Mathematical Geology, Vol. 30, No. 5, 1998

Sedimentation Rates, Observation Span, and the Problem of Spurious Correlation[1]

W. Schlager,[2] D. Marsal,[3] P. A. G. van der Geest,[4] and A. Sprenger[2]

Plots of sedimentation rates vs. time span of observation are routinely used to demonstrate that sedimentation rates decrease if one averages over longer time spans. However, these plots are suspect because they plot a variable, time, against its inverse. It has been shown that even random numbers may yield correlation coefficients of 0.7 or higher under these circumstances. We have circumvented this problem by splitting observed sedimentation rates into time classes and performing regression on the primary variables, thickness and time, separately in each class. An alternative is weighted regression that corrects for the effect of spurious correlation. Regression on the primary variables has been performed on real data from siliciclastic and carbonate rocks. Data were sorted into time classes of 10^{-1}–10^{2} yr, 10^{2}–10^{5} yr, and 10^{5}–10^{8} yr. Sedimentation rates decrease systematically as the time windows increase. The experiment indicates that the decrease of sedimentation rates with increase of time is not simply an effect of the mathematical transformation. It is a physical phenomenon, probably related to the fact that sedimentation is an episodic process and that the sediment record is riddled with hiatuses on all scales.

KEY WORDS: part-and-whole correlation, ratio correlation, regression, Cantor function.

INTRODUCTION

It is widely accepted that a systematic relationship exists between sedimentation rates and the length of the observation span; numerous studies found that rates decrease as the length of the observation interval increases (e.g., Reineck, 1960; Sadler, 1981; Enos, 1991). The phenomenon is normally explained by the distribution of hiatuses and condensed intervals. Sedimentation is a pulsating or episodic process that rarely ever runs at a steady pace. As a consequence, the stratigraphic record is replete with hiatuses and condensed levels, i.e., intervals of extremely slow sedimentation (Barrell, 1917; Schindel, 1982; Plotnick, 1986). Averages from accumulations deposited over long time spans are likely to in-

[1] Received 9 April 1997; accepted 29 October 1997.
[2] Vrije Univ./Earth Sciences, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands.
[3] D-8399 Ering/Inn, Germany.
[4] Shell Int. Expl. & Prod. B.V., Volmerlaan 6, 2288 GD Rijswijk, The Netherlands.

0882-8121/98/0500-0547$15.00/1 © 1998 International Association for Mathematical Geology

clude numerous, long hiatuses either because the stratigraphic resolution is insufficient to reveal them, or because one is not interested in resolving details. Conversely, rates calculated from accumulations of short intervals must be based on high-resolution stratigraphy. Otherwise the short interval could not be defined; thus, significant hiatuses are likely to be recognized and excluded from the averaging process.

The motivation for this study is the nature of the data on which the notion of scaled sedimentation rates is based. The trend has been postulated for time intervals of 12 orders of magnitude (e.g., Sadler, 1981). In only a small part of this range, namely the "human scale" of minutes to tens of years, the sedimentation process has been measured directly. The vast majority of sedimentation rates are calculated by dividing the thickness of the deposit by the time required for its deposition. This time is estimated from the age difference between the bottom and the top of the formation.

The approach is fraught with what is known in statistics as ratio correlation or part-and-whole correlation (Aitchison, 1986). In a classical paper, Pearson (1897) pointed to a peculiar property of compound variables, such as ratios, in correlation. He showed that two variables that have no correlation between themselves become correlated when divided by a third uncorrelated variable. Pearson (1897) introduced the term "spurious correlation" for the "amount of correlation which would still exist between the indices, were the variables on which they depend distributed at random." Pearson's point applies to graphs of sedimentation rates against time because these plots measure a variable, time, against a composite variable that contains the inverse of time.

A considerable literature has developed on the subject. It shows both by experiment and theoretical considerations that spurious correlations sensu Pearson (1897) may easily attain correlations of $r = \pm 0.7$ and higher (Atchley, Gaskins, and Anderson, 1976; Atchley and Anderson, 1978; Kenney, 1982; Jackson and Somers, 1991). Authors rejecting the notion of "spurious correlation" point out that it is a logical consequence of the formulation of the variables and that the plots may be legitimate provided it is really the relationship of ratios that one is after (e.g., Prairie and Bird, 1989). The interest in ratios per se is particularly obvious in sociology where the individual is at the center of concern and risk for the individual (e.g., crime rate) rather than the total number of incidents (number of crimes) is the object of study (Kasarda and Nolan, 1979).

Studies on scaling of sedimentation rates and the completeness of the geologic record have paid little attention to the problem of ratio correlation, even though Anders, Krueger, and Sadler (1987) mention it in passing. We believe that the question needs to be addressed when studying sedimentation rates (and other time-distance rates) in the geologic record, just as it is being discussed in the context of rates of evolution (e.g., Gingerich, 1983; Gould, 1984; Gingerich, 1984). In this report, we first describe two approaches that circumvent the problem of spurious correlation. Subsequently, we apply these techniques to recently published data on sedimentation rates in siliciclastics and carbonates (McNeill, 1989; Enos, 1991; Bosscher and Schlager, 1993; Ebren, 1996).

MATHEMATICAL APPROACH

Spurious Correlation

In most applications, correlation analysis refers to the correlation between two variables that are not correlated out of physical necessity. Thickness L and the time interval T of a stratigraphic formation represent two such variables. Consider a sample $S = (T, L)$ that consists of N observed pairs of thickness and time (T_i, L_i). The index $i = 1, 2, 3 \cdots, N$ signifies that at location i a layer of thickness L_i was deposited during time interval T_i.

From sample S a correlation coefficient $r = r(T, L)$ may be calculated that is completely based on observation because all elements of S were measured directly. The situation is quite different when we consider the plot of sedimentation rate, R, against time. In stratigraphy, sedimentation rate is normally defined as thickness per unit time,

$$R = L/T$$

Using this definition, the original sample $S(T, L)$ may be transformed into sample $S(T, L/T) = S(T, R)$. After transformation, every point (T_i, L_i) is now located on a hyperbola with abscissa T and ordinate $L/T = R$. The transformation produces a nonlinear correlation with correlation coefficient $r(T, R)$. This correlation originates solely from the *a posteriori* observation of L and the *a priori* stipulation that T is related to R by $R = L/T$. It is an example of what Pearson (1897) termed "spurious correlation" of variables with a common term. If the primary variables T, L are correlated, the points $(T_i, L_i/T_i)$ are located on different hyperbolae and the observed trend is composed of self-induced correlation plus (physically meaningful) correlation of the primary variables.

The nonlinearity of the correlation of (T, R) and the dependence of the variable R on the variable T imply that the usual (linear) correlation tests for randomness do not apply to the correlation of deposition rate R vs. deposition time T. In particular, it cannot be assumed that a numerically large correlation coefficient $r(T, R)$ implies physically meaningful correlation between T and R.

Trend Analyses of Sedimentation Rates vs. Time

Consider a sample $S = (T_i, L_i)$ of size N where L_i is stratigraphic thickness and T_i is the associated time interval of deposition. We first examine the correlation coefficient $r(T, L)$ for randomness ($r = 0$). If the two variables are correlated, the method of least squares is applied to the dataset $S(T, L)$ to obtain the parameters B, b_1, b_2, b_3, b_4 of the polynomials

$$L = BT \tag{1}$$

for linear approximation, or

$$L = b_1T + b_2T^2 + b_3T^3 + b_4T^4 \tag{2}$$

for nonlinear approximation. By definition, L vanishes at $T = 0$, thus all regression curves pass through the origin and the polynomials contain no constant term. Dividing Equations (1) and (2) by T yields the sedimentation rate $R = L/T$. B is the (constant) sedimentation rate for the linear case. The rate R for the nonlinear case is given by

$$R = b_1 + b_2T + b_3T^2 + b_4T^3 \tag{3}$$

This formula of the third degree or less permits the calculation of R for every value of T in the specified range of T (determined by $T_{\min}$ and $T_{\max}$ in sample S).

Practical application of the above technique is facilitated by the following measures. If the range of T is large, the sample may be split into several parts and the rate B for linear approximation calculated separately for each part. To obtain a good fit in nonlinear approximation, not all terms of the polynomial may be required. For instance, a good regression formula for our data was obtained by setting $b_2 = b_3 = 0$, such that $R = b_1 + b_4T^3$. However, the improvement over linear regression remained small and was considered not significant for our purpose.

An equivalent approach to regression analysis on the primary variables is weighted regression on the rate data, $S(T, L/T)$, a technique applied when the data are of variable quality or correlated (see Draper and Smith, 1982, p. 108). The approach is best illustrated by comparison with the case of standard linear least-squares regression.

Consider again a set of thickness/time observations (T_i, L_i). In ordinary least-squares approximation with the regression line passing through the origin, stratigraphic thickness, L_i, may be modeled as

$$L_i = BT_i + \epsilon_i \tag{4}$$

where B is a factor of proportionality and ϵ_i an error term, i.e., the amount of deviation from the regression line. The regression line is fitted by determining the sum of squares of ϵ,

$$SS = \sum (L_i - BT_i)^2 \tag{5}$$

and setting the derivative dSS/dB to zero, minimizing the deviation from the fitted regression line.

$$dSS/dB = -2\sum T_i(L_i - BT_i) = 0 \tag{6}$$

$$B = \sum (T_i L_i)/\sum T_i^2 \tag{7}$$

The case of weighted regression for the rate data (L_i/T_i) is developed in analogy to the linear case with weighting applied when forming the sum of squares. The Equations (8)–(11) are analogues of Equations (4)–(7) presented above.

$$L_i/T_i = B + \epsilon_i/T_i \tag{8}$$

$$SS = \sum W_i(L_i/T_i - B)^2 \tag{9}$$

$$dSS/dB = -2 \sum W_i(L_i/T_i - B) = 0 \tag{10}$$

$$B = \sum W_i(L_i/T_i)/\sum W_i = \sum (T_iL_i)/\sum T_i^2 \tag{11}$$

where $W_i = (1/T_i^2)^{-1}$. This weighting is justified as follows: In the primary relationship $L_i = BT_i + \epsilon_i$, the error term ϵ_i has a variance of σ^2. In Equation (8), the analogous equation for rates, this error term is transformed to ϵ_i/T_i, with a variance of σ^2/T_i^2. The weighting removes the effect of this transformation by multiplying with the inverse of $1/T_i^2$.

RESULTS

We used the first approach, least-squares regression on the primary variables, L and T, to test for time-scale dependence of sedimentation rates. The test was carried out on recently published data of well-constrained rates in the range of months to millions of years. Since all these data are published, we present the exact source rather than listing the numbers again in a table. Siliciclastic data were drawn from Enos (1991, p. 64–69; $N = 363$). Carbonate data are from McNeill (1989, p. 26, 91, 166; $N = 43$); Enos (1991, p. 70–73; $N = 96$); Bosscher and Schlager (1993, p. 347; $N = 18$ from Cenozoic only); and Ebren (1966, p. 197; $N = 5$). Data are partly based on direct observation of sedimentation processes as well as interpretation of the stratigraphic record in the range of years to tens of millions of years. Thickness and time were either read directly from the stratigraphic records or were recalculated from published rates and time intervals.

The first step in the test was a plot of thickness vs. time. This produced a distinctly positive correlation—thickness clearly increases as the length of the observation span increases. This trend, an essential requirement for the analysis described below, simply reflects the bias of sedimentary geologists in selecting objects for study: the preferred objects are areas with net sediment accumulation because only they provide a record that can be analyzed.

The second step tested for scale dependence in two datasets, one from siliciclastics, one from carbonate rocks. Both populations covered a time range of 10^{-1}–10^8 years. Each population was subdivided into three subsets with observation spans of 10^{-1}–10^2 yr, 10^2–10^5 yr, 10^5–10^8 yr, respectively (see

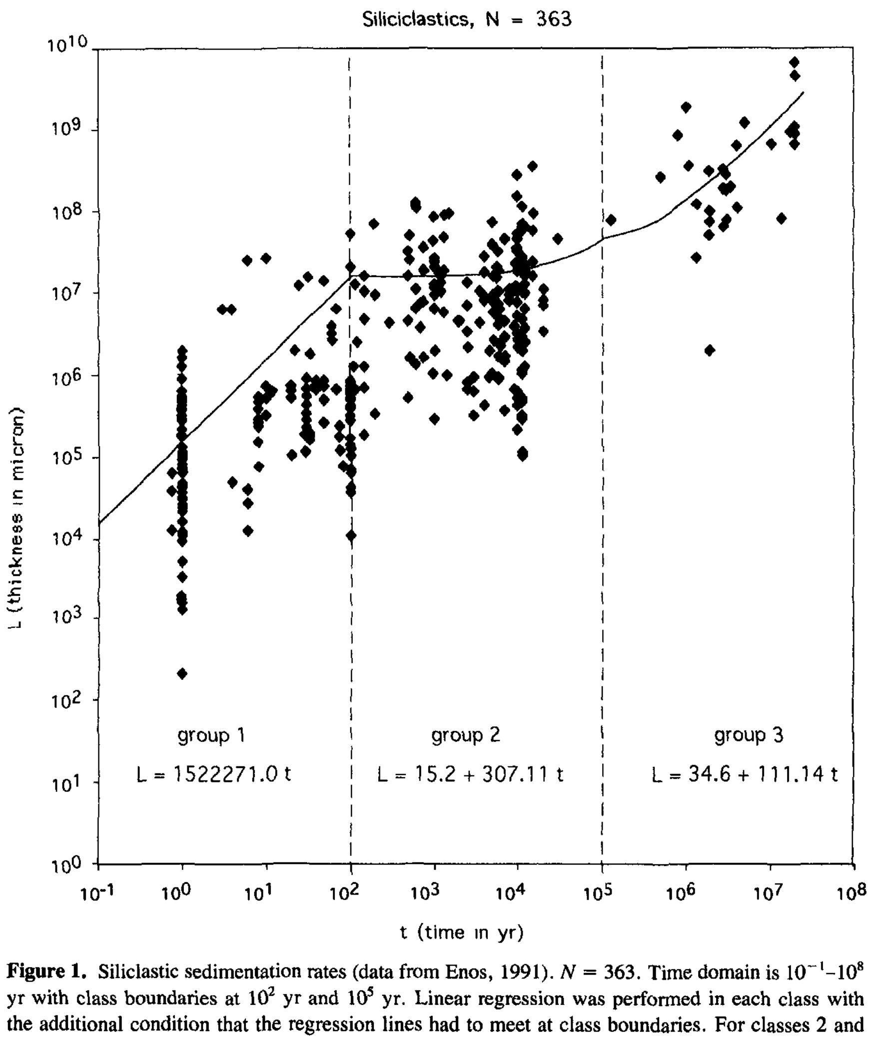

Figure 1. Siliclastic sedimentation rates (data from Enos, 1991). $N = 363$. Time domain is 10^{-1}–10^8 yr with class boundaries at 10^2 yr and 10^5 yr. Linear regression was performed in each class with the additional condition that the regression lines had to meet at class boundaries. For classes 2 and 3, this condition implies that the regression equation has a constant term, which makes the lines appear curved in this log-log plot.

discussion section for justification). Following subdivision in classes, linear least-squares regression was performed in each class. In both siliciclastic and carbonate sediments rates decrease as the length of the observation span increases (Figs. 1 and 2). The rates given by the regression equations on Figs. 1 and 2 are summarized below:

time class	siliciclastic rates (micron/yr)	carbonate rates (micron/yr)
10^{-1}–10^2 yr	152271.0	48906.5
10^2–10^5 yr	307.1	614.0
10^5–10^8 yr	111.1	41.6

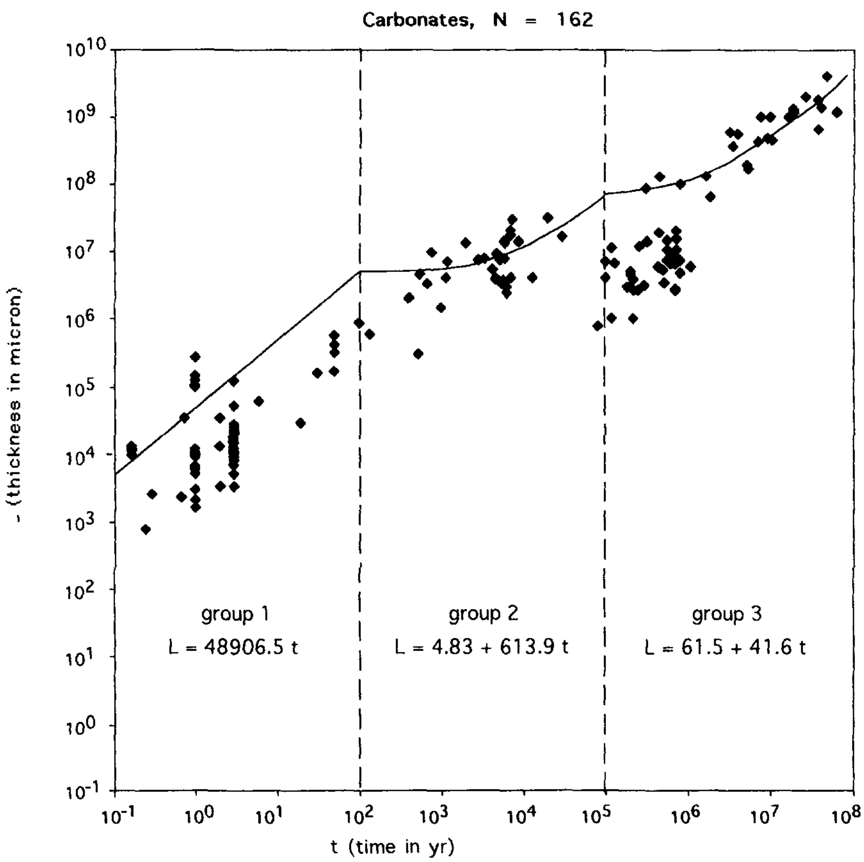

Figure 2. Carbonate sedimentation rates after Enos (1991), McNeill (1989), Bosscher and Schlager (1993), and Ebren (1996). N = 162. Total time domain is 10^{-1}–10^8 yr with class boundaries at 10^2 yr and 10^5 yr. Regression analysis as for siliciclastic data in Fig. 1.

The slope of this decrease in sedimentation rates is similar for siliciclastics and carbonates. In both populations, the center points of the first and the last subset differ by 6 orders of magnitude on the time axis while sedimentation rates decrease by 3 orders of magnitude over this time range. Thus, an arithmetic mean taken over the entire range of our data suggests that sedimentation rates decrease with the inverse of the square root of time. Considering the large scatter and the limited number of data points this should be viewed as a very crude estimate of the first-order trend (see Discussion section).

DISCUSSION

Regression analysis clearly indicates that sedimentation rates in the two datasets decrease as the time span of observation increases. The datasets are small, $N_{sil} = 363$, $N_{carb} = 162$ but the time intervals are particularly well constrained. This is important because time estimates generally introduce more uncertainty into calculations of geologic sedimentation rates than thickness measurements.

Class boundaries at 10^2 yr and 10^5 yr were set to give each class a width of three orders of magnitude. However, classes also approximately coincide with different ways of measuring time. The class of 10^{-1}–10^2 yr represents the range of direct human experience. Class 2, 10^2–10^5 yr, is largely beyond direct human experience and already requires special dating techniques. However, the techniques are very accurate in this time range. The third class, 10^5–10^8 yr is the domain of classical biostratigraphic and radiometric techniques with relatively large uncertainties.

The data are not very evenly spread along the time axis. Particularly obvious is a gap in both datasets around 10^5 yr. Data in this range may be scarce because the recent past was characterized by glacial lowstands of sea level without sedimentation on the shelves. Furthermore, resolution of 10^5 yr is near the limit of classical stratigraphic techniques and results of advanced methods, such as mass-spectrometric U-Th dating, are just beginning to appear in the literature.

Compaction has not been taken into account for the present calculations. We believe that compaction effects are not large enough to introduce a significant error for the present argument. A crude estimate will serve to illustrate this point. In the extreme, compaction leads to a volume reduction that is equal to the total porosity of the original sediment, approximately 40% for clean sands and 60–80% for muds. In most sedimentary rocks, compaction has not gone to completion and particularly in mudstones, 10–20% porosity are present even when these rocks enter the realm of metamorphosis. Consequently, to correct for compaction one would have to multiply the observed rates by, at the most, a factor of 1.6 for sands and 2.5 for muddy sediments. This is much less than the scatter of thicknesses observed at any point along the time axis (commonly 1 or 2 orders of magnitude). Furthermore, sedimentation rates calculated by regression in class 1 differ from those in class 3 by more than three orders of magnitude. Thus, we believe that compaction is of minor importance for the present discussion on the existence of a time-related trend in sedimentation rates. Compaction may become important if one considers a narrower time window (e.g., 10^2–10^6 yr) and if the goal is a precise quantification of the time-rate trend.

The most important result of our test is that it confirms the decrease of sedimentation rates with increasing time span of observations as real. This trend, examined in the range of months to tens of millions of years, has physical meaning and is not just an artifact of ratio correlation. The numerical value of this trend given above—rates proportional to the inverse of the square root of time—is similar to the trend in the rate–time plots of Sadler (1981). Our trend should be considered as a very crude estimate, based on limited data and very few, wide classes. Furthermore, it should be noted that in siliciclastics and carbonates, the decrease between class 1 and 2 is significantly greater than between class 2 and 3. The overall similarity to the rate/time plots of Sadler (1981) may be explained by the use of logarithms, which reduces the effect of spurious correlations (Hills, 1978; Atchley and Anderson, 1978); it may indicate furthermore that the dispersion of time, the common term, is small compared to the dispersion of thickness, the unique term (Kenney, 1982, 1991).

The time-scale dependence of sedimentation rates over so many orders of magnitude clearly points to a fundamental property of the sediment record. This property is elegantly summarized by the conceptual model of Plotnick (1986), who considers the stratigraphic record as an irregular Cantor set, i.e., a fractal function with gaps on all scales. We estimate that this model may be valid for time intervals from fractions of seconds to billions of years. The lower limit may be set for detrital grains by the time required to deposit individual grains. For all depositional systems, the upper limit of the Cantor model is given by the maximum time span for which undisturbed sediment successions exist.

ACKNOWLEDGMENTS

Hemmo Bosscher and Edward G. Purdy were important partners in our early discussions on this subject. W.S. acknowledges partial support by the VU Industrial Associates Program in Sedimentology.

REFERENCES

Aitchison, J., 1986, The statistical analysis of compositional data: Chapman and Hall, New York, 416 p.

Anders, M. H., Krueger, S. S., and Sadler, P. M., 1987, A new look at sedimentation rates and the completeness of the stratigraphic record: Jour. Geology, v. 95, p. 1-14.

Atchley, W. R., Gaskins, C. T., and Anderson, D., 1976, Statistical properties of ratios. I. Empirical results: System. Zoology, v. 25, no. 1, p. 137-148.

Atchley, W. R., and Anderson, D., 1978, Ratios and the statistical analysis of biological data: System. Zoology, v. 27, no. 1, p. 71-78.

Barrell, J., 1917, Rhythms and the measurements of geologic time: Geol. Soc. America Bull., v. 28, no. 12, p. 745-904.

Bosscher, H., and Schlager, W., 1993, Accumulation rates of carbonate platforms: Jour. Geol., v. 101, no. 2, p. 345-355.

Dott, R. H., 1988, Something old, something new, something borrowed, something blue—a hindsight and foresight of sedimentary geology: Jour. Sedimentary Petrology, v. 58, no. 2, p. 358-364.

Draper, N. R., and Smith, H., 1981, Applied regression analysis: Wiley & Sons, New York, 709 p.

Ebren, P., 1996, Impact des variations rapides du niveau marin sur le developpement des atolls au Quaternaire: Mururoa (Polynesie Francaise): Unpubl. doctoral dissertation, Univ. de Provence, Marseille, p. 1-309.

Enos, P., 1991, Sedimentary parameters for computer modelling, *in* Franseen, E. K., Watney, W. L., Kendall, C. G. St. G., and Ross, W., Eds., Sedimentary modelling: computer simulations and methods for improved parameter definition: Kansas Geol. Survey Bull., v. 233, p. 63-101.

Gingerich, P. D., 1983, Rates of evolution: effects of time and temporal scaling: Science, v. 222, no. 4620, p. 159-161.

Gould, S. J., and Gingerich, P. D., 1984, Smooth curve of evolutionary rate: a psychological and mathematical artifact. Discussion and reply: Science, v. 226, no. 4677, p. 994-995.

Hills, M., 1978, On ratios—a response to Atchley, Gaskins, and Anderson: System. Zoology, v. 27, no. 1, p. 61-62.

Jackson, D. A., and Somers, K. M., 1991, The spectre of "spurious" correlations: Oecologia, v. 86, no. 1, p. 147-151.

Kasarda, J. D., and Nolan, P. D., 1979, Ratio measurement and theoretical inference in social research: Social Forces, v. 58, no. 2, p. 212-227.

Kenney, B. C., 1982, Beware of spurious self-correlations: Water Resources Res., v. 18, no. 4, p. 1041-1048.

Kenney, B. C., 1991, Comments on "Some misconceptions about the spurious correlation problem in the ecological literature" by Y. T. Prairie and D. F. Bird: Oecologia, v. 86, no. 1, p. 152.

McNeill, D. F., 1989. Magnetostratigraphic dating and magnetization of Cenozoic platform carbonates from the Bahamas: unpubl. doctoral dissertation, Univ. of Miami, Coral Gables, p. 1-210.

Pearson, K., 1897, On a form of spurious correlations which may arise when indices are used in the measurement of organs: Proc. Roy. Soc. London, v. 60, p. 489-502.

Prairie, Y. T., and Bird, D. F., 1986, Some misconceptions about the spurious correlation problem in the ecological literature: Oecologia, v. 81, no. 1, p. 285-288.

Plotnick, R. E., 1986, a fractal model for the distribution of stratigraphic hiatuses: Jour. Geology, v. 94, no. 6, p. 885-890.

Reineck, H.-E., 1960, Über Zeitlücken in rezenten Flachsee-Sedimenten: Geol. Rundschau, v. 49, no. 1, p. 149-161.

Sadler, P., 1981, Sediment accumulation rates and the completeness of stratigraphic sections: Jour. Geology, v. 89, no. 4, p. 569-584.

Schindel, D. E., 1982, Resolution analysis: a new approach to the gaps in the fossil record: Paleobiology, v. 8, no. 4, p. 340-383.