MINIMUM-ENTROPY DECONVOLUTION WITH THE EXTRINSIC POWER NORM

Lawrence C. Morley

Abstract

Synthetic and real examples of minimum-entropy deconvolution show that maximization of "extrinsic power" gives good results for the debubble problem. The optimization technique used is found to be both rapidly convergent and robust. These results give encouragement to planned applications to multiple suppression.

Introduction

Geophysicists like to see order in their data. This is not a naive attempt on their part to deny nature's chaotic tendencies. It simply reflects the fact that geology is well represented by relatively large homogeneous regions separated by narrow "catastrophic" boundaries.

Recently, this idea has found expression in the "minimum entropy" (ME) concept and has been applied to the debubbling of source signatures from marine data (Wiggins,1978;Claerbout, this report and SEP-15, p. 109-122) The term minimum entropy, as it has been used up until now in deconvolution, is a strictly generic one. It merely reflects a general desire to tailor the histogram of the data so as to increase large amplitudes and decrease small ones - thereby increasing the overall apparent order of the data. The usual approach calls for selecting an appropriate functional of the data and optimizing it via a linear filter in an iterative fashion. A most important (but not fully understood) constraint on this whole procedure is the length of the filter. In fact, a finite filter length is the only thing that keeps the algorithm from eventually carrying things to the extreme - extracting the maximum data value and zeroing everything else.

There is nothing sacred about the functional or inequality used. The fundamental entity in any ME algoritm is the gradient with respect to the data values. The form of this gradient determines whether or not the descent is robust.

A review of some of the methods proposed to date is presented in Table I. The ***varimax norm*** has a gradient proportional to x^3. The recognition that it might be advantageous to decrease small amplitudes as well as drive up big ones led to the development of the ***parsimony*** and ***bit count*** approaches.

Since bit count has a gradient that is unduly concerned with small values it is necessary to clamp the minimum absolute value of the data to some non-zero value - forcing the introduction of another arbitrary parameter. Both the parsimony and the extrinsic power norms avoid this problem by having a gradient which goes to zero with the data values.

Theory

Before presenting any results we will give a brief description of the algorithm. The reader is directed to Claerbout's ME essays in this report for a more complete discussion.

Our deconvolution model is x = y*f where y is the observed seismogram, f the inverse filter, and x the deconvolved trace. We choose the norm

$$S=\Sigma p_i \ln p_i-(\Sigma p_i)\ln \bar{p} \geq 0 \qquad (1)$$

where p_i represents the squared seismogram, x_i^2, or, alternatively, the envelope of x_i. This is termed the "extrinsic power" norm. Maximizing this parameter minimizes the degree of homogeneity in the data. The norm is zero iff all data amplitudes are identical. It has the additional property that if a trace is subdivided into time bins, the aggregate value of S for the whole trace will be the sum of all the S values associated with each bin.

In the lexicon of thermodynamics, any system parameter which behaves in this way (e.g. - volume) is termed an extrinsic parameter. This contrasts with "intrinsic" system parameters, such as pressure, which remain constant for each individual subsystem.

To maximize S we proceed by steepest descent, perturbing x by an amount

$$\delta x_j \;\alpha\; \frac{\partial S}{\partial x_j} = \sum_t \frac{\partial S}{\partial p_t} \frac{\partial p_t}{\partial x_j} \tag{2}$$

In matrix form we have

$$\delta x \;\alpha\; Gx \tag{3}$$

where

$$G = \mathrm{diag}\left[\frac{\partial S}{\partial p_k}\right] \tag{4}$$

We would, however, like to constrain δx so that energy is conserved. To first order this condition is

$$x^T \delta x = 0 \tag{5}$$

The first order change in S due to a change δf in f is

$$\delta S = S(x+\delta x) - S(x) = \sum_{jt}\sum \frac{\partial S}{\partial p_t} \frac{\partial p_t}{\partial x_j} \delta x_j \quad \alpha \quad \delta x^T Gx \tag{6}$$

We wish to maximize δS subject to (5). The appropriate Lagrangian is

$$L(\delta f, \lambda) = (\delta f)^T Y^T Gx + \lambda(\delta f)^T Y^T x \tag{7}$$

Y, in equation (7), is a matrix of shifted rows of y so that $x = Yf$. $\delta x = Y\delta f$ since $\delta Y = 0$. Zeroing the gradient of (7) with respect to δf and premultiplying by f^T gives

$$\lambda = -\langle x,Gx\rangle/\langle x,x\rangle \tag{8}$$

Thus, the constrained gradient of S with respect to the data values is $(G + \lambda I)x$ and our basic equation (3) becomes

$$Y\delta f = (G - I\langle x,Gx\rangle/\langle x,x\rangle)x \tag{9}$$

This equation is now solved in a least-squares sense for a finite length δf using standard Wiener-Levinson techniques.

In practical terms, the main difference between the following results and parsimony is not in the choice of gradient. The parsimony (S_2) gradient and the extrinsic power gradient are in fact identical. The main difference is in the selection of a descent magnitude. The magnitude of the descent in the following examples was chosen to be "$\alpha\ \delta f$." α was initially set to 1 and changed each iteration by a factor between 0.5 and 1.5 according to the formula

$$\alpha_{new} = \alpha_{old}(N + 2signag(dx,dx^-))/2N \tag{10}$$

$signag(dx,dx^-)$ is the number of sign agreements (Hamming distance) between the new and old constrained gradients, and N is the number of time points in dx.

Results

Figure 1 contains the results of ME deconvolution of a synthetic with both 0 and 20% peak-to-peak noise. The noisy descent appears to be remarkably robust. It was necessary to clamp the diagonal of the data autocorrelation matrix (Y) by .1% to 1% before performing the noisy division $YY^T\delta f = Y^T(G + \lambda I)X$.

The next example (figure 2) is of the debubbling of a USGS marine dataset. Figures (2a) and (2b) are the original and predictive deconvolved data, respectively. (2c) and (2d) are both ME deconvolved with the algorithm described above, but in (2d) a multiplicative ramp was applied to the first breaks to attenuate the strong high- frequency component on the middle traces. There is a remarkable difference between (b) and (d) in terms of the number and simplicity of up-shallow events. On the initial data we can see high frequencies on the first breaks of the centre traces which do not appear to fit the convolutional model. It is generally recognized that ME deconvolution is most sensitive to (high-amplitude) first arrivals. In most cases this is good because the first arrivals have a high signal/noise ratio. In this case it is a decided drawback.

As figure 2c demonstrated, MED can be expected to run into trouble whenever high frequency and high amplitude are well correlated. In fact, any data with sharp spectral peaks will only have those peaks reinforced by ME deconvolution if some pre-whitening is not first performed. The Gulf model tank data (figure 3c) is a good example of this. Disaster could only be avoided by first running predictive decon (b) on the original data (a). The data of figure 3b were then input to the ME decon program described above to obtain figure 3c. One problem in doing this deconvolution was the selection of a good length for the inverse filters. This is a problem common to both predictive and ME deconvolution. In both deconvolutions good results were only obtained with relatively short filters (1/2 timing line in length).

Conclusion

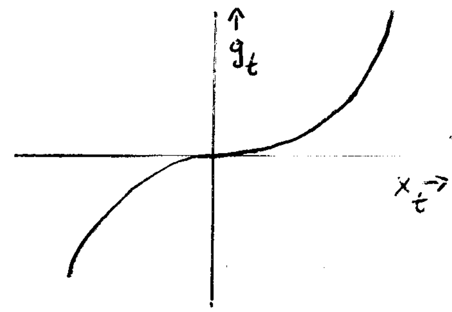
These results indicate that the extrinsic power inequality provides a useful gradient for the minimum-entropy debubble problem. An effort is now being made to apply this to marine multiple suppression using the Noah's model - following up on a suggestion by Claerbout (this report). Unfortunately, conclusive results were not available at report time.

REFERENCE

Wiggins,R.A.,1978,Minimum entropy deconvolution, Geoexploration,vol.16,p.21-35.

VARIMAX:

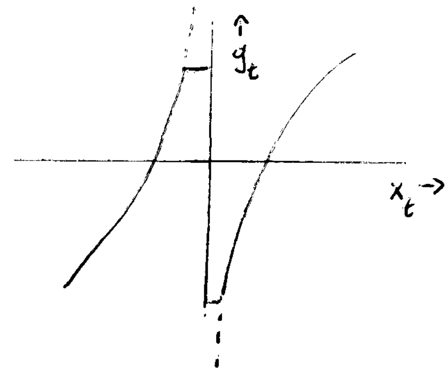
$$\max V = \frac{\Sigma x_i^{\,4}}{(\Sigma x_i^{\,2})^2} \geq 1$$



BITCOUNT:

$$\min S = \frac{1}{N}\Sigma \ln p_i - \ln\left(\frac{1}{N}\Sigma\, p_i\right) \leq 0$$

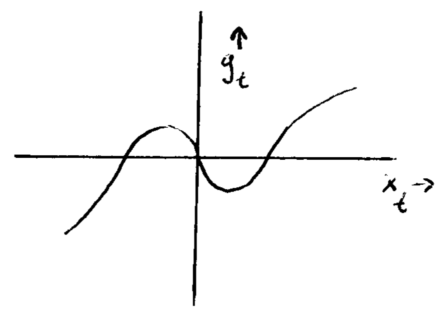
$$(p_i = x_i^{\,2})$$



PARSIMONY:

$$\min S_n = \ln \Sigma|x|^n - \frac{\Sigma x^n \ln|x|^n}{\Sigma|x|^n} \leq 0$$

$$\frac{\partial S_n}{\partial x_i} = \frac{n|x|^{n-1}\operatorname{sgn}(x_i)}{\Sigma|x|^n}\left[\frac{\Sigma|x|^n \ln\,|x|^n}{\Sigma|x|^n} - \ln|x|^n\right]$$



EXTRINSIC POWER:

$$\max\left[\Sigma p \ln p - (\Sigma p_i)\ln\left(\frac{1}{N}\Sigma p\right)\right] \geq 0$$

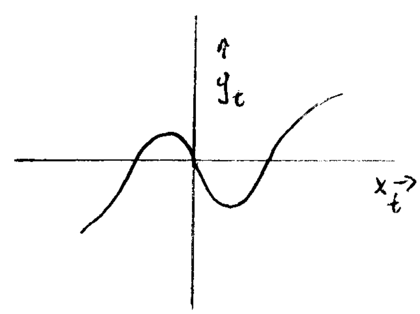


TABLE 1

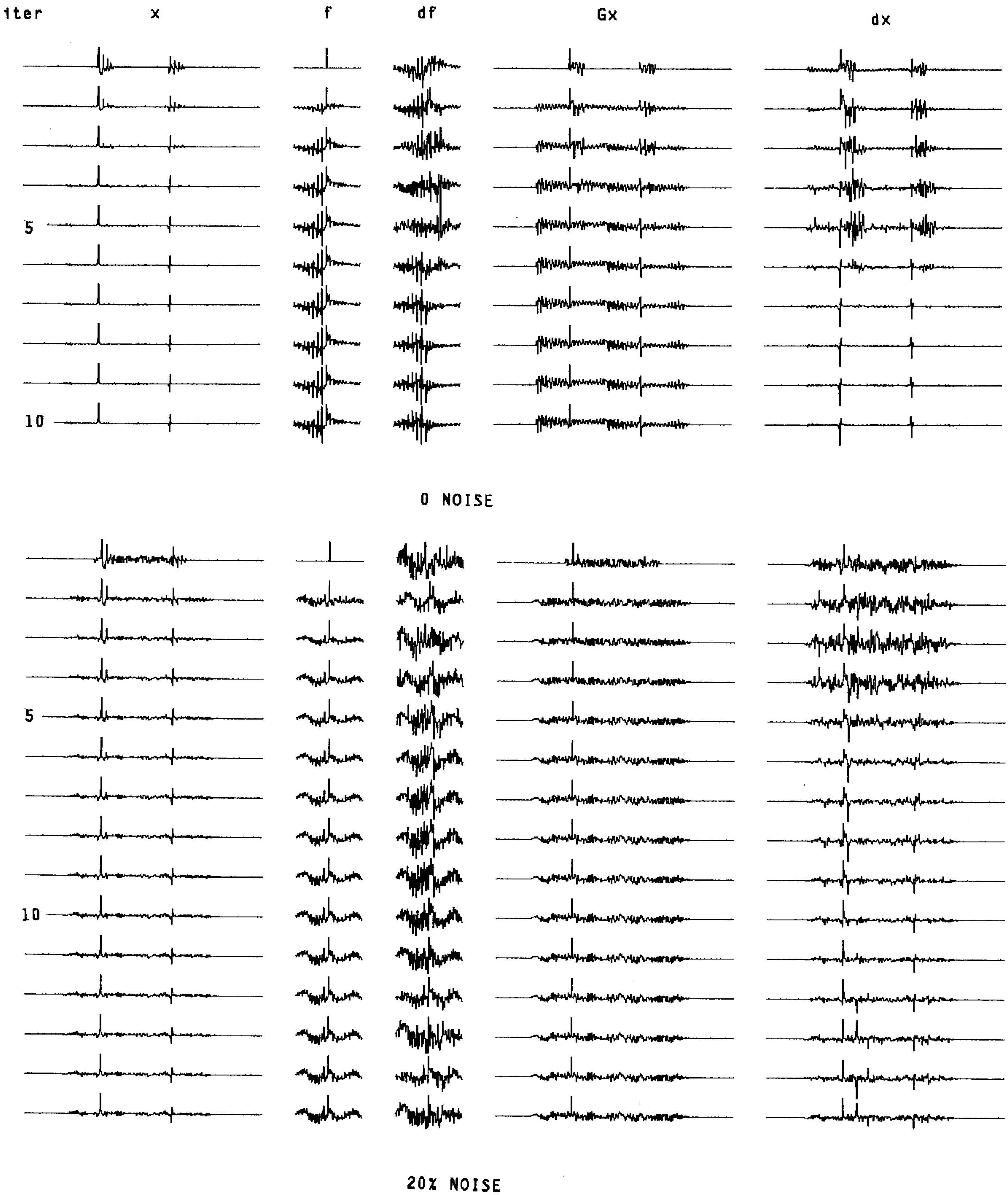


FIGURE 1

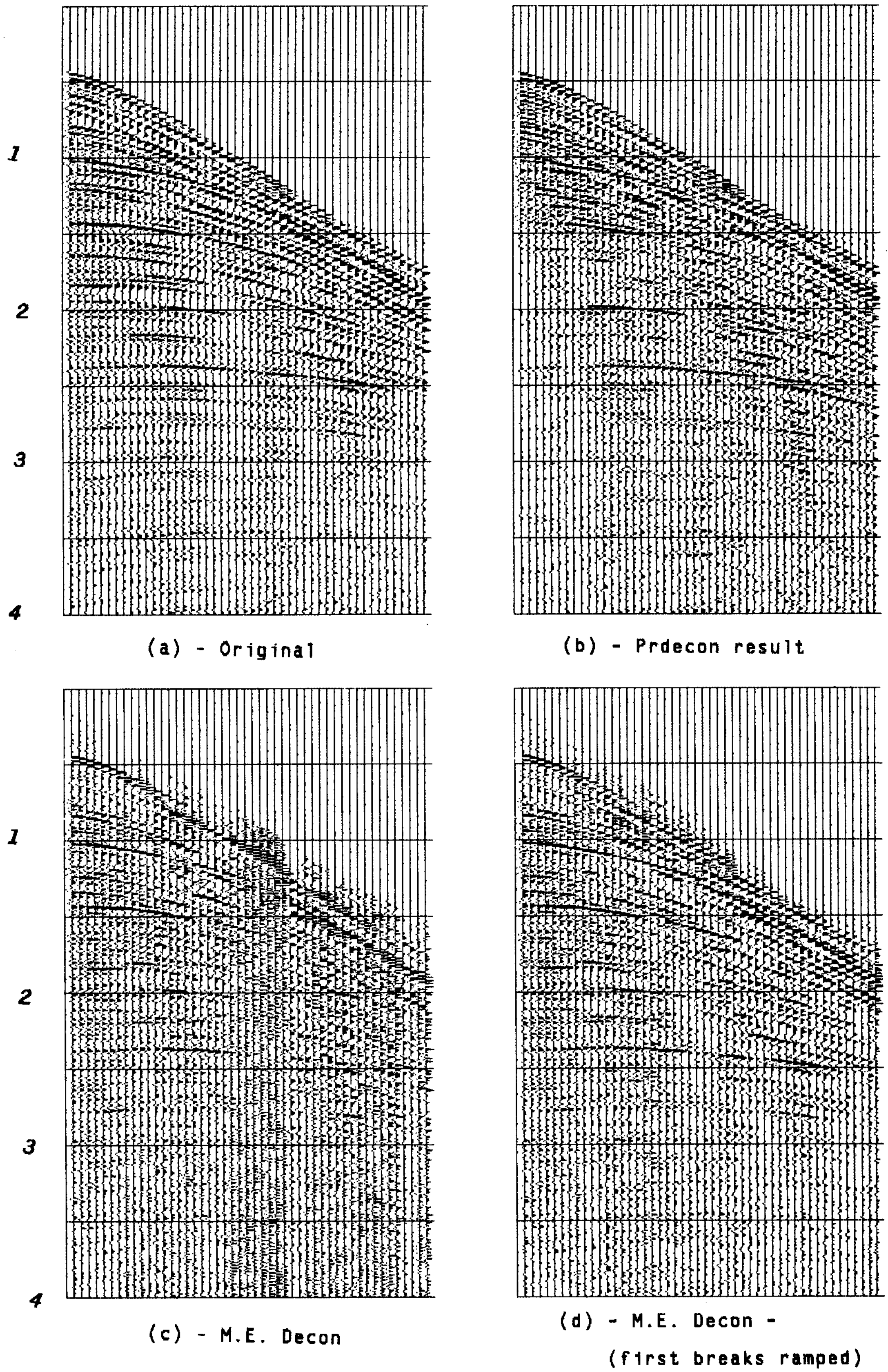
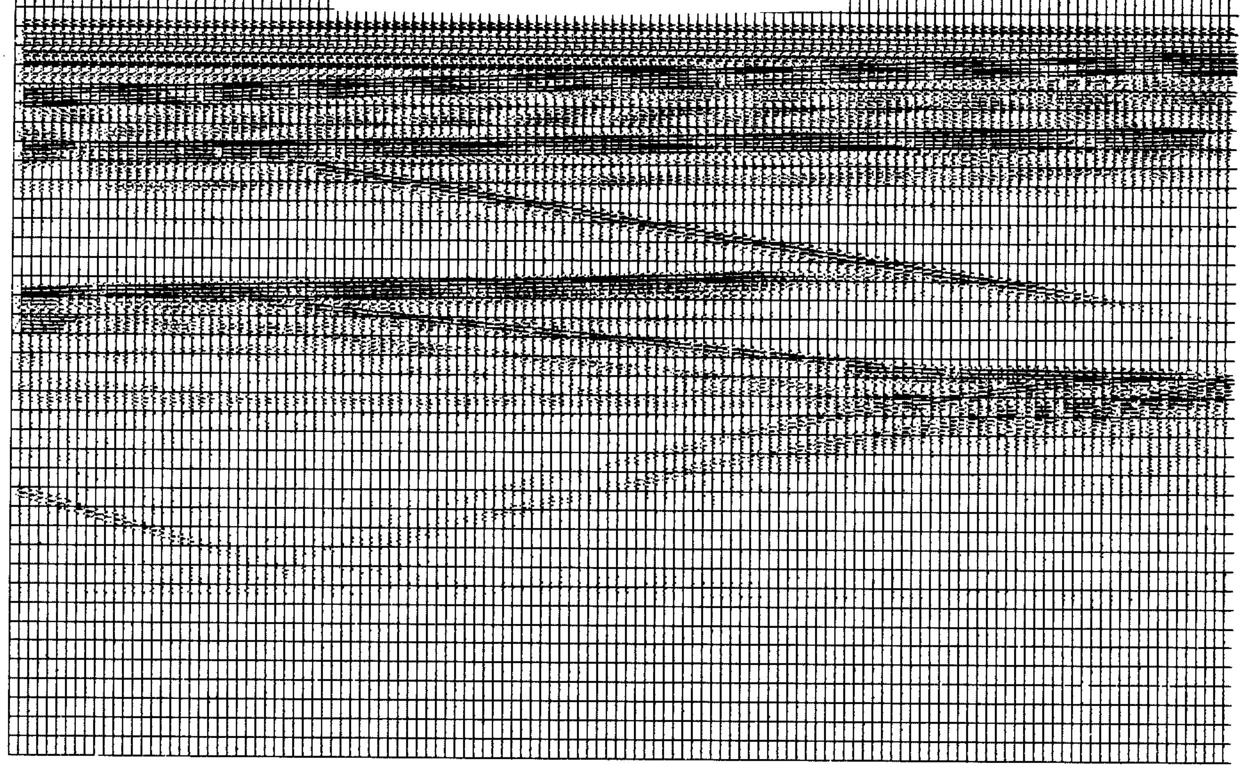
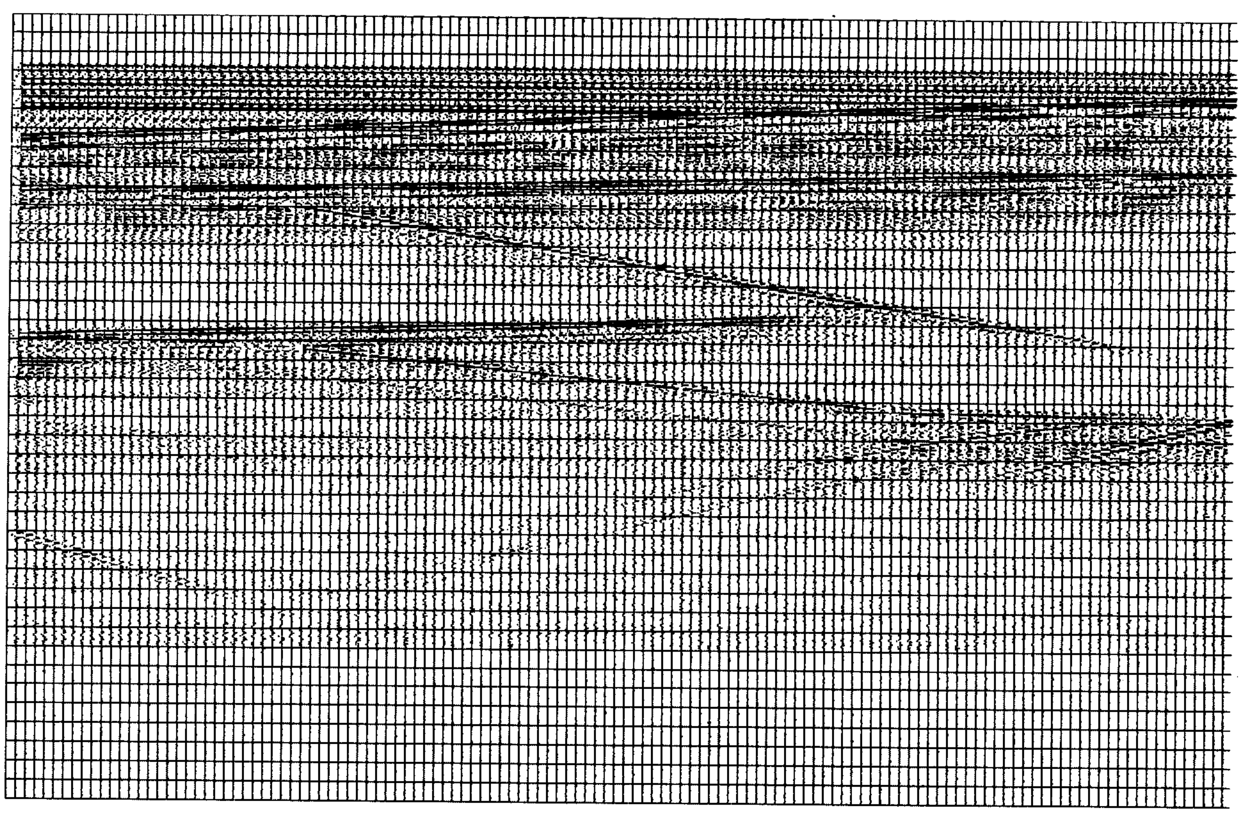

1
2
3
4
(a) - Original
(b) - Prdecon result
1
2
3
4
(c) - M.E. Decon
(d) - M.E. Decon -
(first breaks ramped)

FIGURE 2

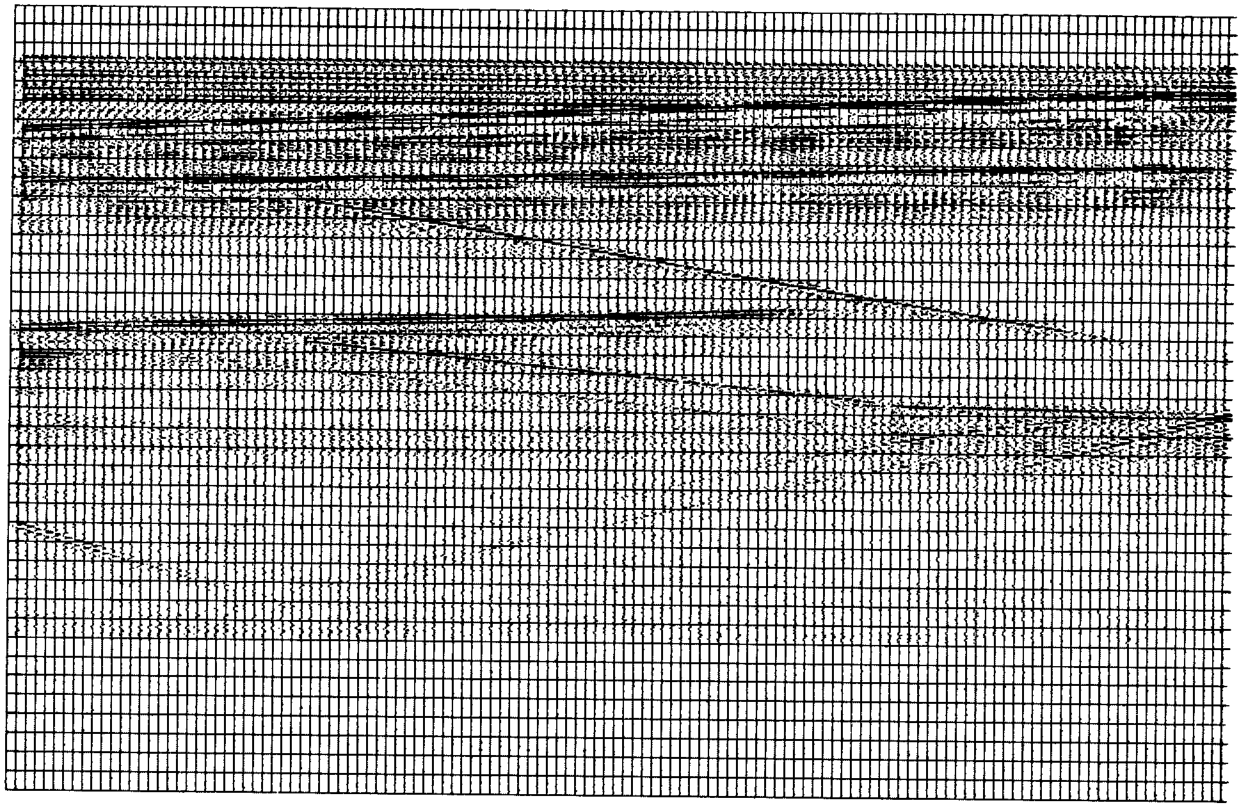
FIGURE 3 - GULF WATER TANK DATA



(3a) - original data



(3b) - predictive decon



(3c) - min. entropy decon on (b)