Fractal Dimensions for Rainfall Time Series

M.C. Breslin and J.A. Belward
Department of Mathematics,
The University of Queensland, QLD 4072, Australia

Abstract Fractals are objects which have a similar appearance when viewed at different scales. Such objects have detail at arbitrarily small scales, making them too complex to be represented by Euclidean space. They are assigned a dimension which is non-integer. Some natural phenomena have been modelled as fractals with success; examples include geologic deposits, topographical surfaces and seismic activity. In particular, time series data has been represented as a curve with dimension between one and two. There are many different ways of defining fractal dimension. Most are equivalent in the continuous domain, but when applied in practice to discrete data sets lead to different results. Three methods for estimating fractal dimension are evaluated for accuracy. Two standard algorithms, Hurst's rescaled range analysis and the box counting method, are compared with a recently introduced method which has not yet been widely used. It will be seen that this last method offers superior efficiency and accuracy, and it is recommended for fractal dimension calculations for time series data. We have applied these fractal analysis techniques to rainfall time series data from a number of gauge locations in Queensland. The suitability of fractal analysis for rainfall time series data is discussed, and how the theory might aid our interpretation of rainfall data.

1. INTRODUCTION

Fractal analysis provides a unique insight into a wide range of natural phenomena. Fractal objects are those which exhibit 'self-similarity'. This means that the general shape of the object is repeated at arbitrarily smaller and smaller scales. Coastlines have this property: a particular coastline viewed on a world map has the same character as a small piece of it seen on a local map. New details appear at each smaller scale, so that the coastline always appears rough. Although true fractals repeat the detail to an infinitely small scale, examples in nature are self-similar up to some finite limit.

The fractal dimension measures how much complexity is being repeated at each scale. A shape with a higher fractal dimension is more complicated or 'rough' than one with a lower dimension, and fills more space. These dimensions are fractional: a shape with fractal dimension of $D = 1.2$, for example, fills more space than a one-dimensional curve, but less space than a two-dimensional area. The fractal dimension succinctly tells much information about the geometry of an object. Very realistic computer images of mountains, clouds and plants can be produced by simple recursions with the appropriate fractal dimension.

Time series of many natural phenomena are fractal. Small sections taken from these series, once scaled by the appropriate factor, cannot be distinguished from the whole signal. Being able to recognise a time series as fractal means being able to link information at different time scales. We call such sets 'self-affine' instead of self-similar becuase they scale by different amounts in each axis direction.

In the next section we very briefly review two of the most well-used methods used for calculating fractal dimensions of graphs, and call attention to a more recent method. We then discuss the interpretation of fractal dimension. In section 4 we dimension estimates for rainfall data from gauge stations covering most of Queensland.

2. A REVIEW OF THREE FRACTAL DIMENSION METHODS

There are many methods available for estimating the fractal dimension of data sets. These lead to different numerical results, yet little comparison of accuracy has been made between them in the literature. We compared a more recent algorithm, the variation method, to two methods which stand out as the most popular for assigning fractal dimensions to time series, the box-counting method and rescaled range analysis.

2.1. Box-Counting

The box counting algorithm is intuitive and easy to apply. It can be applied to sets in any dimension, and has been used on images of everything from river systems to the clusters of galaxies.

A fractal curve is a curve of infinite detail, by virtue of its self-similarity. The length of the curve is indefinite, increasing as the resolution of the measuring instrument increases. The fractal dimension determines the increase in detail, and therefore length, at each resolution change. For a fractal, the length L as a function of the resolution of the measurement device δ is

$$L(\delta) \propto \delta^{-D}, \quad (1)$$

where D is an exponent known as the fractal dimension. (For ordinary curves $L(\delta)$ approaches a constant value as δ decreases.)

Box-counting algorithms measure $L(\delta)$ for varying δ by counting the number of non-overlapping boxes of size δ required to cover the curve, as illustrated in Figure 1. These measurements are fitted to (1) to obtain an estimate of the fractal dimension, known as the box dimension.

A fractal dimension can be assigned to a set of time series data by plotting it as a function of time, and calculating the box dimension. (1) will hold over a finite range of box-size; the smallest boxes will be of width τ, where τ is the resolution in time, and height a, where a is the resolution of the data.

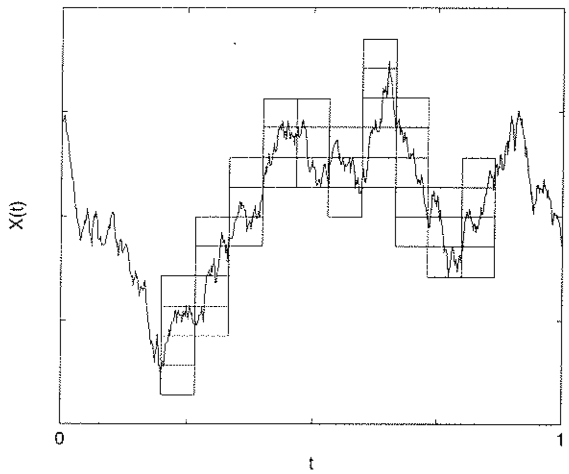

Figure 1: An example of a box counting cover for a record of Brownian motion.

2.2. R/S Analysis

Rescaled range (or R/S) analysis is really a tool to search for long-term memory or correlation in time series. It has been used to show that many natural phenomena previously assumed to be governed by random Gaussian processes exhibit long-range statistical dependence (see section 3.1). A statistic H, $0 < H < 1$, is used as a measure of the degree of correlation, with $H = 1$ indicating perfect correlation between increments. The long range dependence is closely related to self-similarity of the process. If the time series is fractal, the statistic H is related to the fractal dimension by

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

Details of the theory can be found in Mandelbrot [1969], Pang et al. [1996] and Feder [1988].

Estimates of the statistic H are obtained simply, and the technique has been widely used for analysing geological deposits, solar activity and seismic activity (Ruzmaikin [1994], Komm [1995], Ogata [1991], Pang et al. [1996]). Unfortunately, this method overestimates H for $H < 0.72$, and underestimates H for $H > 0.72$ (North et al. [1994], Wallis et al. [1971]), with the error increasing as the dataset size decreases. Very large bias is also introduced where periodicity or non-stationarity of means is present in the data (Mandelbrot [1969], North et al. [1994]). This makes R/S analysis unsuitable as a method for much climatic data, and any datasets of limited size.

2.3. Variation Method

Dubuc et al. [1989] present a method which they claim gives more accurate results than the standard box-counting applications, as well as being more robust and efficient. The method uses coverings built out of intervals rather than boxes. The following is a very simple description of the method. For full details, see Dubuc et al. [1989], and also Dubuc et al. [1996].

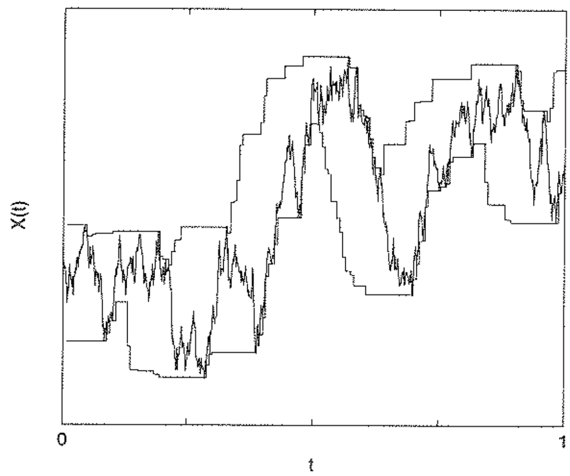

Figure 2: An example of a variation cover for a record of Brownian motion.

An example of a covering constructed by the variation method is shown in Figure 2. It is constructed by calculating the 'oscillation' at points along the curve. If we label the curve $X(t)$, the oscillation at a point $X(t_0)$ is simply

$$v(X(t_0),\epsilon) = \underset{\tau\in(t_0-\epsilon,t_0+\epsilon)}{max} X(\tau) - \underset{\tau\in(t_0-\epsilon,t_0+\epsilon)}{max} minX(\tau). \quad (3)$$

This corresponds to the height of cover shown in Figure 2. ϵ gives the scale at which we measure the oscillation, similar to the size of boxes in the box cover. As ϵ decreases, so does the cover.

To calculate the fractal dimension we find the area of the cover, $V(\epsilon)$, and calculate the rate at which the area tends to 0 as ϵ tends to 0. It turns out that a log-log plot of $V(\epsilon)/\epsilon^2$ vs. $1/\epsilon^2$ gives the fractal dimension as its slope. The area $V(\epsilon)$ is called the variation of X.

2.4. Numerical Comparison of Methods

In order to compare fractal dimensions of data taken from different stations, which may be expected to be close in value, a method is required which produces estimates which are as accurate as possible given the sample size. The rainfall data sets each had approximately 1000 points.

The three methods outlined above were tested on curves with known fractal dimension. The results shown in Figures 3 and 4 are from tests done on Weierstrass-Mandelbrot (W-M) curves; the W-M function is a function whose dimension can be stipulated. For a description of this curve, see Feder [1988].

Figure 3 shows results produced by the three methods for different true fractal dimensions; each dataset contained 1000 points. The plot gives an indication of the difficulty in interpreting results produced by R/S analysis. It also shows that the variation method performs slightly better than the box-counting method on sets of any dimension.

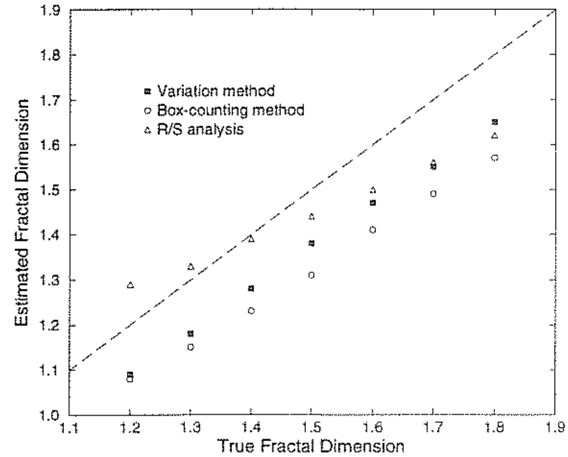

Figure 3: Estimates produced by the three methods of the fractal dimension of the Weierstrass-Mandelbrot curve with varying true fractal dimension. Each sample of the curve contained 1000 points. The dotted line represents ideal behaviour.

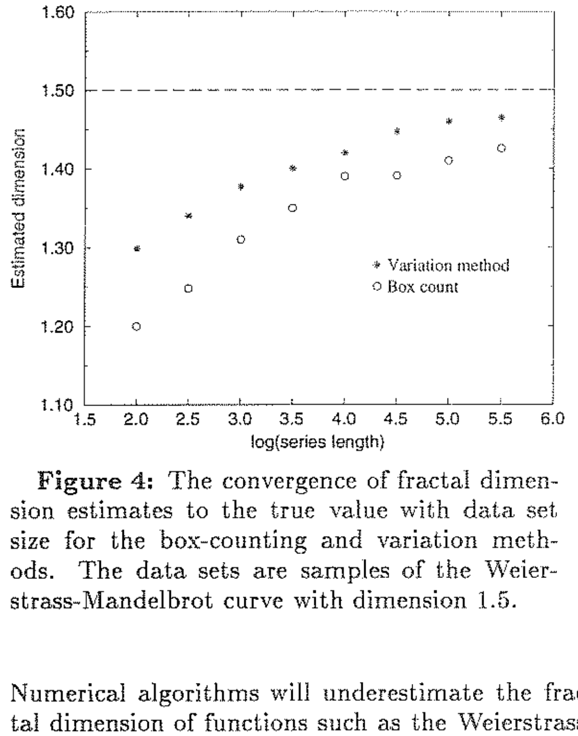

Figure 4: The convergence of fractal dimension estimates to the true value with data set size for the box-counting and variation methods. The data sets are samples of the Weierstrass-Mandelbrot curve with dimension 1.5.

Numerical algorithms will underestimate the fractal dimension of functions such as the Weierstrass-Mandelbrot function, since the calculation is done on a finite sample of points joined by straight lines, an object of reduced complexity. Figure 4 compares the convergence of dimension estimates given by the variation and box-counting methods for the Weierstrass-Mandelbrot curve. Similar rates are obtained with Brownian motion samples. The plot gives an indication of the dataset size required to achieve a specified accuracy; this will be a good indication for sets of fractal dimension approximately 1.5.

3. INTERPRETATION OF FRACTAL DIMENSION

We have already mentioned that the fractal dimension of an object is a measure of complexity and degree of space filling. When the object is a series in time, the dimension also tells us something about the relation between increments. It is a useful and meaningful insight into series of natural processes.

3.1. Fractional Brownian Motion

A particle undergoing Brownian motion moves by jumping step-lengths which are given by independent Gaussian random variables. For one-dimensional motion the position of the particle in time, $X(t)$, is given by the addition of all past increments. The function $X(t)$ is a self-affine fractal, whose graph has dimension 1.5.

Fractional Brownian motion generalises $X(t)$ by allowing the increments to be correlated. Ordinary Brownian motion can be defined by

$$X(t) - X(t_0) \sim \xi |t - t_0|^H, \; t \geq t_0, \qquad (4)$$

where $H = 1/2$, ξ is a normalised independent Gaussian process and $X(t_0)$ is the initial position (Wiener [1923], Feder [1988]). Replacing the exponent $H = 1/2$ in (3) with any other number in the range $0 < H < 1$ defines a fractional Brownian motion (fBm) function $X_H(t)$. The exponent H here corresponds to the statistic H that R/S analysis calculates.

The correlation function of future increments with past increments for the motion $X_H(t)$ can be shown to be (Vicsek [1992])

$$C(t) = 2^{2H-1} - 1. \qquad (5)$$

Clearly, $C(t) = 0$ for $H = 1/2$; increments in ordinary Brownian motion are independent. For $H > 1/2$, $C(t)$ is positive for all t. This means that after a positive increment, future increments are more likely to be positive. This is known as persistence. When $H < 1/2$, increments are negatively correlated, which means an increase in the past makes a decrease more likely in the future. This is called antipersistence.

Now it is true for self-affine functions such as $X_H(t)$ that the fractal dimension, D, is related to H by (Feder [1988])

$$D = 2 - H. \qquad (6)$$

We can then identify persistence or antipersistence in data sets whose graphs are fractal. Persistent time series show long term memory effects. An increasing trend in the past is likely to continue in the future because future increments are positively correlated to past ones. Similarly, a negative trend will persist. This means that extreme values in the series tend to be more extreme than for uncorrelated series. In the context of climatic data, droughts or extended rain periods are more likely for persistent data.

3.2. Simulating fractional brownian motion

The successive random addition algorithm introduced by Voss [1985] can be used to generate samples of fractional Brownian motion. An example of this process for 2000 points and a fractal dimension estimated by the variation method to be 1.38 is shown in Figure 5a, and Figure 5b shows rainfall data with the same fractal dimension estimate. The similarity is obvious.

The rainfall data in Figure 5b is a rainfall 'record', defined to be the cumulative sum of deviations from the mean.

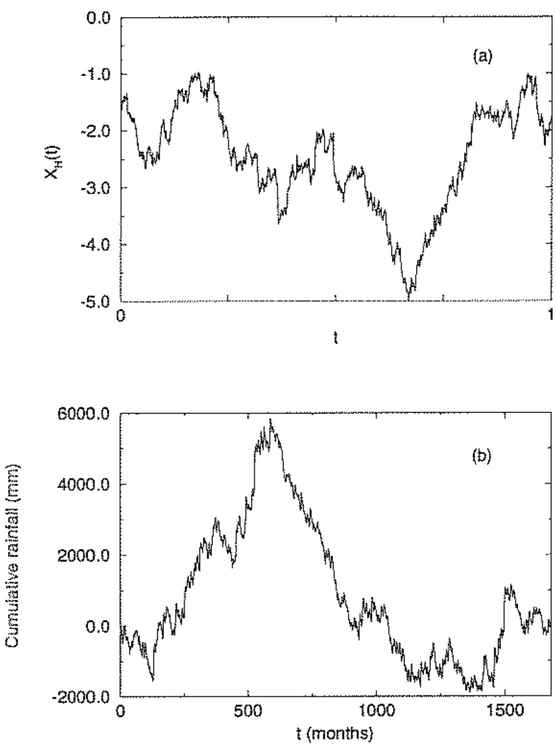

Figure 5: (a) A sample of fractional brownian motion, and (b) A plot of a rainfall record (defined in section 3.2) from Brisbane. Both graphs have fractal dimension estimates of D=1.38.

4. RESULTS FOR RAINFALL TIME SERIES

The data sets analysed were series of monthly rainfall totals, collected at 52 gauge stations in Queensland. The average length of the series was 1200 values.

The fractal dimensions of the records (section 3.2) of the sets were calculated using the variation method. A sample of the results are shown in Figure 7. For legibility, the numbers shown are the first two decimal places of the estimate only: dimensions are always between one and two. A plot of dimension vs. annual rainfall for the 52 stations is shown in Figure 6; it shows a weak correlation between dimension and total amount of rain.

Whether or not the rainfall series are self-similar can be determined from the goodness of the fits of the log-log plots. In simple terms, if the points of the plot line up on a straight line it means that the amount of detail at each new scale is the same, ie. the data set shows self-similarity. Figure 8 shows plots for stations Brisbane and Kowanyama. Data from all stations produced good fits; the average standard deviation for the least squares fit used to obtain the lines was $s = 0.02$. The plots imply that rainfall amounts over the range one month to ~ 100 years are self-similar.

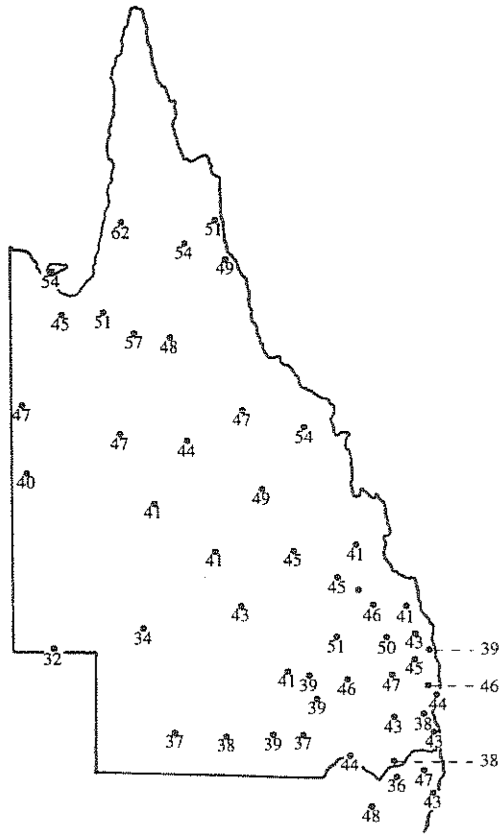

Figure 7: Fractal dimension estimates obtained for monthly rainfall data in Queensland.

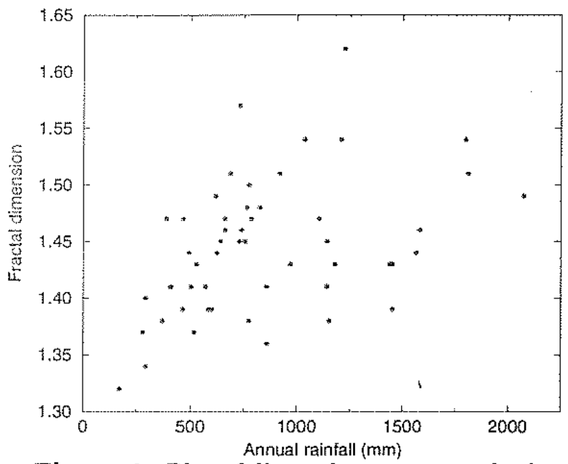

Figure 6: Plot of dimensions vs. annual rainfall.

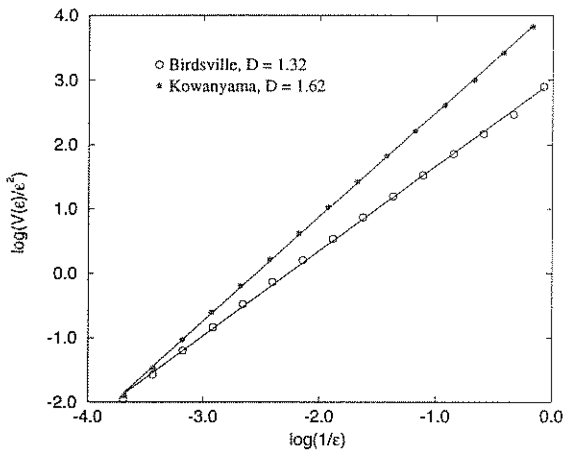

Figure 8: Variation log-log plots for two rainfall data sets. The fractal dimension estimate is given by the slope.

It was noted in section 3.1 that the fractal dimension of a time series gives a measure of the correlation between increments. Dimensions between 1 and 1.5 indicate persistence, or existence of clear trends in a series, while dimensions between 1.5 and 2 indicate antipersistence, or that the signal is very noisy. This gives some insight into the results in Figure 7.

5. CONCLUSION

The Queensland rainfall data sets show self-similarity over scales ranging from one month to $\sim$ 100 years. That is, monthly fluctuations have the same statistical behaviour as fluctuations on a decade scale. The ability to perform scale shifts and thus extend limited data is an attractive prospect. Investigations need to be made into the limit of scales to which self-similarity extends. Also worth considering is the possibility of simulating rainfall using fractional Brownian motion as a model.

The fractal dimnesion is a convenient description of a rainfall time series; it describes the irregularity or randomness in the series, and whether or not there are long term memory effects present. It is simple to calculate; the variation method gives robust and reliable estimates. For these reasons, the fractal dimension should be a useful tool. The spatial distribution of results for Queensland data sets indicates that all rainfall does not have the same dimension, and therefore the same fractal properties. The challenge is to explain the distribution of values in physical terms.

6. REFERENCES

Dubuc, B., J.F. Quiniou, C. Roques-Carmes, C. Tricot, and S.W. Zucker, Evaluating the fractal dimensions of profiles, *Phys. Rev. A*, 39(3), 1500–1512, 1989.

Dubuc, B., and S. Dubuc, Error bounds on the estimation of fractal dimension, *SIAM J. Numer. Anal.*, 33(2), 602–626, 1996.

Komm, R.W., Hurst analysis of Mt. Wilson rotation measurements, *Solar Phys.*, 156, 17–28, 1995.

Falconer, K., *Fractal Geometry*, John Wiley, 288 pp., Chichester, 1990.

Feder, J., *Fractals*, Plenum Press, 288 pp., New York, 1988.

Mandelbrot, B.B., and J.R. Wallis, Some long-run properties of geophysical records, *Water Resources Res.*, 5(2), 321–340, 1969.

North, C.P., and D.I. Halliwell, Bias in estimating fractal dimension with the Rescaled-Range (R/S) technique, *Math. Geol.*, 26(5), 531–555, 1994.

Ogata, Y., and K. Abe, Some statistical features of the long-term variation of the global and regional seismic activity, *Int. Statist. Rev.*, 59(2), 139–161, 1991.

Pang, J., and C.P. North, Fractals and their applicability in geological wireline log analysis, *J. Petrol. Geol.*, 19(3), 339–350, 1996.

Ruzmaikin, A., J. Feynman, and P. Robinson, Long-term persistence of solar activity. *Solar Phys.*, 149, 395–403, 1994.

Takayasu, H., *Fractals in the Physical Sciences*, Manchester University Press, 170 pp., Manchester, 1989.

Vicsek, T., *Fractal Growth Phenomena*, World Scientific, 488 pp., Singapore, 1992.

Voss, R.F., Random Fractal Forgeries, in *Fundamental Algorithms in Computer Graphics*, edited by R.A. Earnshaw, pp. 805-835, Springer, Berlin, 1985.

Wallis, J.R., and N.C. Matalas, Small sample properties of H and K–estimators of the Hurst coefficient *h*, *Water Resources Res.*, 6, 1583–1594, 1971.

Wiener, N., Differential Space, *J. Math. Phys. Mass. Inst. Technol.*, *2*, 131-174, 1923.

7. APPENDIX

The 52 gauge station data sets are sets compiled by NOAA, and were obtained from the LDEO (Columbia University) Climate Data Catalog on the world wide web (http://ingrid.ldgo.columbia.edu/).