Stanford Exploration Project, Report 105, September 5, 2000, pages 1–49

Coherent noise attenuation using Inverse Problems and Prediction Error Filters

Antoine Guitton[1]

ABSTRACT

Two iterative methods that handle coherent noise effects during the inversion of 2-D prestack data are tested. One method approximates the inverse covariance matrices with PEFs, and the other introduces a coherent noise modeling operator in the objective function. This noise modeling operator is a PEF that has to be estimated before the inversion from a noise model or directly from the data. These two methods lead to Independent, Identically Distributed (IID) residual variables, thus guaranteeing a stable convergence of the inversion schemes and permitting coherent noise filtering/separation.

INTRODUCTION

Preserving the amplitudes requires inverse theory when an operator is not unitary. Unfortunately, seismic operators are usually not unitary. Seismic operators can be regarded as the adjoint of forward "modeling" operators (Claerbout, 1992). Lailly (1983) was the first to recognize that the standard migration operator is the adjoint to the corresponding forward operator and used it in the first iteration of full waveform inversion (Tarantola, 1987). Because the adjoint is not the inverse of the operator, the amplitude of the input data is not preserved. An approximate inverse can be computed using least-squares inversion. Thorson (1984) replaces the standard hyperbolic Radon transform with a linear, least-squares inversion of the velocity stack equations. However, the least-squares operator involves the computing of the inverse of the Hessian that is often very difficult to derive. Forgues and Lambare (1997) and Chavent and Plessix (1999) calculated an approximate inverse of the Hessian matrix associated with the migration operator. When the calculation of the least-squares inverse is not feasible (size of the matrices, use of operators instead of matrices), an iterative scheme is preferred.

However, with a least-squares inversion, major difficulties arise when the data are contaminated by noisy events. By noisy event I mean

- Abnormally large or small data components, or *outliers*, where long-tailed probability density functions (pdf) should be used as opposed to short-tailed Gaussian pdf.
- *Coherent noise* that the seismic operator is unable to model (for example, a hyperbolic Radon transform can't properly model ellipses).

[1]**email:** antoine@sep.stanford.edu

The noise will (1) spoil any analysis based on the result of the inversion and (2) affect the amplitude recovery of the input data. From a more statistical point of view, the residual, which measures the quality of the data fitting, will be corrupted by high amplitude (outliers in the data) or highly correlated events (coherent noise in the data) that will attract much of the solver's efforts, thus degrading the model **m**. The first I in IID stands for Independent, meaning that no coherent events are present in the residual (hyperbolas, ellipses, lines, etc.). The second I and D stand for Identically Distributed, meaning that the residual components have similar energy. For example, if seismic data have not been multiplied by t^2 to correct for spherical divergence effects, the variables in the gather will not be Identically Distributed (Figure 1). A possible solution to one particular noise problem is to attribute long-tailed pdfs

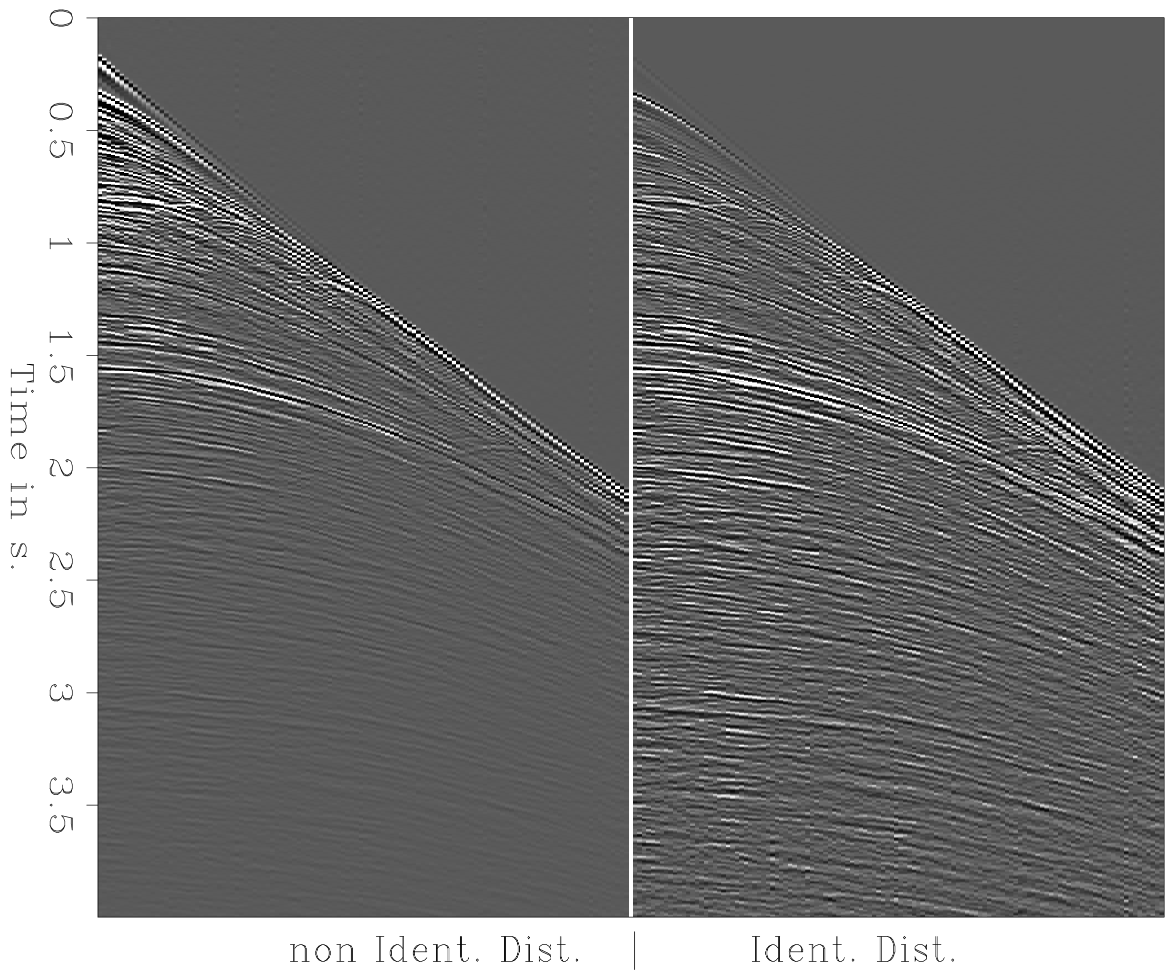

Figure 1: Left: the data have not been multiplied by t^2 thus giving non Identically Distributed variables. Right: after t^2 correction, the data are Identically Distributed. antoine1-iid [ER]

to the residual variables. This long-tailed pdfs lead to the minimization of the l^1 norm of the data residual

$$f(\mathbf{m}) = |\mathbf{Hm} - \mathbf{d}|_1, \tag{1}$$

where **H** is the seismic operator, **m** the model and **d** the input seismic data. Because the l^1 norm is less sensitive to outliers, it will give a better fitting of the data (Claerbout and Muir, 1973). The minimization of such objective functions is a cumbersome problem because the l^1 norm is not differentiable everywhere. However, some alternatives exist if we use a hybrid $l^1 - l^2$ norm such as the Huber norm (Huber, 1973; Guitton and Symes, 1999) or an Iteratively Reweighted Least Squares (IRLS) algorithm (Nichols, 1994; Bube and Langan, 1997) with an appropriate weighting function. These methods have proved efficient (Guitton, 2000a). The use of long-tailed pdfs is particularly effective at attenuating outliers. However, they are usually not effective at attenuating coherent noise because it is not generally distinguishable by its histogram, but by its moveout patterns. In addition, hybrid solvers are either difficult to tune or expensive to use.

What will I do?

The two proposed methods are based on the need to have IID residual components. A typical inverse problem arises when we want to minimize the objective function for the fitting goal

$$\mathbf{0} \approx \mathbf{Hm} - \mathbf{d}, \tag{2}$$

where $\mathbf{m}$ is a mapping of the data (unknown of the inverse problem), $\mathbf{H}$ an operator and $\mathbf{d}$ the seismic data. The residual $\mathbf{r}$ is defined as the difference between input data $\mathbf{d}$ and estimated data $\tilde{\mathbf{d}} = \mathbf{Hm}$,

$$\mathbf{r} = \tilde{\mathbf{d}} - \mathbf{d}.$$

My research is focused on the attenuation/separation of the **coherent noise** only. The first strategy relates to fundamentals in inverse theory as detailed in the **General Discrete Inverse Problem** (Tarantola, 1987) and approximates the inverse covariance matrices with PEFs. The second strategy proposes to introduce a coherent noise modeling part in Equation 2. The noise operator will be a PEF. In the **first strategy** the coherent noise is **filtered**. In the **second strategy** the coherent noise is **subtracted** from the signal. The two methods should (1) give IID residual components, (2) stabilize the inversion, and (3) preserve the “real” events amplitudes as long as the noise and the signal operators have been carefully chosen.

Why two methods?

The two methods achieve the same goal, but each has its own pros and cons. I will show that the filtering method is easier to implement: the PEF estimation can be done iteratively directly from the residual as the iterations go on. In contrast, for the subtraction method, the coherent noise operator (a PEF) should be pre-estimated and kept constant as the iterations go on. This *a priori* information can be sometimes rather difficult to have. The convergence of the subtraction scheme is far better than the convergence of the filtering method, however. Ideally, the nature of the coherent noise should guide us in this choice.

THE INVERSE PROBLEM

In seismic processing we often transform data into equivalent data using linear operators. Among these operators, we have the Fourier transform, the Radon transform, the migration operator etc. Some of these operators are **unitary** (Fourier transform), meaning that the input data are perfectly recoverable using the adjoint. Mathematically speaking, unitarity implies

$$\mathbf{H}^{\mathbf{T}}\mathbf{H} = \mathbf{I}, \tag{3}$$

where $\mathbf{H}$ is the operator, $\mathbf{H}^{\mathbf{T}}$ the adjoint, and $\mathbf{I}$ the identity matrix. Unfortunately, most of the operators are not unitary, meaning that one can't go back and forth between the model $\mathbf{m}$ and the data $\mathbf{d}$ without losing information or resolution. Mathematically speaking, non-unitarity implies

$$\mathbf{H}^{\mathbf{T}}\mathbf{H} \neq \mathbf{I}. \tag{4}$$

This loss of information can be overcome using inverse theory. The goal of inverse theory is to find a model $\mathbf{m}$ that optimally represents the input data $\mathbf{d}$ given an operator $\mathbf{H}$ and given a definition of optimality (minimum energy residual-l^2 for example):

$$f(\mathbf{m}) = \|\mathbf{Hm} - \mathbf{d}\|_2. \tag{5}$$

The classical approach: least-squares criterion

The least-squares criterion comes directly from the hypothesis that the pdf of each observable 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 objective function)

$$f(\mathbf{m}) = (\mathbf{Hm} - \mathbf{d})^{\mathbf{T}}\mathbf{C}_{\mathbf{d}}^{-1}(\mathbf{Hm} - \mathbf{d}) + (\mathbf{m} - \mathbf{m}_{\mathbf{prior}})^{\mathbf{T}}\mathbf{C}_{\mathbf{m}}^{-1}(\mathbf{m} - \mathbf{m}_{\mathbf{prior}}), \tag{6}$$

where $\mathbf{C}_{\mathbf{d}}$ and $\mathbf{C}_{\mathbf{m}}$ are the *covariance operators*, and $\mathbf{m}_{\mathbf{prior}}$ a model given *a priori*. The covariance matrix $\mathbf{C}_{\mathbf{d}}$ combines experimental errors and modeling uncertainties. Modeling uncertainties describe the difference between what the operator can predict and the data. Thus the covariance matrix $\mathbf{C}_{\mathbf{d}}$ is often called the *noise covariance* matrix. Assuming (1) uniform variance of the model and of the noise, (2) covariance matrices are diagonal , i.e., uncorrelated model an data components, and (3) no *prior* model $\mathbf{m}_{\mathbf{prior}}$, the objective function becomes

$$f(\mathbf{m}) = (\mathbf{Hm} - \mathbf{d})^{\mathbf{T}}(\mathbf{Hm} - \mathbf{d}) + \epsilon^{2}\mathbf{m}^{\mathbf{T}}\mathbf{m}, \tag{7}$$

where ϵ is a function of the noise and model variances. The previous assumptions leading to Equation 7 are quite 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 are not independent. Minimizing the objective function in Equation 7 is equivalent to having the two fitting goals for $\mathbf{m}$

$$\mathbf{0} \approx \mathbf{Hm} - \mathbf{d} \tag{8}$$

$$\mathbf{0} \approx \epsilon\mathbf{Im}. \tag{9}$$

The first inequality expresses the need for the operator $\mathbf{H}$ to fit the input data $\mathbf{d}$. The second inequality is often called the regularization (or model styling) term. The minimization of Equation 8, when the operator $\mathbf{H}$ is linear, may be done using any kind of linear method such as the steepest descent algorithm or faster conjugate gradients/directions methods (Paige and Saunders, 1982). From now on, I will refer to Equation 8 as the "simplest" approach. When the assumptions leading to Equation 8 are respected, the convergence towards $\mathbf{m}$ is easy to achieve. In particular, the components of the residual $\mathbf{r} = \mathbf{Hm} - \mathbf{d}$ become IID. This IID property implies that no coherent information is left in the residual and that each variable of the residual has similar intensity (or power). The main factor that may alter this property is the presence of noise in the data that violates assumptions about both the uniform distribution and the need of independent noise components.

PROPOSED SOLUTIONS TO ATTENUATE COHERENT NOISE

Any dataset may be regarded as the sum of signal and noise

$$\mathbf{d} = \mathbf{s} + \mathbf{n}.$$

I assume that the coherent noise $\mathbf{n}$ is made of the inconsistent part (or modeling uncertainties part) of the data $\mathbf{d}$ for any given operator $\mathbf{H}$. My goal is to define new strategies that would lower the influence of the noise $\mathbf{n}$, giving IID residual components.

METHOD 1: A filtering method

Equation 6 introduces two matrices that are difficult to compute: the data covariance matrix $\mathbf{C_d}$ and the model covariance matrix $\mathbf{C_m}$. I concentrate my efforts on the data 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 by diagonal operators. IID residual components is equivalent to having a residual with a white spectrum. Thus coherent noise will add "color" to the spectrum of the residual. The goal of the covariance matrices is to **absorb** this spectrum. As Jon Claerbout (1999) asserts:

> Clearly, the noise spectrum is the same as the data covariance only if we accept the theoritician'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".

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 PEF whitens data from which it was estimated, it approximates the inverse power spectrum of the data. Thus a PEF (squared) estimated from the residual or the model accomplishes the role of the inverse covariance matrices $\mathbf{C_d^{-1}}$ and $\mathbf{C_m^{-1}}$ in Equation 6. The fitting goals in Equation 8 become, omitting the regularization term,

$$0 \approx \mathbf{A_r}(\mathbf{Hm} - \mathbf{d}), \tag{10}$$

where $\mathbf{A_r}$ is a PEF estimated from the residual and $\mathbf{A_m}$ from the unknown model. Thanks to the Helical boundary conditions (Claerbout, 1998), the PEF may be computed in more than one dimension (2-D, 3-D). This gives us a lot of flexibility to calculate the residual spectrum. An important task is to develop a strategy to estimate the residual PEF $\mathbf{A_r}$. I propose the following algorithm:

Algorithm 1

1. Compute the current residual **r = Hm - d**.
2. Estimate a PEF $\mathbf{A_r}$ from the residual.
3. Minimize the objective function (l^2 norm)

$$f(\mathbf{m}) = (\mathbf{Hm}-\mathbf{d})^{\mathbf{T}}\mathbf{A_r^T}\mathbf{A_r}(\mathbf{Hm}-\mathbf{d}), \tag{11}$$

4. Go to 1 after a certain number of iterations in step (3).

Notice that the first PEF is estimated from the data (if no *prior* coherent noise model exists). Then the residual PEF is re-estimated iteratively. This optimization scheme is very similar to IRLS algorithms where weighting functions are re-computed after a certain number of iterations. Because I re-compute the PEF iteratively, my goal is to have the best estimate of the residual's multivariate spectrum. This problem is then piece-wise linear. With this strategy, the residual should be IID. Notice that the minimization of the objective function can be done with our favorite fast conjugate gradients method.

METHOD 2: A subtraction method

Instead of removing the noise by filtering, we can remove it by 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 is based on the idea that if we can model the coherent noise with another operator, then the residual components become IID. Let us consider that we have

$$\mathbf{d} = \mathbf{s} + \mathbf{n}$$

and that there exists an operator $\mathbf{L}$ such that

$$\mathbf{L} = [\mathbf{H} \quad \mathbf{L_n}].$$

We assume that $\mathbf{H}$ is the modeling operator for the signal $\mathbf{s}$ and that $\mathbf{L_n}$ is the modeling operator for the coherent noise $\mathbf{n}$. Following this decomposition, we can write

$$\mathbf{m} = \begin{bmatrix} \mathbf{m_s} \\ \mathbf{m_n} \end{bmatrix}$$

where $\mathbf{m_n}$ is the noise-model and $\mathbf{m_s}$ is the signal-model. Starting from

$$\mathbf{0} \approx \mathbf{Lm} - \mathbf{d}, \tag{12}$$

the fitting goal then becomes

$$\mathbf{0} \approx \mathbf{H}\mathbf{m_s} + \mathbf{L_n}\mathbf{m_n} - \mathbf{d}. \tag{13}$$

Because we have to find $\mathbf{m_s}$ and $\mathbf{m_n}$, this system is clearly under-determined and some regularization is needed. Thus, we end up with the following fitting goals

$$\begin{aligned} \mathbf{0} &\approx \mathbf{H}\mathbf{m_s} + \mathbf{L_n}\mathbf{m_n} - \mathbf{d} \\ \mathbf{0} &\approx \epsilon\mathbf{I}\mathbf{m_s} \\ \mathbf{0} &\approx \epsilon\mathbf{I}\mathbf{m_n}. \end{aligned}$$

Because there should be a different operator $\mathbf{L_n}$ for each different coherent noise pattern, the cost of this method increases considerably. Fortunately, we can use multi-dimensional PEFs to estimate the coherent noise operator. This estimation is possible if we assume that the coherent noise is predictable, i.e., made up of the superposition of local plane wave segments (Claerbout, 1992). If we can estimate PEFs from the coherent noise, then the inverse PEF should be our coherent noise modeling operator $\mathbf{L_n} = \mathbf{A_n^{-1}}$. Computing the inverse of multi-dimensional PEFs is now possible *via* the helix. In addition, with the helical boundary conditions, computing the inverse of multi-dimensional PEFs is as easy as computing the inverse of 1-D filters. We have then

$$\begin{aligned} \mathbf{0} &\approx \mathbf{H}\mathbf{m_s} + \mathbf{A_n^{-1}}\mathbf{m_n} - \mathbf{d} \\ \mathbf{0} &\approx \epsilon\mathbf{I}\mathbf{m_s} \\ \mathbf{0} &\approx \epsilon\mathbf{I}\mathbf{m_n}, \end{aligned} \tag{14}$$

where $\mathbf{A_n}$ is the noise PEF. This approach is similar to Tamas Nemeth's approach (1996). The difference emerges in the choice of the operators $\mathbf{L_n}$ and $\mathbf{H}$. Whereas Nemeth (1996) imposes one operator $\mathbf{L_n}$ to model the noise, we estimate a PEF $\mathbf{A_n}$ and use it in the fitting goals (Equation 14). Because PEFs (with appropriate dimensions) whiten the spectrum of many different plane-waves, this strategy is more flexible (no assumptions regarding the moveout of the noise). This method should give IID residual variables as long as we are able to estimate PEFs for the coherent noise. This is the main difficulty and challenge of this method. The minimization of the objective function in a least-squares sense for the fitting goals in Equation 14 can be done again with a fast conjugate gradients method.

I did not develop any specific algorithm to solve this inverse problem. I assume that we have a strategy that allows us to estimate the operator $\mathbf{A_n}$. We can then minimize the objective function for the fitting goals given in Equation 14 in a least-squares sense, for example.

RESULTS

In this section I show some preliminary results from testing the two proposed strategies. The main operator $\mathbf{H}$ is the hyperbola superposition operator. The adjoint $\mathbf{H^T}$ is the hyperbolic Radon transform (Nichols, 1994; Guitton, 2000b). The model space $\mathbf{m}$ is called the velocity space. The input data $\mathbf{d}$ are CMP gathers. The process of computing the model $\mathbf{m}$ minimizing Equation 11 is called *velocity inversion*.

Filtering method

The following results illustrate the first strategy. The left panel of Figure 2 displays the input CMP gather for the velocity inversion. These data have been pre-whitened with a 1-D PEF (deconvolution). Notice that this CMP is made up of nearly horizontal events, hyperbolas and a slow velocity, low frequency event crossing the gather. The later is the coherent noise we want to get rid of. I iterated 30 times to obtain the final model $\mathbf{m}$ (Figure 6). The residual PEF is 2-D with 25 coefficients on the time axis and 2 coefficients on the space axis. This PEF is re-estimated every ten iterations (see algorithm 1).

In Figure 2, I compare the input data with the remodeled data ($\tilde{\mathbf{d}} = \mathbf{Hm}$) after least-squares inversion with (Equation 10) and without PEF (Equation 8). Notice how close the two results are (the velocity of the linear event is not scanned in the velocity inversion). Figure 3 shows a comparison of the residuals ($\mathbf{r} = \tilde{\mathbf{d}} - \mathbf{d}$). As expected, the residual of the least-squares inversion with PEF estimation gives a white residual (right panel) as opposed to the "simplest" inversion residual contaminated with the linear noise (left panel). Notice that the use of the helical boundary conditions for the PEF estimation has left its footprint on the edges of the residual panel. Figure 4 displays the two spectra for the two residuals.

A comparison of the two model space (Figure 6) shows that (1) both results are difficult to interpret and (2) the inversion scheme with PEF gives a more satisfying panel. As a more striking comparison, Figure 8 shows the output of the least-squares inversion with or without PEF as a function of the number of iterations. After 100 iterations, the "simplest" inversion (Equation 8) gives a velocity panel infested with artifacts, for it tries to fit the linear event left in the residual. In contrast, with the proposed scheme, the change in the number of iteration does not affect the final result: the inversion becomes stable. Figure 7 displays the convolution of one of the inverse PEF estimated during the iterations with a panel of white noise. It demonstrates that the PEF is effectively after the linear event we want to attenuate.

Subtraction method

Now we have the fitting goals in Equation 14. For the time being, I drop the two regularization terms in Equations 14 and focus my analysis on the data fitting part. For the noise modeling operator $\mathbf{A_n}$, I computed a 2-D PEF directly from the data (which gives a very approximate coherent noise PEF). The size of this PEF is 25×2. The convolution of the inverse PEF with a panel filled with white noise is shown in Figure 10. It shows that the PEF predicts both signal (thin lines) and coherent noise (linear event) that will cause crosstalks with the hyperbolic Radon operator. This PEF or coherent noise operator is kept constant during the iterations. I iterated 30 times. Figure 9 displays the model space $\mathbf{m_s}$ on the left, the modeled noise in the middle ($\mathbf{A_n^{-1} m_n}$), and the residual on the right. As expected, because the PEF is not a perfect coherent noise operator, some signal is trapped in the linear event (middle, Figure 9). Figure 5 shows the spectrum of the residual with the "simplest" inversion along with the spectrum of the residual for the subtraction scheme.

After 100 iterations of the subtraction scheme, we see (Figure 12) that the model space does not vary too much, as opposed to the "simplest" approach (Equation 8). This method is stable with respect to the number of iterations.

Comparison study

Figure 11 compares the convergence of the filtering, subtraction and "simplest" methods. The subtraction scheme converges significantly better than the two other schemes. Figure 13 displays the three velocity panels corresponding to each approach (30 iterations). The two proposed methods increase the resolution of the velocity space compared to the "simplest" scheme. The table below intends to summarize some practical issues for each strategy when coherent noise is present in the input data:

Methods	Convergence	PEF estimation	Regularization needed ?	Residual	Stability (iterations)
Filtering	Slow	Easy	Not all the time	IID	Stable
Subtraction	Fast	Difficult	Yes	IID	Stable
Simplest	Slow	NA	Not all the time	not IID	Unstable

The "PEF estimation" columns relates to the difficulty of estimating the residual PEF in Equation 10 and the coherent noise PEF in Equation 14.

DISCUSSION

1. In the filtering method, PEFs are recomputed iteratively from the data residual. I think this solution is the method of choice since the PEF (squared) is the inverse noise covariance matrix $\mathbf{C_d^{-1}}$. For the subtraction method, however, the final result is driven by the orthogonality between the coherent noise operator and the signal operator (Nemeth, 1996). If the two operators can model similar parts of the data, the separation will not be efficient. Nemeth proposes introducing some regularization (Equation 14) to mitigate this difficulty. We could perhaps compute a *prior* **coherent noise model** from which we estimate the PEFs. This approach is related to Spitz's idea (Spitz, 1999), according to which a noise model is utilized to estimate the signal PEF. In any case, the strategy of computing the coherent noise operator (PEF) is of a vital importance for the quality of noise separation.

2. The PEF estimation is one problem, but choosing the signal operator **H** is another. As said before, the two approaches perform noise attenuation (filtering method) or noise separation (subtraction method) along with a conventional signal processing step (velocity analysis here). The processing step should be chosen in agreement with the expected signal in the data. Basically, the processing operator **H** should **mitigate the crosstalks** between the signal and the coherent noise. Coherent noise comes in different flavors and **H** should reflect this heterogeneity. Harlan (1986) gives some guidelines alternating between migration, Slant-Stack, and offset-local Stack as a function of the coherent

noise and of the signal. We should keep these guidelines in mind when dealing with different datasets, different problems.

3. Because the data are not time-stationary, the coherent noise operator should be a function of time and space. This difficulty can be overcome using **non-stationary filters**. In particular, estimating space varying filters with coefficients smoothed along a radial direction proved efficient (Crawley, 1999). Nonetheless, Clapp and Brown (2000) experienced stability problems, making the computing of the inverse PEFs potentially unsafe.

4. The two proposed methods have the advantage of performing noise attenuation (filtering method) or noise separation (subtraction method) along with a geophysical process (velocity inversion in this case). The two algorithms can be used at the same time. The fitting goals become

$$\begin{aligned} \mathbf{0} &\approx \mathbf{A_r}(\mathbf{H}\mathbf{m_s} + \mathbf{A_n^{-1}}\mathbf{m_n} - \mathbf{d}) \\ \mathbf{0} &\approx \epsilon \mathbf{A_{m_s}}\mathbf{m_s} \\ \mathbf{0} &\approx \epsilon \mathbf{A_{m_n}}\mathbf{m_n}. \end{aligned} \tag{15}$$

$\mathbf{A_r}$, $\mathbf{A_{m_s}}$, $\mathbf{A_{m_n}}$ and $\mathbf{A_n}$ are PEFs to be estimated. Note that if (1) $\mathbf{H}\mathbf{m_s} = \mathbf{s}$ (the signal), (2) $\mathbf{A_n^{-1}}\mathbf{m_n} = \mathbf{n}$ (the coherent noise), (3) $\mathbf{A_{m_n}} = \mathbf{A_r} = \mathbf{N}$ (the coherent noise PEF), (4) $\mathbf{A_{m_s}}\mathbf{m_s} = \mathbf{S}\mathbf{s}$ ($\mathbf{S}$ the signal PEF) and (5) $\mathbf{A_{m_n}}\mathbf{m_n} = \mathbf{N}\mathbf{n}$, then Equation 15 is exactly Abma's Equation (1995). Equation 15 gives a simple generalization of the methods proposed above and should be tested.

5. A more general **robust inversion scheme** can be derived combining the two proposed methods along with a robust norm. In particular, it would be interesting to use the hybrid $l^2 - l^1$ Huber norm more routinely. Thus we would solve the noise problem in its totality, handling both outliers and coherent noise effects at the same time.

6. The **extension to 3-D data** should be feasible. A problem arises in the choice of the operators and in the PEF estimation. For this method to be really efficient in 3-D, more 3-D operators should be used. The PEF's estimation in 3-D is theoretically a simple extension of the 1-D case. The shape of the PEF may be more difficult to anticipate, however. In addition, the irregular geometry, intrinsic to 3-D acquisitions, can make the estimation of the PEF very difficult.

CONCLUSION

The two methods introduced in this paper have proved efficient. The goal is attained since the residual has IID components. This property of the residual insures a stable convergence of the iterative scheme along with the filtering/subtraction of the coherent noise present in the data. The next "natural step" would be to test these methods first on a complete 2-D line and then on 3-D data if the 2-D experiment is conclusive enough.

AKNOWLEDGMENTS

I would like to thank Jon Claerbout for insightful discussions. Thanks also to Biondo Biondi for suggesting some improvements to this paper.

REFERENCES

Abma, R., 1995, Least-squares separation of signal and noise with multidimensional filters: Ph.D. thesis, Stanford University.

Bube, K. P., and Langan, R. T., 1997, Hybrid l1/l2 minimization with applications to tomography: Geophysics, **62**, no. 04, 1183–1195.

Chavent, G., and Plessix, R., 1999, An optimal true-amplitude least-squares prestack depth-migration operator: Geophysics, **64**, no. 2, 508–515.

Claerbout, J. F., and Fomel, S., 1999, Geophysical Estimation with Example: Class notes, http://sepwww.stanford.edu/sep/prof/index.html.

Claerbout, J. F., and Muir, F., 1973, Robust modeling with erratic data: Geophysics, **38**, 820–844.

Claerbout, J. F., 1992, Earth Soundings Analysis, Processing versus Inversion: Blackwell Scientific Publication.

Claerbout, J., 1998, Multidimensional recursive filters via a helix: Geophysics, **63**, no. 05, 1532–1541.

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

Crawley, S., 1999, Interpolation with smoothly nonstationary prediction-error filters: SEP–**100**, 181–196.

Forgues, E., and Lambare, G., 1997, Resolution of multi-parameter ray+borne inversion: 61st Mtg. Eur. Assoc. Expl Geophys, Extended Abstracts, Session:P115.

Guitton, A., and Symes, W. W., 1999, Robust and stable velocity analysis using the Huber function: 69th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1166–1169.

Guitton, A., 2000a, Huber solver versus IRLS algorithm for quasi L1 inversion: SEP–**103**, 255–271.

Guitton, A., 2000b, Prestack multiple attenuation using the hyperbolic Radon transform: SEP–**103**, 181–201.

Harlan, W. S., 1986, Signal-noise separation and seismic inversion: Ph.D. thesis, Stanford University.

Huber, P. J., 1973, Robust regression: Asymptotics, conjectures, and Monte Carlo: Ann. Statist., **1**, 799–821.

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Soc. Indust. Appl. Math., Conference on inverse scattering, 206–220.

Nemeth, T., 1996, Imaging and filtering by Least-Squares migration: Ph.D. thesis, The university of Utah.

Nichols, D., 1994, Velocity-stack inversion using $\mathbf{L_p}$ norms: SEP–**82**, 1–16.

Paige, C. C., and Saunders, M. A., 1982, LSQR: an algorithm for sparse linear equations and sparse least squares: ACM Transactions on Mathematical Software, **61**, 43–71.

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

Tarantola, A., 1987, Inverse Problem Theory: Elsevier Science Publisher.

Thorson, J. R., 1984, Velocity stack and slant stack inversion methods: Ph.D. thesis, Stanford University.

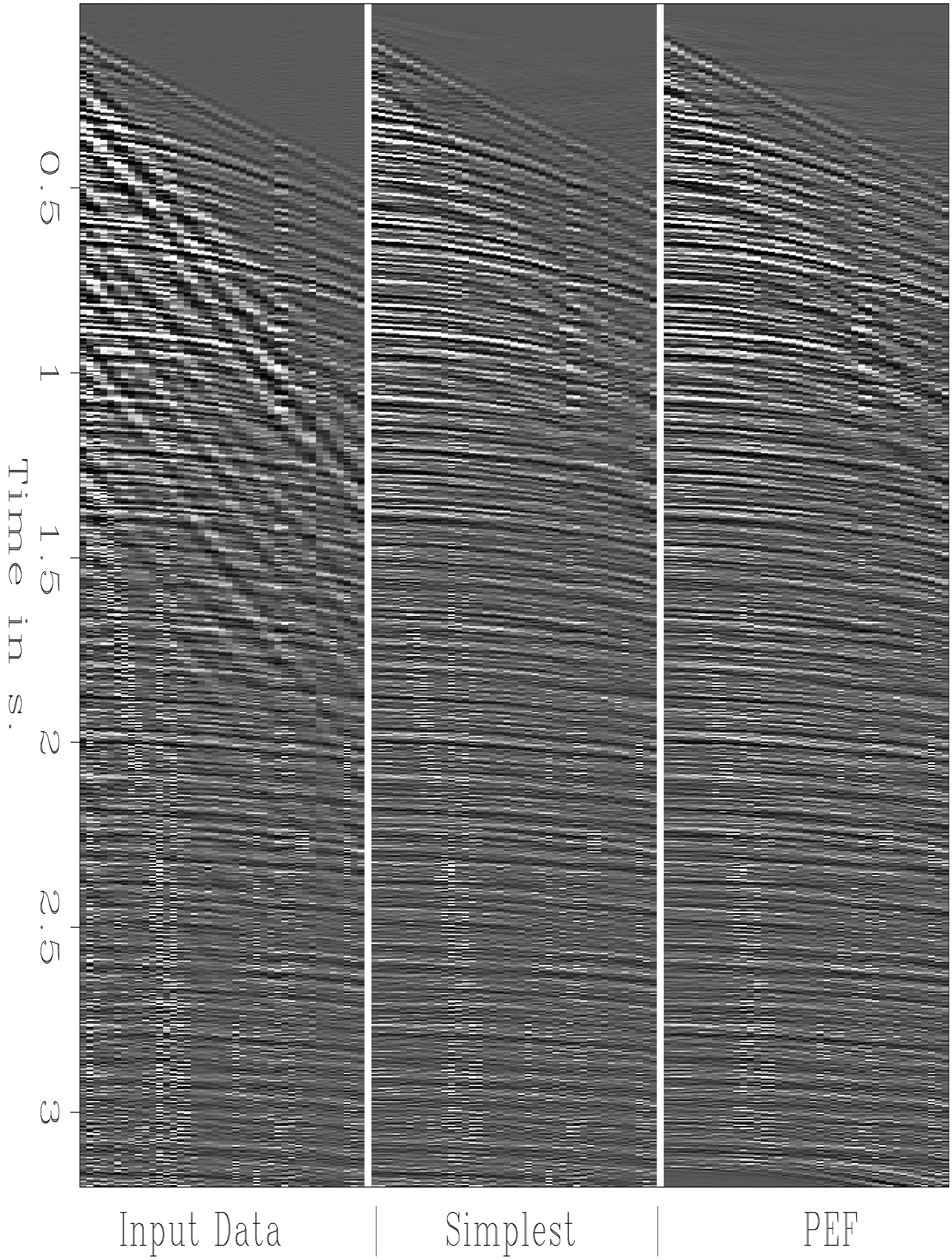

Figure 2: **Filtering method**. Input (left) and remodeled data after inversion. The maximum offset is 2.2km. Middle: fitting goal of Equation 8. Right: fitting goal of Equation 10 antoine1-compdatF [ER]

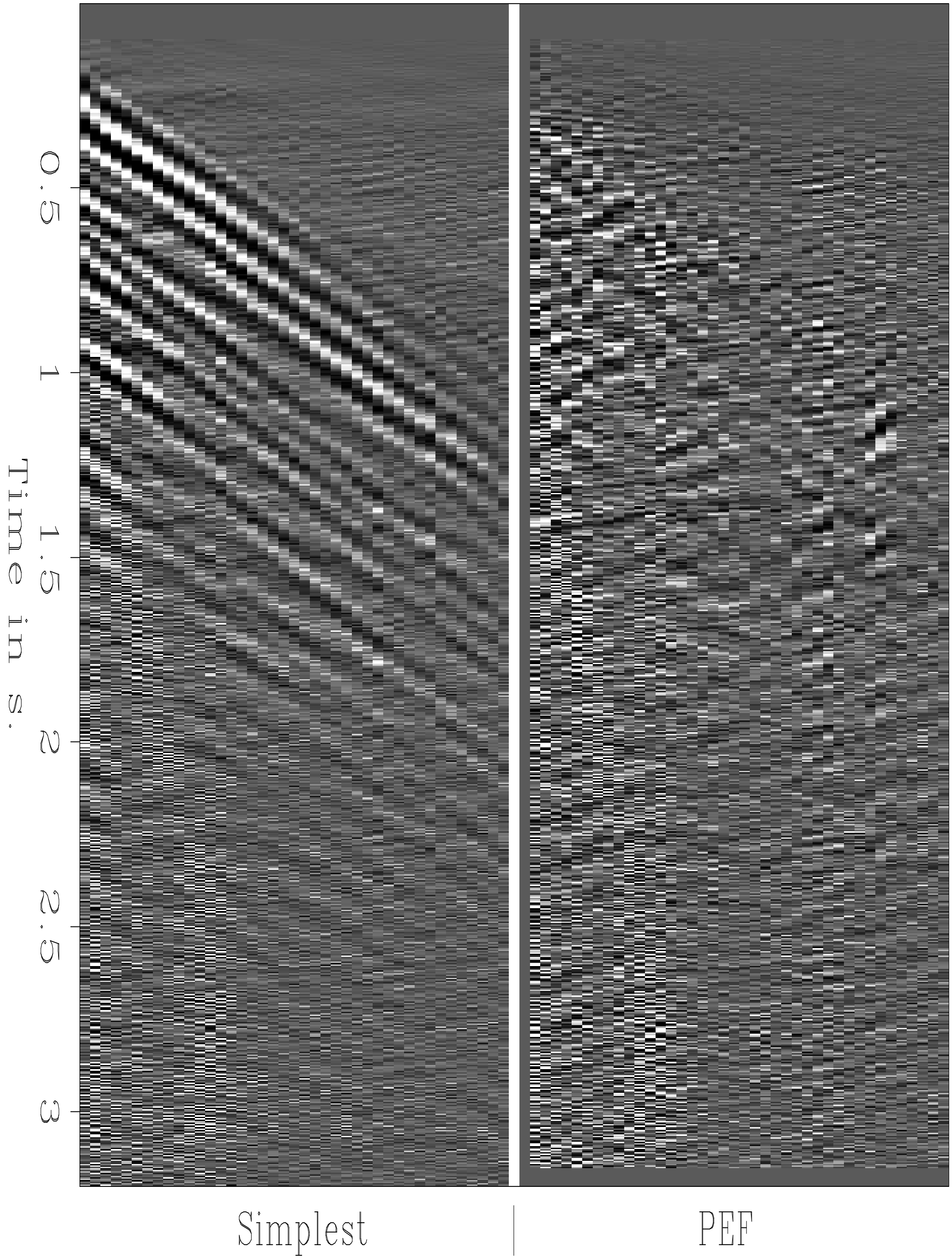

Figure 3: **Filtering method**. Residuals $\mathbf{r} = \tilde{\mathbf{d}} - \mathbf{d}$ after inversion. antoine1-compresF [ER]

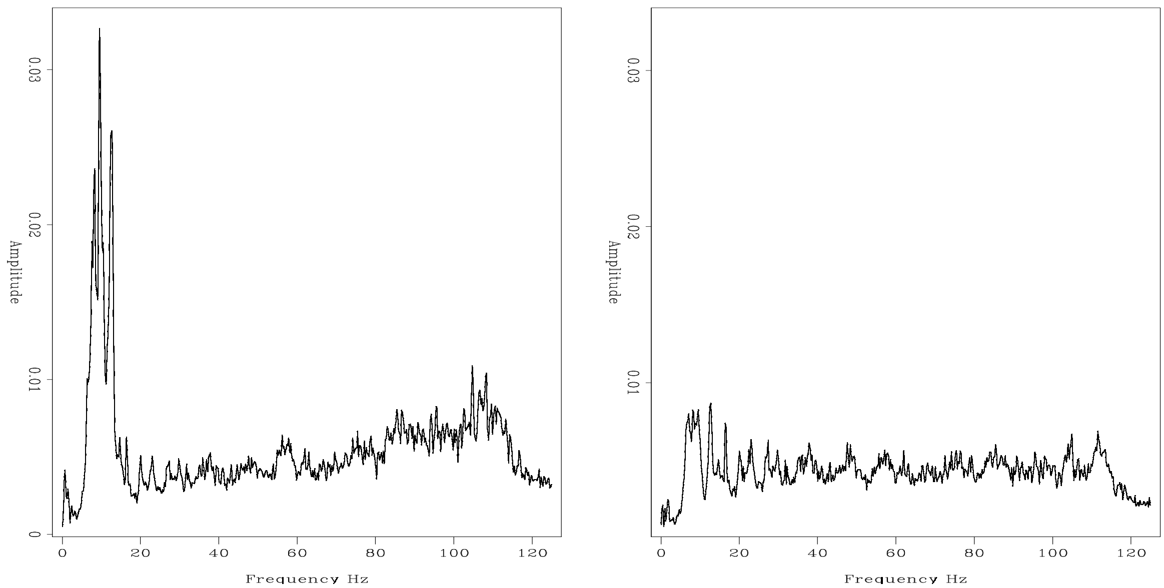

Figure 4: **Filtering method**. Spectrum of the "simplest" inversion residual with (left) and without PEF (right). antoine1-compsepcresF [ER]

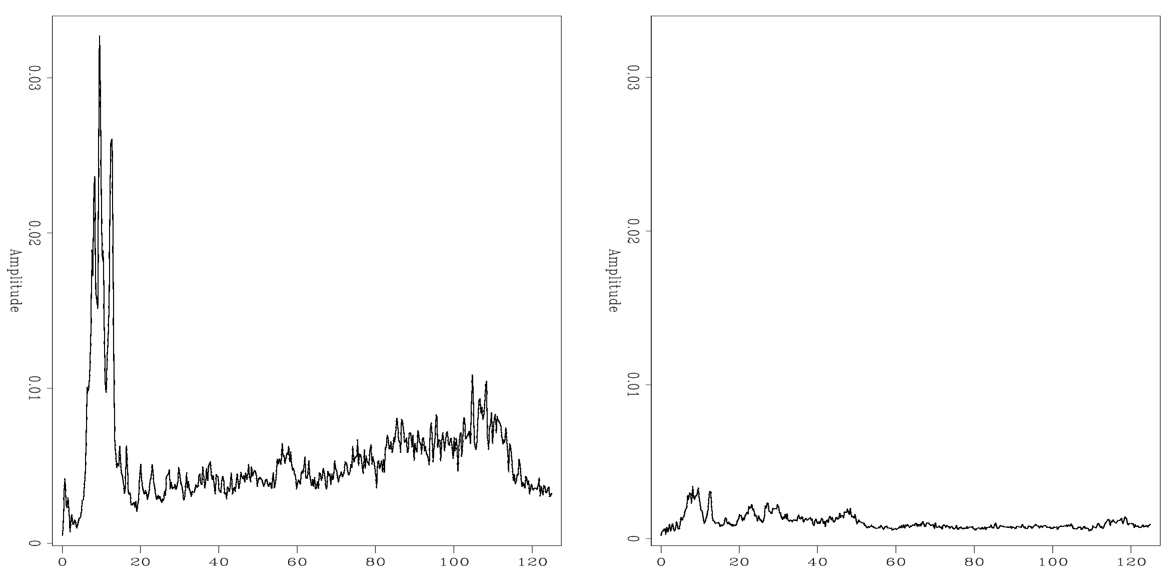

Figure 5: **Subtraction method**. Spectrum of the "simplest" inversion residual (left) and of the subtraction scheme residual (right). antoine1-compspecS [ER]

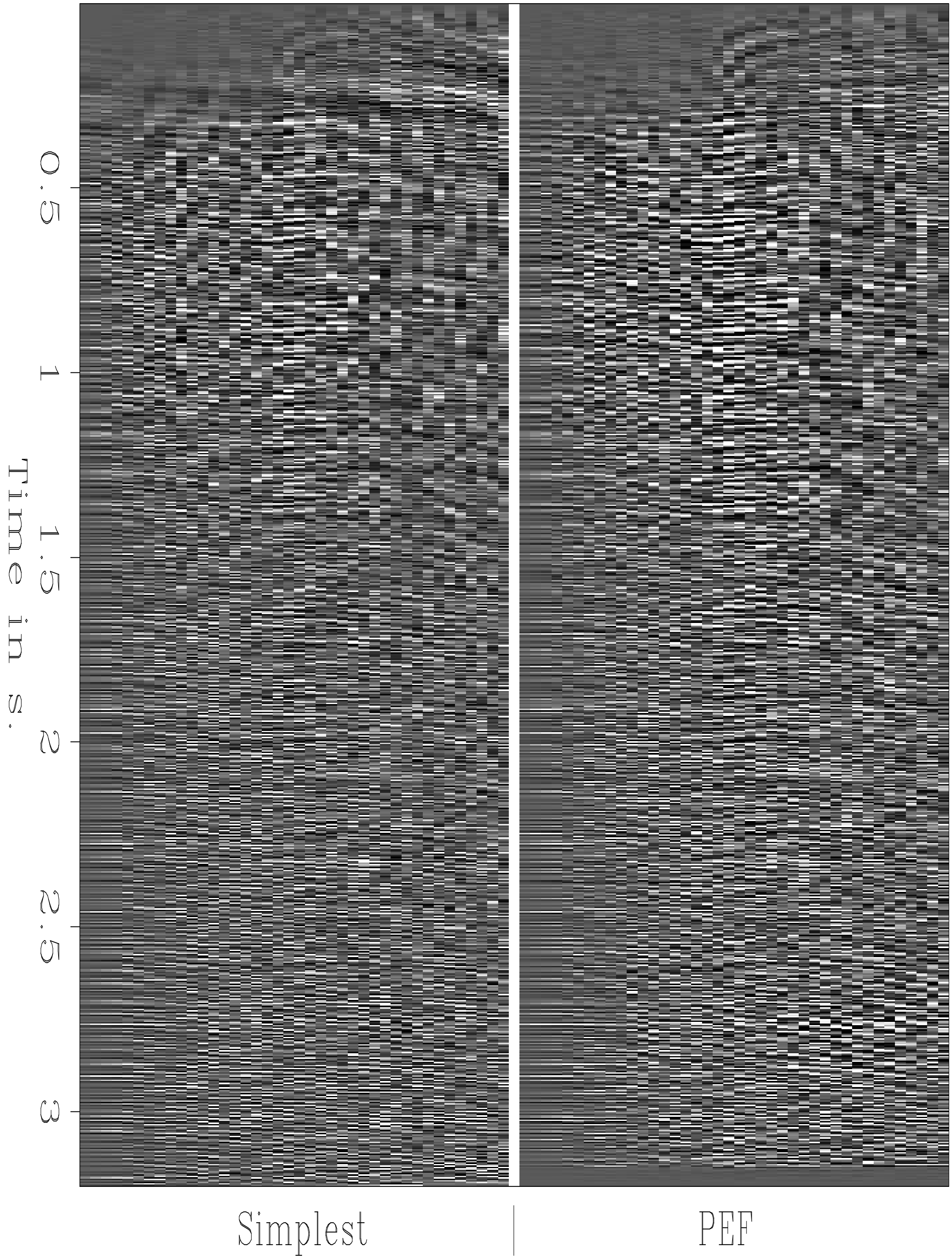

Figure 6: **Filtering method**. Velocity domain antoine1-compmodF [ER]

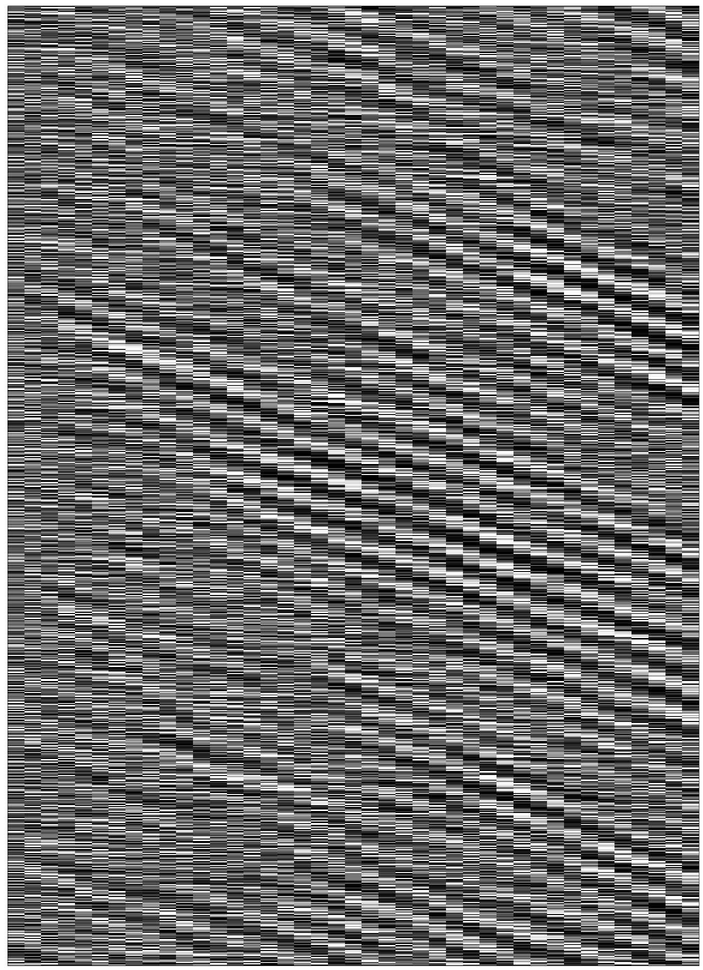

Figure 7: **Filtering method**. Convolution of noise with one of the inverse PEF estimated during the iterations. The coherent noise appears (dipping events). antoine1-impulseF [ER]

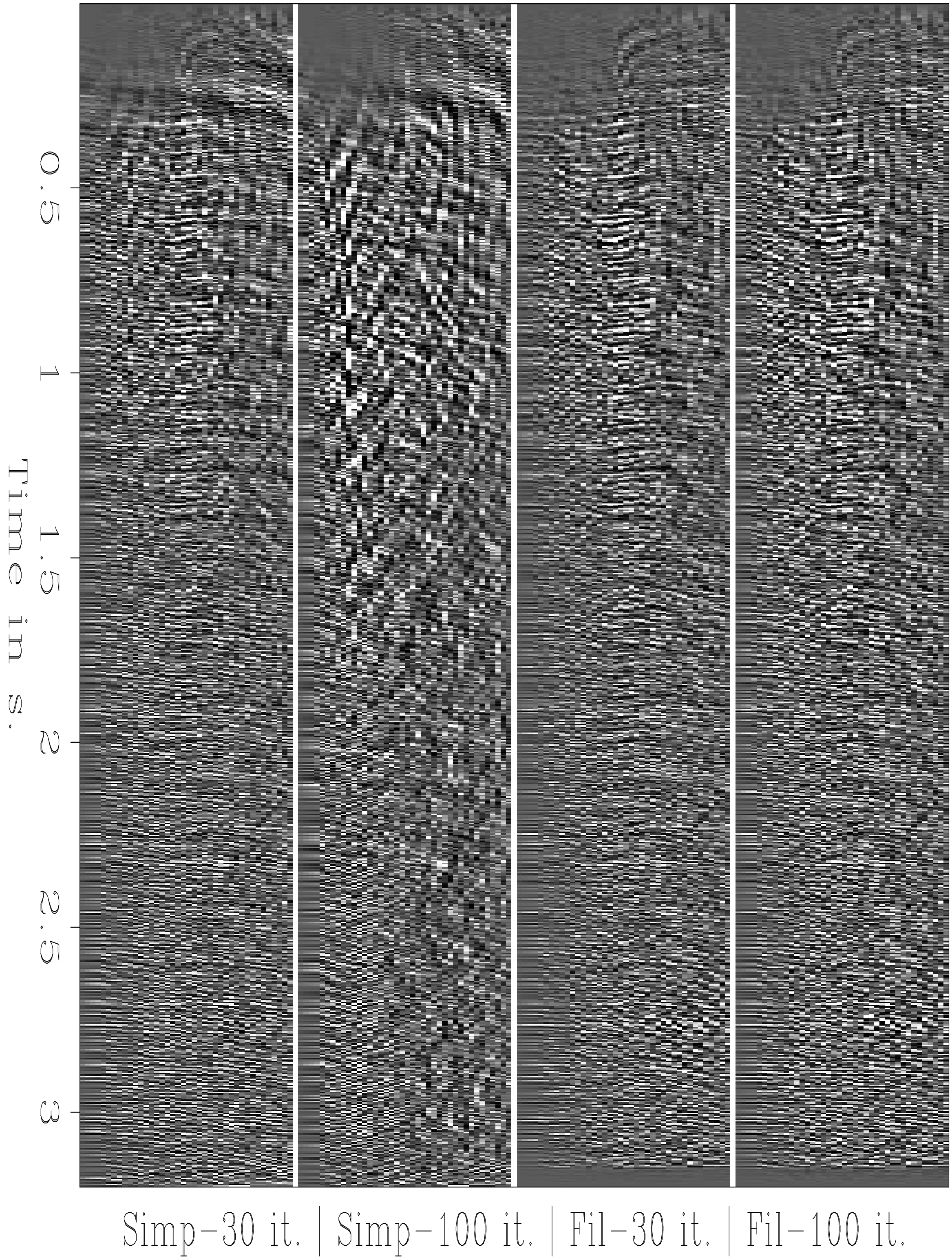

Figure 8: **Filtering method**. Stability of the filtering scheme (the two right panels) as opposed to the stability of the “simplest” approach (the two left panels) to the number of iterations. antoine1-compstabmod [ER]

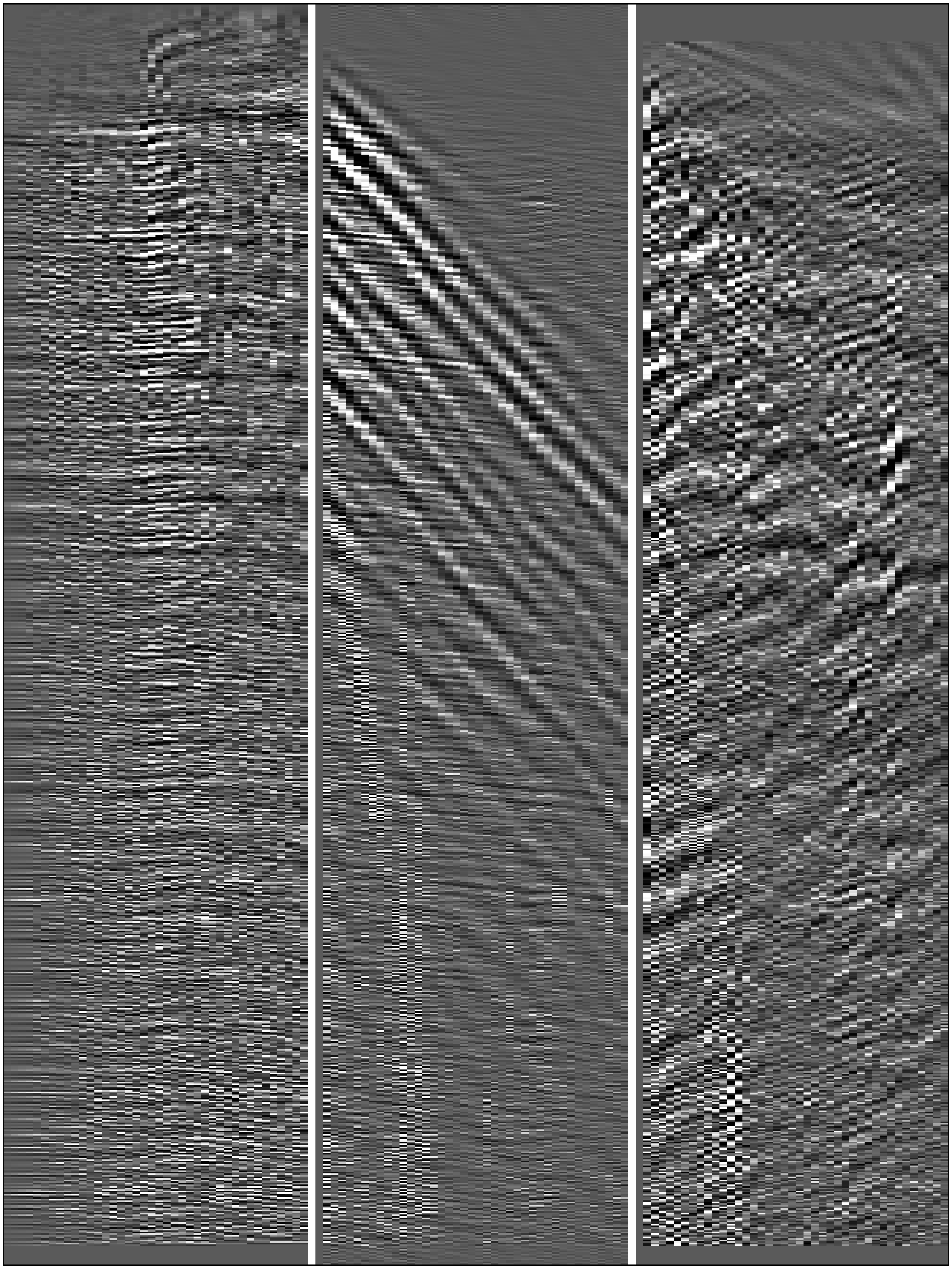

Figure 9: **Subtraction method**. Left: Model Space $\mathbf{m_s}$. Middle: Modeled noise $\mathbf{A_n^{-1} m_n}$. Some signal is trapped in the coherent noise due to crosstalks between **H** and the coherent noise PEF $\mathbf{A_n}$. Right: Data residual $\mathbf{r} = \tilde{\mathbf{d}} - \mathbf{d}$. antoine1-compevS [ER]

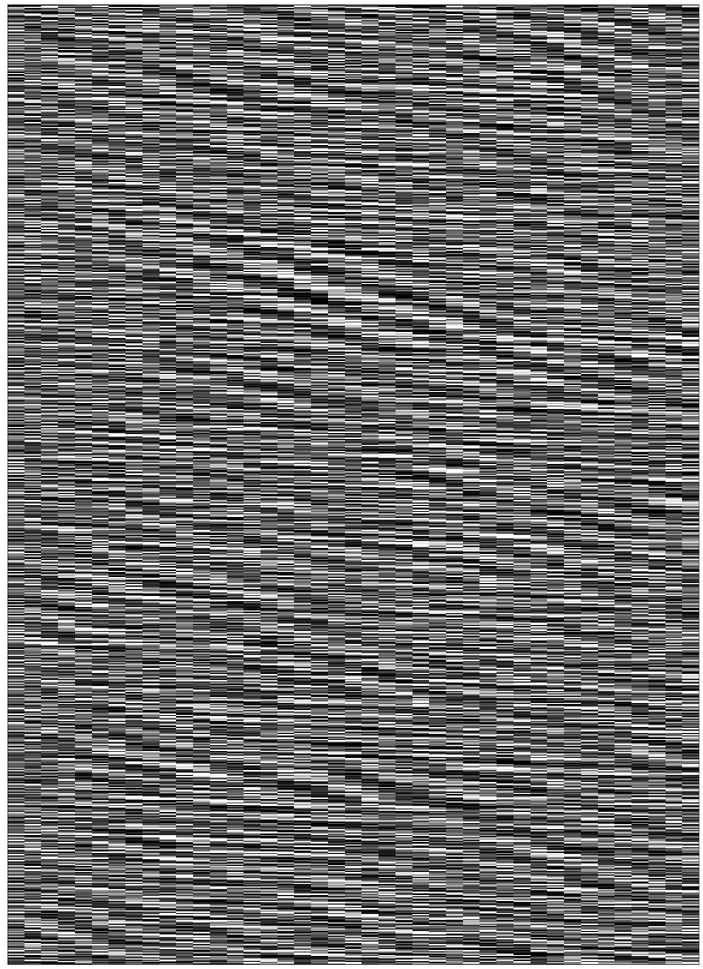

Figure 10: **Subtraction method**. Convolution of noise with the inverse PEF estimated from the data and used as the coherent noise PEF. Notice that both signal (straight lines) and noise (dipping events) are predictable, causing crosstalks with the hyperbolic Radon transform. antoine1-impulseS [ER]

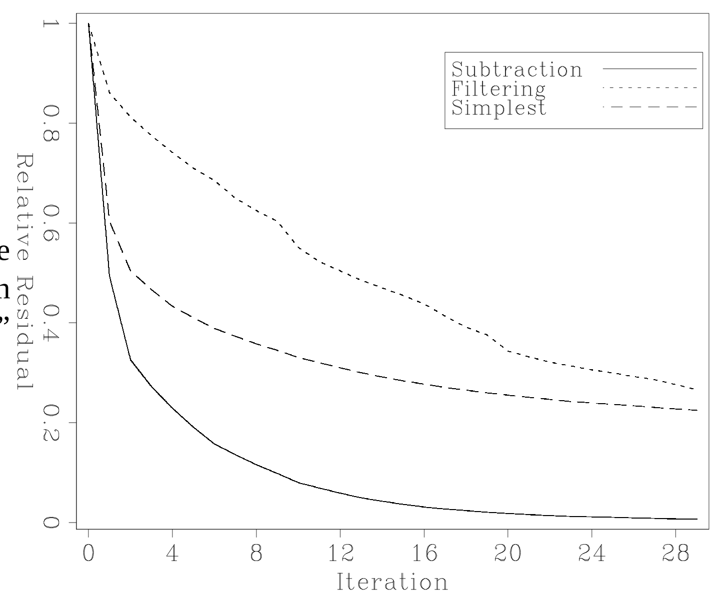

Figure 11: Convergence of the two proposed methods along with the convergence of the "simplest" scheme. antoine1-compiterS [ER]

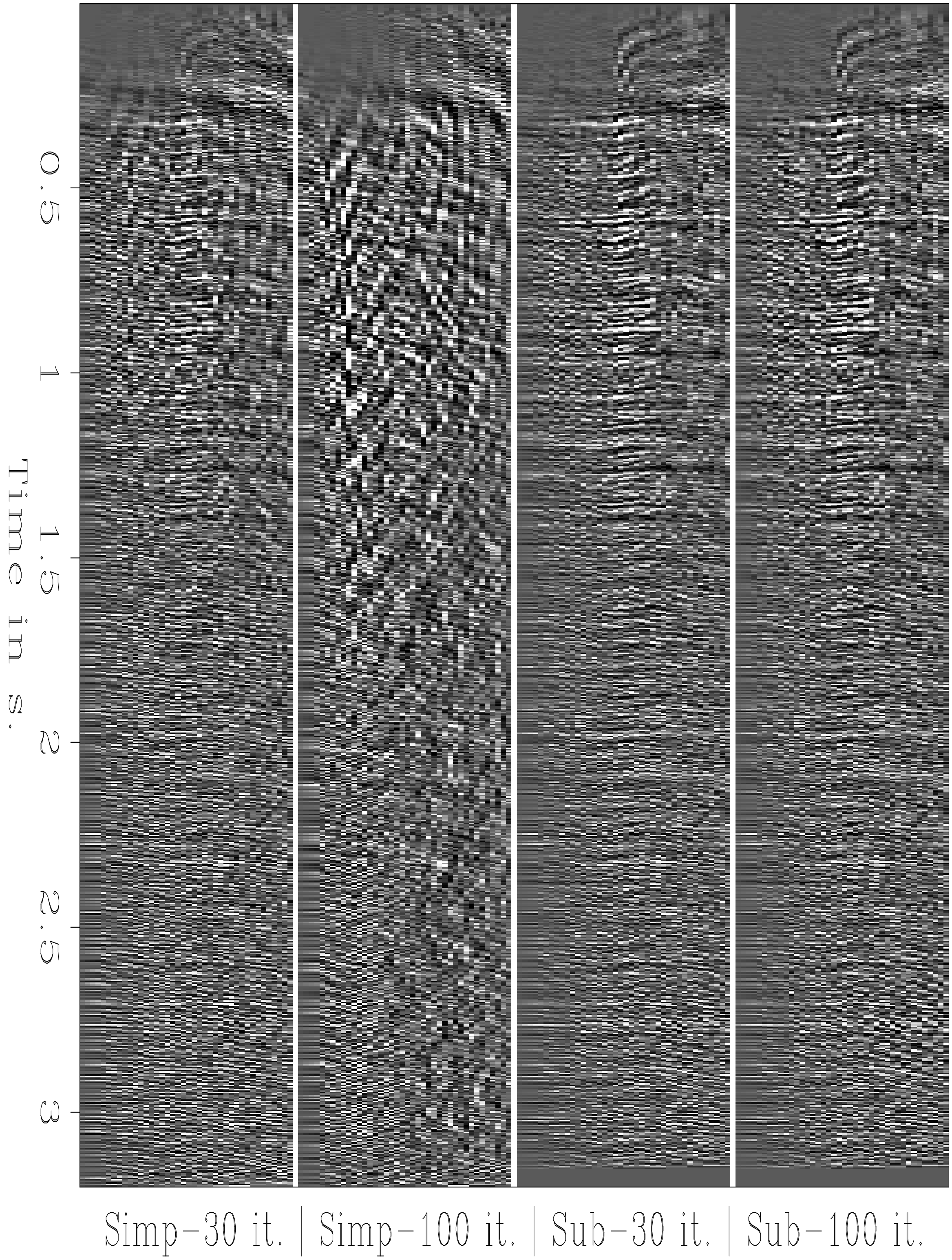

Figure 12: **Subtraction method**. Stability of the subtraction scheme (the two right panels) as opposed to the stability of the "simplest" approach (the two left panels) to the number of iterations. antoine1-compstabmodSub [ER]

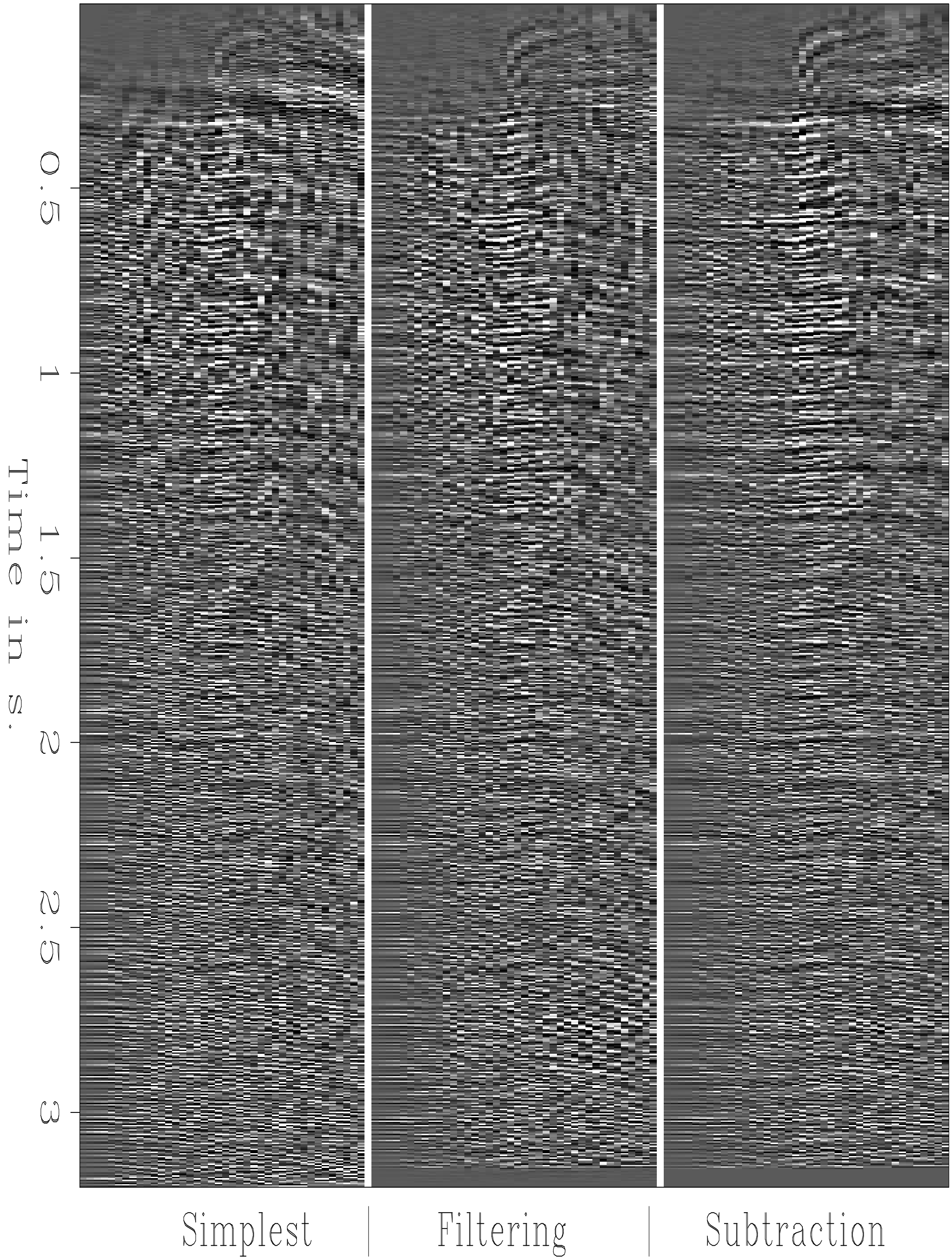

Figure 13: **Comparison study**. The two proposed schemes give a better velocity panel than the “simplest” inversion. antoine1-compmod [ER]