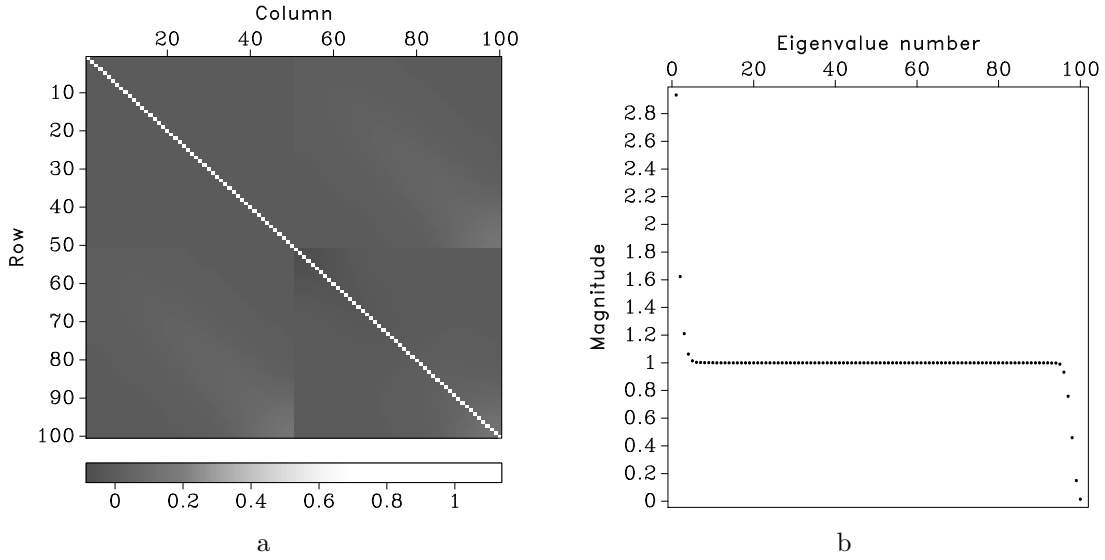


Figure 7: Matrix inverted in shaping regularization applied to nonstationary line fitting (a) and the distribution of its eigenvalues (b). The smoothing radius is 15 samples (moderate smoothing). The condition number is $\kappa \approx 206$. Eigenvalues are well clustered.

contains events with variable frequencies. A prediction-error filter is estimated by setting $N = 2$, $s_1(x) = m(x-1)$ and $s_2(x) = m(x-2)$. I use triangle smoothing with a 5-sample radius as the shaping operator $\mathbf{S}$. The deconvolved signal (bottom plot in Figure 8) shows the nonstationary reverberations correctly deconvolved.

A three-point prediction-error filter $\{1, a_1, a_2\}$ can predict an attenuating sinusoidal signal

$$m(x) = \rho^x \, cos(\omega \, x) \; , \tag{8}$$

where ω is frequency, and ρ is the exponent factor, provided that the filter coefficients are defined as

$$a_1 = -2 \, \rho \, \cos(\omega) \; ; \tag{9}$$

$$a_2 = \rho^2 \; . \tag{10}$$

This fact follows from the simple identity

$$\begin{aligned} m(x) + a_1 \, m(x-1) + a_2 \, m(x-2) &= \\ \rho^x \, [\cos(\omega \, x) - 2 \, \cos(\omega) \, \cos(\omega \, x - \omega) + \cos(\omega \, x - 2 \, \omega)] &= 0 \; . \end{aligned} \tag{11}$$

According to equations 9–10, one can get an estimate of the local frequency $\omega(x)$ from the non-stationary coefficients $b_1(x) = -a_1(x)$ and $b_2(x) = -a_2(x)$ as follows:

$$\omega(x) = \arccos \left(\frac{b_1(x)}{2 \sqrt{-b_2(x)}} \right) \; . \tag{12}$$

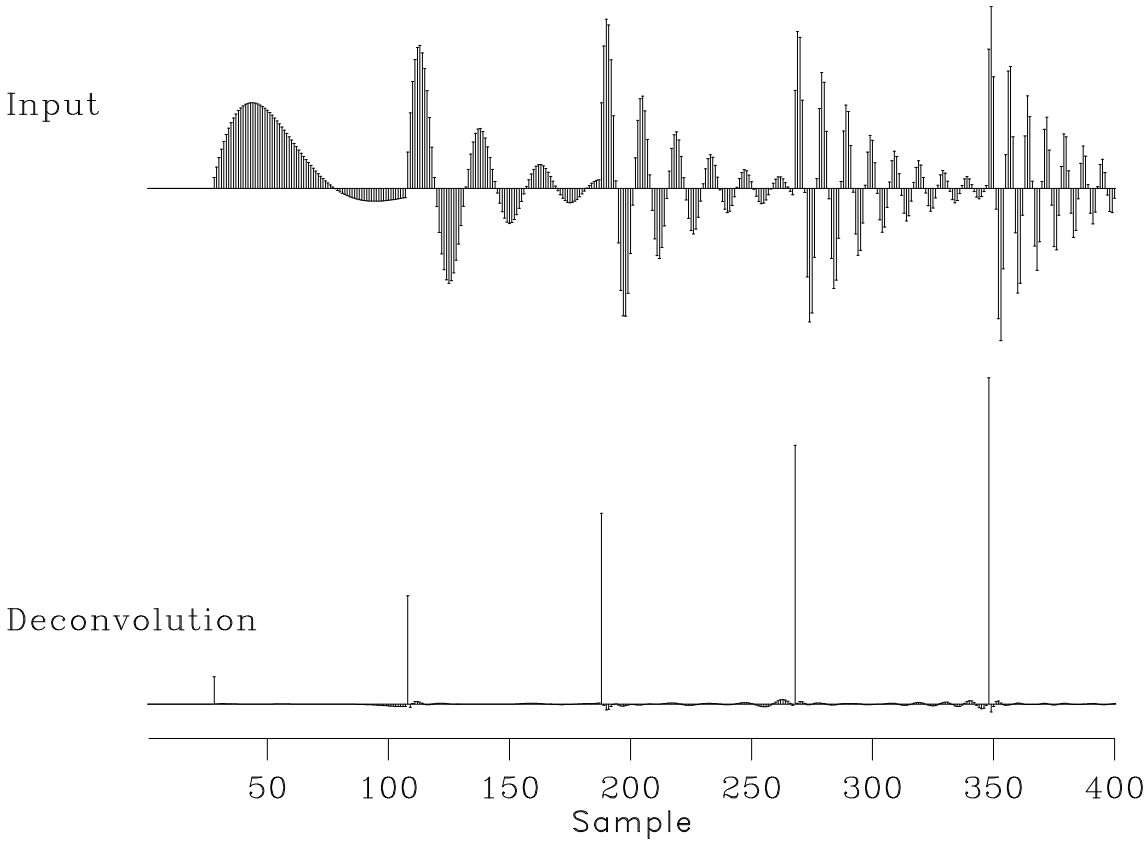


Figure 8: Benchmark test of nonstationary deconvolution from Claerbout (2008). Top: input signal, bottom: deconvolved signal using nonstationary regularized regression.

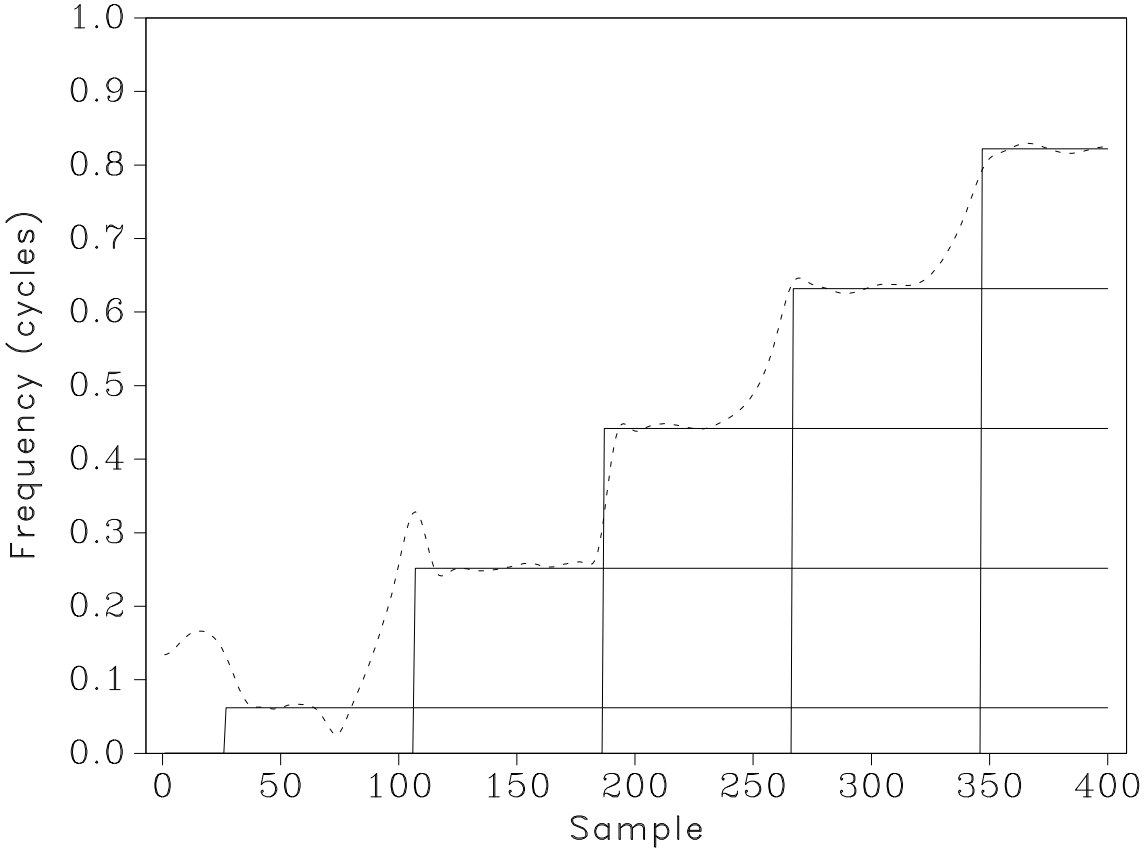


Figure 9: Frequency of different components in the input synthetic signal from Figure 8 (solid line) and the local frequency estimate from non-stationary deconvolution (dashed line).

Figure 9 shows frequency of the different components in the input non-stationary signal from Figure 8 and the local frequency estimate using equation 12. A reasonably accurate match between the true nonstationary frequencies and their estimate from nonstationary regression can be observed. The local frequency attribute (Fomel, 2007a) provides a different approach to the same problem.

ADAPTIVE MULTIPLE SUBTRACTION

In this section, I apply regularized nonstationary regression to the problem of adaptive subtraction and evaluate its performance using benchmark synthetic tests.

One possible variation of non-stationary regression applied to adaptive subtraction is an introduction of the signal prediction filter (Spitz, 1999). The signal prediction applies to the residual error $e(\mathbf{x})$ in equation 2, thus modifying the objective function (Guitton, 2005).

Abma test

Abma et al. (2005) present a comparison of different adaptive subtraction algorithms used for multiple suppression. I use Abma's benchmark examples to illustrate an application of nonstationary regression to the adaptive subtraction problem. Figure 10 shows the first test: the input data (Figure 10a) contain a horizontal "signal" event and dipping "noise" events. We are provided with a model of the noise (Figure 10b). However, the model events are slightly shifted and have a different amplitude behavior. This test simulates a typical situation in adaptive surface-related multiple elimination, where there are phase and amplitude distortions in the multiple model, caused by an imperfect prediction. To handle the variability in noise amplitudes, I design a non-stationary matching filter with 13 coefficients and 3-sample smoothing radius and apply it to match the noise model to the data. Subtracting matched noise (Figure 10d) produces the desired and nearly perfect signal estimate (Figure 10c). The variability of filter coefficients is illustrated in Figure 11, which displays the zero-lag coefficient and the mean coefficient of the non-stationary matching filter.

Another benchmark from Abma et al. (2005) is shown in Figure 12. This time, the noise part of the data is a curved event which has a predictive model (Figure 12b) with some amplitude and phase differences. Non-stationary regularized regression correctly predicts the noise signal (Figure 12d) using match filtering and produces an accurate signal estimate (Figure 12c). The variability of non-stationary filter coefficients is shown in Figure 13.

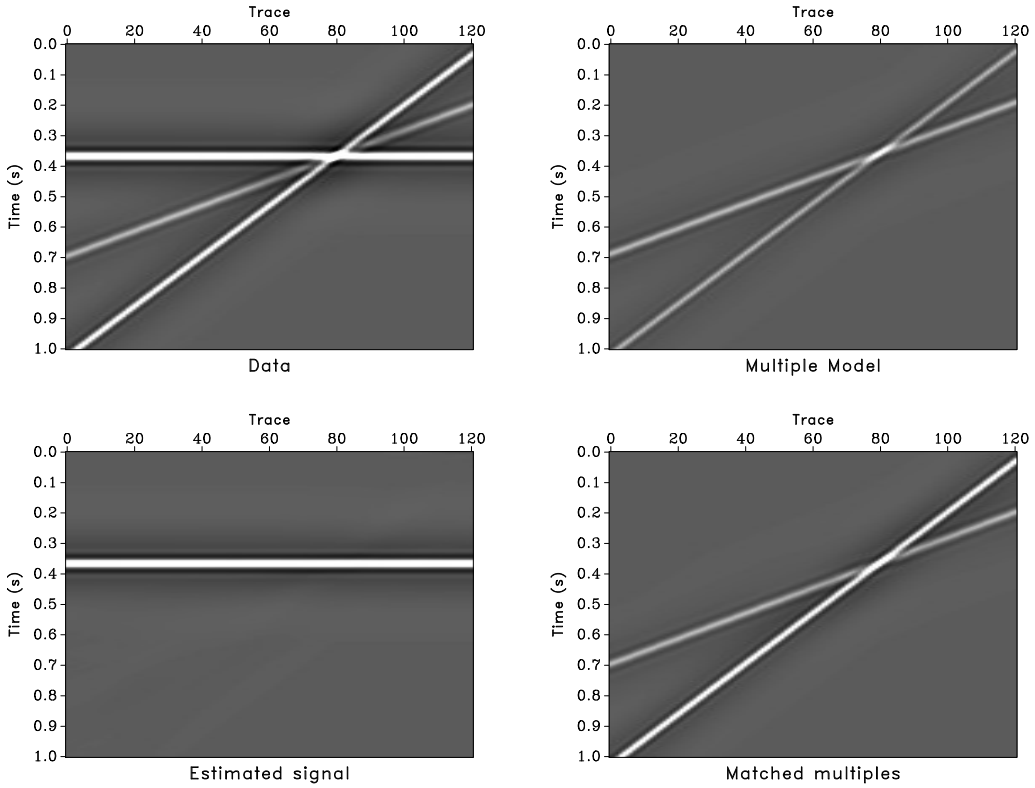


Figure 10: Benchmark test on adaptive subtraction from Abma et al. (2005). a – Input synthetic data. b – Model of the noise containing amplitude and phase differences with respect to the corresponding part of the data. c – Extracted signal. d – Estimated noise.

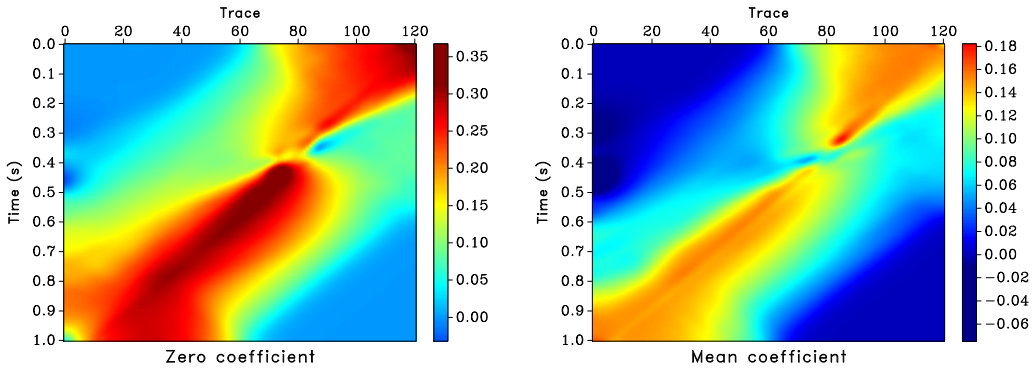


Figure 11: Variability of non-stationary match filter coefficients for the example shown in Figure 10. a – Zero-lag coefficient. b – Mean coefficient.

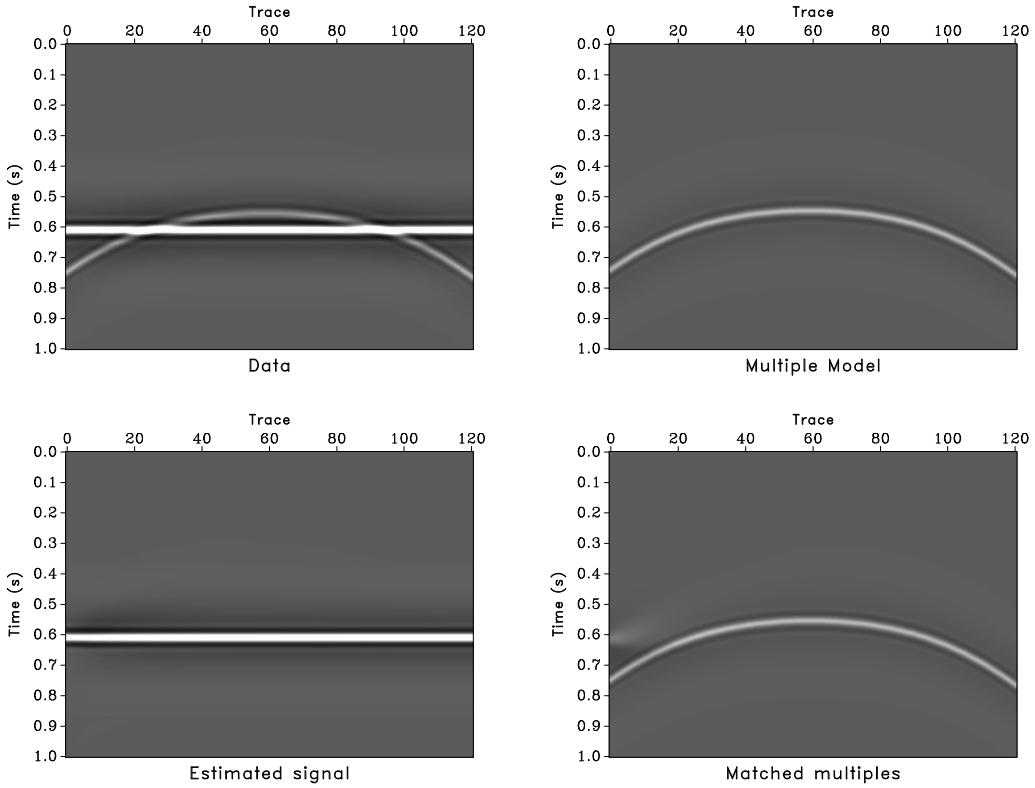


Figure 12: Benchmark test on adaptive subtraction from Abma et al. (2005). a – Input synthetic data. b – Model of the noise containing amplitude and phase differences with respect to the corresponding part of the data. c – Extracted signal. d – Estimated noise.

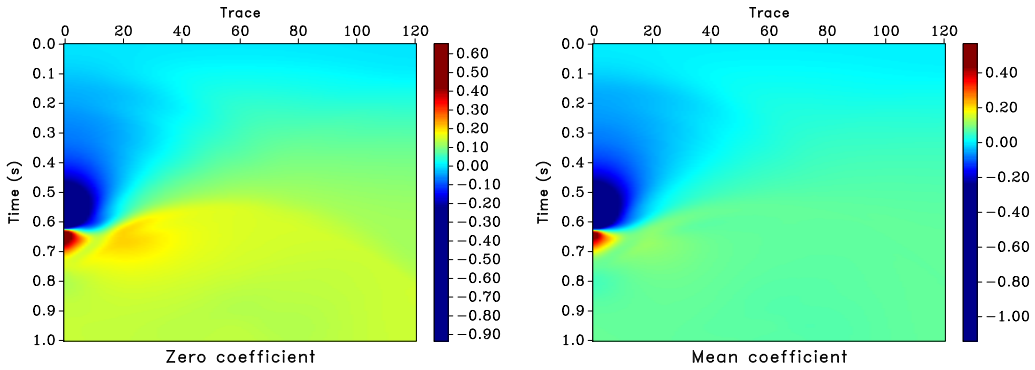


Figure 13: Variability of non-stationary match filter coefficients for the example shown in Figure 12. a – Zero-lag coefficient. b – Mean coefficient.

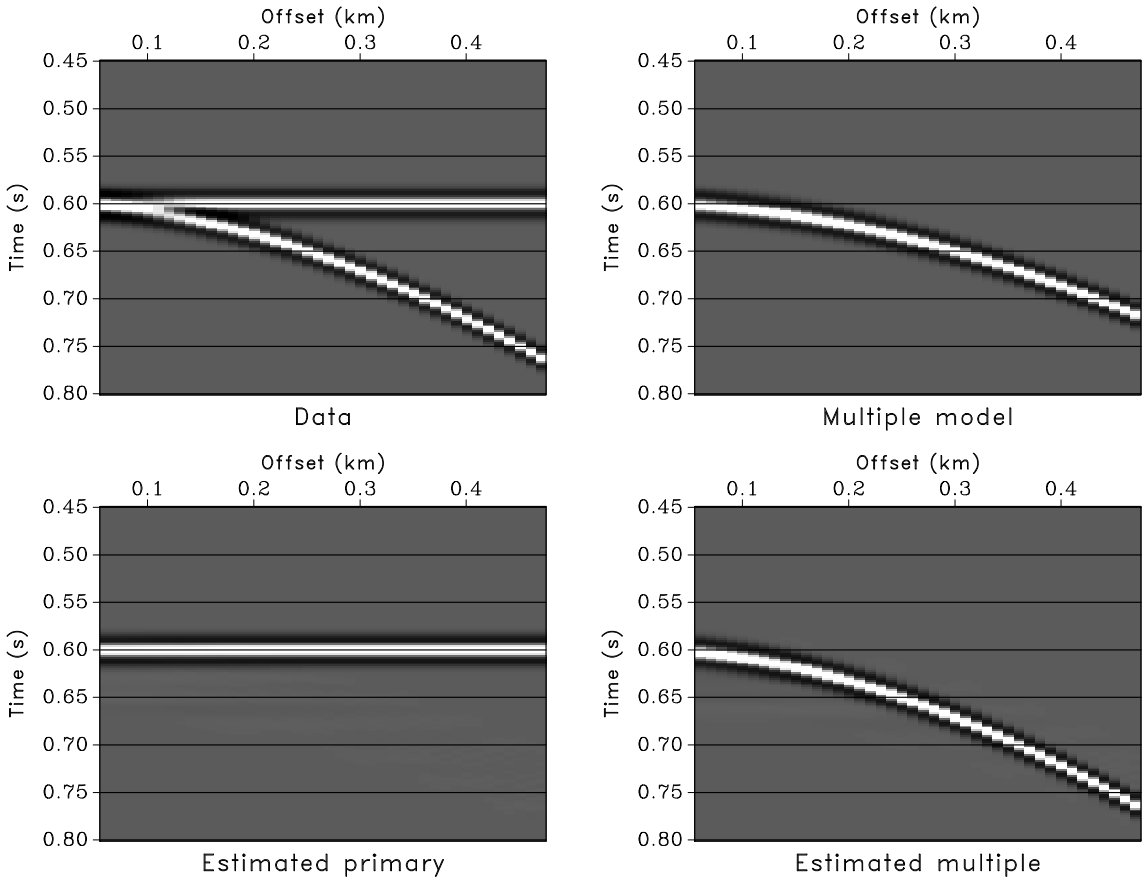


Figure 14: Benchmark test on multiple attenuation from Spitz (2007). a – Input synthetic data containing a signal (primary reflection) and noise (multiple reflection) events. b – Model of the noise containing phase differences with respect to the corresponding part of the data. c – Extracted signal. d – Estimated noise.

Spitz test

The next benchmark example (Figure 14) reproduces a test created by Spitz (2007), which simulates a seismic gather after migration or normal moveout correction. The signal (primary reflection) event is horizontal while the noise (multiple reflection) event exhibits curvature and overlaps with the signal at small offsets (Figure 14a). Multiple prediction (Figure 14b) contains a curved event but with incorrect curvature. As in the previous examples, non-stationary regularized regression correctly predicts the noise signal (Figure 14d) using match filtering and produces an accurate signal estimate (Figure 14c).

Pluto test

Finally, Figure 15 shows an application of the nonstationary matching technique to the Pluto synthetic dataset, a well-known benchmark for adaptive multiple subtraction. Matching and subtracting an imperfect model of the multiples created by the surface-related multiple elimination approach of Verschuur et al. (1992) leaves a clean estimate of primary reflections. Figure 16 shows a comparison between the multiple model obtained by surface-related prediction and the multiple model generated by nonstationary matching. The matching filter non-stationarity is depicted in Figure 17, which shows the variability of filter coefficients with time and space.

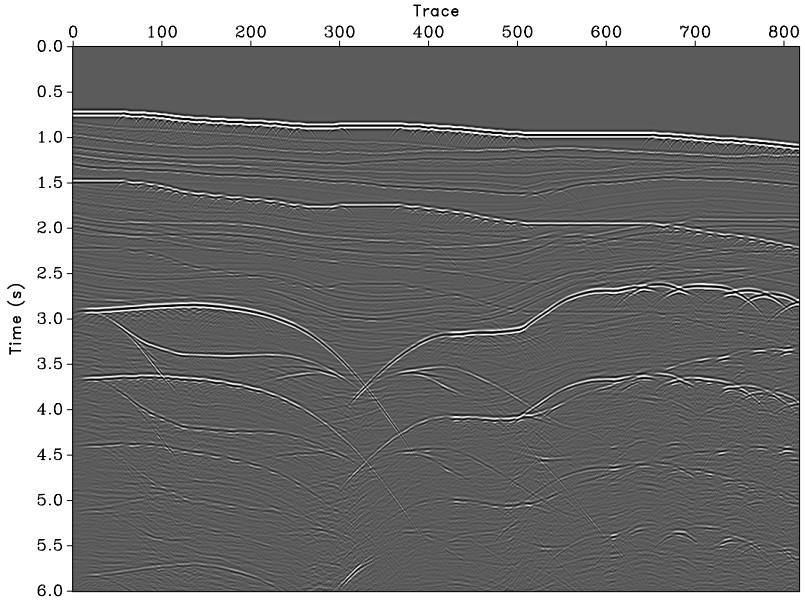
CONCLUSIONS

I have presented a general method for regularized nonstationary regression. The key idea is the application of shaping regularization for constraining the variability of nonstationary coefficients. Shaping regularization has clear advantages over conventional Tikhonov's regularization: a more intuitive selection of regularization parameters and a faster iterative convergence because of better conditioning and eigenvalue clustering of the inverted matrix.

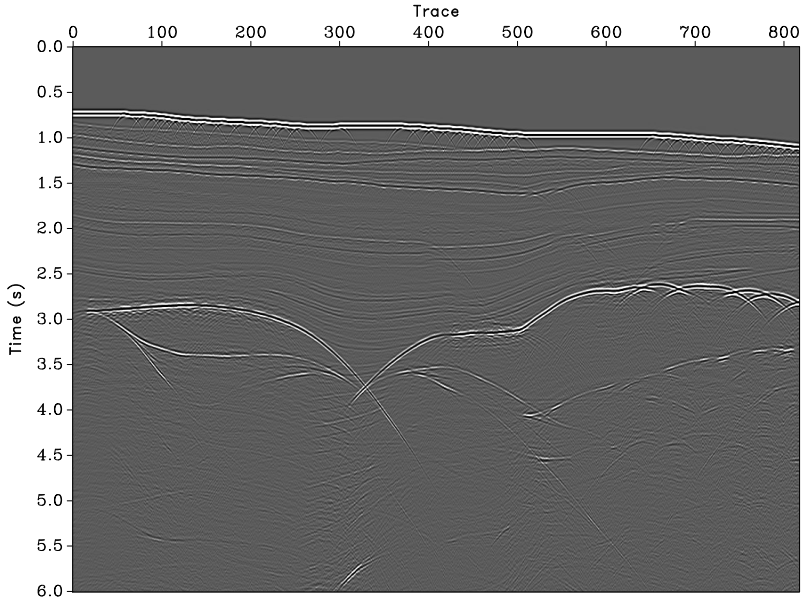
I have shown examples of applying regularized regression to benchmark tests of adaptive multiple subtraction. When the signal characteristics change with time or space, regularized nonstationary regression provides an effective description of the signal. This approach does not require breaking the input data into local windows, although it is analogous conceptually to sliding (maximum overlap) spatial-temporal windows.

ACKNOWLEDGMENTS

I thank StatoilHydro for a partial financial support of this work. I would like also to thank Dave Hale, Long Jin, Dmitry Lokshtanov, and Gary Sitton for inspiring

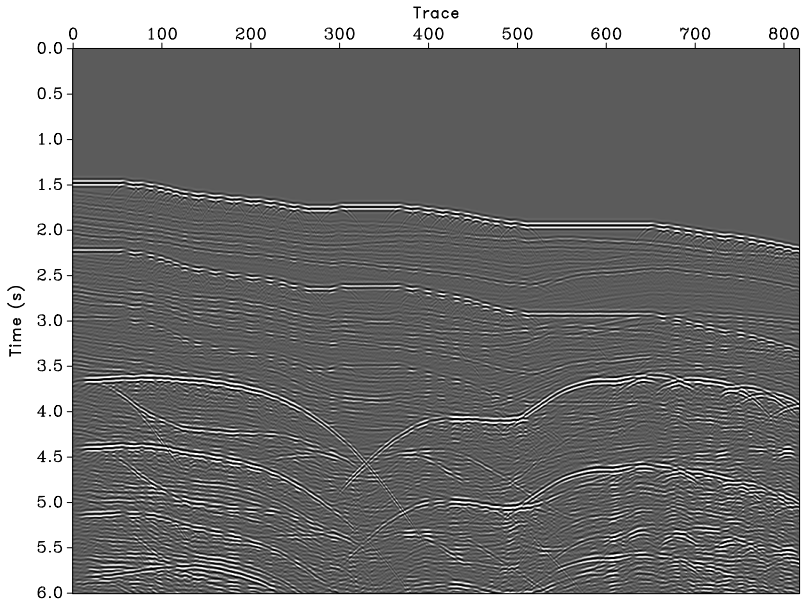


a

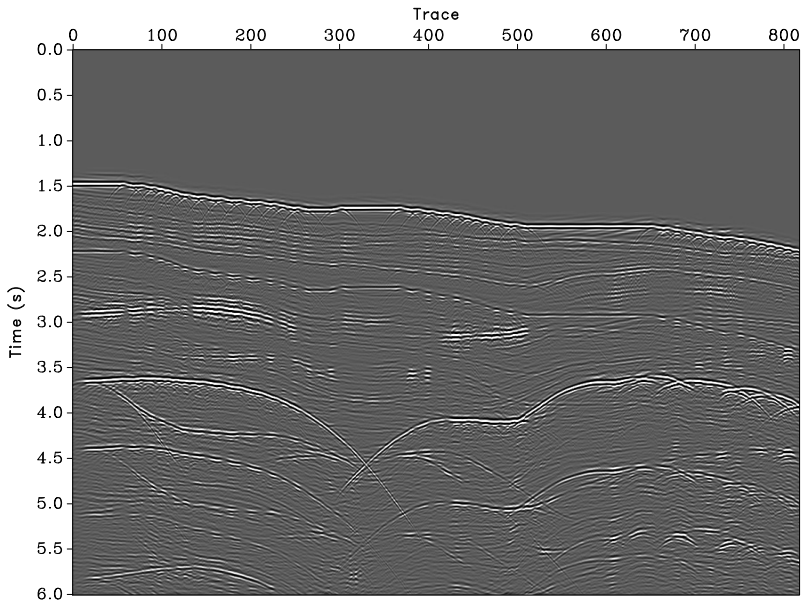


b

Figure 15: Adaptive multiple subtraction in the Pluto synthetic dataset. (a) Input data. (b) Extracted Signal. Surface-related multiples are successfully subtracted.

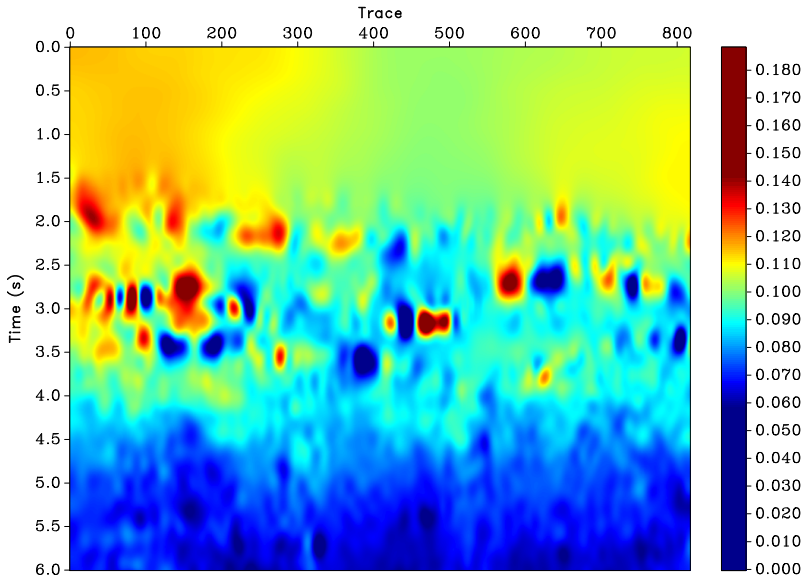


a

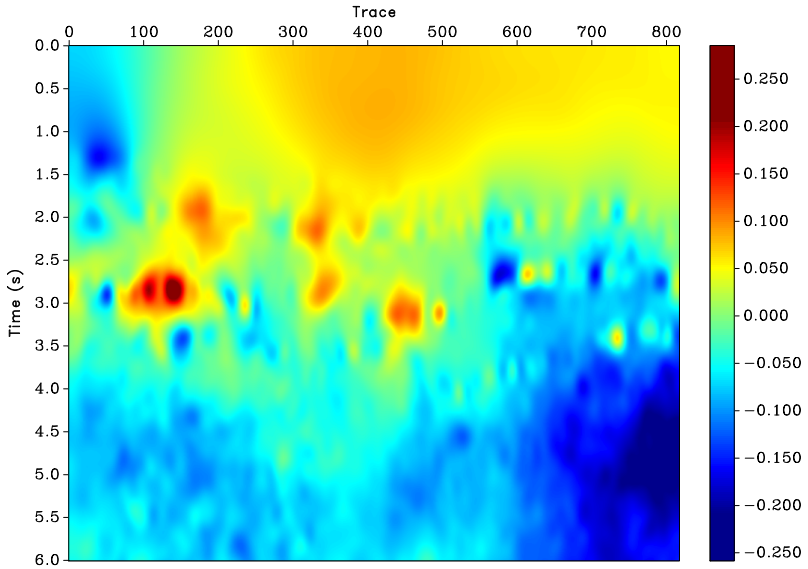


b

Figure 16: Multiple model from surface-related prediction (a) and estimated multiples (b) for the Pluto synthetic dataset.



a



b

Figure 17: Variability of non-stationary match filter coefficients for the Pluto test. (a) Zero-lag coefficient. (b) Mean coefficient.

discussions and Ray Abma for providing benchmark datasets for adaptive subtraction. The Pluto synthetic was generated by the SMAART consortium.

This publication is authorized by the Director, Bureau of Economic Geology, The University of Texas at Austin.

REFERENCES

Abma, R., N. Kabir, K. H. Matson, S. Michell, S. Shaw, and B. McLain, 2005, Comparisons of adaptive subtraction methods for multiple attenuation: The Leading Edge, **24**, 277–280.

Claerbout, J., 2008, Image estimation by example: Environmental soundings image enhancement: Stanford Exploration Project.

Crawley, S., J. Claerbout, and R. Clapp, 1999, Interpolation with smoothly nonstationary prediction-error filters: 69th Ann. Internat. Mtg, Soc. of Expl. Geophys., 1154–1157.

Curry, W., 2003, Interpolation of irregularly sampled data with nonstationary, multiscale prediction-error filters: 73rd Ann. Internat. Mtg., Soc. of Expl. Geophys., 1913–1916.

Denisov, M. S., M. A. Polubojarinov, and D. B. Finikov, 2006, Robust methods of adaptive seismic multiple suppression: EAGE St. Petersburg, Expanded Abstracts, European Association of Geoscientists & Engineers, B033.

Engl, H., M. Hanke, and A. Neubauer, 1996, Regularization of inverse problems: Kluwer Academic Publishers.

Fomel, S., 2007a, Local seismic attributes: Geophysics, **72**, A29–A33.

———, 2007b, Shaping regularization in geophysical-estimation problems: Geophysics, **72**, R29–R36.

Golub, G. H., and C. F. Van Loan, 1996, Matrix computations: The John Hopkins University Press.

Guitton, A., 2004, Amplitude and kinematic corrections of migrated images for nonunitary imaging operators: Geophysics, **69**, 1017–1024.

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

Guitton, A., and D. Verschuur, 2004, Adaptive subtraction of multiples using the L1-norm: Geophys. Prosp., **52**, 27–38.

Haykin, S., 2001, Adaptive filter theory: Prentice Hall.

Lu, W., and F. Mao, 2005, Adaptive multiple subtraction using independent component analysis: The Leading Edge, **24**, 282–284.

Margrave, G. F., 1998, Theory of nonstationary linear filtering in the Fourier domain with application to time-variant filtering: Geophysics, **63**, 244–259.

Monk, D. J., 1993, Wave-equation multiple suppression using constrained gross-equalization: Geophys. Prosp., **41**, 725–736.

Rickett, J., A. Guitton, and D. Gratwick, 2001, Adaptive Multiple Subtraction with Non-Stationary Helical Shaping Filters: 63rd Mtg., Eur. Assn. Geosci. Eng., Session: P167.

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

———, 2007, Perturbed prediction-driven multiple elimination: A PEF scheme: 77th Ann. Internat. Mtg, Soc. of Expl. Geophys., 2649–2653.

Tikhonov, A. N., 1963, Solution of incorrectly formulated problems and the regularization method: Sovet Math. Dokl., 1035–1038.

Valenciano, A. A., B. Biondi, and A. Guitton, 2006, Target-oriented wave-equation inversion: Geophysics, **71**, A35–A38.

van Borselen, R., G. Fookes, and J. Brittan, 2003, Target-oriented adaptive subtraction in data-driven multiple removal: The Leading Edge, **22**, 340–343.

van der Vorst, H. A., 2003, Iterative Krylov methods for large linear systems: Cambridge University Press.

Verschuur, D. J., A. J. Berkhout, and C. P. A. Wapenaar, 1992, Adaptive surface-related multiple elimination: Geophysics, **57**, 1166–1177.

Wang, Y., 2003, Multiple subtraction using an expanded multichannel matching filter: Geophysics, **68**, 346–354.