Non-Gaussian Signal Analysis

by Jon F. Claerbout

Any geophysicist who has looked at well logs, or just anyone who has inspected sedimentary rocks in a road cut could make the following observations about the rock properties (be they velocity, density, porosity or subjective appearance): As a function of depth the properties change randomly in a largely unpredictable manner. Locally the properties may (1) fluctuate about a fixed average, (2) fluctuate about a steady gradient gradually going from one rock type to a somewhat different type, or (3) jump abruptly as one rock type is overlain by an entirely different type. Furthermore, the fluctuations themselves may be large or small at different depths in the sedimentary column. Some examples are shown in Figure 1.

Our knowledge of the stratigraphic sequence is often via reflection seismographs where the fluctuating earth properties are seen as an echo time series modeled by the methods of statistical time series analysis. Unfortunately, much of the subject of time series analysis is only optimum in the presence of stationary Gaussian random signals and noises. Since these presumptions are rather doubtful, we need to consider the consequences and alternatives.

An earlier paper by Francis Muir and I (Robust Modeling with Erratic Data, Geophysics, vol. 38, no. 5, p. 826-844, 1973) considered the non-Gaussian character of much geophysical data. To sharpen the distinction between non-Gaussian signals and non-Gaussian noise we

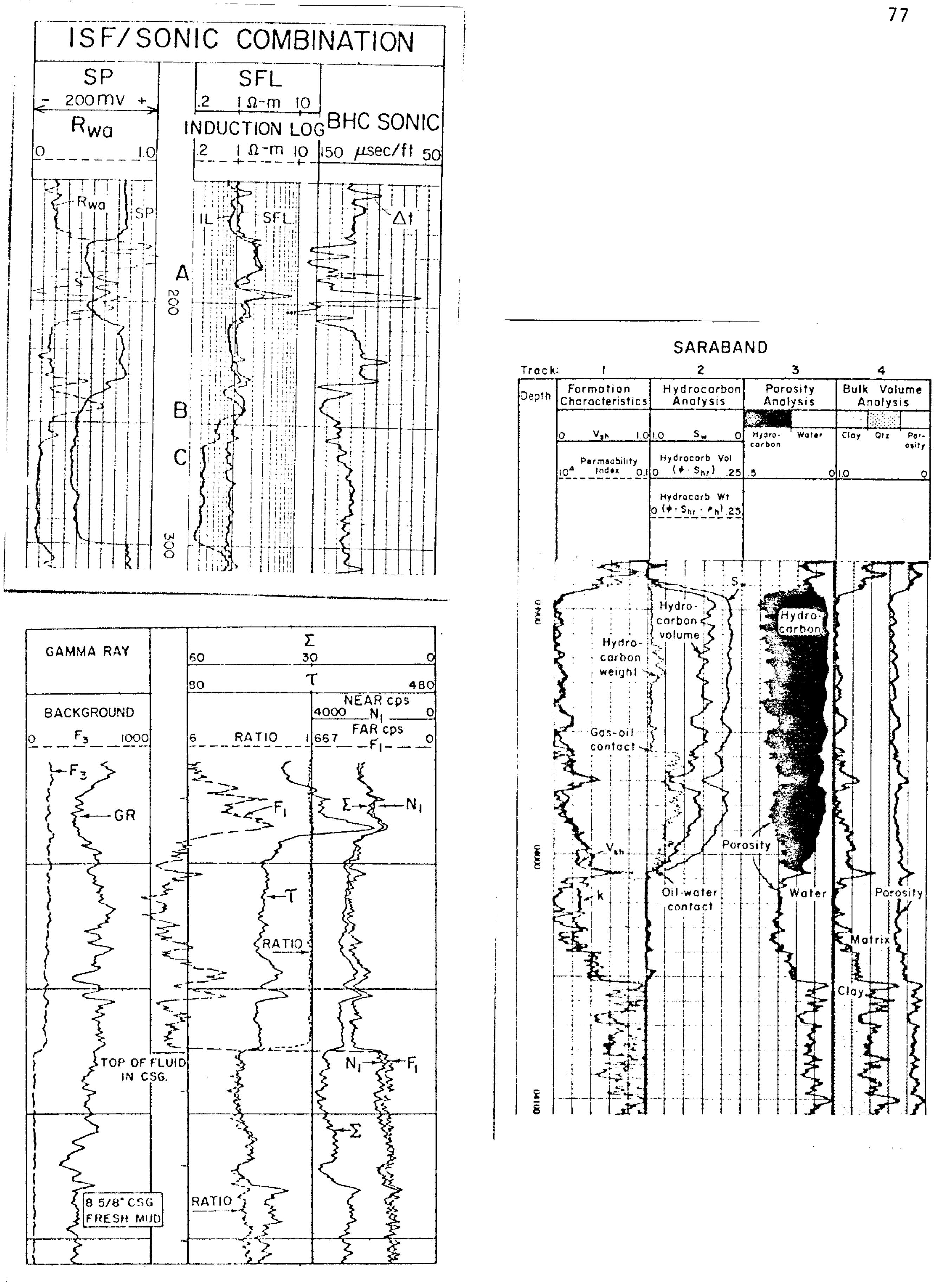

Figure 1. Sample well logs from the 1974 Schlumberger Log Interpretation, Volume II-Applications.

consider the equation set

$$\begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \\ d_7 \end{bmatrix} - \begin{bmatrix} x_1 & 0 & 0 \\ x_2 & x_1 & 0 \\ x_3 & x_2 & x_1 \\ x_4 & x_3 & x_2 \\ x_5 & x_4 & x_3 \\ 0 & x_5 & x_4 \\ 0 & 0 & x_5 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} \qquad (1)$$

In our Robust Modeling paper we indicated that infinite errors in a few of the components the data vector d could be tolerated if the L_1 norm of e (the sum of the absolute values of the error vector components) was minimized. However, infinite errors in the x vector cannot be handled in that way. Now we want to consider that the x vector contains the information about the stratigraphic record. In reflection seismology the x vector would be the derivative of the logarithm of the acoustic impedance (density times velocity) with respect to two-way vertical travel time. The acoustic impedance might resemble the logs shown in Figure 1. The reflectivity would be roughly the derivative of these logs. Note that the variance of such a curve is constantly changing and that geological unconformities would often show up as outliers in x in equation (1). In our earlier work with outliers in the data vector d we found that the median is a more suitable average than the mean. Fortunately, if random properties arise from a symmetrical probability function (such as the Gaussian) the median equals the mean. Hence, for a large data sample from a Gaussian population the median has the same use as the mean. In statistical jargon one says that under the Gaussian assumption the median is a consistent estimator of the mean.

Let us first reduce our time series problem (1) to the simpler projection problem

$$\begin{bmatrix} \\ y \\ \\ \end{bmatrix} \approx \begin{bmatrix} \\ x \\ \\ \end{bmatrix} c \tag{2}$$

Now the problem is to estimate a "best" value for c . The usual least-squares implicitly-Gaussian approach is to dot both sides of (2) by the transpose of the x vector obtaining

$$\hat{\hat{c}} = \frac{x \cdot y}{x \cdot x} \tag{3}$$

Now if it is supposed that x contains a few wild points then we could consider the estimator

$$\hat{c} = \text{median}\,(y_i / x_i) \tag{4}$$

It happens that under Gaussian assumptions $\hat{c}$ is a consistent estimator of $\hat{\hat{c}}$. Specifically, let x and y be zero-mean, correlated Gaussian random variables. Then it is shown in Papoulis, Probability, Random Variables, and Stochastic Processes, McGraw-Hill, 1965, pages 197-198, that y/x is a random variable with a Cauchy distribution centered at $\rho\, \sigma_y / \sigma_x$ which is the same as $E(xy) / E(x^2)$. Thus, we expect that for Gaussian x and y , $\hat{c}$ is as good as $\hat{\hat{c}}$. The big question is whether the median estimator $\hat{c}$ offers any significant advantages over $\hat{\hat{c}}$ in the non-Gaussian case. Consider the synthetic seismogram

$$.000 \quad 1.000 \quad .500 \quad .250 \quad .125 \quad 1.063 \quad 1.531 \quad .766 \quad .383 \quad .191 \quad .096 \qquad (5)$$

The average human observer quickly suspects a $1 / (1 - .5Z)$ filter response and quickly determines a signal excitation

$$0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 1 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \qquad (6)$$

The excitation (6) could easily be the result of a coin toss experiment, so it could obviously be thought of as a realization of a random process. Since only the integers 0 and 1 are observed, the probability function is more bimodal than Gaussian.

The classical Wiener-Levinson-Robinson deconvolution of the series (5) gives the filter $1 / (1 - .7Z)$ and the excitation function

$$0 \quad 1 \quad -.21 \quad -.11 \quad -.05 \quad .97 \quad .77 \quad -.33 \quad -.16 \quad -.08 \quad -.04 \qquad (7)$$

The reason this doesn't turn out to be the desired result is the large unit lag autocorrelation of the excitation (6). Let us see how a systematic analysis of (5) can lead to the intuitive result (6). Figure 2 shows a scattergram of the points of (5) in the (x_t , x_{t+1}) plane. Suppose we create a straight line for each point in the scattergram by passing each line through the origin and the data point. Next we

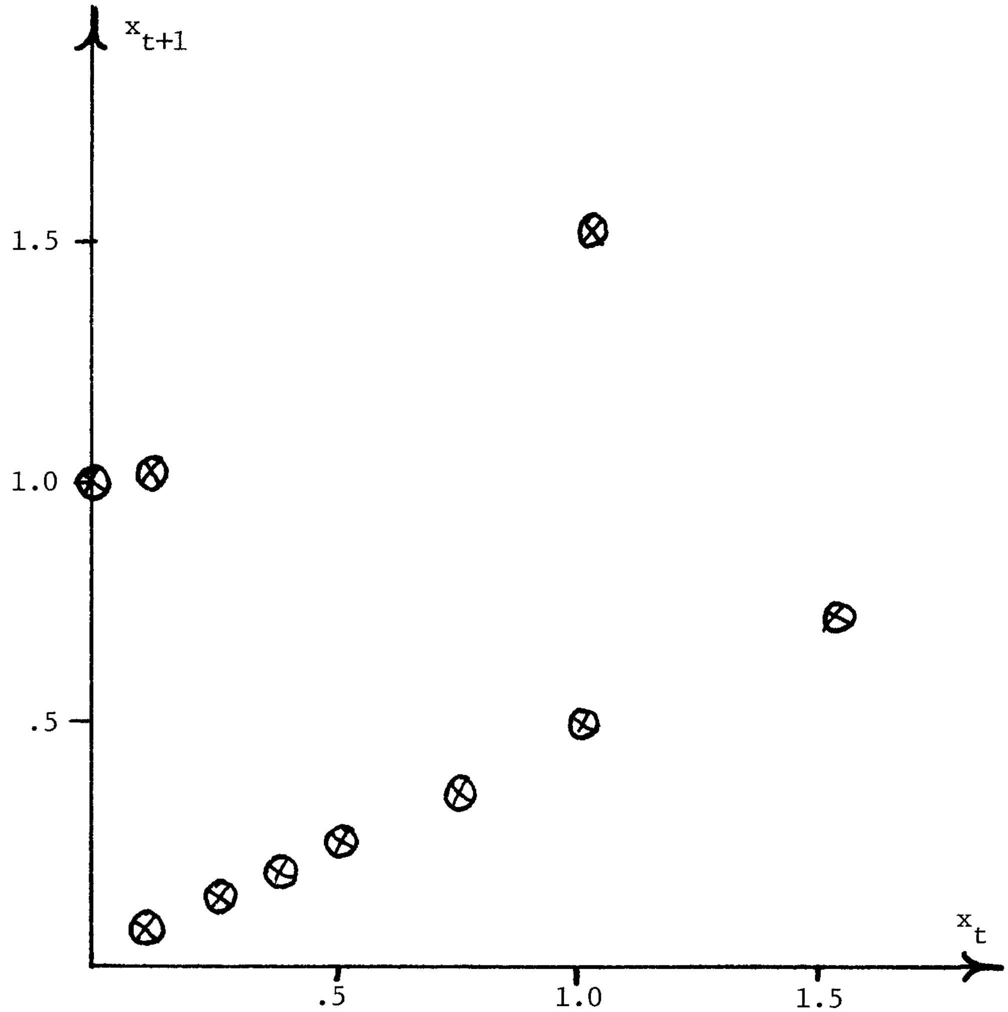

Figure 2. Scattergram of the time series (5) in the (x_t, x_{t+1}) plane. All but three points lie along a line of slope equal one half.

compute the slope of each line and arrange the slopes in an ascending numerical sequence. Then we pick the middle point (median) of the sequence which turns out to be a slope of one half, implying a filter of $1/(1-.5Z)$. A little noise obviously would not make a serious problem.

The Levinson recursion mainly amounts to a succession of vector orthogonalizations. This is particularly apparent when the computation is organized in Burg form (see my book, p. 133-136 or p. 160-161, for example). With the definition of y'

$$y_i' = y_i - x_i \text{ Median} (y_i/x_i) \tag{8}$$

We are doing a kind of orthogonalization. To illustrate, in Euclidean vector space example (8) would be

$$y' = y - x \frac{x \cdot y}{x \cdot x} \tag{9}$$

The fact that y' is orthogonal to x is shown by taking the dot product of (9) with x . With (8) the orthogonalization works a little differently. Instead of the sum of $x_i y_i'$ vanishing, it will turn out that it is the median which vanishes. Note that the vanishing of the median of $x_i y_i'$ is equivalent to saying that $x_i y_i'$ has as many positive terms as negative terms. That is like saying that the number of sign agreements between x_i and y_i' equals the number of sign disagreements. Equivalent statements of the condition are that the median of y_i'/x_i or x_i/y_i' should vanish. Thus, we would like to show that the definition of y' given by (8) implies any of the following three "orthogonality" conditions

$$0 = \text{Median}\,(\, y_i' \; x_i \,) \tag{10a}$$

$$0 = \sum_i \text{sgn}(y_i') \; \text{sgn}(x_i) \tag{10b}$$

$$0 = \text{Median}\,(\, y_i' \,/\, x_i \,) \tag{10c}$$

The easiest to show is (10c). Insert (8) into (10c)

$$0 \stackrel{?}{=} \text{Median}\,\{\, [\, y_i - x_i \,\text{Median}\,(\, y_i \,/\, x_i \,)\,] \,/\, x_i \,\}$$

$$0 = \text{Median}\,[\, y_i \,/\, x_i - \text{Median}\,(\, y_i \,/\, x_i \,)\,] \tag{11}$$

But (11) is obviously true.

Fortunately sorting numbers into numerical order requires only something like $N \ln N$ comparisons. This means that we need not restrict ourselves to medians, but we can select any percentile we choose. For example, we could look at the 40th percentile and the 60th percentile. If they straddle a correlation coefficient of zero, then zero could be used, otherwise the smaller of the two correlation coefficients could be selected. Another possibility is to restrict the correlation coefficients to be zero if the estimate turns out positive. This would result in prediction error filter coefficients which are all positive.