Geophys. J. Int. (2003) **153,** 201–212

Rescaled-range and power spectrum analyses on well-logging data

Chun-Feng Li

Department of Geosciences, University of Tulsa, Tulsa, OK 74104-3189, *USA. E-mail: chunfeng-li@utulsa.edu*

Accepted 2002 October 28. Received 2002 October 8; in original form 2001 December 7

SUMMARY

Fractal theory and related developments have significantly enlightened our understanding of many natural phenomena and led to an upsurge in research efforts in many different disciplines. This has unfortunately caused some ambiguities, misunderstandings, and even misuses in the published literature concerning some new concepts in fractal theory and their actual implementations. As new applications and ideas arise almost daily, it is important to present the methods in a common framework to minimize further confusion. In this paper, we scrutinize and compare two of the most frequently applied methods for characterizing fractal time-series, namely the rescaled-range (R/S) and the power spectrum method, and describe their proper applications in wire line well-logging data. We outline the problems related to applications of these two methods. Based on numerical analysis on fractional Gaussian noise (fGn) and fractional Brownian motion (fBm), we demonstrate the clear necessity to distinguish between fGn-like and fBm-like data series. For fBm-like data series, the use of their successive increments is proposed against using various other alternatives in R/S analysis in order to make the estimated Hurst exponent (Hu) comparable with the global scaling exponent (H) from power spectrum analysis. We argue that well-logging data are generally analogous to fBm, and thus we use their incremental series rather than raw well data themselves to estimate Hu when applying R/S. We also reveal a connection between the transient zone length in the R/S technique and the shortest-wavelength that can be included for a linear regression in the power spectrum method.

Key words: borehole geophysics, Fourier transforms, fractals, geostatistics, numerical techniques.

1 INTRODUCTION

Careful studies have revealed some problems in the current application of rescaled-range analysis (R/S) and the power spectrum in determining the Hurst exponent (Hu) and the fractal dimension (D). In particular, it appears to be evident that R/S and power spectrum methods rarely give similar estimates of Hu, and in many cases differ significantly if not performed with particular care. Individually, both R/S and the spectrum method have been criticized for their inaccuracies (Fox 1989; Hough 1989; Lo 1991; Bassingthwaighte & Raymond 1994; North & Halliwell 1994). However, discussions on the obvious discrepancies between the two are often safely ignored. It is our view that neglecting this issue could hamper true understanding of Hu and D, and could make less meaningful estimates. This paper, through a detailed mathematical analysis and case studies, is meant to clarify some of the ambiguities involved in their applications and make results from different authors and/or techniques more consistent and comparable.

The concept of the Hurst exponent was first introduced by Hurst (1951) via his well-known empirical relationship, called the rescaled-range technique, in studying long-term discharge variations of the River Nile, and later publicized and solidified by Mandelbrot & Wallis (1969) and Feder (1988). Since then it has gained popularity and been applied in many subjects, including network traffic, bioengineering, physical simulation and of course, geology and geophysics. The key point of this technique is the following empirical relationship:

$$R/S = (a\tau)^{Hu}, \tag{1}$$

where a is a constant, R is the difference between the maximum and minimum accumulated departure from the mean within time span τ, S is the standard deviation serving as a scaling factor and Hu is the Hurst exponent coined by Mandelbrot (1982). The accumulated departure with respect to the mathematical mean $\langle\xi\rangle$ of a time-series ξ_i is

$$X(t) = \sum_{i}^{N}(\xi_i - \langle\xi\rangle), \tag{2}$$

where t is time, i is the time sample index and N is the number of time samples. The range R is defined as

$$R(\tau) = \max_{1\le t\le\tau} X(t) - \min_{1\le t\le\tau} X(t) \tag{3}$$

and $S(\tau)$ is the standard deviation

$$S(\tau) = \left[\frac{1}{\tau} \sum_{t=1}^{\tau} (\xi_t - \langle \xi \rangle)^2 \right]^{1/2}. \quad (4)$$

Eq. (1) clearly reveals a scaling-invariant property of the data with respect to the period τ. Specifically, the rescaled range R/S is linearly proportional to $a\tau$ on a log–log plot, the slope of which is equal to Hu.

Among various definitions for a fractal time-series, one of the commonly used ones is

$$F(\sigma t) = \sigma^H F(t). \quad (5)$$

It describes a self-affine process, i.e. a subset of a fractal process is a scaled version of the fractal set itself. We call H in eq. (5) the global scaling exponent and σ is a scaling operator. For a self-affine time-series defined by eq. (5), it can be shown that its power spectral density $S(f)$ has a power-law dependence on frequency f:

$$S(f) \propto f^{-\beta}, \quad (6)$$

where we call β the spectrum power. Turcotte (1997) showed that the following relationship between H and β exists:

$$\beta = 2H + 1. \quad (7)$$

We are now confronted with a dilemma as to whether the global scaling exponent (H) derived from eq. (7) is equal to the Hurst exponent (Hu) from the R/S technique. In many studied cases no distinctions are made between H and Hu, and they are often used interchangeably (Hewett 1986; Shiomi *et al.* 1997; Dolan *et al.* 1998; Taqqu & Teverovsky 1998), while in a few situations H is differentiated from Hu (Turcotte 1997). This conflict has posed confusion and misinterpretation in estimating Hu. It has been observed that estimating Hu and H for the same data series, but using two different techniques above, can give significantly different results. This prompts us to consider in what sense H and Hu can be considered equivalent.

2 IS H EQUAL TO Hu?

To answer this question it is helpful to discuss two basic fractal processes that are closely related: the fractional Brownian motion (fBm) and fractional Gaussian noise (fGn). fGn is the successive steps or increments of fBm; in other words, the accumulative summation of fGn gives fBm. While fBm is non-Gaussian and non-stationary, fGn is Gaussian and stationary.

It is straightforward to prove that for fBm with spectrum power $1 \le \beta_{\mathrm{fBm}} \le 3$, the corresponding fGn has $-1 \le \beta_{\mathrm{fGn}} \le 1$. An fBm series $X(t)$ is nowhere differentiable, but in a distribution when the subsequent interval $\Delta X = X(t_i) - X(t_{i-1})$ is sufficiently small compared with the total length of X, from the definition of the derivative we can write

$$G(t) = \Delta X = X(t_i) - X(t_{i-1}) \cong \Delta t X'(t), \quad (8)$$

where Δt is a time sampling interval. Taking the first derivative of $X(t)$ in the time domain is equivalent to multiplying $X(f)$, which is the Fourier transform of $X(t)$, by $-i2\pi\ f$ (Bracewell 1965), $i = (-1)^{1/2}$. By taking the Fourier transform of both sides of (8), we have

$$G(f) \cong \Delta t[-2\pi i f X(f)]. \quad (9)$$

Here we see that the non-differentiability of fBm can actually be relaxed in the frequency domain. The power spectra are given by

$$|G(f)|^2 \cong |\Delta t[-i2\pi f X(f)]|^2 = -4\Delta t^2 \pi^2 f^2 |X(f)|^2. \quad (10)$$

Here $|\ G(f)|^2$ is the power spectrum of the fGn and $|X(f)|^2$ is the power spectrum of the fBm. Because $|X(f)|^2 \propto f^{-\beta}$, we have $|G(f)|^2 \propto f^{-\beta+2} = f^{-(\beta-2)}$. This proves that

$$\beta_{\mathrm{fGn}} \cong \beta_{\mathrm{fBm}} - 2. \quad (11)$$

Or from eq. (7),

$$H_{\mathrm{fGn}} \cong H_{\mathrm{fBm}} - 1. \quad (12)$$

The proof of eqs (11) and (12) reveals that β and H are indicators of the roughness of the data. The larger H is the smoother the data. By taking a first-order derivative we increase the roughness of the data but decrease the value of H by 1. Now eq. (7) can be extended to:

$$\begin{aligned} \beta_{\mathrm{fBm}} &= 2H_{\mathrm{fBm}} + 1 \\ \beta_{\mathrm{fGn}} &= 2H_{\mathrm{fGn}} + 1 \cong 2(H_{\mathrm{fBm}} - 1) + 1 = 2H_{\mathrm{fBm}} - 1. \end{aligned} \quad (13)$$

While H varies from 1 to -1 from fBm to fGn, we realize from eqs (1)–(4) that Hu will always lie between 0 and 1 (Mandelbrot & Van Ness 1968; Mandelbrot & Wallis 1969). Thus, in general, H should not be considered as equivalent to Hu. Ignoring the difference between H and Hu can lead to confusing results. What has often been assumed is to make $Hu_{\mathrm{fBm}} = Hu_{\mathrm{fGn}} = H_{\mathrm{fBm}}$, and because of this, eq. (3) is often unjustifiably simplified to

$$\begin{aligned} \beta_{\mathrm{fBm}} &= 2Hu + 1 \\ \beta_{\mathrm{fGn}} &= 2Hu - 1. \end{aligned} \quad (14)$$

In reality, $Hu_{\mathrm{fGn}} \approx H_{\mathrm{fBm}}$ but $Hu_{\mathrm{fBm}} \neq Hu_{\mathrm{fGn}}$ and $Hu_{\mathrm{fBm}} \neq H_{\mathrm{fBm}}$. Eq. (14) is to keep Hu unchanged from $1 \le \beta \le 3$ to $-1 \le \beta \le 1$ and is actually a sacrifice, realizing that the meaning of Hu has been altered from $1 \le \beta \le 3$ to $-1 \le \beta \le 1$, because theoretically the relation $\beta = 2H + 1$ holds to be true for $-\infty < \beta < +\infty$. In order for eq. (14) to be meaningful, one has to assume that Hu of fBm is Hu from fGn, the successive increments of fBm. Without this common consensus, results of R/S analysis from different authors are incompatible and significant differences exist between H and Hu. To avoid these problems we have to clearly identify whether the time-series under study is analogous to fBm or fGn. If it is analogous to a fBm with spectrum power $1 \le \beta \le 3$, we have to find the subsequent increments of this time-series and then apply R/S to the new incremental series. Otherwise, as will be shown later in the next section, H will not equal Hu and Hu is almost certainly larger than 0.85, making Hu a very limited quantity to be used in characterizing fBm-like time-series. This is equivalent to saying that, in a strict sense, Hu is only better suited for capturing the properties of noise, such as fGn with Hu between 0 and 1, but not for random walks such as fBm.

Because of the delicate relationship between Hu and H, extra caution needs to be taken in relating Hu and H to the fractal dimension (D). For self-affine time-series, Voss (1985) derived the following approximate relationship between the spectrum power β and the fractal dimension D for fBm ($1 \le \beta \le 3$):

$$\beta = 5 - 2D. \quad (15)$$

From eq. (7), we have the much celebrated relationship

$$D = 2 - H. \quad (16)$$

A direct analytical relationship between D and Hu has not been observed. However, for fGn, because $Hu_{\rm fGn} \approx H_{\rm fBm}$, eq. (16) can be extended to Hu so we can have

$$D \approx 2 - Hu. \tag{17}$$

This relationship cannot be extended for fBm because $Hu_{\rm fBm} \neq H_{\rm fBm}$. In order for eq. (17) to be applicable for fBm, we first need to obtain the associated fGn from fBm, calculate Hu from fGn, and take it as Hu of fBm.

We wish to point out that practically there is nothing wrong with applying R/S to fBm directly, but the Hu so obtained is not comparable to H from the power spectrum. In geophysical applications such as reservoir simulations, we value H as a better quantity than Hu because H captures the complexities of the data more accurately, while Hu is confined between 0 and 1. It is more realistic to say that R/S approximates the true scaling exponent H in eq. (5), but care must be taken when we relate H to Hu because these two quantities are not always interchangeable. The R/S technique can be used as a secondary method to roughly estimate H, mainly to validate the results from the spectrum method.

3 NUMERICAL SIMULATIONS

Using a frequency domain algorithm, we simulate 21 fBm series, each with 4096 data points. The input H for fBm ranges from 0 to 1 with a constant interval of 0.05. We could also simulate 21 fGn series using the same algorithm, but for the purpose of this article, we obtain the 21 fGn series directly from the 21 fBm series by taking successive increments. We then apply the R/S method to both fGn and fBm series to estimate Hu. It is expected that the H value from the power spectrum should be very close to the input H, for the simulation is itself based on a spectrum algorithm. It is nevertheless meaningful for us to check the robustness of our power spectrum method by estimating H on simulated fBm with the spectrum method. We also apply a third independent method called the variance of increment method to estimate H and validate our simulation.

Table 1 shows the results of analysis on simulated data. A rescaled-range analysis is applied on both fGn and fBm. We can immediately realize that R/S analysis on fBm almost surely gives an estimate of $Hu > 0.85$, as predicted. On the other hand R/S-estimated Hu from fGn is close to the estimate from the power spectrum and the variance of the increment analysis on fBm (Fig. 1). This further proves our earlier statement that, in order to make the estimate of Hu from R/S comparable to H from the power spectrum, we should apply R/S to the corresponding fGn series of fBm rather than to fBm itself. The final two columns of Table 1 show the spectrum power β of fBm and fGn. We can see that the relationships between β and Hu for fBm, $\beta_{\rm fBm} \approx 2Hu + 1$, and for fGn, $\beta_{\rm fGn} \approx 2Hu - 1$ are approximately true for our simulations, when the correct meaning of Hu is assigned.

Fig. 1 shows clearly that the rescaled range tends to overestimate the lower H values (for $H \lesssim 0.7$) and underestimate high H values (for $H \gtrsim 0.7$). This agrees well with results from other studies (Mandelbrot & Wallis 1969; Bassingthwaighte & Raymond 1994). Fig. 1 also shows an obvious linear correlation between Hu and H. An empirical relationship between Hu and H for fBm/fGn based on the least-squaresx regression is

$$H = 1.23Hu - 0.16. \tag{18}$$

Eq. (18) provides a simple way to roughly estimate the true scaling exponent H from a rescaled-range analysis for fGn/fBm. It is now straightforward to determine the crossover point where $H = Hu$. We find that when $H = Hu \approx 0.7$, the two methods should be able to give nearly identical results.

The initial 'transient zone' for the R/S technique persists when the input H is less than or equal to 0.5 (Fig. 2). Visually, the length of the transient zone is increasing from a lag spacing of 31 data points to approximately 63 data points when H increases from 0 to 0.5. The gradient of this zone is always higher than that of the rest of the data.

We find that when input H becomes larger than 0.5, the transient zone can no longer be clearly discernible from the rest, and it is safe to include this zone to calculate Hu. Meanwhile, however, another

Table 1. Results from R/S and power spectrum analyses on simulated fBm/fGn.

fBm no	Input H	R/S on fGn (Hu)	R/S on fBm (Hu)	Power spectrum (H)	Variance of increment (H)	β of fBm	β of fGn
1	0.0	0.173	0.900	0.012	0.054	1.024	−0.783
2	0.05	0.155	0.890	0.043	0.034	1.087	−0.763
3	0.1	0.226	0.938	0.114	0.139	1.228	−0.572
4	0.15	0.261	0.952	0.186	0.186	1.372	−0.452
5	0.2	0.309	0.947	0.206	0.219	1.411	−0.397
6	0.25	0.285	0.939	0.248	0.164	1.497	−0.327
7	0.3	0.323	0.970	0.324	0.272	1.648	−0.171
8	0.35	0.365	0.965	0.370	0.373	1.740	−0.104
9	0.4	0.443	0.995	0.411	0.371	1.822	0.000
10	0.45	0.508	0.996	0.463	0.529	1.927	0.117
11	0.5	0.569	1.004	0.504	0.531	2.009	0.197
12	0.55	0.580	0.991	0.541	0.561	2.081	0.261
13	0.6	0.676	0.999	0.631	0.576	2.261	0.407
14	0.65	0.670	0.993	0.654	0.575	2.308	0.485
15	0.7	0.764	0.999	0.700	0.756	2.399	0.579
16	0.75	0.786	0.997	0.772	0.702	2.543	0.714
17	0.8	0.784	1.000	0.807	0.688	2.614	0.763
18	0.85	0.789	0.966	0.849	0.637	2.699	0.855
19	0.9	0.837	1.005	0.922	0.714	2.844	1.003
20	0.95	0.863	1.000	0.986	0.618	2.971	1.097
21	1.0	0.886	1.000	1.012	0.542	3.024	1.188

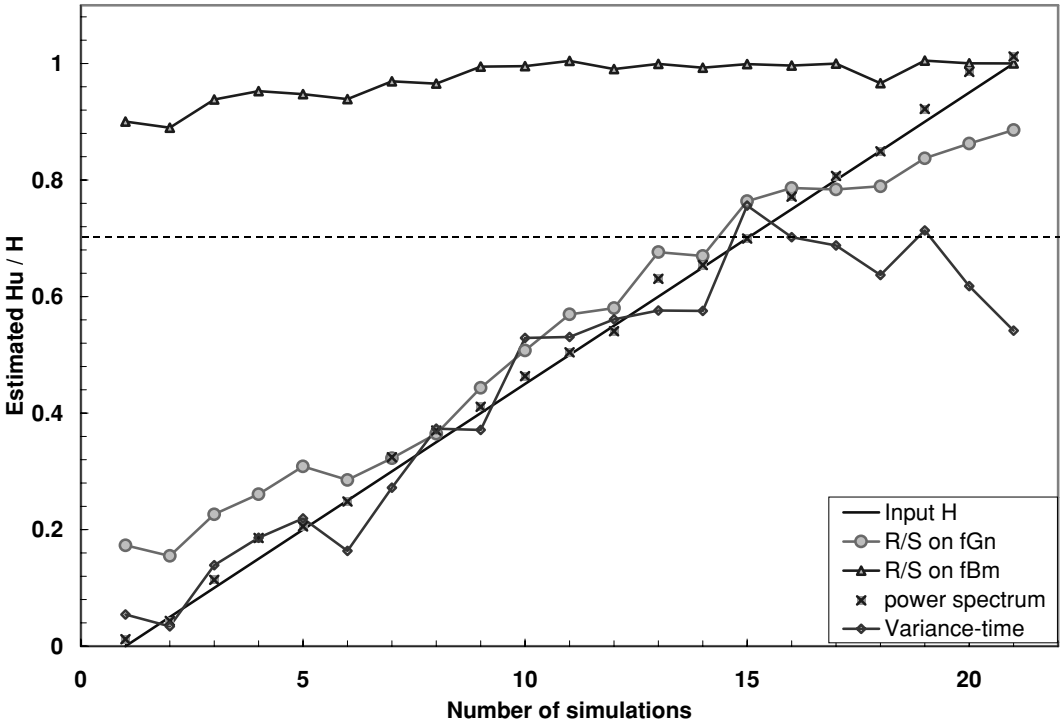

Figure 1. Results from *R*/*S* and power spectrum analyses on simulated fBm/fGn. The dashed horizontal line is $H = 0.7$. Note that results from *R*/*S* analysis on fBm is erroneously too high, but applying *R*/*S* on fGn gives good estimates of true *H* values.

problem may arise with increasing H. Fig. 3 shows the *R*/*S* plot for input $H = 0.85$. It may appear that there is an initial transient zone with large lag spacing. In fact, the low gradient zone at large lag spacing is another artefact that can be identified as the 'final tightening' in *R*/*S* analysis (Mandelbrot & Wallis 1969; Goggin *et al.* 1992). It is now vital how we treat these two parts. We may pick up the low gradient section in Fig. 3 to estimate *Hu* and take the high gradient section as the transient zone; it is expected that the result will be totally wrong if done this way. When the estimated *Hu* is large and when the high gradient zone extends to a wide range of lag spacing, we have to be very cautious in determining whether this high gradient zone can indeed be considered as the transient zone. It is most likely in these situations that we need to use the initial high gradient zone to estimate *Hu*, rather than using the low gradient part. On this point we agree with North & Halliwell (1994).

The variance of the increment method agrees excellently with the input H all the way from 0 to approximately 0.75, before making a sharp drop afterwards. This departure needs to be further analysed to understand its causes. It appears that both *R*/*S* and the variance of increment techniques tend to give very biased estimates for simulations at high H. Therefore, we should always be careful in accepting estimations from various techniques. Two or more independent techniques should be applied whenever possible for validating the results.

4 WELL-LOGGING DATA ANALYSIS

In the last section we focus primarily on fGn and fBm. Next, we study how our findings in earlier sections can contribute to solving real problems. For well-logging data even larger discrepancies between *R*/*S* and the power spectrum analysis could occur owing to the more complex nature of the real data. The noticeable intermittency and non-stationarity of well log data could bias the results of *R*/*S* or the power spectrum analysis in a profound way (North & Halliwell 1994). This brought to our attention that we need to be very cautious in extending the results of fGn/fBm to the real well data.

From the well log literature, we find that the Hurst exponents reported vary widely from 0 (Dolan *et al.* 1998) to greater than 1 (Goggin *et al.* 1992; Pang & North 1996). In cases where the *R*/*S* technique is applied alone the estimated *Hu* is consistently much larger then 0.5 (Hewett 1986; Leary 1991; Pang & North 1996), and in cases where the power spectrum technique is applied only the estimated *Hu* is consistently much less than 0.5 (Holliger 1996; Shiomi *et al.* 1997). This does not appear to be a coincidence. Only a few cases can be found where both *R*/*S* and the power spectrum are applied to give close estimations of H (Dolan & Bean 1997; Dolan *et al.* 1998). The wide range of *Hu* could indicate that different well-logging data at different regions could be either persistent ($Hu > 0.5$) or antipersistent ($Hu < 0.5$). And we also know that *R*/*S* tends to give biased result if *Hu* is not close to 0.7 (Mandelbrot & Wallis 1969; Feder 1988; Bassingthwaighte & Raymond 1994). However, this large discrepancy between these two techniques indicates technical problems in their applications. In one case, both Leary (1991) and Holliger (1996) analysed sonic log data from the Cajon Pass borehole located in southern California. Leary (1991) estimated *Hu* to be 0.7 using the *R*/*S* method, while Holliger (1996) obtained a much lower estimate of 0.11 from the power spectrum. The difference is so large that it prevents any correlation between the results. In particular, Holliger (1996) noticed that the high *Hu* estimate from Leary (1991) does 'not allow for the generation of synthetic data sets that bear any close resemblance to the observed sonic-log data'.

We see two main causes for these problems. First, fGn is assumed for what is actually an fBm process; that is, the *R*/*S* method is applied directly to time-series, which should be considered to be of fBm type. This leads to an erroneously high estimate of *Hu*. It is clear

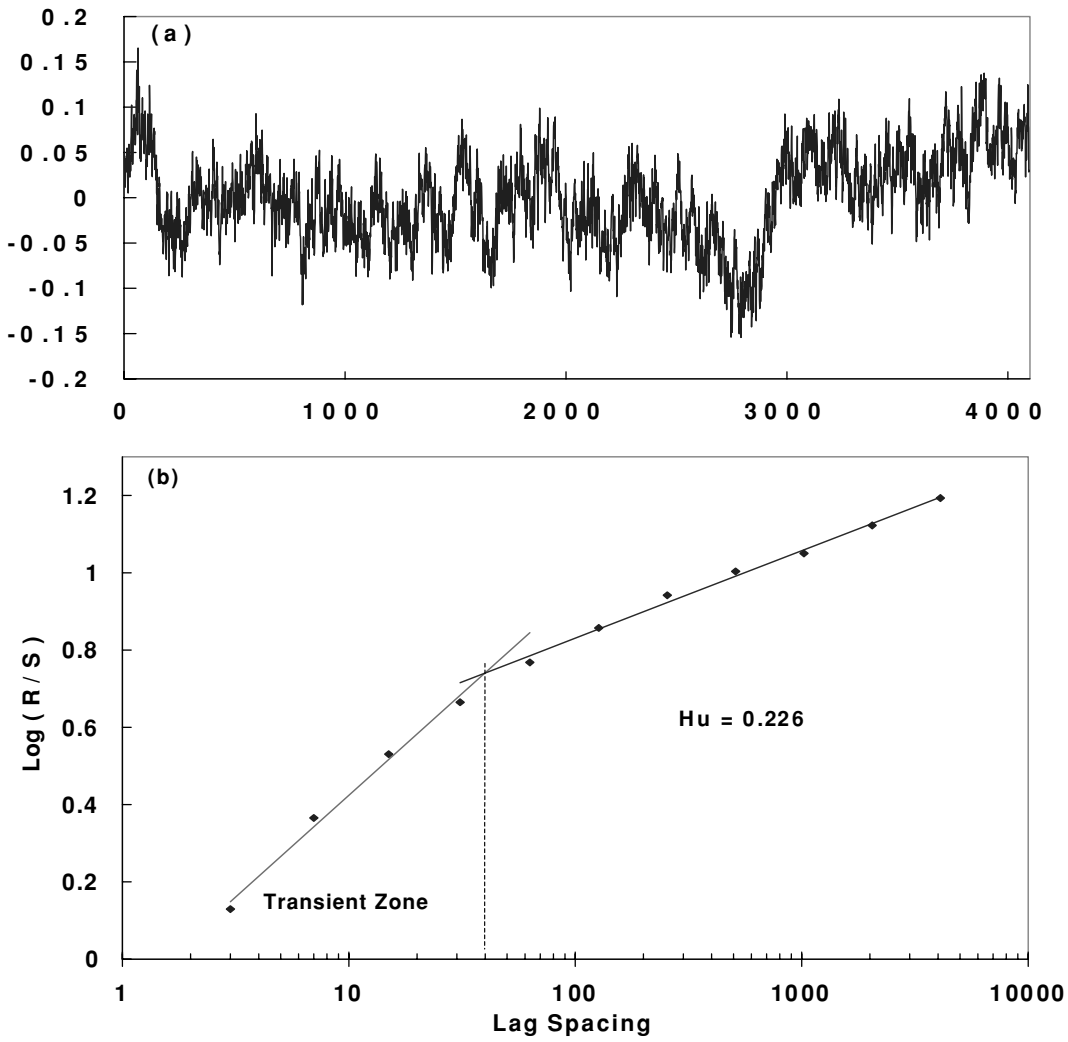

Figure 2. Plots showing the transient zone in the *R*/*S* analysis. (a) Simulated fBm no 3 with input $H = 0.1$. (b) *R*/*S* analysis on the simulation.

that from various studies that the spectrum powers β of raw well data are consistently greater than 1 (Dolan & Bean 1997; Dolan *et al.* 1998; Shiomi *et al.* 1997). Based on these observations and our own studies on wavelet-based multiscale analysis, we argue that raw well logs are fBm-type data. Obviously, if *R*/*S* is applied directly to raw well-log data instead of their incremental series, the estimated *Hu* will always be very large (>0.85), and become incomparable to the power spectrum analysis.

The second major causes for the discrepancy is that the estimated *Hu* is biased when *R*/*S* is applied to some modified forms of raw time-series. These modifications include trend removal and normalization. *R*/*S* is known to be sensitive to short-term dependences (Hipel & McLeod 1978; McLeod & Hipel 1978; Lo 1991), and this drawback can sometimes produce a profound bias (Moody & Wu 1996). The non-Gaussianity and the non-stationarity of raw well data can cause major problems in *R*/*S* analysis (North & Halliwell 1994; Taqqu & Teverovsky 1998). Taqqu & Teverovsky (1998) noted that for non-Gaussian series, *R*/*S* analysis becomes so inaccurate that it should only be used to obtain a very rough idea of the intensity of the long-range dependence. Trend removal is used to reduce the effect of short-wavelength variability by removing a moving average (MA) or low-order polynomial fitting of the data (Holliger 1996). This also effectively modifies the raw well data to be approximately Gaussian and stationary. Normalization is another way to reduce the effect of short-wavelength variability (Hewett 1986; Leary 1991). Well-logging data are generally non-Gaussian and non-stationary. By normalization we tend to obtain a normalized series of zero mean and unit variance. In these respects, we see the values for trend removal and normalization because these operations are likely to produce more accurate estimates of *Hu* from *R*/*S*.

However, we have learned in earlier sections that the appropriate way for estimating *Hu* of fBm-type data is to apply *R*/*S* to the incremental series. Would the data after trend removal or normalization be equivalent to an incremental series in terms of *Hu* estimation? In the following two case studies based on real well log examples, we find that *Hu* from well data after trend removal and normalization is lower compared with *Hu* obtained directly from the raw data. We will also illustrate that, by trend removal and normalization, we actually try to estimate *Hu* from successive increments of raw well data. Even so, we conclude that *R*/*S* on successive increments of raw well data should be used in estimating *Hu*. There are two main reasons for doing this. First, it is more objective to obtain increments than trend removal or normalization because unlike trend removal or normalization, there is only one incremental series for each raw data set. Secondly, estimating *Hu* on incremental series is easier and more accurate.

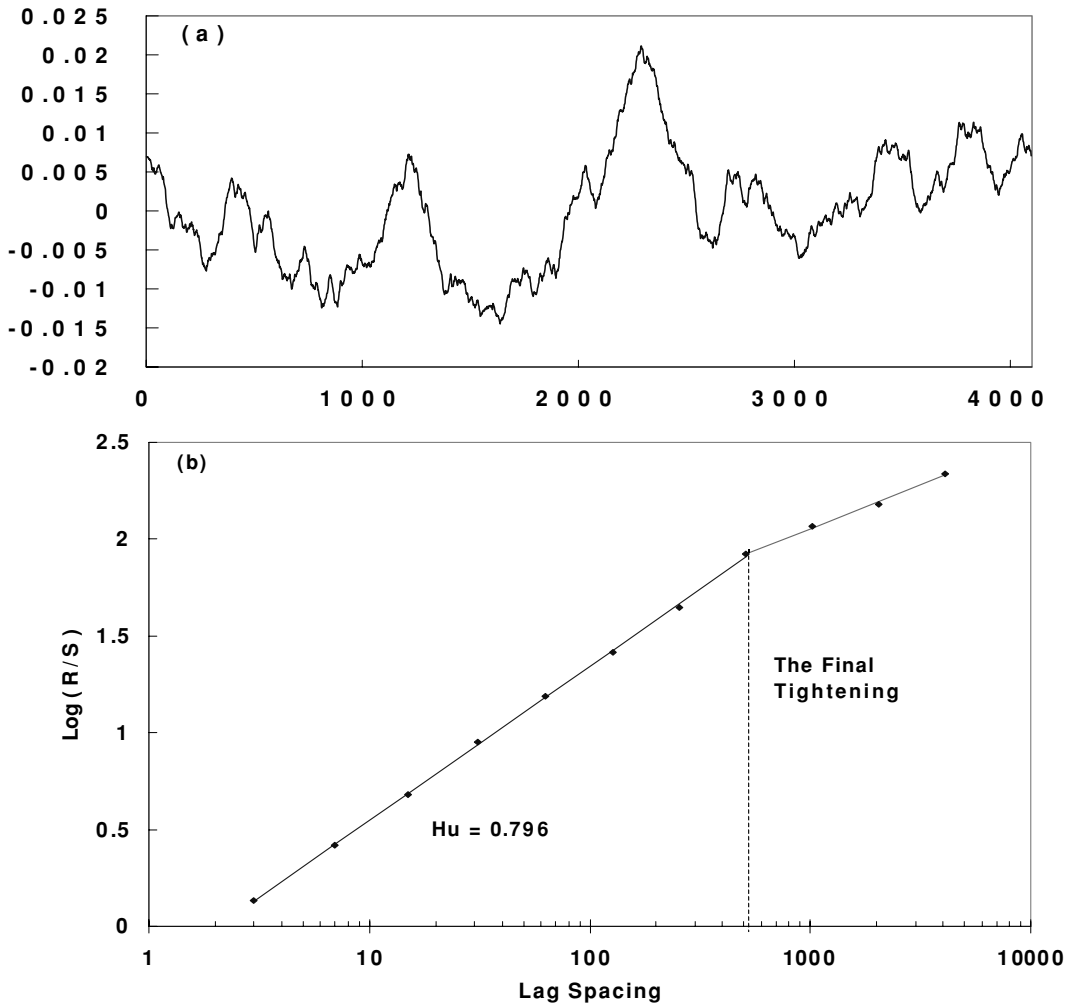

Figure 3. Plots showing the final tightening in the *R*/*S* analysis. (a) Simulated fBm no 18 with input $H = 0.85$. (b) *R*/*S* analysis on the simulation.

4.1 Gamma-ray data

The data, sampled at a constant rate of 0.5 ft for a total interval of 1984 ft, is from a well located in Panoma gas field, Grant County, southwest Kansas. Formation rocks penetrated by this well include shale, sandstone, siltstone, gypsum, anhydrite and dolomite. Fig. 4 shows the raw gamma-ray data (a), normalized data (b), data after moving average trend removal (c) and successive increments (d) from the raw data. To the right of these data series in Fig. 4 are their respective histograms. The raw data does not have a normal distribution, but all the others in the panel have distributions close to normal distributions. Unlike the raw gamma-ray data, we observe that modified data sets can be considered to be Gaussian/normal and stationary. Note here that normalization is not simply to extract the mean from the data and then divide the data by its standard deviation, because this will have little effect on removing short-wavelength variability. In fact, we estimate the time-varying variance and find the trend of the variance of non-stationary time-series data. We then normalize the data using the estimated variance.

Fig. 5 shows the rescaled range analysis on all data sets in Fig. 4. First, we see no transient zone for raw gamma-ray data, but distinct transient zones of various lengths (n) exist for modified data sets (b, c and d). Incremental series (d) gives the shortest transient zone. All observable transient zones are overlapping with the *R*/*S* plot of the raw data; in other words, transient zones from different data sets (b, c and d) have the same gradient on the *R*/*S* plot. For the raw gamma-ray data (Fig. 6a), the blindness of *R*/*S* to the transient zone may have an adverse effect on the *Hu* estimation if *R*/*S* is applied directly to raw data. This may also help to explain why various modifications to the raw data are necessary. Secondly, *R*/*S* analyses on modified data sets (b, c and d) give similar estimates of *Hu* regardless of how the raw data are modified. In this example, $Hu_{\text{increment}}$, $Hu_{\text{normalized}}$ and Hu_{MA} are 0.518, 0.378 and 0.54, respectively. On the other hand, *Hu* estimated from raw data is 0.91, which is substantially higher. Finally, we find that the incremental data (d) gives the best estimate of *Hu* because it has the shortest transient zone and the highest correlation coefficient ($R = 0.998$) from least-squares regression. In contrast, data with normalization (b) gives the lowest correlation coefficient of 0.955 and data with moving average trend removal (c) has the fewest data points for regression and has a correlation coefficient of 0.993. This leads to our conclusion that, while all the modified data give close estimates of *Hu*, the incremental series from the raw data gives the best estimate of *Hu*.

Unlike simulated fGn/fBm, well data suffer from the moving average effect of the logging tools (Shiomi *et al.* 1997). This effect will induce an abrupt drop of power at large wavenumbers (or frequencies) on the power spectrum plot (Fig. 6). The location of this breaking point is dependent on the resolution of the logging tools.

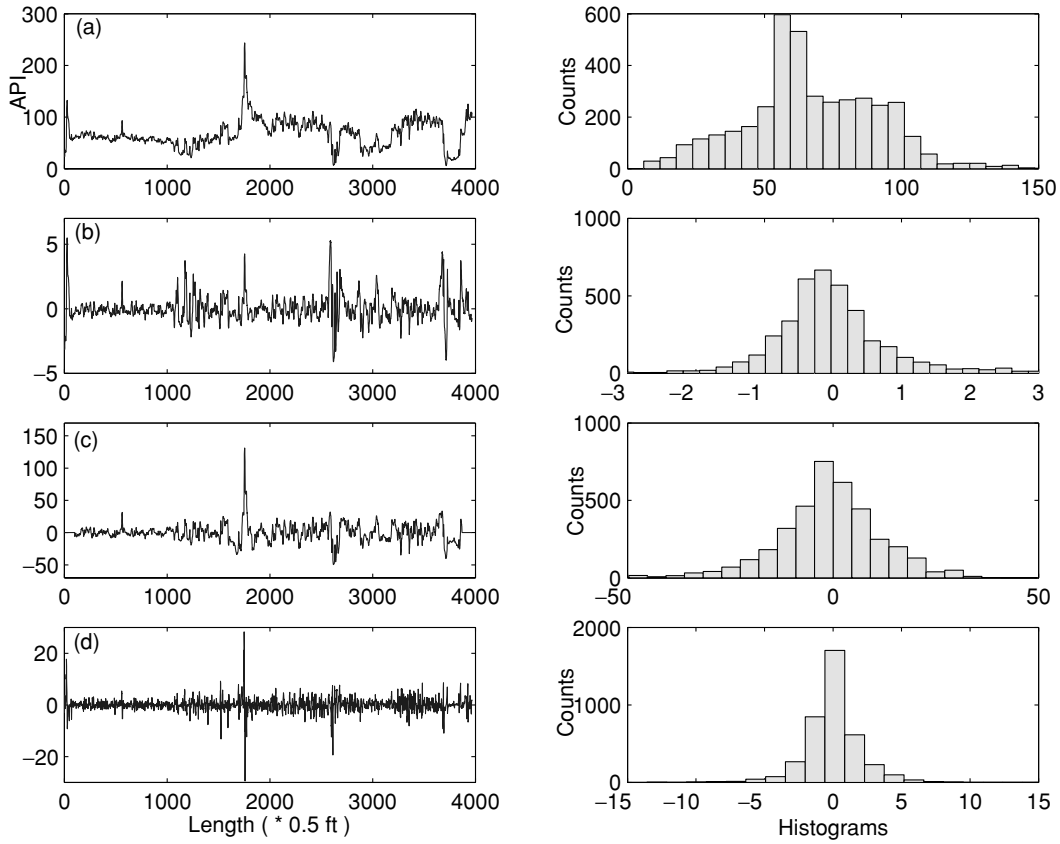

Figure 4. Gamma-ray data and its modified versions. (a) Raw gamma-ray logging data. (b) Normalized gamma-ray data. (c) Data after moving average removal with a window length of 100 ft. (d) Data after taking successive increments. To the right of the data sets are their respective histograms.

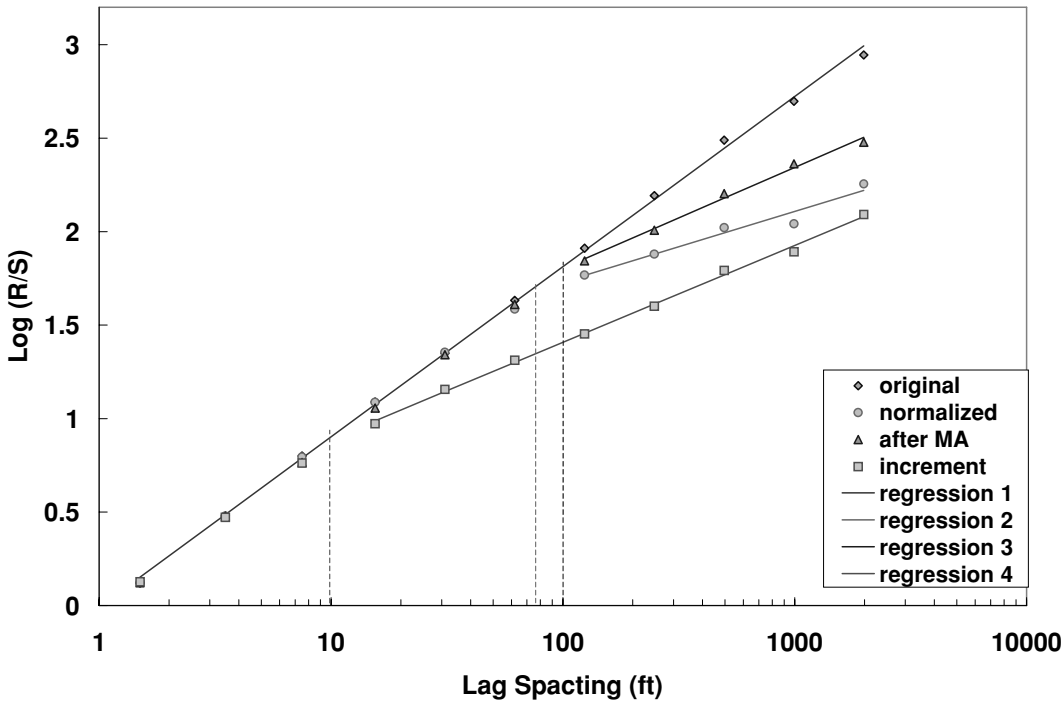

Figure 5. *R*/*S* analysis on gamma-ray data. The vertical dashed lines indicate the lengths of transient zones.

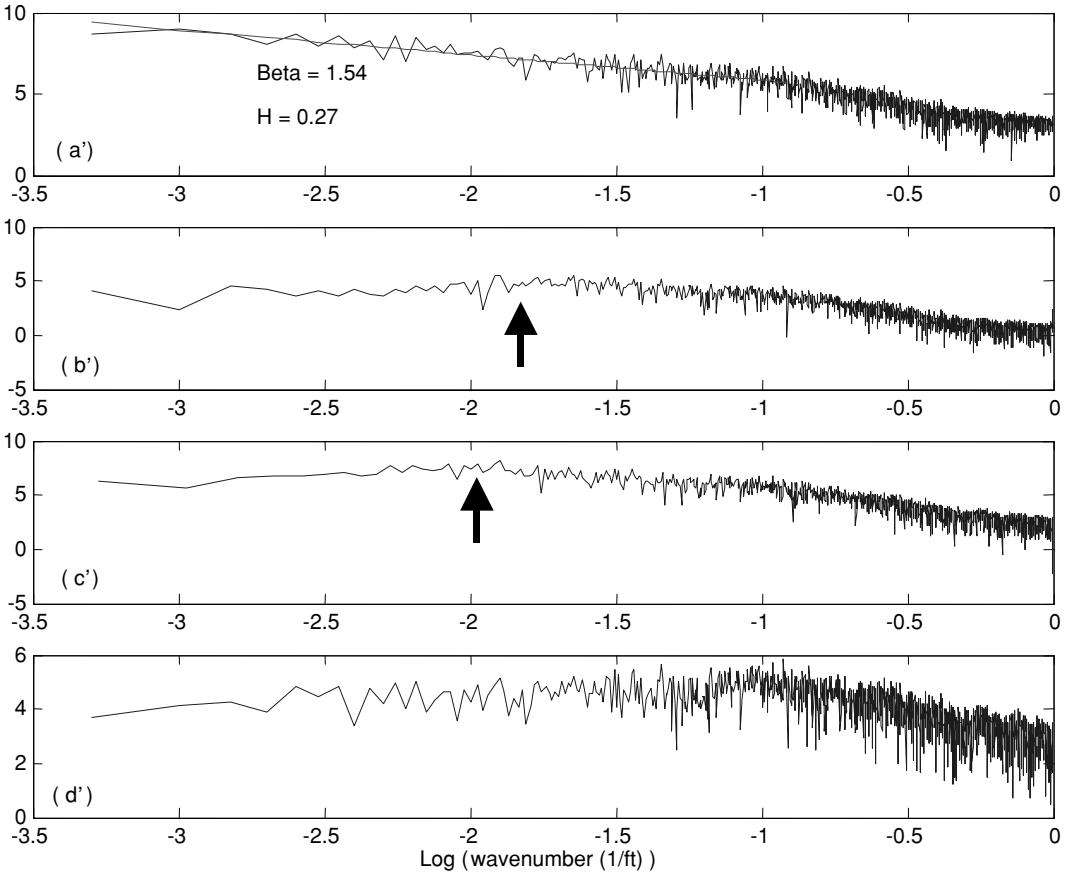

Figure 6. Power spectra of gamma-ray data and modifications. Drop-offs arising from the moving average effect of logging tools are obvious. The straight line in panel a*'* is the linear least-squares fit at long wavelengths for estimating *H*. Solid arrows in b*'* and c*'* indicate the wavelengths that are equal to the transient zone lengths in the *R*/*S* analysis as shown in (b) and (c) in Fig. 5.

For all plots in Fig. 6 we observe that sudden drop-offs at high wavenumbers occur at a fixed location that is independent of the way the data are modified. Suitable lengths for regression in the spectrum plot are consistent with the regression lengths in the *R*/*S* method. This consistency reveals close relationships between these two methods for a monoscaling analysis. This can be observed by comparing Fig. 5 with Fig. 6. Transient zone lengths in Fig. 5 for data after normalization (b), moving average trend removal (c) and the increment (d) occur at a lag spacing of 70, 94.6 and 9.3 ft, respectively. Accordingly, on power spectrum plots, these lengths correspond roughly to the shortest wavelengths that can be possibly included for accurate linear regressions (Fig. 6).

Comparing between *R*/*S* and the power spectrum method we can draw the following conclusions.

(1) When making least-squares regressions at long-wavelength ranges to calculate *H*, we realize that (a) and (d) have wider ranges of wavelengths for linear regressions than (b) and (c). Instead of having two distinct segments as in (a) and (d), (b) and (c) appear to have three linear segments. We find that an incremental data series is still a better choice for estimating *H*/*Hu* owing to its broadest range of frequencies/wavenumbers for linear regression on the power spectrum, even though all three types of modified data sets (b, c, d) have similar gradients for linear regression.

(2) The rolling-off effect caused by the logging tool appears to be a major contributor to the existence of a transient zone for incremental data series. The transient zone can also exist for simulated fGn/fBm, but apparently their causes are different from the causes in real well data. From (b) and (c), normalization and trend removal also introduce long transient zones, in addition to that caused by the resolution limitations of logging tools.

(3) Compared with spectrum method, *R*/*S* is less sensitive to the moving average effects caused by logging tools. In cases where the roll-off on spectra are not severe, and where the linear section for regression has a gradient that is of the same sign as that of the roll-off part, *R*/*S* is unable to differentiate between transient zones caused by modifications on the raw data and those caused by logging tools. This reveals the insensitivity of *R*/*S* to minor gradient changes on the power spectrum plots. Overall we find that the spectrum method is less problematic than *R*/*S*. This statement is in agreement with the results from some other studies (Dolan *et al.* 1998; Taqqu & Teverovsky 1998).

4.2 Sonic data

To test whether our findings from gamma-ray data are general enough to be applicable to other types of logging data, we studied a sonic log with a length of 4249.5 ft, also sampled at a constant interval of 0.5 ft. This well is located in Big Bow oil field, Stanton county, southwest Kansas. The rock formation is characterized mainly by interbedded deposits of carbonates and shale. Identical data processing and analysing procedures as for the gamma-ray case are carried out. Fig. 7 shows raw sonic log data and all modified versions, alone

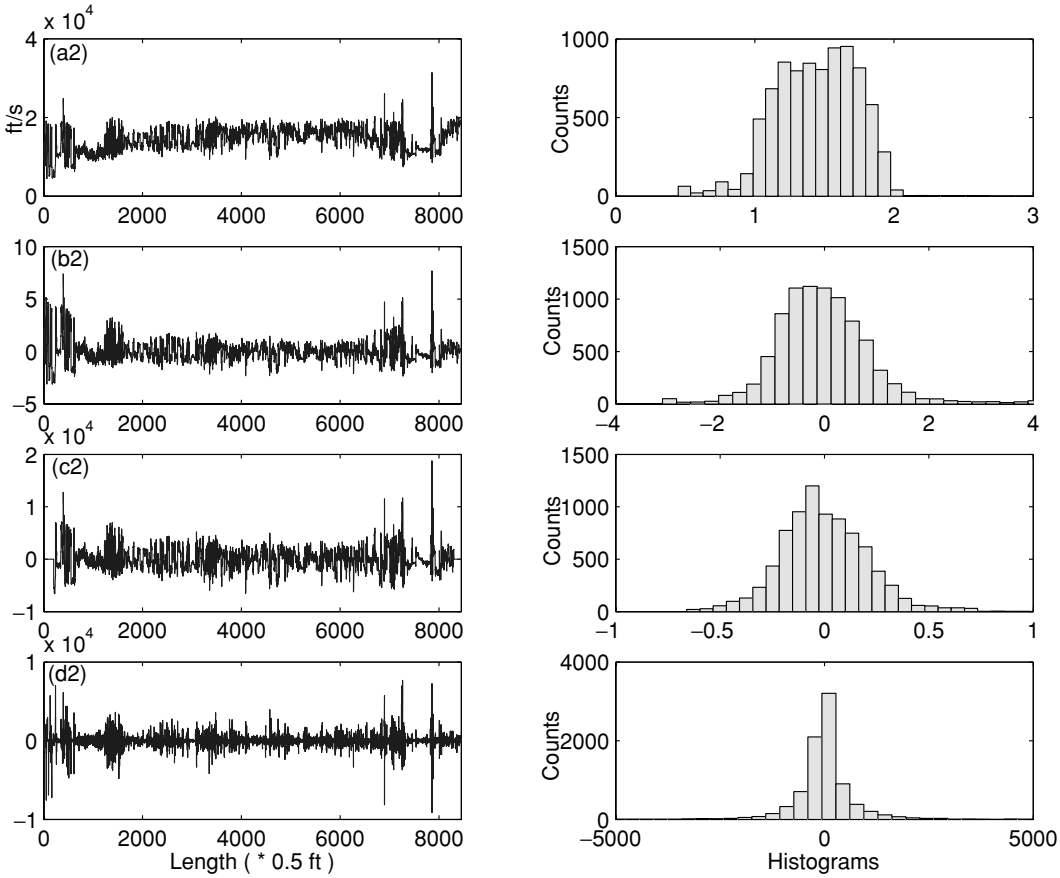

Figure 7. Sonic velocity data and its modified versions. (a2) Raw sonic velocity logging data. (b2) Normalized sonic velocity data. (c2) Data after moving average removal data with a window length of 200 ft. (d2) Data after taking successive increments. To the right of the data sets are their corresponding histograms.

with their respective histograms to the right. The raw sonic log is obviously not Gaussian, but is modified to a Gaussian or Gaussian-like series by normalization (b2), moving average trend removal (c2) or by taking the successive increments (d2). Both b2 and c2 have very short ranges of lag spacing for regression (Fig. 8), compared with a2 and d2. Although in last example the normalization gives an estimate of Hu (=0.38) closer to H (=0.27) from the power spectrum, for this sonic data series, Hu (=0.51) estimated from normalization is too high compared with H (=0.017) estimated from the spectrum. Meanwhile, Hu (=0.25) from the incremental series becomes closer to H. This clearly indicates that various types of modifications can significantly change the outcomes of R/S analyses. Since different authors could resort to different ways of performing the normalization or moving average, the R/S results are unlikely to be stable and consistent in different cases. This could make the results from R/S analysis meaningless under some situations. To avoid these problems, we consider that the incremental series is a superior choice for R/S analysis. Analysis on incremental data is more objective and stable. It also has the shortest transient zone, which leads to more accurate least-squares fitting.

The transient zone lengths for b2, c2, d2 are 212.7, 212.7 and 11.2 ft, respectively (Fig. 8). Once we find the wavelengths equal to these transient zone lengths on the power spectrum plots for b2, c2 and d2, respectively (Fig. 9), we see immediately that transient zone lengths are roughly equal to the shortest wavelengths that can be included for successful linear regressions in Fig. 9. This shows again the underlying ties between these two methods. The combination of these two methods helps identify the transient zones and the final tightening, and helps determine which linear part on the R/S plot should be considered for regression in cases where a fake transient zone might exist, such as in Fig. 3(b).

It is important to make some effort to identify the transient zone and exclude it from the Hu estimation. Otherwise the result could be totally different. For R/S analysis on normalized data (b2) in this example, it could also be performed with a regression that spans the whole lag spacing range yet still maintains a high correlation coefficient of 0.99 (Fig. 10). Such a high correlation coefficient seems to warrant a good regression without the necessity of identifying a transient zone, but the estimated Hu is too high ($Hu = 0.75$). We therefore strongly urge that serious efforts should be taken to identify the possible existence of a transient zone that appears to be always existent for various modified data.

The fractal dimensions, along with estimated H and Hu for two well data sets are shown in Table 2. First, we calculate D using $D = 2 - H$ or $D = 2 - Hu$. Then we calculated both the box-counting dimension (D_b) and the regularization dimension (D_r) independently to double check our results. Although D_b and D_r are not quite equal to each other (as expected), it is still clear to us that they are closer to the estimated D using Hu from the incremental data than to D estimated directly from the raw logging data. Taking the sonic data, for example, the fractal dimension calculated from successive increments of the sonic log using $D = 2 - Hu$ is 1.75, which falls in

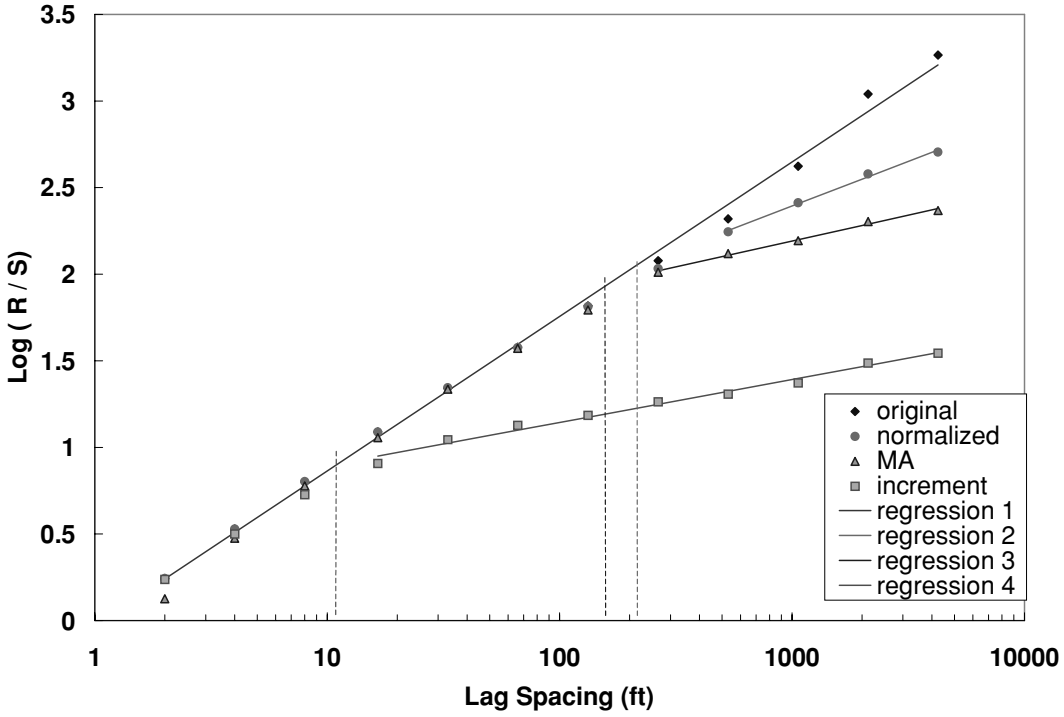

Figure 8. *R*/*S* analysis on sonic velocity data. Vertical dashed lines indicate the lengths of transient zones.

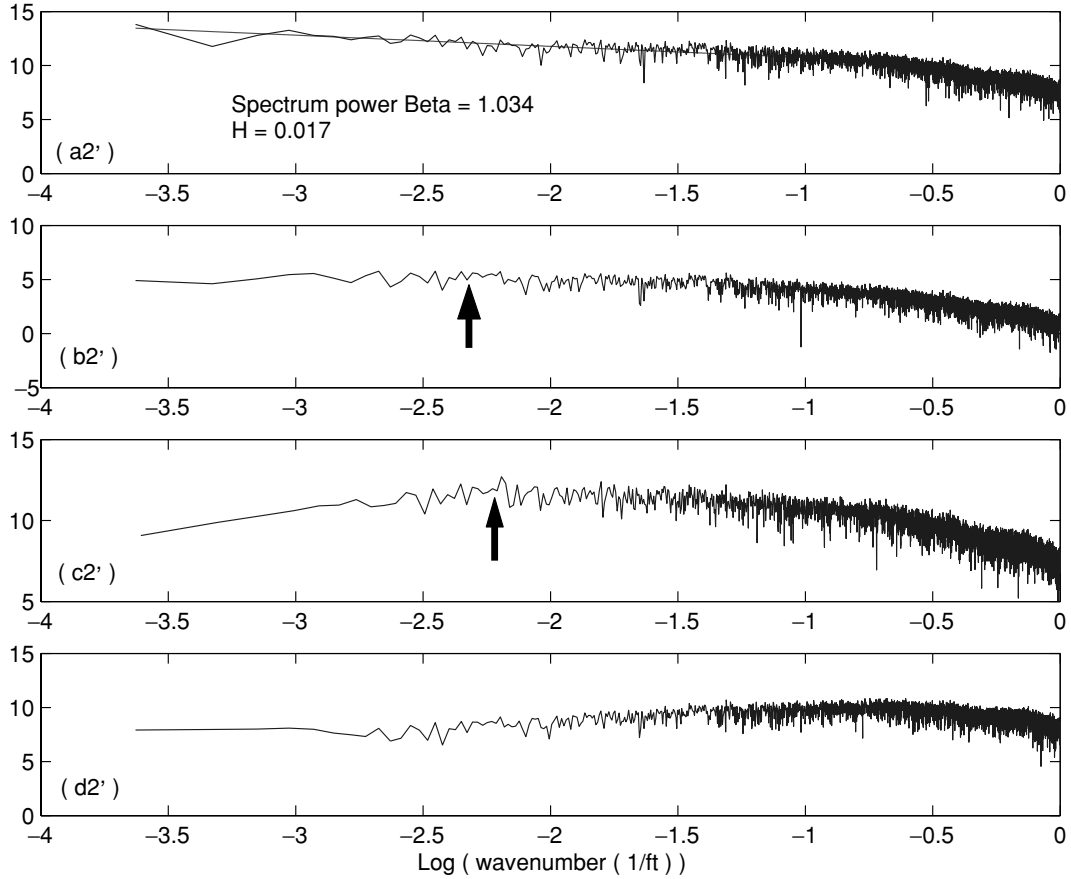

Figure 9. Power spectra of sonic velocity data and its modifications. The straight line in panel a2*'* is the linear least-squares fit at long wavelengths for estimating *H*. Solid arrows in b2*'* and c2*'* indicate the wavelengths that are equal to the transient zone lengths in *R*/*S* analysis as shown in b2 and c2 in Fig. 8.

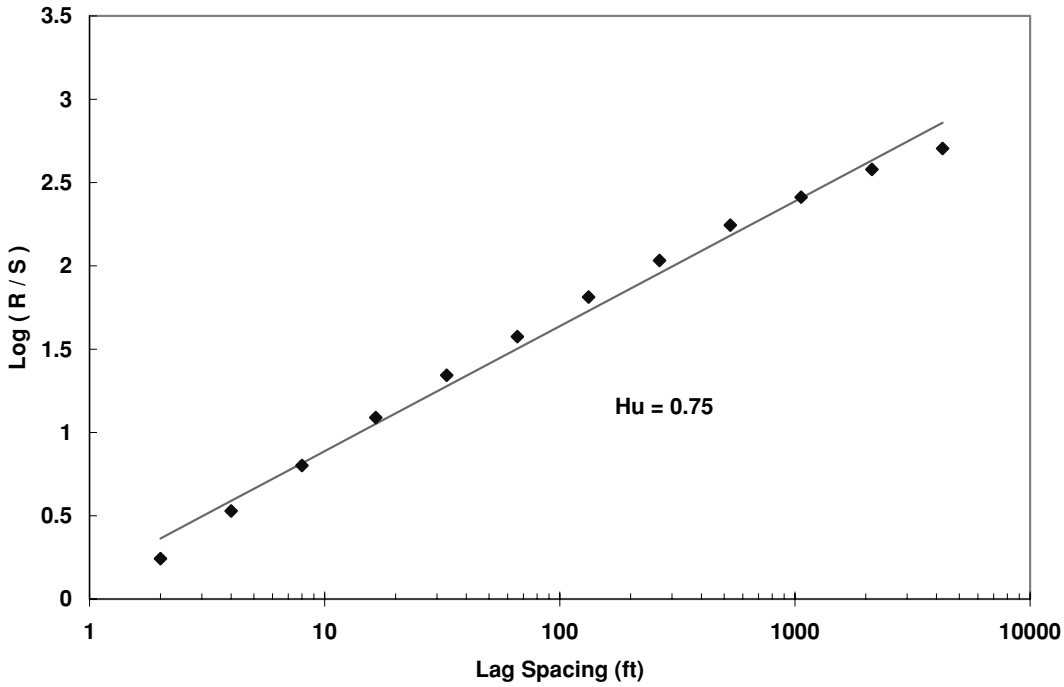

Figure 10. An example showing that, although a good linear regression ($R = 0.99$) can be made without identifying the existence of a transient zone, the estimated *Hu* is too high compared with H.

Table 2. *Hu*, H and D calculated from two well-logging data. (D is the fractal dimension.)

		Raw	Normalized	MA	Increment	Spectrum (H)
GR	*Hu*	0.91	0.38	0.54	0.52	0.27
	$D = 2 - Hu$ (or H)	1.09	1.62	1.46	1.48	1.73
	Box counting D_b	1.37				
	Regularization D_r	1.65				
	Transient zone lengths (ft)	0	70	94.6	9.3	
Sonic	*Hu*	0.89	0.51	0.3	0.25	0.017
	$D = 2 - Hu$ (or H)	1.11	1.49	1.7	1.75	1.983
	Box counting D_b	1.60				
	Regularization D_r	1.94				
	Transient zone lengths (ft)	0	212.7	170.3	11.2	

between $D_b = 1.60$ and $D_r = 1.94$. However, using $D = 2 - Hu$ based on the raw data, the estimated D is only 1.11, much smaller than D_b and D_r. D_b and D_r are also reasonably close to the fractal dimension D (=1.983) from $D = 2 - H$. This once again demonstrates that *R*/*S* analysis directly on the raw data is not appropriate for determining the fractal dimension.

5 CONCLUSIONS AND SUGGESTIONS

In this paper, we examined the paradoxical results often found in the literature concerning the Hurst exponent estimation, and provided an acceptable framework for more accurate analyses. Our arguments were made based on theoretical considerations, analyses from simulations of a group of fGn/fBm and studies on real data examples. We gained a better understanding of the nature of the well log data and argued that raw well-logging data should be considered as fBm-type time-series, which have a spectrum power $\beta \geq 1$. Consequently, for fBm-like data, in order to make the results from rescaled-range and power spectrum techniques consistent and comparable, and to make meaningful fractal dimension estimates, we need to use their incremental series, rather than the raw time-series themselves for *R*/*S* analysis. *R*/*S* analyses applied directly on the raw well data always tend to give erroneously high estimates of *Hu* (>0.85), making the technique itself less useful. On the other hand, *R*/*S* analyses on the incremental series give estimates of *Hu* close to H from power spectrum analyses on the raw data.

Other types of data modification, such as normalization or trend removal, could also be used to improve the analysis, and these modified new data series including the incremental series can give similar estimates of *Hu*. However, the increments have the shortest transient zones that make *R*/*S* analyses more accurate. It is also more objective to use the incremental series, because for each raw data series there exists only one unique incremental series that is also very easy to obtain. Data from normalization or trend removal could vary a lot depending on the specific data processing procedures. The degree of accuracy can also be viewed on the power spectrum plot where the increments have the broadest wavenumber/frequency ranges for linear regression.

Since these modifications to raw well data introduce transient zones in the *R*/*S* analysis, it is very important to identify this zone and exclude it from calculating *Hu*; otherwise it will also result in erroneously high estimates of *Hu*. Comparing with the power spectrum analysis, we find that the transient zone lengths are closely connected to the regression lengths at long-wavelength region in the power spectrum. This reveals the close tie between these two independent data analysis methods.

Fractal models have been used for reservoir simulation (Hewett 1986; Hewett & Behrens 1990; Goggin *et al.* 1992). The decision in choosing the correct Hu/H is very important. For example, a realization based on $Hu = 0.11$ (Holliger 1996) will be quite different from a realization using $Hu = 0.7$ (Leary 1991), even for the same sonic data in Cajon Pass, CA. While we argue that raw well-logging data are of fBm type, in practice, for a new unknown data series it is wise to use the power spectrum method first to determine whether the spectrum power β is larger or smaller than 1. If $\beta \geq 1$, the data need to be considered as fBm-like and we use its increments for *R*/*S* analysis. There might be cases where the calculated $\beta < 1$ even for fBm-type data. This is most probably caused by an insufficient number of data points in the calculation. Both *R*/*S* and spectrum methods require a sufficient number of registered points for an accurate estimation. At least two independent methods are necessary for an accurate scaling analysis because they can provide a double-checking mechanism.

Finally, we realize that the Hurst exponent (*Hu*) should not be confused with the global scaling exponent (*H*) in practice. While these two quantities have some connections, they are fundamentally different and can only be considered equivalent under some constraints. The many discrepancies and conflicts in reported studies are a result of a lack of genuine understanding of the nature of both the data and the analysing techniques.

ACKNOWLEDGMENTS

The author wishes to thank the Geophysical Society of Tulsa and the University of Tulsa for their financial support during this research. C. Bean and J. Pang are kindly thanked for having greatly improved the manuscript with careful reviews and constructive criticism. Thanks are also due to C. Liner who reviewed part of an earlier draft.

REFERENCES

Bassingthwaighte, J.B. & Raymond, G.M., 1994. Evaluating rescaled range analysis for time series, *Ann. Biomed. Eng.,* **22**, 432–444.

Bracewell, R., 1965. *The Fourier Transform and its Applications,* p. 381, McGraw-Hill, New York.

Dolan, S.S. & Bean, C.J., 1997. Some remarks on the estimation of fractal scaling parameters from borehole wire-line logs, *Geophys. Res. Lett.,* **24**, 1275–1278.

Dolan, S.S., Bean, C.J. & Riollet, B., 1998. The broad-band fractal nature of heterogeneity in the upper crust from petrophysical logs, *Geophys. J. Int.,* **132**, 489–507.

Feder, I., 1988. *Fractals,* p. 283, Plenum, New York.

Fox, C.G., 1989. Empirically derived relationships between fractal dimension and power law form frequency spectra, *Pageophys.,* **131**, 211–239.

Goggin, D.J., Chandler, M.A., Kocurek, G. & Lake, L.W., 1992. Permeability transects of Eolian sands and their use in generating random permeability fields, *SPE Form. Eval.,* **7**, 7–16.

Hewett, T.A., 1986. Fractal distribution of reservoir heterogeneity and their influence on fluid transport, SPE paper 15386, presentation at the *61st Annual Technical Conference and Exhibition of SPE,* New Orleans, LA.

Hewett, T.A. & Behrens, R.A., 1990. Conditional simulation of reservoir heterogeneity with fractals, *SPE Form. Eval.,* 217–225.

Hipel, K. & McLeod, A., 1978. Preservation of the rescaled adjusted range. 2. Simulation studies using Box–Jenkins models, *Water Res. Res.,* **14**, 509–516.

Holliger, K., 1996. Upper-crustal seismic velocity heterogeneity as derived from a variety of, *P*-wave sonic logs, *Geophys. J. Int.,* **125**, 813–829.

Hough, S.E., 1989. On the use of spectral methods for the determination of fractal dimension, *Geophys. Res. Lett.,* **16**, 673–676.

Hurst, H.E., 1951. Long-term storage capacity of reservoirs, *Trans, Am. Soc. Civil Eng.,* **116**, 770–808.

Leary, P., 1991. Deep borehole log evidence for fractal distribution of fractures in crystalline rock, *Geophys. J. Int.,* **107**, 615–627.

Lo, A.W., 1991. Long term memory in stock market prices, *Econometrica,* **59**, 1279–1313.

Mandelbrot, B.B., 1982. *The Fractal Geometry of Nature,* Freeman, San Francisco.

Mandelbrot, B.B. & Van Ness, J.W., 1968. Fractional Brownian motions, fractional noises and applications, *SIAM Rev.,* **10**, 422–437.

Mandelbrot, B.B. & Wallis, J.R., 1969. Some long-run properties of geophysical records, *Water Res. Res.,* **5**, 321–340.

McLeod, A. & Hipel, K., 1978. Preservation of the rescaled adjusted range. 1. A reassessment of the Hurst phenomenon, *Water Res. Res.,* **14**, 491–508.

Moody, J. & Wu, L., 1996. Improved estimates for the rescaled range and Hurst exponents, in *Neural Networks in the Capital Markets,* eds Refenes, A., Abu-Mostafa, Y. & Ewigend, A., World Scientific, London.

North, C.P. & Halliwell, D.I., 1994. Bias in estimating fractal dimension with the rescaled-range (*R/S*) techniques, *Math. Geol.,* **26**, 531–555.

Pang, J. & North, C.P., 1996. Fractals and their applicability in geological wireline log analysis, *J. Petrol. Geol.,* **19**, 339–350.

Shiomi, K., Sato, H. & Ohtake, M., 1997. Broad band power law spectra of well log data in Japan, *Geophys. J. Int.,* **130**, 57–64.

Taqqu, M.S. & Teverovsky, V., 1998. On estimating the intensity of long-range dependence in finite and infinite variance time series, in *A Practical Guide to Heavey Tails: Statistical Techniques and Applications,* pp. 177–217, eds Adler, R., Feldman, R. & Taqqu, M.S., Birkhäuser, Boston, MA.

Turcotte, D.L., 1997. *Fractals and Chaos in Geology and Geophysics,* Cambridge University Press, Cambridge.

Voss, R.F., 1985. Random fractals: characterization and measurement, *Proc. NATO ASI Scaling Properties of Disordered Media,* Geilo, Norway.