Characterization of long-range correlations in complex distributions and profiles

Ali Reza Mehrabi, Hossein Rassamdana, and Muhammad Sahimi
Department of Chemical Engineering, University of Southern California, Los Angeles, California 90089-1211
(Received 17 June 1996; revised manuscript received 25 February 1997)

Characterizing long-range correlations in complex distributions, such as the porosity logs of field-scale porous media, and profiles, such as the fracture surfaces of rock and materials, is an important problem. We carry out an extensive analysis of such distributions represented by synthetic and real data to determine which method provides the most efficient and accurate tool for characterizing them. The synthetic data and profiles are generated by a fractional Brownian motion (FBM) and the real data analyzed are a porosity log of an oil reservoir and time variations of the pressure fluctuations in three-phase flow in a fluidized bed. The FBM is generated by three different numerical methods and the data are analyzed by seven different techniques. Our analysis indicates that the size of the data array greatly influences the accuracy of characterization of its long-range correlations. We also find that if the size of the data array is large enough, the commonly used rescaled-range (R/S) method of analyzing FBM series fails to provide accurate estimates of the Hurst exponent, although it can provide a reasonably accurate analysis of a data array that is generated by a fractional Gaussian noise. In contrast, the maximum entropy and wavelet decomposition methods offer highly accurate and efficient tools of characterizing long-range correlations in complex distributions and profiles. New methods that are somewhat similar to the R/S method are also suggested. [S1063-651X(97)05407-X]

PACS number(s): 47.55.Mh, 05.40.+j, 47.53.+n

I. INTRODUCTION

Field-scale porous media (FSPM), such as oil reservoirs and groundwater aquifers, are highly heterogeneous at many length scales. Their heterogeneities manifest themselves at three different scales, which are (i) microscopic, which is at the level of pores and grains; (ii) macroscopic, which is at the level of core plugs; and (iii) megascopic, which includes the entire reservoir or aquifer. Modeling flow and transport in such porous media depends crucially on the characterization of their structure and in particular the distributions of their heterogeneities. However, although the characterization of laboratory-scale (macroscopic) porous media has been done in great detail and a reasonable understanding of them has been obtained [1,2], the same is not true about FSPM, whose characterization is plagued by a lack of sufficient data and hampered by the wide variations in the data that are collected at various locations throughout the system.

Two important characteristics of FSPM are their porosity logs, often collected by *in situ* methods, and their permeability distributions, which are usually obtained by collecting a number of core plugs at various depths along a vertical (or horizontal) well. In a pioneering work, Hewett [3] proposed that the porosity logs and permeability distributions of FSPM obey fractal statistics. More specifically, he provided evidence that the porosity logs in the direction perpendicular to the bedding may obey the statistics of fractional Gaussian noise (FGN), while those parallel to the bedding follow a fractional Brownian motion (FBM), which is a stochastic process $B_H(\mathbf{r})$ [4] with the properties

$$\langle B_H(\mathbf{r}) - B_H(\mathbf{r}_0)\rangle = 0\,, \tag{1}$$

$$\langle [B_H(\mathbf{r}) - B_H(\mathbf{r}_0)]^2\rangle \sim |\mathbf{r}-\mathbf{r}_0|^{2H}, \tag{2}$$

where $\mathbf{r}=(x,y,z)$ and $\mathbf{r}_0=(x_0,y_0,z_0)$ are two arbitrary points and H is the Hurst exponent. A remarkable property of the FBM is that it generates correlations whose extent is *infinite* (i.e., it is as large as the linear size of the system). For example, consider the one-dimensional case and define an incremental correlation function $C(x)$ of the ''future'' increments $B_H(x)$ with the ''past'' increments $B_H(-x)$ by (the meaning of past and future becomes clear if we replace x with a time variable)

$$C(x) = \frac{\langle -B_H(-x)B_H(x)\rangle}{\langle B_H(x)^2\rangle}; \tag{3}$$

then one finds that $C(x)=2(2^{2H-1}-1)$, *independent* of x. Moreover, the type of correlations can be tuned by varying H. If $H>1/2$, then $C(x)>0$ and the FBM displays *persistence*, i.e., a trend (for example, a high or low value) at x is likely to be followed by a similar trend at $x+\Delta x$. If $H<1/2$, then $C(x)<0$ and the FBM generates *antipersistence*, i.e., a trend at x is not likely to be followed by a similar trend at $x+\Delta x$. For $H=1/2$ the trace of the FBM is similar to that of a random walk and the *increments* are uncorrelated.

A convenient way of representing a distribution function is through its spectral density $S(\boldsymbol{\omega})$, the Fourier transform of its variance. For a d-dimensional FBM it can be shown that

$$S(\boldsymbol{\omega}) = \frac{1}{\left(\sum_{i=1}^{d}\omega_i^2\right)^{H+d/2}}, \tag{4}$$

where $\boldsymbol{\omega}=(\omega_1,\ldots,\omega_d)$. Fractional Brownian motion is stationary, but not ergodic. Its variance for a large enough array is *divergent*. It is not differentiable, but by smoothing it over

an interval one can obtain its ''derivative'' numerically, which is the FGN whose spectral density in, e.g., one dimension, is given by

$$S(\omega)=\frac{1}{\omega^{2H-1}}. \tag{5}$$

We should point out that if v is a random variable that obeys the statistics of the FBM, then the increments $v(\mathbf{r}+\Delta\mathbf{r})-v(\mathbf{r})$ are Gaussian random variables.

It is clear that the Hurst exponent H determines the nature of the correlations. Moreover, given H, one can generate the porosity logs and the permeability distributions for use in the simulation of flow and transport in FSPM if such properties follow the statistics of a FBM or a fGn. Thus, given a porosity log or a permeability distribution of a porous medium, how can one accurately determine the associated value of H? The most widely used method of estimating H is the rescaled-range (R/S) analysis (described below), first developed by Hurst *et al.* [5] and analyzed exhaustively by Mandelbrot and Wallis [6] and Wallis and Matalas [7]. According to Mandelbrot and Wallis [6], this method is very robust. However, in the course of analyzing extensive field data on the porosity logs and permeability distributions of an oil reservoir in the Middle East, it appeared to us that, while the R/S method may be robust, it is not very accurate in the sense that it yields roughly the same value of H regardless of the data. This prompted us to undertake a systematic analysis of the data, as well as synthetic data generated by a FBM, by a variety of methods to see which method provides the most accurate description of them and an estimate of the associated value of H. The purpose of this paper is to report the results of this analysis. This issue has already been investigated by a few authors [8–11]; however, Refs. [8,9] concerned themselves mainly with the accuracy of the R/S method, whereas our goal here is to test various methods of analyzing the data. Reference [10] did not study the problem systematically, while Schmittbuhl *et al.* [11] did not investigate several important issues and methods that we discuss in this paper (see below).

The problem that we study is, however, more general and not restricted to analyzing complex distributions, such as porosity logs. In recent years, the description of scale-invariant properties of disordered systems by fractal analysis has become very fashionable. The concept of self-similarity of fractal systems [12] has been extended [13] to anisotropic systems through the notion of self-affinity, which implies that the scale invariance of an anisotropic fractal system is preserved only if the scale factors are direction dependent. Examples include the profiles that are generated by a FBM or a FGN. There are also a large number of natural systems that are self-affine [14], an example of which is fracture surfaces of materials and natural rock [1,2,15–17]. If we consider a two-dimensional cut, or a roughness profile, of such systems, then the average height of the profile scales with its length L as L^H, where H is again the Hurst exponent. Therefore, if we digitize such a profile to estimate its Hurst exponent, we will be faced with the same task as analyzing a complex distribution and describing its long-range correlations.

The plan of this paper is as follows. In Sec. II we discuss numerical generation of the synthetic data by a FBM and its sensitivity to the accuracy of the method and the size of the data array. In Sec. III we describe the methods that we use to analyze the data, while Sec. IV presents the results of our analysis of the synthetic and real data and a discussion of their implications. The paper is summarized in Sec. V.

II. GENERATION OF SYNTHETIC DATA

Since the real data that we analyze are one dimensional, the synthetic data are also generated by a one-dimensional FBM for which there are a number of numerical methods. As one goal of our paper is to see which method is the most accurate, we used three different numerical methods for generating a one-dimensional FBM, a brief description of each is given below. For each method, each value of H, and each size of the FBM array, we generated 100 different realizations of the FBM, and the results that are discussed below represent the averages over all the realizations. The typical standard deviation among all of the realizations was about $\pm 5\%$ of the average values. We used 20 values of H, equally spaced in [0,1].

A. Fast Fourier transformation

Equation (4) provides a convenient method of generating an array of numbers that obey the FBM statistics, using a fast Fourier transform (FFT) technique. In this method, one first generates random numbers, uniformly distributed in [0,1), and assigns them to the sites of a d-dimensional lattice, which in our case is a linear chain. The Fourier transform of the resulting d-dimensional array of the numbers is then calculated numerically. The Fourier-transformed numbers are then multiplied by $\sqrt{S(\boldsymbol{\omega})}$ and the results are then inverse Fourier transformed back into the real space. The numbers so obtained obey the FBM statistics with the desired long-range correlations. To avoid the problem associated with the periodicity of the numbers arising as a result of their Fourier transforming, one has to generate the array with a much larger lattice size than the actual size that is used in the analysis and use the central part of the array. In the discussion of our results, when we refer to the size of the synthetic data array we mean the size of the central part of the FBM array that we used in our study.

B. Successive random additions

In the successive random addition (SRA) method [18] one starts with the two end points on [0,1] and assigns a zero value to them. Then Gaussian random numbers Δ_0 are added to these values. In the next stage, new points are added at a fraction r of the previous stage by interpolating between the old points (by either linear or spline interpolation) and Gaussian random numbers Δ_1 are added to the new points. Thus, given a sample of N_i points at stage i with resolution λ, stage $i+1$ with resolution $r\lambda$ is determined by first interpolating the $N_{i+1}=N_i/r$ new points from the old points and then Gaussian random numbers Δ_i are added to all of the new points. At stage i with $r<1$, the Gaussian random numbers have a variance [see Eq. (2)]

$$\sigma_i^2 \sim r^{2iH}. \tag{6}$$

This process is continued until the desired length of the data array is reached. Here we use a value of $r=1/2$ to generate the FBM's with different values of H.

C. The Weierstrass-Mandelbrot algorithm

In the Weierstrass-Mandelbrot (WM) method [18] one first divides the interval [0,1] into $n-1$ equally spaced subintervals, where n is the size of the data array that we wish to generate, and assigns zero value to all the points in the interval. Then, to point i at a distance x_i from the origin one adds a random number generated by the Weierstrass function defined by

$$\mathcal{W}(x_i)=\sum_{j=-\infty}^{\infty} C_j r^{jH}\sin(2\pi r^{-j}x_i+\phi_j), \tag{7}$$

where C_j and ϕ_j are random numbers distributed according to Gaussian and uniform distributions, respectively, and r is a measure of the distance between the frequencies, which is usually chosen to be small, so that the distance is small; we used $r=0.9$. The variance of C_j is proportional to r^{2jH} and the random phases ϕ_j are distributed uniformly on $[0,2\pi]$. Usually, the infinite series in Eq. (7) is approximated by a finite number of terms; we used up to 140 terms in $-70\leqslant j\leqslant 70$ to ensure its accuracy. The power spectrum of the data array generated by the WM method is discrete and does not contain all the frequencies. However, its spectral density is still proportional to $\omega^{-(2H+1)}$, in agreement with Eq. (4).

III. METHODS OF ANALYSIS

As mentioned above, we have used several methods of analyzing the synthetic and real data. Our goal is to test the accuracy and efficiency of each method and their possible sensitivity to various factors that may affect their performance, e.g., the size of the data array. What follows is a brief description of all the methods that we have used.

A. Rescaled-range analysis

Suppose that a variable v takes the value $v(l)$ at position l. In the R/S method one calculates $R(L)$ by

$$R(L)=X_{\max}(l,L)-X_{\min}(l,L), \quad 1\leqslant l\leqslant L, \tag{8}$$

where

$$X(l,L)=\sum_{u=1}^{l}[v(u)-\langle v\rangle_L], \quad \langle v\rangle_L=\frac{1}{L}\sum_{l=1}^{L} v(l). \tag{9}$$

One also calculates the quantity $S(L)$ given by

$$S(L)=\left\{\frac{1}{L}\sum_{l=1}^{L}[v(l)-\langle v\rangle_L]^2\right\}^{1/2}. \tag{10}$$

It has been argued that [5–7,9] if the data follow the statistics of the FBM, then

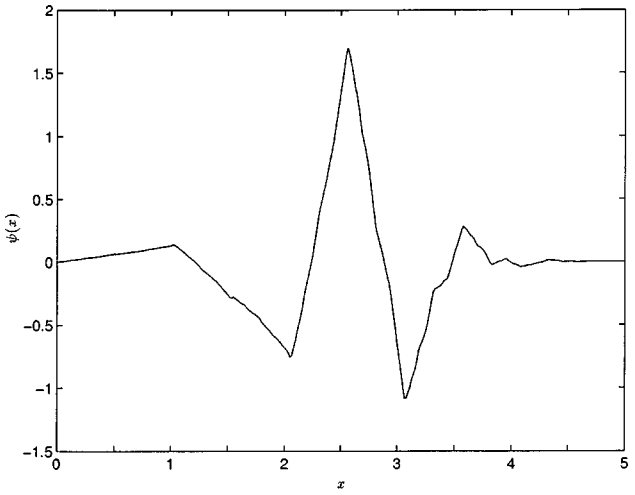

FIG. 1. Daubechies function $\psi(x)$ used in the wavelet decomposition method.

$$\frac{R(L)}{S(L)}\sim L^H, \tag{11}$$

so that a logarithmic plot of $R(L)/S(L)$ versus L yields H.

B. Orthonormal wavelet decomposition method

Wavelet decomposition analysis, which is a space-scale decomposition method, is a suitable tool for analyzing the FBM. Using discrete orthonormal wavelet decomposition of a one-dimensional FBM, we obtain [19]

$$D_j(k)=2^{-j/2}\int_{-\infty}^{\infty} B_H(x)\,\psi(2^{-j}x-k)\,dx, \tag{12}$$

where $D_j(k)$ are the wavelet-detail coefficients of the FBM, ψ is the wavelet function, $k=1,2,\ldots,n$, where n is the size of the data array, and the j's are integers. Thus, in this method one fixes j and varies k to calculate $D_j(k)$. For each j one determines n such numbers and calculates their variance $\sigma^2(j)$. Then it can be shown that, regardless of the wavelet function ψ, one has [19,20]

$$\log_2[\sigma^2(j)]=(2H+1)j+\text{const.} \tag{13}$$

Thus, plotting $\log_2[\sigma^2(j)]$ versus j yields H. One can use a variety of wavelet functions; we used the Daubechies function [21], which is shown in Fig. 1.

C. Covariance analysis

This method is based on the fact that the variance of the variable v that obeys the statistics of the FBM is given by Eq. (2). Thus a logarithmic plot of the variance versus x is sufficient for estimating H. Our main point in using this method is to check the numerical accuracy of the three methods that we use to generate the synthetic data array since the covariance method uses the very definition of a FBM.

D. Lévy analysis

The Lévy distribution is a group of symmetric-stable distributions given by [22]

$$P(x)=\frac{1}{\pi}\int_0^\infty \exp[-(C\omega)^\alpha]\cos(\omega x)d\omega, \tag{14}$$

where $0<\alpha\leqslant 2$ and $C>0$ are the parameters of the distribution. This distribution contains long tails that represent very rare events. The presence of long tails in the distributions of the increments of some natural data has led to the consideration of a general form of complex distributions called the fractal Lévy motions (FLM's) [23], which have recently been used in the analysis of some seismic and porosity log data [24–27]. The FLM has stationary increments with a Lévy-stable distribution given by Eq. (14) and is characterized by the parameter H. The increments $v(x+\Delta x)-v(x)$ of the variable v, which obeys the statistics of the FBM, are Gaussian variables and thus the limit $\alpha=2$ corresponds to the FBM case. To estimate H, C is evaluated first and its dependence on the lag $x=x_2-x_1$ is monitored. Then it can be shown that the coefficients $C(x)$ are related to the lag x by

$$C(x)\sim x^H. \tag{15}$$

The limit $H=1/\alpha$ corresponds to the case of independent increments. For positive (negative) correlations (anticorrelation) in the increments we have $H>1/\alpha$ $(H<1/\alpha)$. Since $P(x)$ cannot be obtained in closed form, the estimation of C and α is difficult and may be subject to some inaccuracies [28]. However, compared to the FBM, this is a more general distribution that allows more flexibility in the interpretation of the data.

E. Spectral methods

Spectral methods of analyzing the FBM use its power spectrum given by Eq. (4). Thus H is estimated from a plot of $\log S(\omega)$ versus $\log\omega$. However, the accurate calculation of the power spectrum is the most important part of the method. If, for example, the data are noisy, their power spectrum would be difficult to calculate accurately and therefore the estimated H may be subject to great uncertainty. The power spectrum can be calculated by a discrete Fourier transformation (DFT), and if the length of the data array is a power of 2, then the FFT method can also be used. The main problem with this method is the spectral leaks from the other frequencies [21], i.e., in addition to its true range of the frequencies, the power spectrum also contains components at other (disallowed) frequencies, which result in an overestimation of $S(\omega)$ and thus an underestimation of H. Therefore, we did not use a DFT for analyzing the data. Instead, we used two other methods of calculating the power spectrum of the synthetic data, which are as follows.

(i) The windowed Fourier transformation method. The problem of the spectral leaks can be overcome by using a windowed Fourier transformation method instead of a DFT or a FFT alone. In this method [21], the data are first divided into windows or sections. To each window is assigned a correponding weight function, and the weighted FFT of the data are then calculated. The window weight function can be chosen from a number of different choices; here we used the Bartlett window, which for the jth window is given by $1-2|(j-m/2)/m|$, where m is the size of the window. The disadvantage of this method is that if the data that contain long-range correlations are grouped into windows, the nature of the correlations may be masked, unless we have very long data arrays and use large windows.

(ii) The maximum entropy method (MEM). This is a method of estimating the power spectrum without using a FFT and thus is subject to far less noise [21]. In this method the power spectrum is approximated by

$$S(\omega)\simeq\frac{a_0}{\left|1+\sum_{k=1}^{M}a_k z^k\right|^2}, \tag{16}$$

where the coefficients a_k are calculated such that Eq. (16) matches the Laurent series $S(\omega)=\sum_{-M}^{M}b_i z^i$. Here z is the frequency in the z transform plane $z\equiv e^{2\pi i\omega\Delta}$ and Δ is the sampling interval in the real space. In practice, to calculate the coefficients a_i one first computes the correlations functions

$$\phi_j=\langle v_i v_{i+j}\rangle\simeq\frac{1}{n-j}\sum_{i=1}^{n-j}v_i v_{i+j}, \tag{17}$$

where n is the number of data points and v_i is the datum at point i. The coefficients a_i are then calculated from

$$\sum_{j=1}^{M}\phi_{|j-k|}a_j=\phi_k, \quad k=1,2,\ldots,M. \tag{18}$$

The advantage of Eq. (16) over the Laurent series is that if $S(\omega)$ contains sharp peaks, then Eq. (16) can easily detect them as the peaks may show up as the poles of the equation, whereas one may have to use a very large number of terms in the Laurent series to detect the same peaks.

F. Roughness-length method

In the roughness-length (RL) method, the data are first grouped into windows of length l. Then one calculates the root-mean-square ($\mathcal{R}_{\rm rms}$) residual roughness defined by [10]

$$\mathcal{R}_{\rm rms}(l)=\frac{1}{n_l}\sum_{i=1}^{n_l}\sqrt{\frac{1}{(n_i-2)}\sum_{j\in W_i}[f(j)-\langle f\rangle]}, \tag{19}$$

where n_l is the number of windows of length l, n_i is the number of data points in the window W_i, and $\langle f\rangle$ is the average of f. $f(j)$ is the residual value at j calculated by subtracting the datum at j from the value that is obtained from the linear trend in the data, obtained either by a linear regression of the data or by connecting the first and the last data points. It can be shown that

$$\mathcal{R}_{\rm rms}(l)\sim l^H, \tag{20}$$

so that H can be estimated from a logarithmic plot of $\mathcal{R}_{\rm rms}(l)$ versus l.

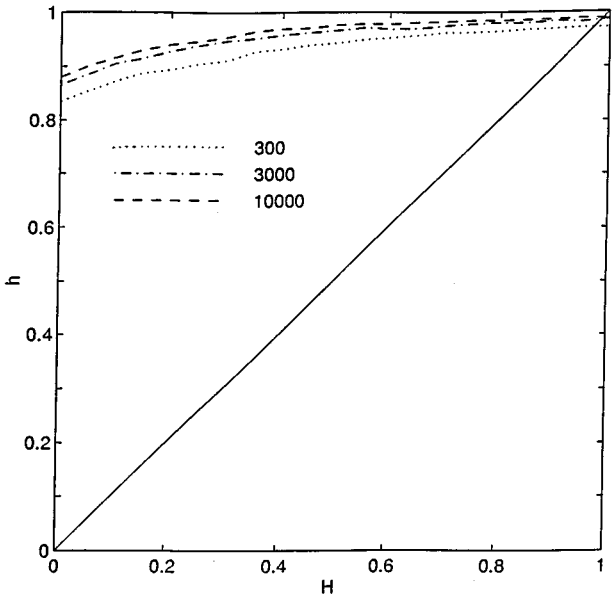

FIG. 2. Estimated Hurst exponent h versus its true value H as predicted by the R/S method. Numbers in the figure denote the size of the data array, which was generated by a fast Fourier transformation method.

IV. RESULTS AND DISCUSSION

In this section we present the results and discuss the accuracy of each method of analysis. In what follows H denotes the value of the Hurst exponent that we *set* in order to generate the synthetic data array and is referred to as the true value, while h is the value of the Hurst exponent that we *estimate* using the above methods of analysis. Thus, if the two values were the same, their plot would be a straight 45° line, which is shown in all of the figures discussed below. We first discuss the results for the synthetic data to pick the most accurate method of data analysis. We then use this method for analyzing the real data.

A. Synthetic data

Since we used three different numerical methods for generating the synthetic data array, we discuss the results for each method separately.

1. Results with the FFT method

Figure 2 presents the results for the R/S method for three sizes of the data array. Even for the largest array the predictions are greatly different from the true values. Note also that for a given array size and regardless of the value of H, roughly the same value of h is predicted. In fact, Fig. 2 makes it clear that with increasing array size one eventually has $h \to 1$, *regardless of the value of H.* Thus the R/S method yields completely wrong results for a large data array that obeys the statistics of a FBM. We will return to this point shortly.

Figure 3 depicts the results with the wavelet decomposition method for three different array sizes. As can be seen, except for $H > 0.8$, as the size of the array increases the predictions appear to converge towards the true values. In contrast, except for $0.15 < H < 0.4$, the predictions of the covariance method, which uses the very definition of the FBM, show large deviations from the true values; these are shown in Fig. 4. Although with increasing size of the data array the predictions appear to converge towards the true values of the Hurst exponent, the convergence is not very fast.

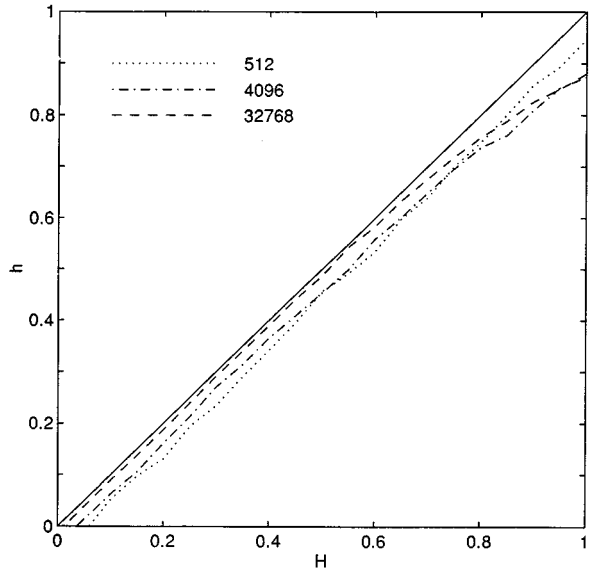

FIG. 3. Same as in Fig. 2, but with the wavelet decomposition method.

Figure 5 shows the predictions of the Lévy method. In this case, the largest data array contained 10^4 points, as the analysis of the data with this method requires large computer times. Although we obtained $\alpha \simeq 2$, as we should for a FBM, except for $0.15 < H < 0.3$, the agreement between the predictions and the true values of H is not good, although there is a convergence trend with increasing size of the data array. These results indicate that a FFT method does not generate a very accurate FBM array, unless the size of the lattice is very large and care is taken to ensure that periodicity effects associated with finite lattices do not have a significant effect. Moreover, for $H > 0.5$ the size of the system has to be larger

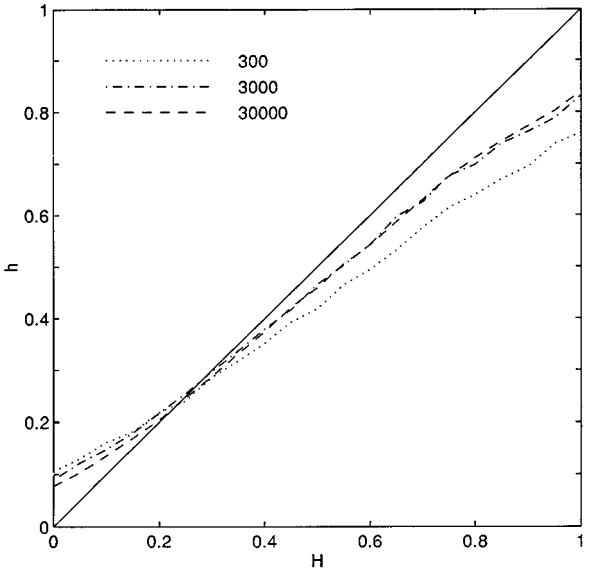

FIG. 4. Same as in Fig. 2, but with the covariance method.

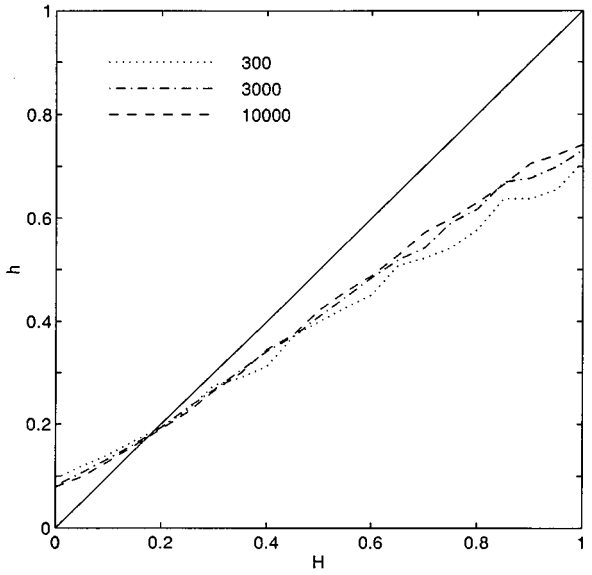

FIG. 5. Same as in Fig. 2, but with the Lévy method of analysis.

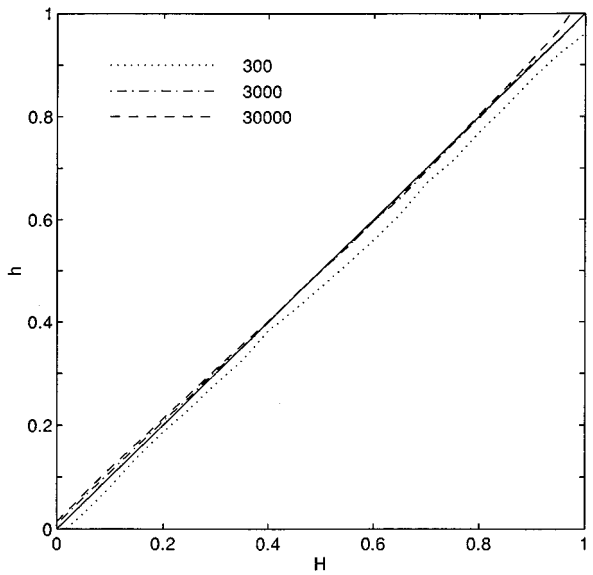

FIG. 7. Same as in Fig. 2, but with the maxiumum entropy method.

than that for $H<0.5$. This assertion is confirmed below where we discuss the results with the SRA and WM methods.

Figure 6 presents the results with the windowed FT method. The agreement between the predictions and the true values is excellent for the two larger sizes of the data array over the entire range of H. The convergence of the predictions towards the true values also appears to be very fast. However, note that since a FFT was used for generating the arrays, it is not really surprising that windowed FT method provides such accurate predictions. Even more accurate predictions are provided by the maximum entropy method, whose predictions are shown in Fig. 7. As can be seen, even a data array as small as 300 points provides predictions that are virtually indistinguishable from the true values. This is a great advange of this method since although the wavelet and windowed FT methods also provide accurate predictions, their accuracy is achieved only when the size of the data array is very large, which in practice does not occur very often. For example, one hardly ever has a porosity log with more than 1000–2000 data. Finally, Fig. 8 presents the predictions of the roughness-length method. It is clear that the convergence of the predictions towards the true values is not systematic since an array of 3000 data points appears to provide more accurate predictions than an array of 10^4 points, if $H>0.5$, with the reverse being true if $H<0.5$.

2. Results with the SRA method

The R/S method also provides very poor estimates of the true values of H if the array is generated by the SRA method, implying that the inaccuracy of the R/S method has nothing to do with the method of generating the synthetic data array.

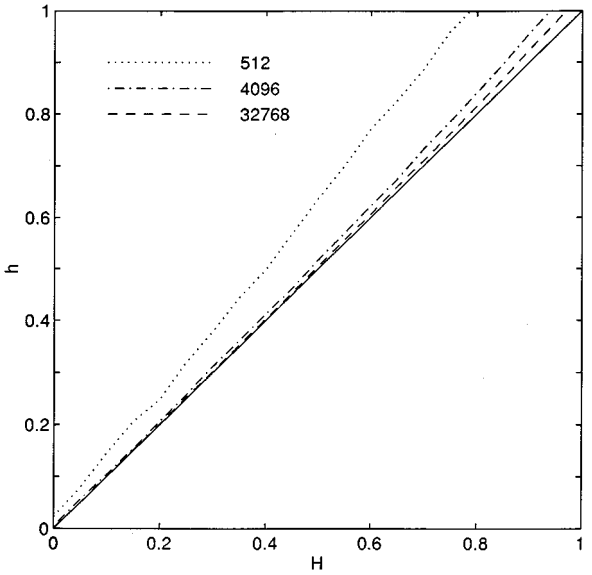

FIG. 6. Same as in Fig. 2, but with the windowed Fourier transformation method.

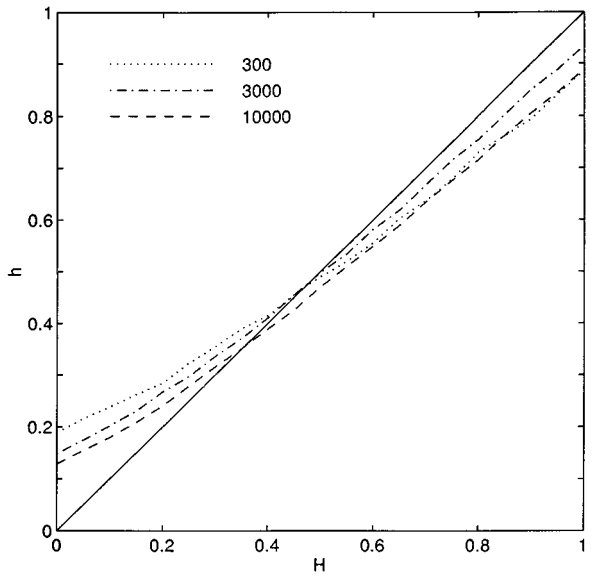

FIG. 8. Same as in Fig. 2, but with the roughness-length method.

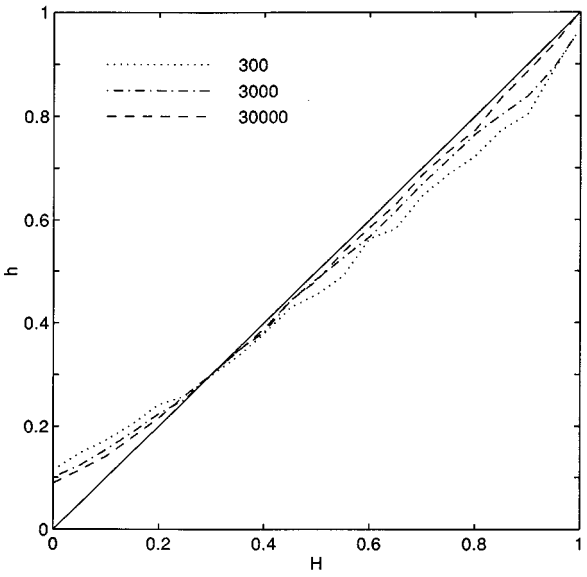

FIG. 9. Estimated Hurst exponent h versus its true value H as predicted by the covariance method. The data arrays were generated by the successive random addition method.

As in Fig. 2, the predicted Hurst exponent is more or less independent of H, with $h \rightarrow 1$ with increasing size of the data array. The predictions with the wavelet method are only slightly more accurate than those obtained with the FFT array (see Fig. 3) and, moreover, for $H > 0.9$ the approach to the true values is not systematic as the array size becomes larger. Unlike the results shown in Fig. 4, the covariance method provides relatively accurate predictions for the true values of H, except when $H < 0.2$; see Fig. 9. This is not really surprising, as the SRA method is in some sense the inverse of the covariance method. This indicates that the accuracy of the covariance method is sensitively dependent upon the accuracy of the data array. Since this method is

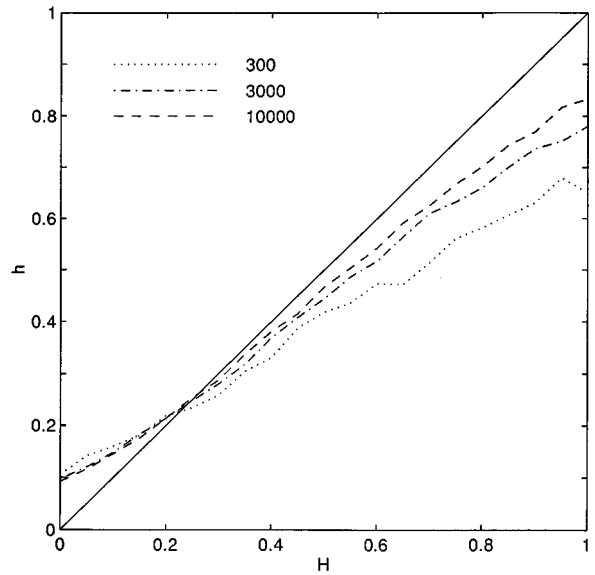

FIG. 10. Same as in Fig. 9, but with the Lévy method of analysis.

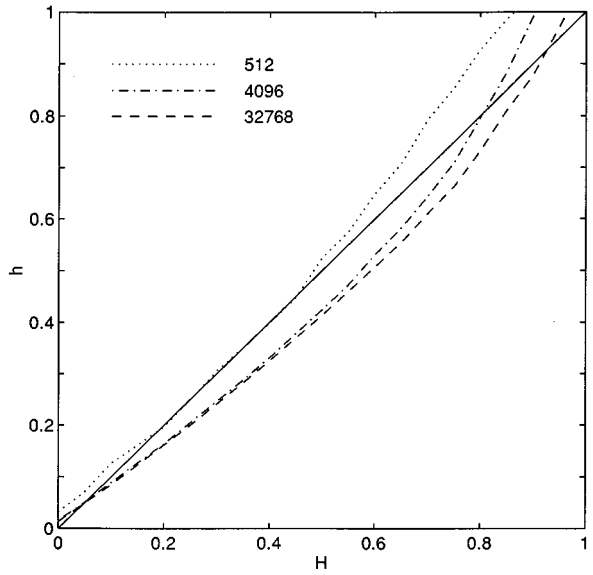

FIG. 11. Same as in Fig. 9, but with the windowed Fourier transformation method.

based on the very definition of the FBM, we may conclude that, for a given array size, the SRA method provides a more accurate technique for generating the FBM array.

Figure 10 depicts the results obtained with the Lévy method of analyzing the data. The agreement between the predictions and the true values of H is better than those presented in Fig. 5, which were obtained using the data arrays generated by the FFT method. However, unlike the results with the FFT method shown in Fig. 6, the windowed FT method does not provide accurate predictions of, and systematic convergence towards, the true values of H if the data array is generated by SRA method; these are shown in Fig. 11. Similar to the results shown in Fig. 7, which were obtained with the data arrays generated by the FFT method, the maximum entropy method provides again highly accurate

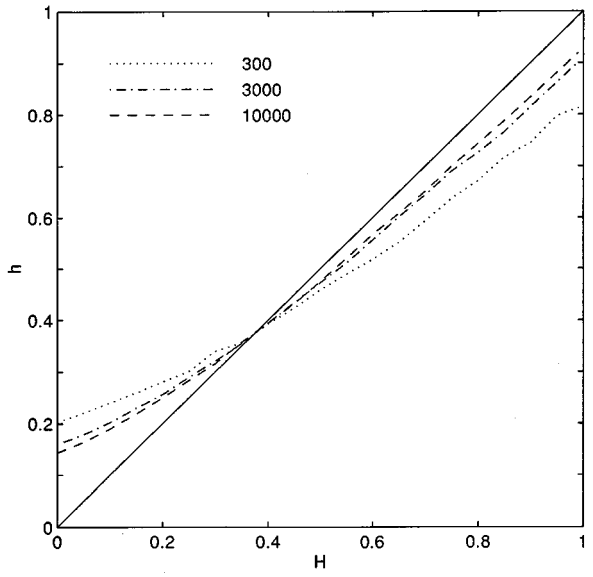

FIG. 12. Same as in Fig. 9, but with the roughness-length method.

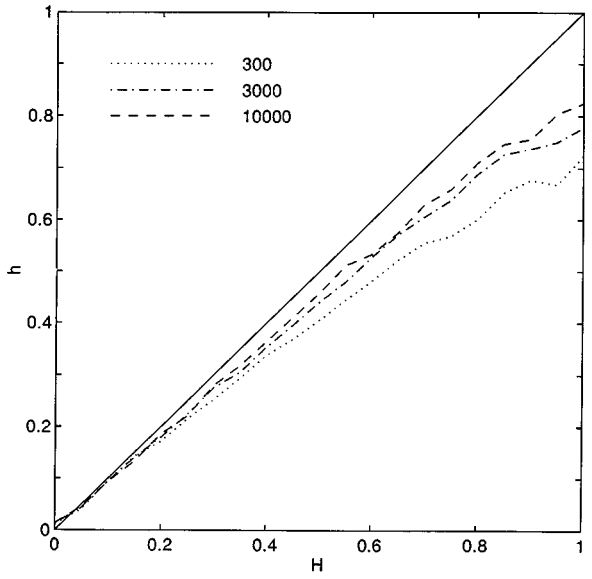

FIG. 13. Estimated Hurst exponent h versus its true value H as predicted by the Lévy method. The data arrays were generated by the Weierstrass-Mandelbrot algorithm.

predictions for the true values of the Hurst exponent when the data array is generated by the SRA method. Again, even an array as small as 300 data points provides very accurate estimates of H. The results with the roughness-length method, which are shown in Fig. 12, indicate that, unlike the case with the FFT method, with the SRA method of generating the data arrays, the convergence towards the true value of H is systematic as the array size increases.

3. Results with the WM method

As mentioned above, the power spectrum of the data array generated by the WM method is discrete and does not contain all the frequencies. Therefore, we analyzed the data array generated by this method only with the four methods that do not require the power spectrum of the data. The analysis of the WM data by the R/S method yields results completely similar to those obtained with the FFT and SRA methods. We thus conclude that the inaccuracy of the R/S method, a commonly used method of analyzing complex data and profiles, has nothing to do with the method of generating the data or its accuracy. In practical terms, this means that this method should not be used for analyzing real data unless one has *a priori* some information about the nature of the correlations in the data.

As with the case of the SRA method (Fig. 9), the covariance method provides accurate estimates of the true Hurst exponent H and the convergence of its predictions towards the true values with increasing data array size is systematic. Thus we conclude that, relatively speaking, the SRA and the WM methods generate more accurate FBM arrays than does the FFT method, unless the size of the data array is large. Compared with the results with the FFT and SRA methods (Figs. 5 and 10), the Lévy method yields more accurate predictions for the true values of H if the data arrays are generated by the WM method; these are shown in Fig. 13. This is also the case with the roughness-length method.

Summarizing our results, the wavelet decomposition, windowed FT, and the maximum entropy methods all provide highly accurate characterizations of long-range correlations in synthetic data. However, in terms of the efficiency of the computations and the required size of the data array, the maximum entropy method offers the best tool for analyzing a given data array.

B. Real data

We now analyze two sets of real data. Since our results with the synthetic data indicate that the maximum entropy method is the most efficient and accurate tool of analyzing complex data and their long-range correlations, we use this method here. The data that we analyze are a porosity log of a carbonate oil reservoir in the Middle East, measured along a vertical well, and the time variations of the pressure fluctuations in three-phase flow in a fluidized bed. Both sets of data have been analyzed previously by the R/S method, leading, as we show here, to erroneous results and conclusions.

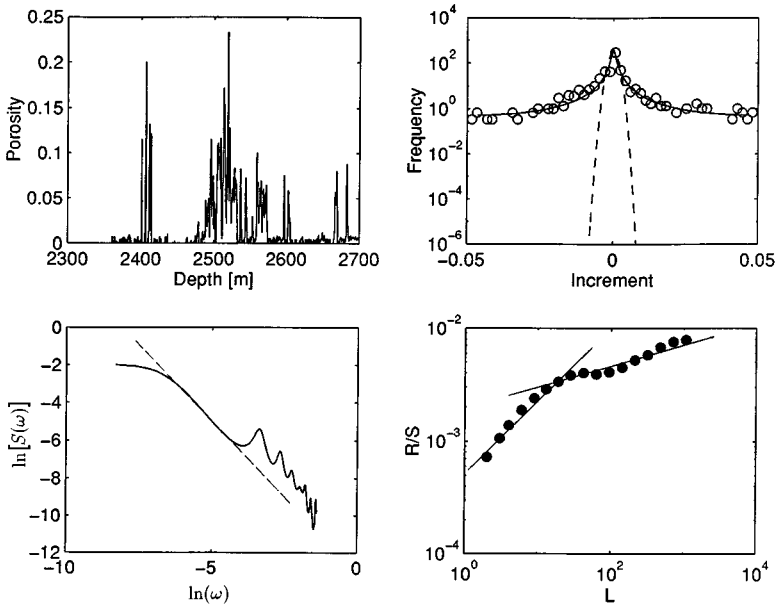

FIG. 14. Porosity log (top left), fit of its increments (top right) by a Lévy distribution (curve) and a Gaussian distribution (dashed curve), and the analysis of the log by the maximum entropy method (bottom left) and by the R/S method (bottom right).

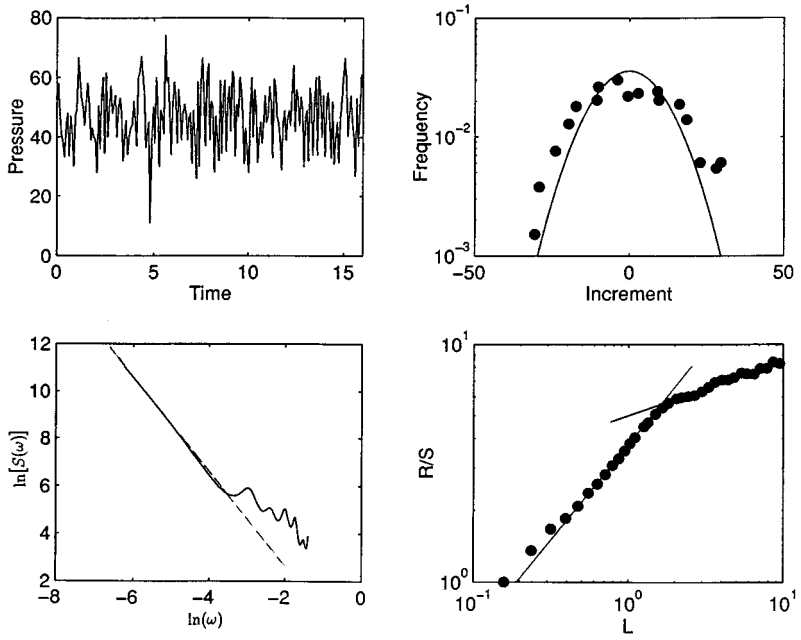

FIG. 15. Pressure fluctuations versus time in three-phase flow in a fluidized bed (top left), fit of its increments by a Gaussian distribution (top right), and its analysis by the maximum entropy method (bottom left) and by the R/S method (bottom right).

1. Porosity log of an oil reservoir

Figure 14 shows the porosity log measured between the depths 2350 and 2700 m. To see whether the data actually obey the statistics of a FBM (or a FGN, its ''derivative'') [3], we formed the increments in the data and constructed their frequency distribution. If the data do obey the FBM statistics, then the frequency distribution of their increments must be Gaussian. Figure 14 demonstrates clearly that this is not the case. Instead, the Lévy distribution with $\alpha \simeq 0.87$ appears to provide an excellent fit of the increments in the data. Most carbonate oil fields in the Middle East have very low porosities [1,2]. Thus, in this porosity log the rare event, characteristic of the Lévy stable distribution, may be indicative of a large fracture or fault whose porosity is much larger than that of the matrix around it. Indeed, the oil reservoir, whose porosity log we analyze, is known to be fractured.

Figure 14 also shows the analysis of the data by the maximum entropy method. As can be seen, at large frequencies, i.e., at short distances, the power spectrum of the data is noisy. This is perhaps due to the fact that the measurement equipment does not have high enough resolution to accurately measure the porosities at two close neighboring points and distinguish between them. At any point, there may also be interference from the neighboring points, which could also give rise to the noise in the data. If we fit the spectrum to Eq. (4) (with $d=1$), ignoring the noisy part, we obtain $H \simeq 0.3 < 1/\alpha$, indicating the existence of long-range negative correlations in the data; with the noisy part inluded we obtain $H \simeq 0.12$. The porosity log is consistent with $H < 1/2$, as a large value of the porosity is followed by a low value, and vice versa. Thus the maximum entropy method also allows one to separate the noisy part of the data and obtain a realistic value of H. Since the coefficients $C(x)$ [see Eq. (14)] obey Eq. (15), we also analyzed the data using this equation. We found $H \simeq 0.34$, consistent with the estimate obtained with maximum entropy method. For comparison, we also analyzed the data by the R/S method, the results of which, shown in Fig. 14, indicate the existence of two distinct regions. For large L's a fit of the data to Eq. (11) yields $H = 0.2$, whereas for small L's we obtain $H \simeq 0.65$.

2. Pressure fluctuations in a fluidized bed

Gas-liquid-solid fluidized beds, in which solid particles are fluidized by a liquid and a gas phase, such as water and air, have wide applications in the chemical, petrochemical, and biochemical industries. The pressure in the bed fluctuates widely and appears to be a stochastic variable. The stochasticity may be caused by a variety of factors, such as the motion of the fluidized particles. Modeling such fluctuations can help predict the behavior of the system, an unsolved problem.

Figure 15 shows typical pressure fluctuations versus time in a three-phase fluidized bed, reported by Fan *et al.* [29]. Unlike the porosity data, the distribution of the increments is roughly Gaussian; see Fig. 15. This means that the data are either completely random or obey the statistics of the FBM. The power spectrum of the data determined by the maximum entropy method is also shown in Fig. 15. As in the case of the porosity log of Fig. 15, there is a noisy part at high frequencies. A fit of the spectrum to Eq. (4), ignoring the noisy part, gives $H \simeq 0.5$, implying that the data are completely random with no correlations between them. On the other hand, if we use the R/S method, we obtain the results that are also shown in Fig. 15, which are in agreement with those obtained by Fan *et al.* [29]. As in the case of the porosity log, there are two distinct segments to the results. If we fit the apparently linear part to Eq. (11), we obtain $H \simeq 0.8$, indicating the existence of long-range positive correlations. However, this is clearly wrong, as a high value of the pressure is followed by a low value and vice versa. Thus, as our analysis with the synthetic data already indicated, the R/S method, which has been used heavily in the past, is not a reliable method of estimating H and characterizing the nature of long-range correlations.

V. SUMMARY AND CONCLUSIONS

We have carried out an extensive analysis of complex distributions and their long-range correlations, in the form of

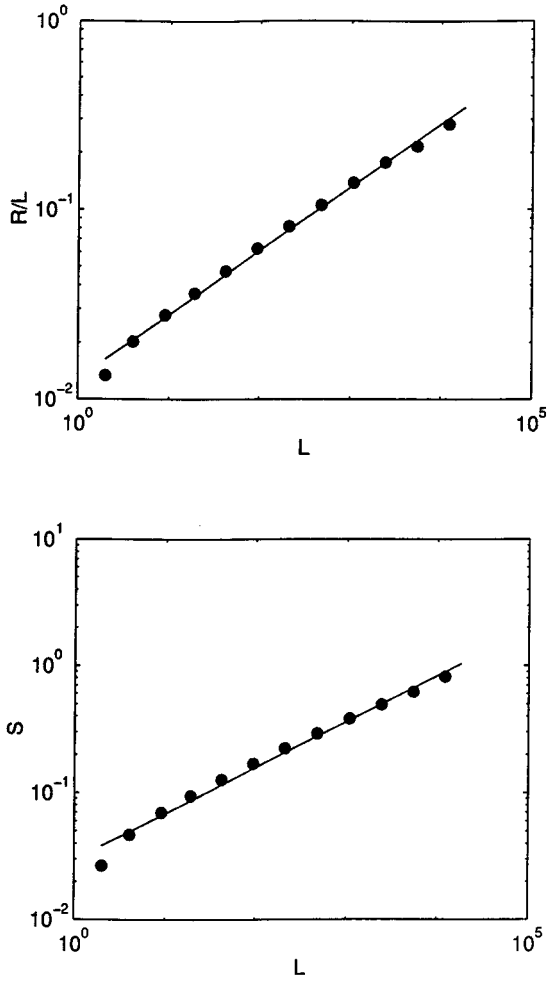

FIG. 16. Logarithmic plots of $R(L)/L$ and $S(L)$ versus L for a FBM array of 32 000 points. The true value of H is 0.33, while the slope of the straight lines is $H \simeq 0.34$ (top) and $H \simeq 0.36$ (bottom).

synthetic and real data. The synthetic data were generated by a fractional Brownian motion using three different methods and the resulting data arrays were analyzed with seven different methods. Our analysis indicates that the size of the data array greatly influences the accuracy of the estimates of the various parameters that one may wish to extract from the data.

A surprising result of our study is that the commonly used rescaled-range method of analysis appears to be unreliable. Mandelbrot and Wallis [6] have already pointed out several weaknesses of this method, some of which are as follows. (i) The plot of log R/S versus log L may not be linear, unless the sample size is large. In our study we do use large data arrays. Even our real data arrays are relatively large. (ii) The estimate h of the Hurst exponent approaches the true value H only asymtotically. However, we find that with increasing size of the data array the estimated Hurst exponent approaches 1 instead of its true value. (iii) The R/S method is *not* robust to cyclic effects. In our work we always generated a large array and analyzed only the central part of it; thus this effect should not be important.

We believe that the theoretical reason for the failure of the R/S method is as follows. Since for a FBM one has [9]

$$B_H(bx) - B_H(0) = b^H [B_H(x) - B_H(0)], \tag{21}$$

where b is any scale factor, then the function $B_H(\Delta x)$ should be proportional to $|\Delta x|^H$. This implies, according to Mandelbrot and Wallis [6], that $R(x) \sim x^H$, or

$$R(L) \sim L^H. \tag{22}$$

Feder [9] states that the true variance for the normalized FBM is $S = 1$; thus, together with Eq. (22), one obtains Eq. (11). However, as mentioned above, it is well known that the variance of a large-enough FBM array is divergent. According to Eq. (10) for large L, S^2 is proportional to the variance of the FBM, which, together with Eq. (2), implies that $S^2(L) \sim L^{2H}$ and thus

$$S(L) \sim L^H. \tag{23}$$

This is precisely what we find in our synthetic data; see Fig. 16. We also find that, contrary to Eq. (22),

$$R(L) \sim L^{1+H}. \tag{24}$$

This can also be seen by inspecting Eqs. (8) and (9). If we replace the variable v by $B_H(x)$, then Eqs. (8) and (9) tell us that $R(L) \sim \Sigma[B_H(x) - \langle B_H(x) \rangle]$. Since the argument of the sum is proportional to L^H, if we replace the sum by an integral, we obtain Eq. (24), and therefore $R(L)/S(L) \sim L$. This means that *if the data array is large enough, one should always obtain a Hurst exponent $H = 1$*, in agreement with our results. To the best of our knowledge, Eq. (24) has never appeared in the literature before. We also mention that Hardy and Beier [30] state that ''[the] R/S analysis fails for [FBM] since it yields $H = 1.0$,'' in complete agreement with our results, although these authors do not provide any reasoning or results in support of their statement. However, we must point out that if we generate a FBM array, calculate the increments in the data array (that is, subtract the neighboring points in the array) to form a new array, and analyze it with the R/S method, the estimated values of the Hurst exponents are in reasonable agreement with their true values. However, the new array is essentially a FGN array and therefore the R/S method is relatively accurate if the data array obey the statistics of a FGN, but not a FBM.

On the other hand, Eqs. (23) and (24) suggest alternative methods of analyzing the FBM data: Make a logarithmic plot of $S(L)$, $R(L)$, or $R(L)/L$ versus L. The slopes of the resulting straight lines are H, $1+H$, and H, respectively. As an example, we show in Fig. 16 the results for the R/L method. The size of the data array is 32 000, the true value of H is 0.34, and the plot yields $H \simeq 0.36$.

We find that the maximum entropy method offers a fast, efficient, and reliable method of analyzing the data and yields accurate estimates of the Hurst exponent even with a small data array. The wavelet decomposition method is also accurate, but its necessary computations and size of the data array for obtaining reliable results are much larger than those of the maximum entropy method. Thus we believe that in practical applications the maximum entropy method should be used and great caution should be taken if any other method is to be employed.

We also point out that, as Voss [31] suggested, the Hurst exponent H is related to the fractal dimension D_f of a FBM profile through $D_f = 2 - H$. This has been exploited by several authors [32–34] for estimating H by using a variational method to determine D_f. Thus the Hurst exponent can also be determined indirectly.

Finally, another important conclusion of our work is that when analyzing a given set of data, one should use two different methods to ensure the accuracy of the analysis since, for example, the spectral methods are more sensitive to the length of the data array, whereas the RL method is more sensitive to the true value of H.

ACKNOWLEDGMENTS

This work was supported in part by the Petroleum Research Fund, administered by the American Chemical Society. A.R.M. is grateful to the John and Alice Tylor Fund for financial support.

[1] M. Sahimi, Rev. Mod. Phys. **65**, 1393 (1993).
[2] M. Sahimi, *Flow and Transport in Porous Media and Fractured Rock* (VCH, Weinheim, 1995).
[3] T.A. Hewett, in *Proceedings of the 61st Annual Conference of the Society of Petroleum Engineers, New Orleans, 1986* (Society of Petroleum Engineers, Richardson, TX, 1986).
[4] B.B. Mandelbrot and J.W. van Ness, SIAM (Soc. Ind. Appl. Math.) Rev. **10**, 422 (1968).
[5] H.E. Hurst, R.P. Black, and Y.M. Simaika, *Long-Term Storage: An Experimental Study* (Constable, London, 1965).
[6] B.B. Mandelbrot and J.R. Wallis, Water Resour. Res. **5**, 228 (1969); **5**, 242 (1969); **5**, 260 (1969); **5**, 321 (1969); **5**, 967 (1969).
[7] J.R. Wallis and N.C. Matalas, Water Resour. Res. **7**, 1583 (1971).
[8] C.P. North and D.I. Halliwell, Math. Geol. **26**, 531 (1994).
[9] J. Feder, *Fractals* (Plenum, New York, 1988).
[10] J.C. Gallat, I.D. Moore, M.F. Hutchinson, and P. Gessler, Math. Geol. **26**, 455 (1994).
[11] J. Schmittbuhl, J.-P. Vilotte, and S. Roux, Phys. Rev. E **51**, 131 (1995).
[12] B.B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, San Francisco, 1982).
[13] B.B. Mandelbrot, Phys. Scr. **32**, 257 (1985).
[14] *Dynamics of Fractal Surfaces*, edited by F. Family and T. Vicsek (World Scientific, Singapore, 1991).
[15] B.B. Mandelbrot, D.E. Passoja, and A.J. Paullay, Nature (London) **308**, 721 (1984).
[16] E. Bouchaud, G. Lapasset, and J. Planes, Europhys. Lett. **13**, 73 (1990).
[17] S.R. Brown and C.H. Scholz, J. Geophys. Res. **90**, 12 575 (1985).
[18] R.F. Voss, in *Fundamental Algorithms for Computer Graphics*, edited by R.A. Earnshaw, Vol. 17 of *NATO Advanced Study Institute, Series E: Applied Science* (Springer-Verlag, Heidelberg, 1985), p. 805.
[19] P. Flandrin, IEEE Trans. Inf. Theory **38**, 910 (1992).
[20] S.G. Mallat, IEEE Trans. Pattern. Anal. Mach. Intell. **PAMI-11**, 674 (1989).
[21] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes*, 2nd ed. (Cambridge University Press, Cambridge, 1992).
[22] M.F. Shlesinger, Annu. Rev. Phys. Chem. **39**, 269 (1988).
[23] M.S. Taqqu, J. Geophys. Res. **92D**, 9683 (1987).
[24] S. Painter and L. Paterson, Geophys. Res. Lett. **21**, 2857 (1994).
[25] M. Sahimi, H. Rassamdana, and A.R. Mehrabi, in *Fractal Aspects of Materials*, edited by Fereydoon Family, B. Sapoval, P. Meakin, and R. Wool, MRS Symposia Proceedings No. 367 (Materials Research Society, Pittsburgh, 1995), p. 203.
[26] S. Painter, G. Beresford, and L. Paterson, Geophysics **60**, 1187 (1995).
[27] S. Painter, Water Resour. Res. **32**, 1183 (1996).
[28] E.F. Famma and R. Roll, J. Am. Stat. Assoc. **63**, 817 (1968); **66**, 331 (1971).
[29] L.T. Fan, D. Neogi, M. Yashima, and R. Nassar, AIChE. J. **36**, 1529 (1990).
[30] H.H. Hardy and R.A. Beier, *Fractals in Reservoir Engineering* (World Scientific, Singapore, 1994), p. 47.
[31] R.F. Voss, Physica D **38**, 362 (1989).
[32] B. Dubuc, J.F. Quiniou, C. Roques-Carmes, C. Tricot, and S.W. Zucker, Phys. Rev. A **39**, 1500 (1989).
[33] K.R. Sreenivasan and A. Juneja, Bull. Am. Phys. Soc. **37**, 1740 (1992); K.R. Sreenivasan and A. Juneja (unpublished).
[34] A. Scotti, C. Meneveau, and S.G. Saddoughi, Phys. Rev. E **51**, 5594 (1995).