Coherent noise attenuation using inverse problems and prediction–error filters

Antoine Guitton[1]

Introduction

In seismic, we can express many of the processing steps as linear operators. These operators perform a mapping of one domain, usually a model of the earth parameterized in terms of velocity, reflectivity, into another domain, usually seismic data sorted into CMP or shot gathers. This mapping is called modelling because it models the seismic data. Usually we desire the opposite of modelling, i.e. given some data, we want to retrieve the model. In many cases the adjoint of the modelling operator is used to estimate the model. For some operators, like the Fourier transform, the adjoint is the exact inverse; for others, the vast majority, the adjoint is not the true inverse but rather an approximation of the inverse.

Nowadays, amplitude-preserving processing is a mandatory task for true-amplitude migration, AVO analysis or 4D interpretation; extracting the modelling part with approximate inverses is then risky. Inversion theory provides us with methods to compute a 'good' inverse that will honour the seismic data. Pioneering work by Tarantola (1987) has shown the usefulness of inversion for earthquake location and tomography. Since then inversion has been at the heart of many seismic processing breakthroughs, such as least-squares migration (Nemeth 1996), high-resolution radon transforms (Thorson & Claerbout 1985; Sacchi & Ulrych 1995) or projection filtering (Soubaras 1994; Abma & Claerbout 1995). A very popular method of inversion is the least-squares approach, which can be related to a Bayesian estimation of the model parameters.

It is well understood that the inversion in a least-squares sense is very sensitive to the noise level present in the data. By noise, I mean abnormally large or small data components, or outliers which are better described by long-tailed probability density functions (PDFs) as opposed to short-tailed Gaussian PDFs, and coherent noise that the seismic operator is unable to model. The noise will spoil any analysis based on the result of the inversion and affect the amplitude recovery of the input data. From a more statistical point of view, if the residual, which measures the quality of the data fitting, is corrupted by outliers or coherent noise in the data, it will not have independent and identically distributed (IID) components. A more 'geophysical way' of saying this is that the residual will not have a white spectrum.

In this paper I show how the residual can be whitened when coherent noise is present in the data. Outliers and noise-burst problems are not addressed here. They can be winnowed out by applying, iteratively, a locally re-weighted regression (Wang, White & Pratt 2000). In the first section I review some basics of inverse theory. Then in the following section I introduce two inversion methods that yield white residuals. The first method proposes approximating the noise covariance operators with prediction–error filters (PEFs). The second method handles the coherent noise by introducing a noise modelling operator within the inversion. These methods are tested with field data.

A short review of inverse problems

In this section I review some basic notions on inversion. The least-squares criterion comes directly from the hypothesis that the PDF of each observable type of data and each model parameter is Gaussian. These assumptions lead to the general discrete inverse problem (Tarantola 1987). Finding $\mathbf{m}$ is then equivalent to minimizing the quadratic function (or cost/objective function),

$$f(\mathbf{m}) = (\mathbf{Hm} - \mathbf{d})^T\mathbf{C}_d^{-1}(\mathbf{Hm} - \mathbf{d}) + \mu(\mathbf{m} - \mathbf{m}_{prior})^T\mathbf{C}_m^{-1}(\mathbf{m} - \mathbf{m}_{prior}), \quad (1)$$

where the index T represents the (Hermitian) transpose, $\mathbf{m}$ is a mapping of the data (the unknown of the inverse problem), $\mathbf{H}$ is a seismic operator, $\mathbf{d}$ is the seismic data, $\mathbf{C}_d$ and $\mathbf{C}_m$ are the data and model covariance operators, respectively, $\mathbf{m}_{prior}$ is a model given *a priori*, and μ is a trade-off parameter (Wang & Pratt 1997) between the amount of data fitting ($\mathbf{Hm} - \mathbf{d}$) and model perturbation ($\mathbf{m} - \mathbf{m}_{prior}$). Thus we can write eqn (1) in the following intuitive form:

$$f(\mathbf{m}) = (\text{data residual})^2 + \mu(\text{model pertubation})^2. \quad (2)$$

The covariance matrix $\mathbf{C}_d$ combines experimental errors and modelling uncertainties. Modelling uncertainties describe the difference between what the operator can predict and what is contained in the data. Thus the covariance matrix $\mathbf{C}_d$ is often called the noise covariance matrix (Sacchi & Ulrych 1995). Often we assume that (i) the variances of the model and of the

[1]*Correspondence: E-mail*: antoine@sep.stanford.edu

noise are uniform, (ii) the covariance matrices are diagonal, i.e. the model and data components are uncorrelated, and (iii) no prior model is known in advance. Given these approximations the objective function becomes

$$f(\mathbf{m}) = (\mathbf{Hm} - \mathbf{d})^T(\mathbf{Hm} - \mathbf{d}) + \varepsilon^2\mathbf{m}^T\mathbf{m}, \quad (3)$$

where $\varepsilon = \mu\sigma_m^2/\sigma_d^2$ is a function of the trade-off parameter μ and of the noise and model variances. The model perturbation (squared) reduces to a damping of the cost function. This damping is used to compensate for numerical instabilities when the parameters **m** are poorly constrained. Following eqn (2) we now have

$$f(\mathbf{m}) = (\text{data residual})^2 + \varepsilon^2(\text{damping})^2 \quad (4)$$

The prior assumptions leading to eqn (3) are too strong when we are dealing with seismic data because the variance of the noise/model may be not uniform and the components of the noise/model may not independent. For simplicity I rewrite the objective function in eqn (3) in terms of 'fitting goals' for **m**,

$$0 \approx \mathbf{Hm} - \mathbf{d} \quad (5)$$

$$0 \approx \varepsilon\mathbf{m}, \quad (6)$$

and estimate **m** accordingly in a least-squares sense. The first equality stresses the need for **Hm** to fit the input data **d**. The second equality is often called the regularization or 'model styling' term. When the assumptions leading to eqn (4) are respected, the convergence towards **m** is easy to achieve. In particular the components of the residual $\mathbf{r} = \mathbf{Hm} - \mathbf{d}$ are IID. This IID property implies that no coherent information remains in the residual and that each variable of the residual has similar intensity (or power). Coherent noise in the data violates the assumptions of both the uniform distribution and the need for independent noise components. The next section shows how IID residual components can be derived in practice with coherent noise-infested seismic data.

Proposed solutions to attenuate coherent noise

Any data set **d** may be regarded as the sum of signal **s** and noise **n** as follows:

$$\mathbf{d} = \mathbf{s} + \mathbf{n}. \quad (7)$$

I assume that the coherent noise **n** consists of the inconsistent part (or modelling uncertainties part) of the data **d** for any given operator **H**. Described below are two techniques designed to handle the coherent noise effects during the inversion.

A filtering method

Equation (1) introduces two matrices that, in general, are not calculated: the noise covariance matrix $\mathbf{C}_d$ and the model covariance matrix $\mathbf{C}_m$. When calculated, these matrices are usually approximated with diagonal operators. In this section I describe a method that computes non-diagonal covariance matrices using prediction error filters.

I concentrate my efforts on the noise covariance matrix only, the computation of the model covariance matrix being beyond the scope of this paper. When coherent noise is present in the data, residual variables are no longer IID and the covariance matrices should not be approximated with diagonal operators. The coherent noise will add colour to the spectrum of the residual. The goal of the covariance matrices is then to absorb this spectrum. Now, as Claerbout and Fomel assert:

> 'Clearly, the noise spectrum is the same as the data covariance only if we accept the theoretician's definition that $E(\mathbf{d}) = \mathbf{Fm}$. There is no ambiguity and no argument if we drop the word "variance" and use the word 'spectrum'.
>
> (See http://sepwww.stanford.edu/sep/prof/index.html, *Geophysical Estimation by Examples*, Class Notes).

This statement is the basis of the first filtering method. It says that the experimental residuals (squared) should be weighted inversely by their multivariate spectrum for optimal convergence. Because a prediction–error filter (PEF) whitens data from which it was estimated (Jain 1989), it approximates the inverse power spectrum. Thus a PEF (squared) with the inverse spectrum of the coherent noise accomplishes the role of the inverse covariance matrix $\mathbf{C}_d^{-1}$ in eqn (1). The fitting goals in eqn (5) become, omitting the regularization term,

$$0 \approx \mathbf{A}_n(\mathbf{Hm} - \mathbf{d}), \quad (8)$$

where $\mathbf{A}_n$ is a PEF that whitens the coherent noise. The cost function becomes

$$f(\mathbf{m}) = (\mathbf{Hm} - \mathbf{d})^T\mathbf{A}_n^T\mathbf{A}_n(\mathbf{Hm} - \mathbf{d}). \quad (9)$$

Comparing eqns (1) and (9), we see that we are approximating the noise covariance matrix as follows:

$$\mathbf{A}_n^T\mathbf{A}_n \approx \mathbf{C}_d^{-1}. \quad (10)$$

Thanks to the helical boundary conditions (Claerbout 1998), this PEF may be computed in more than one dimension, e.g. in 2D or 3D.

A prediction-subtraction method

Instead of removing the noise by filtering, we can remove it by prediction-subtraction. If an operator is unable to model all the information embedded in the data, then the residual is not IID. The second formulation I propose treats the coherent noise as a component of the data. Now, if we can model the coherent noise with another operator, as we do for the signal, then the residual components become IID. Thus I simply add a coherent noise modelling operator to the fitting goal (Nemeth 1996) in eqn (5) with $\varepsilon = 0$ as follows:

$$0 \approx \mathbf{H}\mathbf{m}_s + \mathbf{B}\mathbf{m}_n - \mathbf{d}, \qquad (11)$$

where $\mathbf{H}$ is the signal modelling operator, $\mathbf{m}_s$ is the model for the signal, $\mathbf{B}$ is an operator that models the coherent noise present in the data $\mathbf{d}$, and $\mathbf{m}_n$ is the model for the noise. Thus ideally, each operator $\mathbf{H}$ and $\mathbf{B}$ models a different part of the data space. The relationships between the noise/signal and $\mathbf{m}_n$/$\mathbf{m}_s$ are then

$$\mathbf{n} \approx \mathbf{B}\mathbf{m}_n \text{ and} \qquad (12)$$

$$\mathbf{s} \approx \mathbf{H}\mathbf{m}_s. \qquad (13)$$

The cost function is

$$f(\mathbf{m}_s, \mathbf{m}_n) = (\mathbf{H}\mathbf{m}_s + \mathbf{B}\mathbf{m}_n - \mathbf{d})^T(\mathbf{H}\mathbf{m}_s + \mathbf{B}\mathbf{m}_n - \mathbf{d}). \qquad (14)$$

Now, for the noise operator $\mathbf{B}$, I propose using the inverse of the PEF $\mathbf{A}_n$ that whitens the coherent noise. To explain this choice, Fig. 1b shows the whitening properties of the PEF if we convolve the PEF back with the input data. Consequently if we deconvolve a spike with the same PEF (Fig. 1c), we model the data. Interestingly, if we assume that seismic data are mainly made up of a superposition of plane waves, the PEF can then whiten quite complex noise patterns. Thus the new cost function is as follows:

$$f(\mathbf{m}_s, \mathbf{m}_n) = (\mathbf{H}\mathbf{m}_s + \mathbf{A}_n^{-1}\mathbf{m}_n - \mathbf{d})^T(\mathbf{H}\mathbf{m}_s + \mathbf{A}_n^{-1}\mathbf{m}_n - \mathbf{d}). \qquad (15)$$

For more details on the minimization of the cost function in eqn (15), the reader is referred to Nemeth *et al.* (2000). So far, I have not said anything about the PEF. In the next section I give guidelines for the PEF estimation.

How to estimate the PEF $\mathbf{A}_n$

For the two proposed methods a PEF needs to be estimated in order to attenuate the coherent noise effects in the residual. In the simplest cases we might be able to derive a noise model from which we can estimate the PEF directly.

Now, if a noise model is not known in advance, I propose the following algorithm:

1 Minimize the cost function in eqn (3).
2 Estimate a PEF $\mathbf{A}_n$ from the residual when only coherent noise remains in the residual.
3 Restart the inverse problem ($\mathbf{m} = \mathbf{0}$) and minimize the cost function in either eqn (9) or eqn (15).

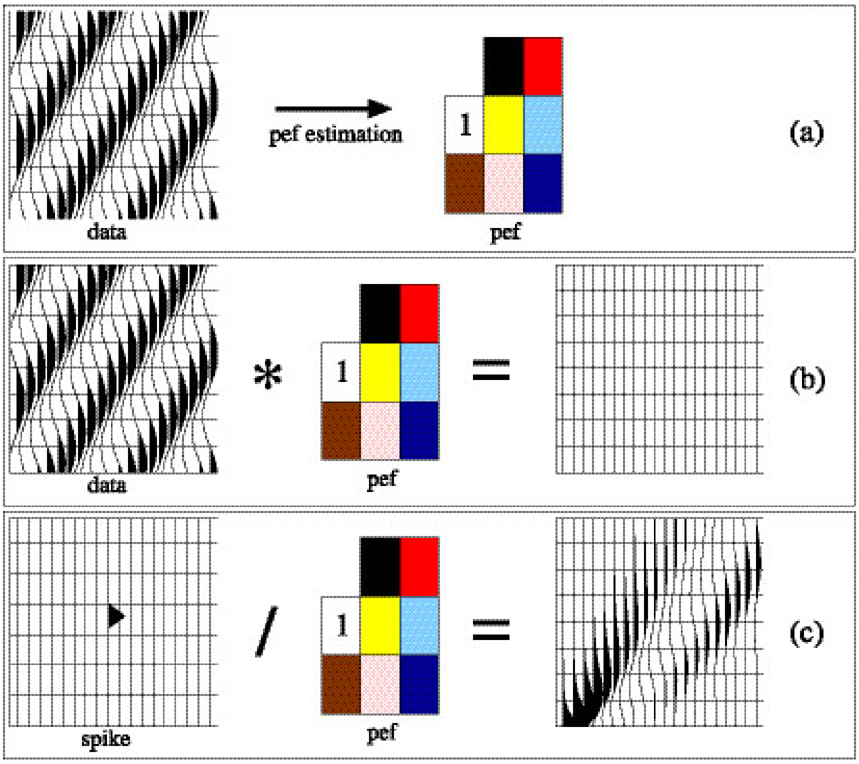

Figure 1 Some basic properties of the prediction–error filters (PEF). (a) A PEF is estimated from a data set. (b) The estimated PEF is convolved with the input data leading to a white residual. (c) A spike is deconvolved with the estimated PEF leading to a pattern close to the input data in (a).

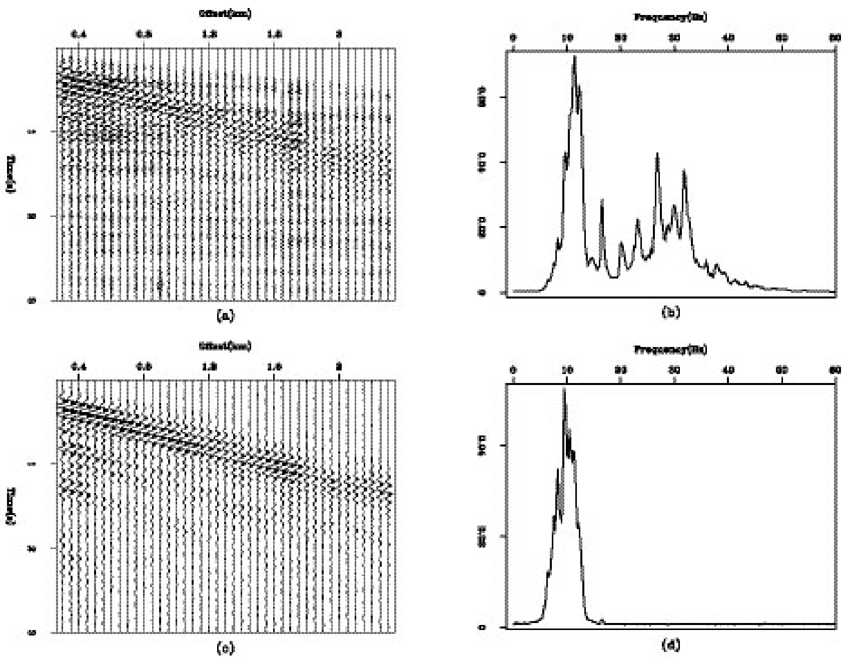

Figure 2 The data and the noise model. (a) A CMP gather. (b) The amplitude spectrum of this gather. (c) A noise model obtained by low-passing the data and adding random zero-mean Gaussian noise. (d) The amplitude spectrum of the noise model. The noise model is used to estimate the PEF $\mathbf{A}_n$.

4 Re-estimate the PEF $\mathbf{A}_n$ from the residual ($\mathbf{Hm} - \mathbf{d}$).
5 Go to step (3).
6 Stop when the residual has a white spectrum.

This nonlinear process tries to refine the PEF estimation in order to obtain the best possible coherent noise attenuation. It is also based on the assumption that the coherent noise will remain in the residual. In the next section, I show results of coherent noise attenuation on a CMP gather contaminated with a low-velocity–low-frequency event.

Coherent noise attenuation results

In this section I show some results from testing the two proposed strategies. The main operator $\mathbf{H}$ is the hyperbola superposition operator, the adjoint $\mathbf{H}^T$ being the hyperbolic radon transform (Thorson & Claerbout 1985).

The model space $\mathbf{m}$ is called the velocity space or velocity spectrum. The input domain of the data $\mathbf{d}$ is the CMP domain. The process of computing the model $\mathbf{m}$ is called velocity inversion.

Figures 2a and b show the input data and their corresponding amplitude spectrum. The signal is made up of a series of hyperbolae, probably all multiples. The goal is to attenuate the low-velocity–low-frequency noise that creates the main peak at 10 Hz in the frequency panel. Firstly, we estimate a noise model by low-passing the data (Fig. 2c,d). This model resembles the noise well enough for us to estimate a PEF directly from it (Brown & Clapp 2000). Then the inversion starts and should give residuals with IID components.

For this simple case, one might be tempted to bandpass the data to attenuate the noise. Nonetheless, as shown in Fig. 3, it is very likely that the noise spectrum partially overlaps the signal spectrum. Thus a simple bandpass filter would attenuate the noise but would also affect the signal, something which we want to avoid. In the next section, the noise attenuation results for both the filtering and the prediction-subtraction techniques are shown.

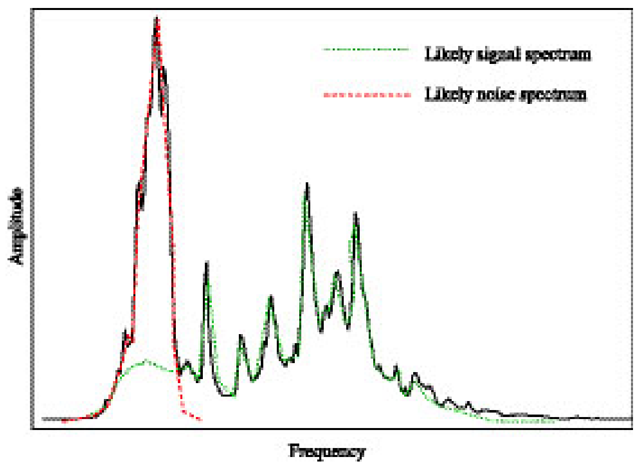

Figure 3 Solid line: average amplitude spectrum of the 2D gather in Fig. 2. Green line: possible signal spectrum. Red line: possible noise spectrum. A simple bandpass filter cannot properly separate the noise from the signal because their corresponding spectra overlap.

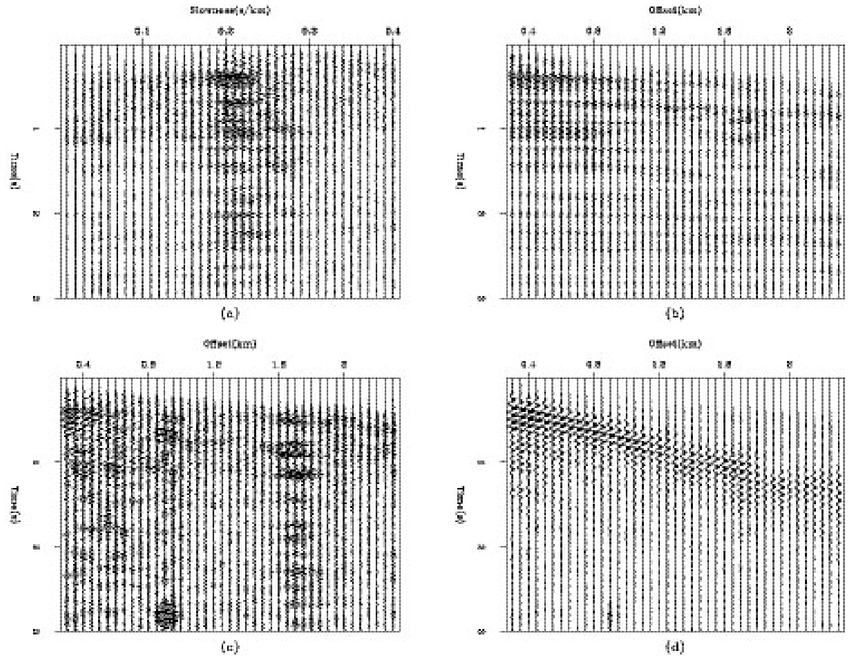

Figure 4 Filtering the coherent noise in real data. (a) An estimated model space $\tilde{\mathbf{m}}$ after inversion; it displays the velocity spectrum of the data. (b) Reconstructed data $\mathbf{H}\tilde{\mathbf{m}}$. (c) The weighted residual ($r = \mathbf{A}_n(\mathbf{H}\tilde{\mathbf{m}} - \mathbf{d})$) after inversion, the residual components should be IID. (d) The difference between the input data in Fig. 2a and the reconstructed data in Fig. 4b; this panel displays the coherent noise filtered by the method.

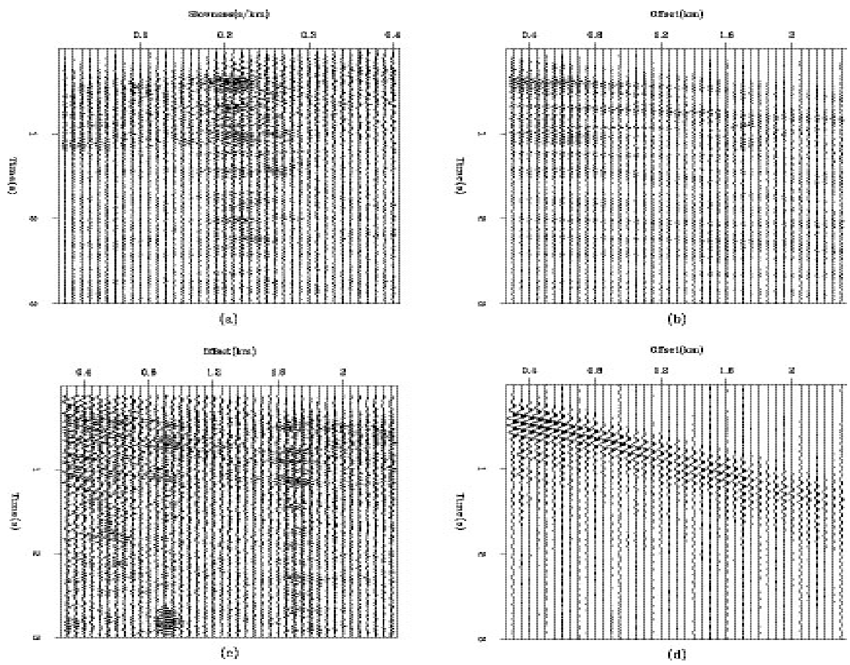

Figure 5 Subtracting the coherent noise in real data. (a) An estimated model space $\tilde{\mathbf{m}}$. (b) The reconstructed data from the model space, e.g. $\mathbf{H}\tilde{\mathbf{m}}$. (c) The residual after inversion, i.e. $\mathbf{r} = \mathbf{H}\tilde{\mathbf{m}}_s + \mathbf{A}_n^{-1}\tilde{\mathbf{m}}_n - \mathbf{d}$; it should have IID components. (d) The estimated coherent noise $\mathbf{A}_n^{-1}\tilde{\mathbf{m}}_n$ after inversion.

Filtering method results

The result of the inversion is displayed in Fig. 4. The residual (Fig. 4c) is not perfectly white, but the coherent noise has been mostly filtered out. Because the noise model does not incorporate them, the remaining artefacts in the residual are amplitude anomalies near offset 0.8 km and time 2.8 s. The remodelled data in Fig. 4b are free of coherent noise, and the velocity panel in Fig. 4a shows quite clearly a well-focused corridor. The coherent noise filtered out during the inversion is shown in Fig. 4d. As anticipated, I was able to perform a

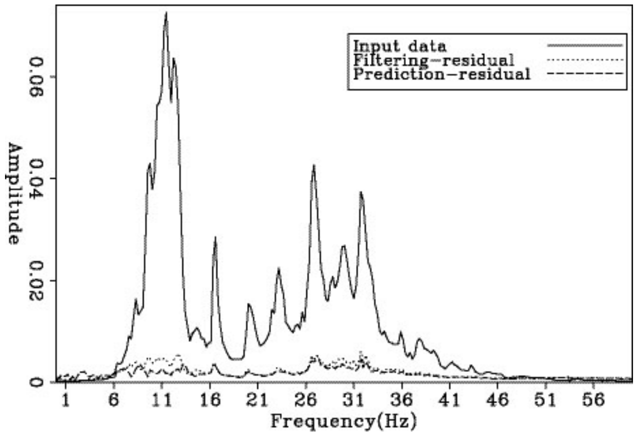

Figure 6 A comparison between the average amplitude spectrum of (1) the input data (solid line) (2) the residual after inversion with the filtering method (dotted line) and (3) the residual after inversion with the prediction method (dashed line). The spectra of the residuals is not perfectly white due to some noise not being accounted for in the noise model (amplitude anomalies near offset 0.8 km and time 2.8 s). None the less, the spectra are fairly flat proving that the residual components are nearly IID.

velocity inversion along with coherent noise filtering. The residual has IID components thanks to the addition of the PEF in the cost function in eqn (9).

Subtraction method results

The coherent noise attenuation resembles that obtained with the filtering approach. Figure 5 shows the result of the inversion: Fig. 5a shows the model space (velocity spectrum); Fig. 5b shows the reconstructed data from Fig. 5a,d shows the estimated noise model after inversion; Fig. 5c shows the residual. The noise model is composed mainly of the coherent noise I am trying to attenuate. However, the residual shows linear events scattered throughout the panel with very low energy. Again, this method proves efficient since the coherent noise has been correctly predicted by the PEF and subtracted during the inversion.

Discussion

The need for IID residuals components led me to design two different inversion schemes. Figures 6 and 7 illustrate the properties of these methods further. Figure 7 shows that the two proposed inversion methods increase the convergence and Fig. 6 shows that the amplitude spectrum of the residual after inversion is almost flat for both methods. Therefore, our goal of getting IID residual components is achieved. I also demonstrated that the noise covariance matrix can be approximated with prediction–error filters.

For more complicated patterns than the one shown in the preceding results, a single PEF might not be enough to model/filter the coherent noise. In that case a bank of non-stationary filters can be estimated (Crawley *et al.* 1999) to allow for the variability of the undesirable events.

In the prediction-subtraction method, the coherent noise attenuation can be improved by adding a model or data space regularization term to the objective function. These regularizations also have the property of limiting the leakage effects when the noise and signal modelling operators predict similar components of the data.

Acknowledgements

I thank the sponsors of the Stanford Exploration Project and Jon Claerbout for insightful discussions.

References

Abma, R. and Claerbout, J. [1995] Lateral prediction for noise attenuation by T-X and F-X techniques. *Geophysics* **60**, 1887–1896.

Brown, M. and Clapp, R. [2000] T-X domain, pattern-based ground-roll removal. 70th meeting, Society of Exploration Geophysicists, Expanded Abstracts, pp. 2103–2106.

Claerbout, J.F. [1998] Multidimensional recursive filters via a helix. *Geophysics* **63**, 1532–1541.

Crawley, S., Claerbout, J. and Clapp, R. [1999] Interpolation with smoothly nonstationary prediction-error filters. 69th meeting, Society of Exploration Geophysicists, Expanded Abstracts, pp. 1154–1157.

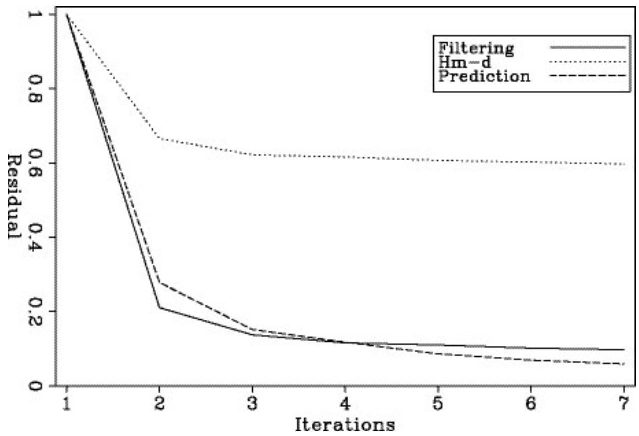

Figure 7 Relative convergence of three inversion schemes. Solid line: filtering technique. Dashed line: prediction technique. Dotted line: minimization of the objective function in eqn (3), with $\varepsilon = 0$. Because the coherent noise is not fitted but rather filtered or predicted, the convergence is greatly improved in the two methods proposed in this paper.

Jain, A.K. [1989] *Fundamentals of Digital Image Processing*. Prentice-Hall Inc.

Nemeth, T. [1996] *Imaging and filtering by least-squares migration*. PhD Thesis, University of Utah.

Nemeth, T., Sun, H. and Schuster, G.T. [2000] Separation of signal and coherent noise by migration filtering. *Geophysics* **65**, 574–583.

Sacchi, M.D. and Ulrych, T.J. [1995] High-resolution velocity gathers and offset space reconstruction. *Geophysics* **60**, 1169–1177.

Soubaras, R. [1994] Signal-preserving random noise attenuation by the F-X projection. 64th meeting, Society of Exploration Geophysicists, Expanded Abstracts, pp. 1576–1579.

Tarantola, A. [1987] *Inverse Problem Theory*. Elsevier Science.

Thorson, J. and Claerbout, J. [1985] Velocity stack and slant stochastic inversion. *Geophysics* **50**, 2727–2741.

Wang, Y. and Pratt, R.G. [1997] Sensitivities of seismic traveltimes and amplitudes in reflection tomography. *Geophysical Journal International* **131**, 618–642.

Wang, Y., White, R.E. and Pratt, R.G. [2000] Seismic amplitude inversion for interface geometry: practical approach for application. *Geophysical Journal International* **142**, 162–172.