Decon in the log domain with variable gain

Jon Claerbout, Antoine Guitton, and Qiang Fu

ABSTRACT

We base deconvolution on the concept of output model sparsity. We improve our method of log spectral parameterization by including time-variable gain. Since filtering does not commute with time variable gain, gain is now done after decon (not before). Results at two survey locations confirm the utility. We resolve a stability issue with a long-needed regularization. An intriguing theoretical aspect shows that log spectral parameterization links penalty functions to crosscorrelation (not autocorrelation) statistics of outputs.

LOG SPACE, SPARSITY, AND GAIN

Because predictive decon fails on the Ricker wavelet, Zhang and Claerbout (2010) devised an extension to non-minimum phase wavelets (Zhang et al., 2011). Then (Claerbout et al. (2011)) replaced the traditional unknown filter coefficients by lag coefficients u_t in the log spectrum of the deconvolution filter. Given data $D(\omega)$, the deconvolved output is

$$r_t \;=\; \mathrm{FT}^{-1}\left[D(\omega)\ \exp\left(\sum_t u_t Z^t \right) \right] \tag{1}$$

where $Z = e^{i\omega}$. The log variables u_t transform the linear least squares (ℓ_2) problem to a non-linear one that requires iteration. Losing the linearity is potentially a big loss, but we lost that at the outset when we first realized we needed to deal with the non-minimum phase Ricker wavelet. We find convergence is typically quite rapid.

The source wavelet, inverse to the decon filter above, corresponds to $-u_t$. The positive lag coefficients in u_t correspond to a causal minimum phase wavelet. The negative lag coefficients correspond to an anticausal filter.

Here for the first time we introduce the complication that seismic data is non-stationary requiring a time variable gain g_t. The deconvolved data is the residual r_t. The gained residual $q_t = g_t r_t$ is "sparsified" (Li et al., 2012) by minimizing $\sum_t H(q_t)$ where

$$q_t \;=\; g_t\ r_t \tag{2}$$

$$H(q_t) \;=\; \sqrt{q_t^2+1} - 1 \tag{3}$$

$$\frac{dH}{dq} \;=\; H'(q) \;=\; \frac{q}{\sqrt{q^2+1}} \;=\; \mathrm{softclip}(q) \tag{4}$$

Traditional decon approaches are equivalent to chosing a white spectral output. Here we opt for a sparse output. In practice they might be much the same, but they do differ. Consider low frequencies. A goal is integrating reflectivity to yield impedance. We wish to restore low frequencies where they enhance sparsity, but not where they merely amplify noise.

Our prefered penalty function $H(q)$ used for finding u_t is the hyperbolic (or hybrid) penalty function (equation (3)). The output q_t best senses sparsity when gain is such that the typical penalty $H(q_t)$ value is found near the transition level between ℓ_1 and ℓ_2 norms, namely, when typical $|q_t| \approx 1$.

MINIMUM PHASE EXTENSION

A minimum phase wavelet can be made from any causal wavelet by taking it to Fourier space, and exponentiating. The proof is straightforward: Let $U(Z) = 1+u_1Z+u_2Z^2+\cdots$ be the Z transform ($Z = e^{i\omega}$) of any causal function u_t. Consider $e^{U(Z)}$. Although we would always do this calculation in the Fourier domain, the easy proof is in the time domain. The power series for an exponential $e^U = 1+U+U^2/2!+U^3/3!+\cdots$ has no powers of $1/Z$ (because U has no such powers), and it always converges because of the powerful influence of the denominator factorials. Likewise e^{-U}, the inverse of e^U, always converges and is causal. Thus both the filter and its inverse are causal. This is the essense of minimum phase.

We seek to find two functions, one strictly causal the other strictly anticausal.

$$U^+ = u_1 Z + u_2 Z^2 + \cdots \tag{5}$$

$$U^- = u_{-1}/Z + u_{-2}/Z^2 + \cdots \tag{6}$$

Notice U, U^2, etc do not contain Z^0. Thus the coefficient of Z^0 in $e^U = 1 + U + U^2/2! + \cdots$ is unity. Thus $a_0 = b_0 = 1$.

$$e^{U^+} = A = 1 + a_1 Z + a_2 Z^2 + \cdots \tag{7}$$

$$e^{U^-} = B = 1 + b_1/Z + b_2/Z^2 + \cdots \tag{8}$$

Define $U = U^- + U^+$. The decon filter is $AB = e^U$ and the source waveform is its inverse e^{-U}.

Consider $U(\omega) = \ln AB$ the log spectrum of the filter. We will be adjusting the various u_t, all of them but not u_0 which is the average of the log spectrum. The other u_t cannot change the average; they merely cause the log spectrum to oscillate.

THE GRADIENT

Having data d_t, having chosen gain g_t, and having a starting log filter, say $u_t = 0$, let us see how to update u_t to find a gained output $q_t = g_t r_t$ with better hyperbolicity.

Our forward modeling operation with model parameters u_t acting upon data d_t (in the Fourier domain $D(Z)$ where $Z = e^{i\omega}$) produces deconvolved data r_t (the residual).

$$r_t = \text{FT}^{-1}\ D(Z)\ e^{\cdots+u_2Z^2+u_3Z^3+u_4Z^4+\cdots} \tag{9}$$

$$\frac{dr_t}{du_\tau} = \text{FT}^{-1}\ D(Z)\ Z^\tau e^{\cdots+u_2Z^2+u_3Z^3+u_4Z^4+\cdots} \tag{10}$$

$$\frac{dr_t}{du_\tau} = r_{t+\tau} \tag{11}$$

This follows because Z^τ shifts the data $D(Z)$ by τ units which shifts the residual the same amount. Output formerly at time t moves to time $t+\tau$. This is not the familiar result that the derivative of an output with respect to a filter coefficient at lag τ is the shifted *input* $d_{t+\tau}$. Here we have the *output* $r_{t+\tau}$. This difference leads to remarkable consequences below.

It is the gained residual $q_t = g_t r_t$ that we are trying to sparsify. So we need its derivative by the model parameters u_τ.

$$q_t = g_t\ r_t = r_t\ g_t \tag{12}$$

$$\frac{dq_t}{du_\tau} = \frac{dr_t}{du_\tau}\ g_t = r_{t+\tau}\ g_t \tag{13}$$

Recall $u_0 = 0$ and hence $\Delta u_0 = 0$. To find the update direction at nonzero lags $\Delta\mathbf{u} = (\Delta u_t)$ take the derivative of the hyperbolic penalty function $\sum_t H(q_t)$ by u_τ.

$$\Delta\mathbf{u} = \sum_t \frac{dH(q_t)}{du_\tau} \qquad \tau \neq 0 \tag{14}$$

$$= \sum_t \frac{dq_t}{du_\tau}\ \frac{dH(q_t)}{dq_t} \tag{15}$$

$$\Delta\mathbf{u} = \sum_t (r_{t+\tau})\quad (g_t H'(q_t)) \qquad \tau \neq 0 \tag{16}$$

This says to crosscorrelate the physical residual r_t with the statistical residual $g_t H'(q_t)$. Notice in reflection seismology the physical residual r_t generally decreases with time while the gain g_t generally increases to keep the statistical variable q_t roughly constant, so $g_t H'(q_t)$ grows in time(!)

In the frequency domain the crosscorrelation (16) is:

$$\Delta U = \overline{\text{FT}(r_t)}\ \ \text{FT}(g_t\,\text{softclip}(q_t)) \tag{17}$$

Equation (17) is wrong at $t = 0$. It should be brought into the time domain and have Δu_0 set to zero. More simply, the mean can be removed in the Fourier domain.

Causal least squares theory in a stationary world says the signal output r_t is white (Claerbout, 2009); the *auto*correlation of the signal output is a delta function. Noncausal sparseness theory (other penalty functions) in a world of echoes (nonstationary gain) says the *cross*correlation of the signal output with its gained softclip is also a delta function (equation (16), upon convergence).

TAKING THE STEP

We adopt the convention that components of a vector $\mathbf{u}$ range over the values of (u_t), likewise for other vectors. Given the gradient direction $\Delta\mathbf{u}$ we need to know the residual change $\Delta\mathbf{r}$ and a distance α to go: $\alpha\Delta\mathbf{r}$ and $\alpha\Delta\mathbf{u}$.

A two-term example demonstrates a required linearization.

$$e^{\alpha\Delta U} = e^{\alpha(\Delta u_1 Z+\Delta u_2 Z^2)} \tag{18}$$
$$e^{\alpha\Delta U} = 1+\alpha(\Delta u_1 Z+\Delta u_2 Z^2)+\alpha^2(\cdots) \tag{19}$$
$$\mathrm{FT}^{-1}\, e^{\alpha\Delta U} = (1,\alpha\Delta u_1,\alpha\Delta u_2)+\alpha^2(\cdots) \tag{20}$$
$$\mathrm{FT}^{-1}\, e^{\alpha\Delta U} = (1,\alpha\Delta\mathbf{u})+\alpha^2(\cdots) \tag{21}$$

With that background, neglecting α^2, and knowing the gradient $\Delta\mathbf{u}$, let us work out the forward operator to find $\Delta\mathbf{q}$. Let "$*$" denote convolution.

$$\mathbf{r}+\alpha\Delta\mathbf{r} = \mathrm{FT}^{-1}(De^{U+\alpha\Delta U}) \tag{22}$$
$$= \mathrm{FT}^{-1}(De^{U}e^{\alpha\Delta U}) \tag{23}$$
$$= \mathrm{FT}^{-1}(De^{U}) * \mathrm{FT}^{-1}(e^{\alpha\Delta U}) \tag{24}$$
$$= \mathbf{r}*(1,\alpha\Delta\mathbf{u}) \tag{25}$$
$$= \mathbf{r}+\alpha\,\mathbf{r}*\Delta\mathbf{u} \tag{26}$$
$$\Delta\mathbf{r} = \mathbf{r}\ *\ \Delta\mathbf{u} \tag{27}$$
$$\Delta q_t = g_t\ \Delta r_t \tag{28}$$

It is pleasing that $\Delta\mathbf{r}$ is proportional to $\mathbf{r}$. This might mean we can deal with a wide dynamic range within r_t. The convolution, a physical process, occurs in the physical domain which is only later gained to the statistical domain q_t. Naturally, the convolution may be done as a product in the frequency domain.

To minimize $H(\mathbf{q}+\alpha\Delta\mathbf{q})$ express it as a Taylor series approximation to quadratic order. Minimizing yields

$$\alpha = -\sum_t \Delta q_t H'_t \ \ / \ \ \sum_t (\Delta q_t)^2 H''_t \tag{29}$$

Update $q_t = q_t+\alpha\Delta q_t$ and $U = U+\alpha\Delta U$, optionally (Newton method) iterate (because the locations of the many Taylor series have changed slightly with the change in $\mathbf{q}$).

ALGORITHM

Pseudo code below finds the best single filter for a group of seismograms. Notice $g(t,x)$ could contain mute patterns, etc.

Lower case letters are used for variables in time and space like `d` $= d(t,x)$, `r` $= r(t,x)$, `g` $= g(t,x)$, `q` $= q(t,x)$, `dq` $= \Delta q(t,x)$. while upper case for functions of frequency `D` $= D(\omega, x)$, `R` $= R(\omega, x)$, `dR` $= \Delta R(\omega, x)$, `U` $= U(\omega)$, `dU` $= \Delta U(\omega)$. Asterisk $*$ means multiply within an implied loop on t or ω.

```
D =   FT(d)
U = 0.            # or other initializations
Remove the mean from U(omega).
Iteration {
     dU = 0
     For all x
          r = IFT( D * exp(U))
          q = g * r
          dU = dU + conjg(FT(r)) * FT(g*softclip(q))
     Remove the mean from dU(omega)
     For all x
          dR = FT(r) * dU
          dq = g * IFT(dR)
     Newton iteration for finding alfa {
          H'  = softclip( q )
          H'' = 1/(1+q^2)^1.5
          alfa= - Sum( dq * H' ) / Sum( dq^2 * H'')
          q = q + alfa * dq
          U = U + alfa * dU
          }
     }
```

UNIQUENESS

As the figures show, our results are excellent, amazing even, but we've had a continuing problem with uniqueness. We find the pseudo-code presented here can spike any of the three lobes of the Ricker wavelet defining the sea floor. This is particularly annoying as it amounts to apparent time shifts and polarity changes. For about a year we ascribed this difficulty to nonlinear problems having many solutions, so we concentrated on controlling the descent. Now it looks like the problem is much simpler.

We now ascribe our uniqueness problem to a familiar problem in linear optimization. We believe we have what amounts to a null space. Tiny changes in initialization or other conditions lead to a wide variety of solutions.

For example, we often found by the third iteration we could see the spiking, and we could see the bubble estimation was well underway. By the tenth iteration it was pretty much settled down, and we would begin to be happy. But the computation was quick, so we were tempted to continue iterating. Maybe about the 150th iteration we would notice that spiking on the center of the Ricker wavelet would begin transition to spiking the first or third lobes of the Ricker wavelet (including the accompanying apparant polarity change). To make matters worse, only slight changes in the gain function g_t would determine the selection of which final lobe.

We wasted a lot of time believing nonlinearity was responsible for multiple solutions. Our early primative attempts at regularization had failed. With the pseudocode above you can have results like in this paper in a dozen iterations, however, the theory below explains the missing regularization that should allow you all the iterations you like.

REGULARIZATION

Regularization is where we impose our prior knowledge to account for the inadequacy of the data to completely define a solution. With years of experience we would look at the standard formulation $0 \approx \epsilon \sum_\tau w_\tau (u_\tau - \bar{u}_\tau)^2$ and theory would guide us to statistical averages to give us ϵ, w_τ and $\bar{u}_\tau$. We have recently understood that the weighting function w_τ should be a matrix $\mathbf{W}$, and we now know what that matrix should be. First, our goals:

1. The shot waveform should resemble a Ricker wavelet near zero lag.
2. The shot waveform should be small or vanishing at larger negative lags. The decon wavelet should not have a long low frequency precursor.

Theoretically, the even part of u_τ controls the amplitude spectrum of the shot waveform. (A parallel analysis is found elsewhere in this report (Claerbout (2012)).) We will not touch that. The phase spectrum is determined by the odd part of u_τ. The near zero lags in u_τ control the near zero lags in the shot waveform and decon filter. We want the odd near-zero lags to be symmetric because the Ricker wavelet is symmetric. Thus the regularization is to minimize the antisymmetric part of the near-zero lags in u_τ.

The larger positive lags in u_τ deal with marine bubble and soil layer reverberation. That is good stuff. Bad are the larger anticausal lags. They should be zero because they are non-physical. They can be handled as an additional regularization, or more simply by windowing Δu_τ.

Code modifications required by regularization

Consider regularization of the form $0 \approx u_\tau - u_{-\tau}$. In matrix form this is $\mathbf{0} \approx \mathbf{r}_m = \mathbf{J}\mathbf{u}$ where the matrix $\mathbf{J}$ is defined below with six vector components in the ordering required by the fast Fourier transform program.

$$
\mathbf{0} \;\approx\; \begin{bmatrix} r_m(1) \\ r_m(2) \\ r_m(3) \\ r_m(4) \\ r_m(5) \\ r_m(6) \end{bmatrix} \;=\; \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 & 0 & -1 \\ 0 & 0 & +1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & +1 & 0 \\ 0 & -1 & 0 & 0 & 0 & +1 \end{bmatrix} \begin{bmatrix} u(1) \\ u(2) \\ u(3) \\ u(4) \\ u(5) \\ u(6) \end{bmatrix} \tag{30}
$$

Note that $\mathbf{J}^* = \mathbf{J}$. The gradient search direction is

$$
\Delta \mathbf{u} \;=\; \frac{\partial\, \mathbf{r}_m^* \mathbf{W} \mathbf{r}_m}{\partial \mathbf{u}_m^*} \;=\; \frac{\partial \mathbf{r}_m^*}{\partial \mathbf{u}^*}\, \mathbf{W}\mathbf{r}_m \;=\; \mathbf{J}^* \mathbf{W} \mathbf{r}_m \tag{31}
$$

where $\mathbf{W}$ is a diagonal matrix of weights. Again for six components, the diagonal contains $(1, w_1, w_2, 0, w_2, w_1)$.

Here are the modifications needed to incorporate ℓ_2 regularization on u_τ:

$$
\begin{aligned}
\text{argmin}_{\mathbf{u}} \quad & \sum_{t,x} H(q_t) + \frac{\epsilon}{2}\, \mathbf{u}^* \mathbf{J}^* \mathbf{W} \mathbf{J} \mathbf{u} && (32) \\
\Delta u_t \;&=\; \text{as before } + \epsilon\, \mathbf{J}^* \mathbf{W} \mathbf{r}_m && (33) \\
\Delta r_t \;&=\; \text{as before} && (34) \\
\alpha \;&=\; - \frac{\left(\sum_{x,t} \Delta q_t H'_t\right) \;+\; \epsilon \;\; (\mathbf{r}_m \cdot \Delta \mathbf{r}_m)}{\left(\sum_{x,t} (\Delta q_t)^2 H''_t\right) \;+\; \epsilon\, (\Delta \mathbf{r}_m \cdot \Delta \mathbf{r}_m)} && (35)
\end{aligned}
$$

In a least squares problem we compute a step size α as minus a ratio $\mathbf{r} \cdot \Delta\mathbf{r}$ over $\Delta\mathbf{r} \cdot \Delta\mathbf{r}$. Adding a least squares regularization to any convex fitting problem we simply add $\epsilon\, (\mathbf{r}_m \cdot \Delta\mathbf{r}_m)$ to the numerator and $\epsilon\, (\Delta\mathbf{r}_m \cdot \Delta\mathbf{r}_m)$ to the denominator.

Actually, another regularization is desireable. We should also request u_τ to be small for large anticausal lags, lags more negative than the range we are considering for the antisymmetry regularization. This might be handled by truncating the gradient rather than as a regularization.

A third regularization can be added to weaken u_τ at its large positive lags in circumstances where we feel we have insufficient data to estimate trace-long filters.

GOALS

A long range goal is to successfully integrate the reflectivity to get the log impedance. This requires good low frequency handling. Recording equipment often suppresses low frequencies for various practical reasons whose validity is likely location dependent. Our decon is pulling back some of these low frequencies but should know to stop before pulling up noise. Figure 2 demonstrates doing gain after non-minimum phase decon makes a valuable first step. To find impedance may require the additional statistical assumption of sparseness, but by solving the physical problem correctly, we have reduced the need for that.

ACKNOWLEDGEMENT

We thank Western Geophysical for the Gulf of Mexico data set. Jon Claerbout and Qiang Fu thank the sponsors of the Stanford Exploration Project. Antoine Guitton thanks Repsol Sinopec Brasil SA and Geo Imaging Solucoes Tecnologicas em Geociencias Ltda. We'd like to thank Yang Zhang for continued interest.

REFERENCES

Claerbout, J., 2012, Polarity preserving decon in "N log N" time: SEP-Report, **147**, 305–312.

Claerbout, J., Q. Fu, and Y. Shen, 2011, A log spectral approach to bidirectional deconvolution: SEP-Report, **143**, 297–300.

Claerbout, J. F., 2009, Image estimation by example.

Li, Y., Y. Zhang, and J. Claerbout, 2012, Hyperbolic estimation of sparse models from erratic data: Geophysics, **77**, V1–V9.

Zhang, Y. and J. Claerbout, 2010, A new bidirectional deconvolution method that overcomes the minimum phase assumption: SEP-Report, **142**, 93–104.

Zhang, Y., J. Claerbout, and A. Guitton, 2011, A new bi-directional sparse/spike deconvolution method that overcomes the minimum phase assumption: 73th EAGE Conference and Exhibition Extended Abstract, F001.

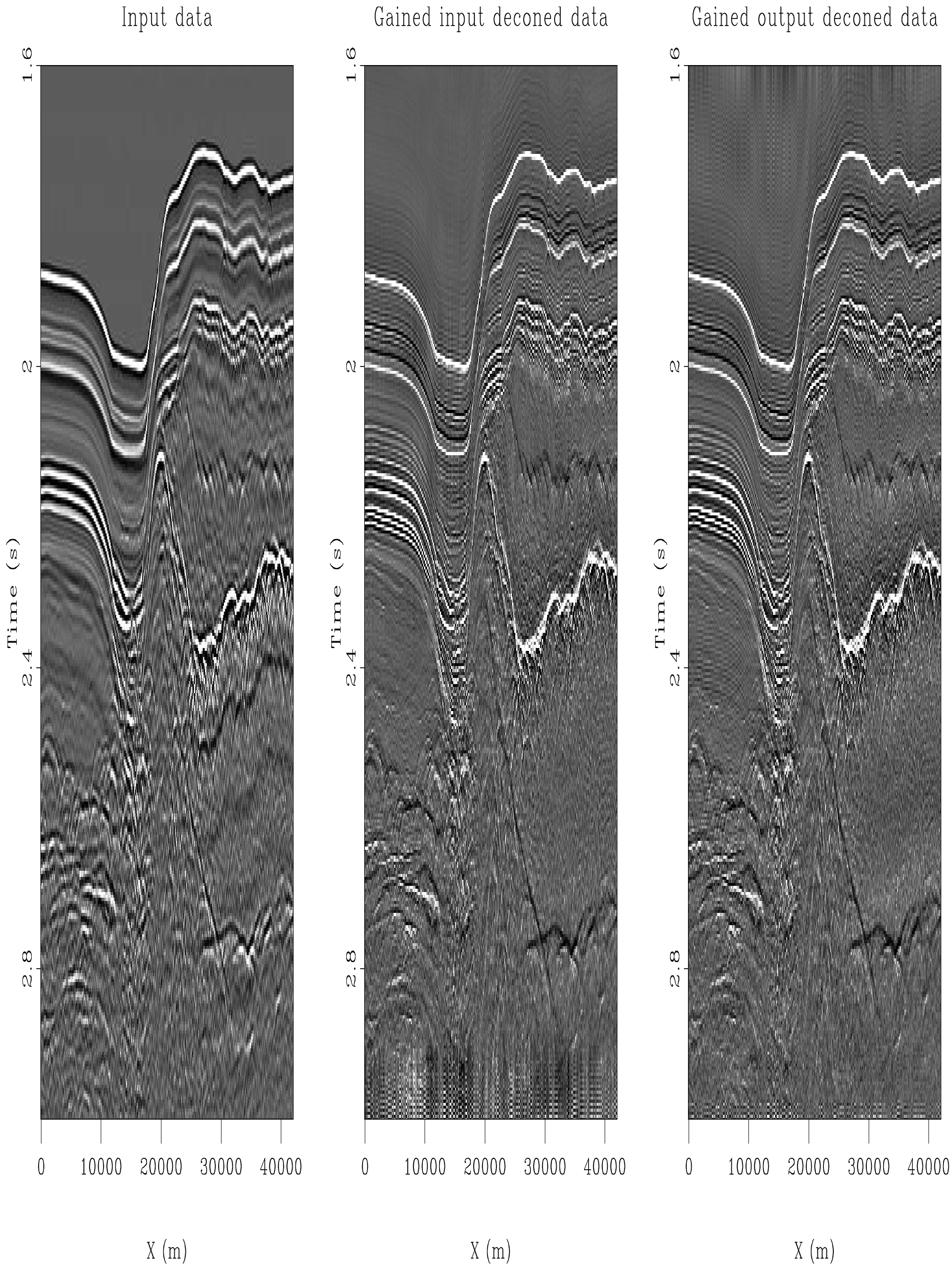

Figure 1: Gulf of Mexico. Decon produces plain white reflections from hard boundaries, and plain black boundaries from soft ones. WB= Water Bottom (white), TS= Top Salt (white), BS= Bottom Salt (black), ME= Mystery Event (black), soft reflector could be rugose salt solution of a former salt layer. [**ER**]

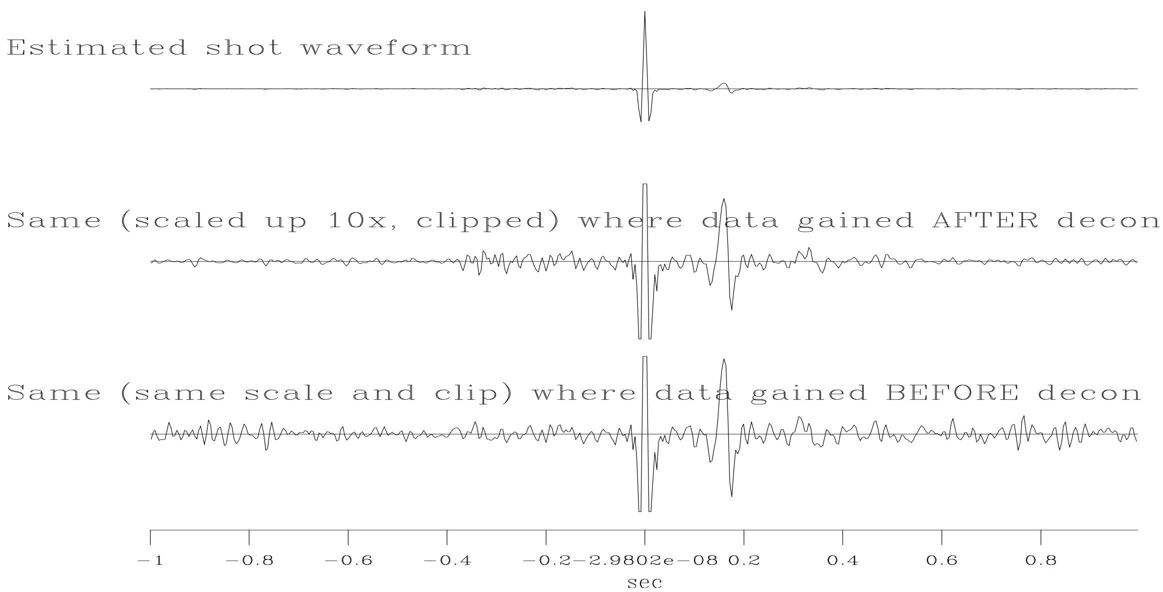

Figure 2: Wavelet from Cabo data. Although causality is not imposed, the estimated shot waveform is near causal (discounting the leading lobe of the Ricker wavelet). The importance of gain (here t^2) after deconvolution instead of before is shown by the lower two traces. There is much less noise when we gain AFTER decon, not BEFORE. Notice also that gain before decon estimates a slightly larger bubble (which is wrong). [**ER**]

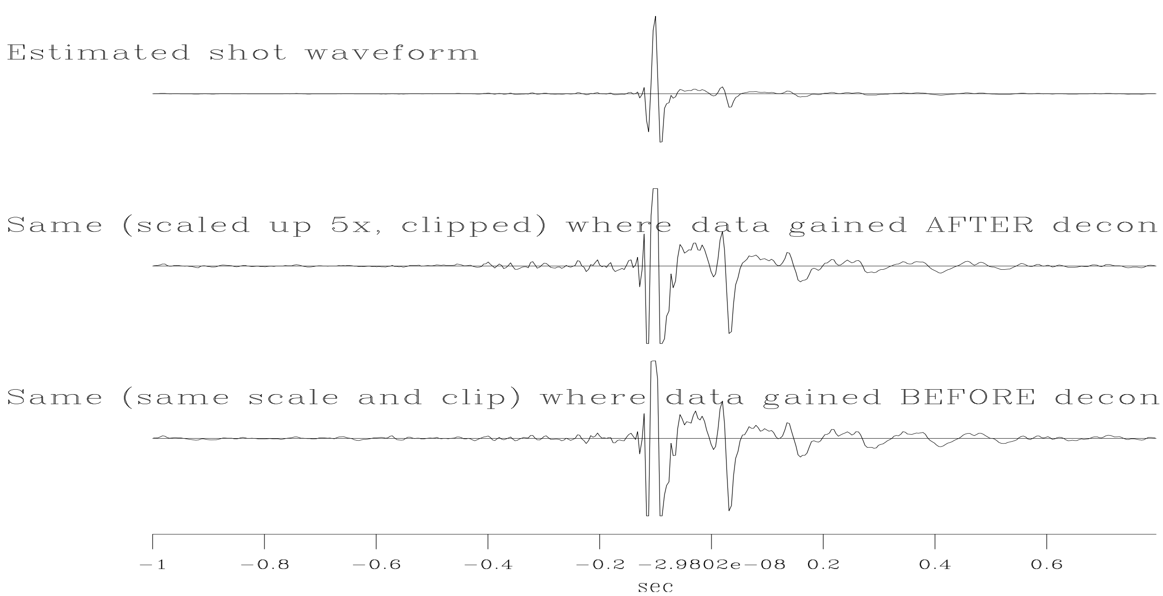

Figure 3: The Gulf of Mexico data produces a very different bubble but the same conclusions as Figure 2. The lack of symmetry in the Ricker wavelet may be related to the unresolved uniqueness issue. (Awaits better regularization.) [**ER**]

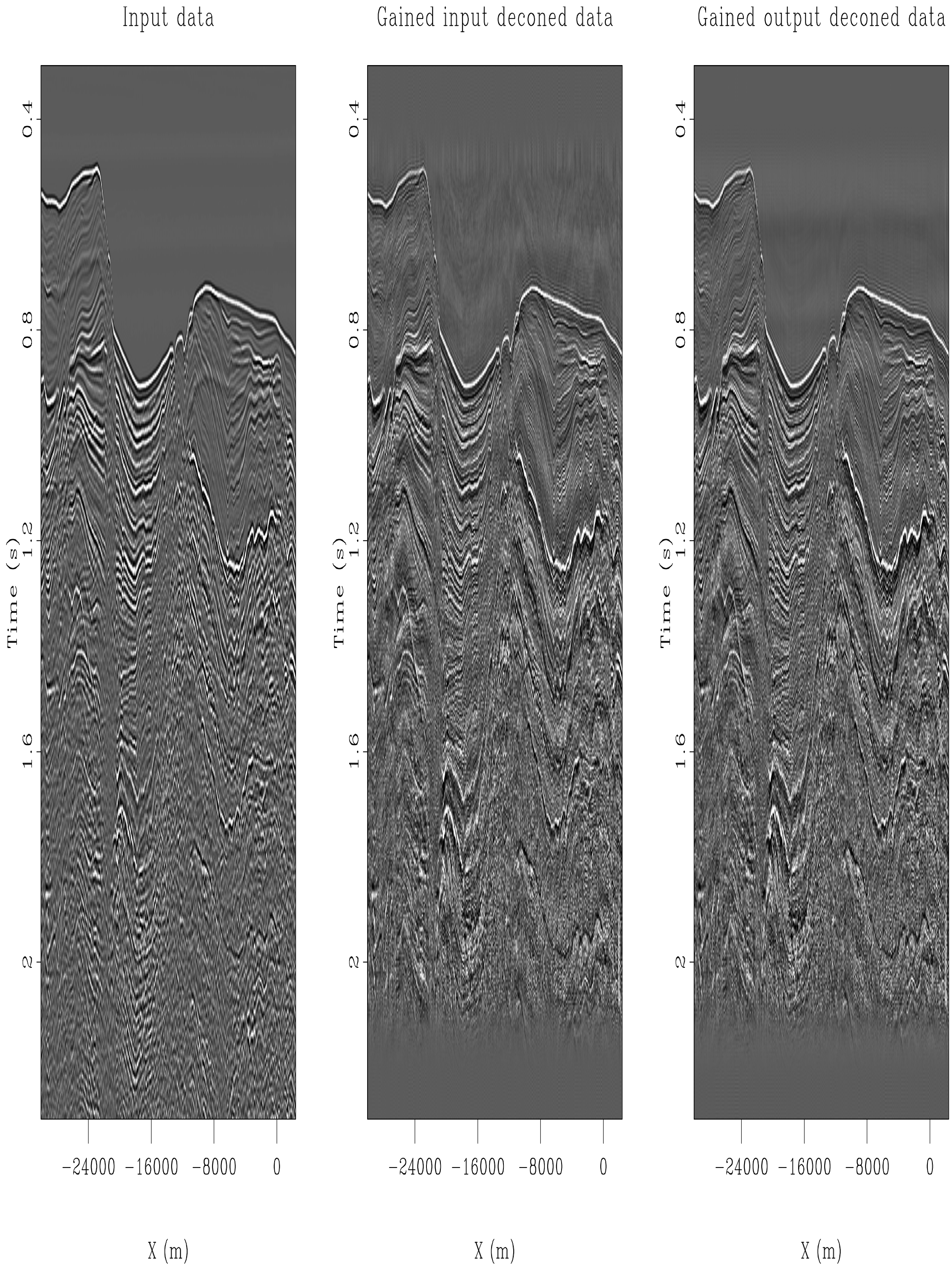

Figure 4: Cabo data. One filter on all traces. Bubble (at about 0.9s) removed. Enhanced high frequency at 1.1s. Gained-input method gave low frequency event precursors especially clear above the event at 1.2s but also visible above the water bottom. The problem is overcome by the gained-output method. (Guitton) [**ER**]