Estimating the Dimension of Weather and Climate Attractors: Important Issues about the Procedure and Interpretation

A. A. TSONIS

Department of Geosciences, University of Wisconsin—Milwaukee, Milwaukee, Wisconsin

J. B. ELSNER

Department of Meteorology, The Florida State University, Tallahassee, Florida

K. P. GEORGAKAKOS

Department of Civil and Environmental Engineering and Iowa Institute of Hydraulic Research, The University of Iowa, Iowa City, Iowa

3 March 1992 and 4 November 1992

ABSTRACT

When the reconstruction of attractors from observables is sought, the Grassberger–Procaccia algorithm for estimating the correlation dimension is often used. An overview of recent developments concerning data requirements and algorithm performance is presented within. In view of these developments the significance of previously estimated dimensions of weather and climate attractors is discussed.

1. Introduction

Lately, ideas from the theory of nonlinear dynamical systems and chaos have been applied to many problems from many different disciplines, including atmospheric sciences. The main goal is the search for low-dimensional chaos and the extraction of the properties of the underlying attractors, if any. The procedure often involves an observable (time series) and a reconstruction of the attractor. The reconstruction is achieved by taking a scalar time series $x(t_i)$ and its successive time shifts (delays) as coordinates of a vector time series given by

$$\mathbf{X}(t_i) = \{x(t_i), x(t_i + \tau), \cdots, x(t_i + (n-1)\tau)\}, \tag{1}$$

where n is the dimension of the vector $\mathbf{X}(t_i)$ (often referred to as the embedding dimension) and τ is an appropriate delay (Packard et al. 1980; Ruelle 1981; Takens 1981). For proper reconstructions the embedding dimension n should be equal or greater to $2D + 1$, where D is the dimension of the manifold containing the attractor. Such an embedding preserves the topological properties of the attractor. More specifically the embedding will be a diffeomorphism—a differentiable mapping with a differentiable inverse—from the true phase space to the delay space. This is Whitney's theorem and, strictly speaking, is valid only when we have an infinite and dense set in our disposal. When we only have a limited dataset the theorem may not be valid. In fact, in those cases the word embedding is used loosely as any topologist will argue.

For an n-dimensional phase space, a "cloud" or a set of points will be observed. From this set the various dimensions and exponents that characterize the underlying attractor can be calculated. The most popular approach is to calculate the correlation dimension. According to this approach (Grassberger and Procaccia 1983a,b), given the cloud of points one finds the number of pairs $N(r, n)$ with distances less than a distance r. In this case, if for significantly small r, we find that

$$N(r, n) \propto r^{d_2}, \tag{2}$$

then the scaling exponent d_2 is the correlation dimension of the attractor for that n. Since the dimension of the underlying attractor is not known, we test Eq. (2) for increasing values of n and check for a saturation value D_2, which will be an estimation of the correlation dimension of the attractor. For more details on the aforementioned procedures and the applications related to weather and climate, see the review articles by Tsonis and Elsner (1989, 1990).

This brief report will discuss three important issues pertaining to the applications of the aforementioned procedures: 1) the choice of τ, 2) how many points are enough, and 3) what do the results mean?

Corresponding author address: Dr. A. A. Tsonis, Department of Geosciences, University of Wisconsin—Milwaukee, P.O. Box 413, Milwaukee, WI 53201.

2. The choice of τ

When we reconstruct the attractor by producing a cloud of points at a given embedding dimension, those points should be independent of each other. Otherwise $x(t + \tau) \approx x(t)$ and the points tend to fall on the diagonal. As a result the estimation of the correlation dimension may be biased (typically underestimated). Therefore, τ must be chosen so as to result in points that are not correlated to previously generated points. Thus, a first choice of τ should be in terms of the decorrelation time of the time series under investigation. The question now arises: How do we define the decorrelation time? A straightforward procedure is to consider the decorrelation time equal to the lag at which the autocorrelation function for the first time attains the value of zero. Other approaches consider the lag at which the autocorrelation function attains a certain value like $1/e$, 0.5 (Schuster 1988), or 0.1 (Tsonis and Elsner 1988). Another suggestion for the choice is to take τ equal to T/n where T is the dominant periodicity (as revealed by Fourier analysis) and n is the embedding dimension. In this way τ gives some measure of statistical independence of the data average over an orbit and it is an appropriate approach if the autocorrelation function is periodic. As it was pointed out, however, by Frazer and Swinney (1986) the autocorrelation function measures the linear dependence among successive points and may not be appropriate when we are dealing with nonlinear dynamics. They argue that what should be used as τ is the local minimum of the mutual information that measures the general dependence among successive points. Evidently, no one of the aforementioned rules has emerged as the undisputed rule for choosing τ, but the mutual information approach appears to have the edge. Nevertheless, a very reassuring practice is experimenting with various τ's (while repeating the aforementioned constraints) in order to address possible effects of the choice of τ. Note that in most (if not all) of the studies reporting on low-dimensional attractors in weather and climate, care has been taken to address the choice of τ. Commonly, the delay parameter has been derived by means of the autocorrelation function.

3. How many points?

Figure 1 shows the correlation dimension d_2 as a function of the embedding dimension n for data representing the pulse of rainfall [a time series of the time Δt (s) between successive raingage signals each corresponding to the collection of 0.01 mm of rain]. Here the calculation of d_2 at higher embedding dimensions is not achieved by shifting successively one time series by a delay parameter but by introducing a new event of size 2200 (i.e., an independent time series of the same meteorological convective character) every time we increase the embedding dimension. Thus, the observed points in an n-dimensional point are independent and the need of defining a suitable τ is overcome. Keeping this example in mind, what follows is an in-depth discussion on some facts concerning the correct application and limitations of the Grassberger–Procaccia algorithm.

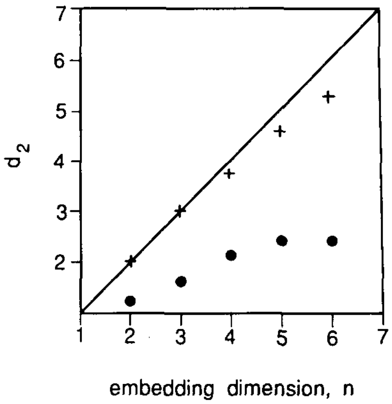


FIG. 1. Correlation dimension d_2 versus embedding dimension n (a) for data representing pulse of storm rainfall (points) and (b) a random sample (crosses).

For a finite dataset one can argue that there is a distance (or radius) r below which there are no pairs of points (depopulation). At the other extreme, when the radius approaches the diameter of the cloud of points, the number of pairs will increase no further as the radius increases (saturation). The scaling region would always be found somewhere between depopulation and saturation. As it has been pointed out by Essex et al. (1987) and Tsonis and Elsner (1990) the scaling region may be completely masked if the number of points $N(r, n)$ is not sufficiently large or when the embedding dimension is increased beyond a critical embedding dimension n_c. This unavoidably brings us to the following question: What is the necessary number of points N for a given embedding dimension?

This problem can be approached by assuming that in all embedding dimensions n less than the dimension of the object in question, the object is space filling like uniformly distributed random numbers in the interval [0, 1]. The painful exercise of determining the minimum number of points was first tackled by Smith (1988), who concluded that this number is equal to 42^m, where m is the smallest integer above the dimension of the object, which under the aforementioned assumption is the dimension in which the random numbers are embedded in. Thus, for $m = 4$, if N is not at least equal to 3 111 696 no accurate estimate of d_2 can be obtained. Such a restrictive figure put in doubt all the reports claiming low-dimensional attractors since none of these reports used such great numbers of data points. In fact, Smith's result effectively "killed" all hopes for estimating the dimension of low-dimensional attractors irrespective of the availability of data, since even supercomputers could not handle such vast samples. But accurate estimates for $m = 4$ can be obtained with as little as 5,000 points (see Tsonis and Elsner 1990). This number is significantly lower than 3 111 696. Why this great discrepancy? The only explanation is that the 42^m conclusion is in error. In fact, it has been recently demonstrated (Nerenberg and Essex 1990) that Smith's procedure to obtain the 42^m estimate is flawed and that the data requirements are not as nearly as extreme. In fact, the minimum number of points $N_{\min}$ required to produce no more than an error A (typically $A = 0.05n$) is

$$N_{\min} = \frac{\sqrt{2}[\Gamma(n/2+1)]^{1/2}}{(A \ln k)^{(n+2)/2}} \times \left[\frac{2(k-1)\Gamma(n+4/2)}{[\Gamma(1/2)^2\Gamma(n+3/2)]}\right]^{n/2} \frac{n+2}{2}, \quad (3)$$

where n is the embedding dimension, and $\Gamma(x)$ is the gamma function. The parameter k indicates how wide the scaling interval is. Recall that d_2 is estimated via a relation $d_2 = [\ln C(r') - \ln C(r)]/[\ln(r') - \ln(r)]$. The parameter k is then defined as $k = r'/r$. For a faster use for $A = 0.05n$ (or in other words for a 95% confidence) and for $k = 4$ the Eq. (3) can be approximated for $n < 20$ by

$$N_{\min} \propto 10^{(2+0.4n)}. \quad (4)$$

Thus, for $n = 4$, $N_{\min} \sim 10^{3.6} \sim 4000$ points as mentioned above.

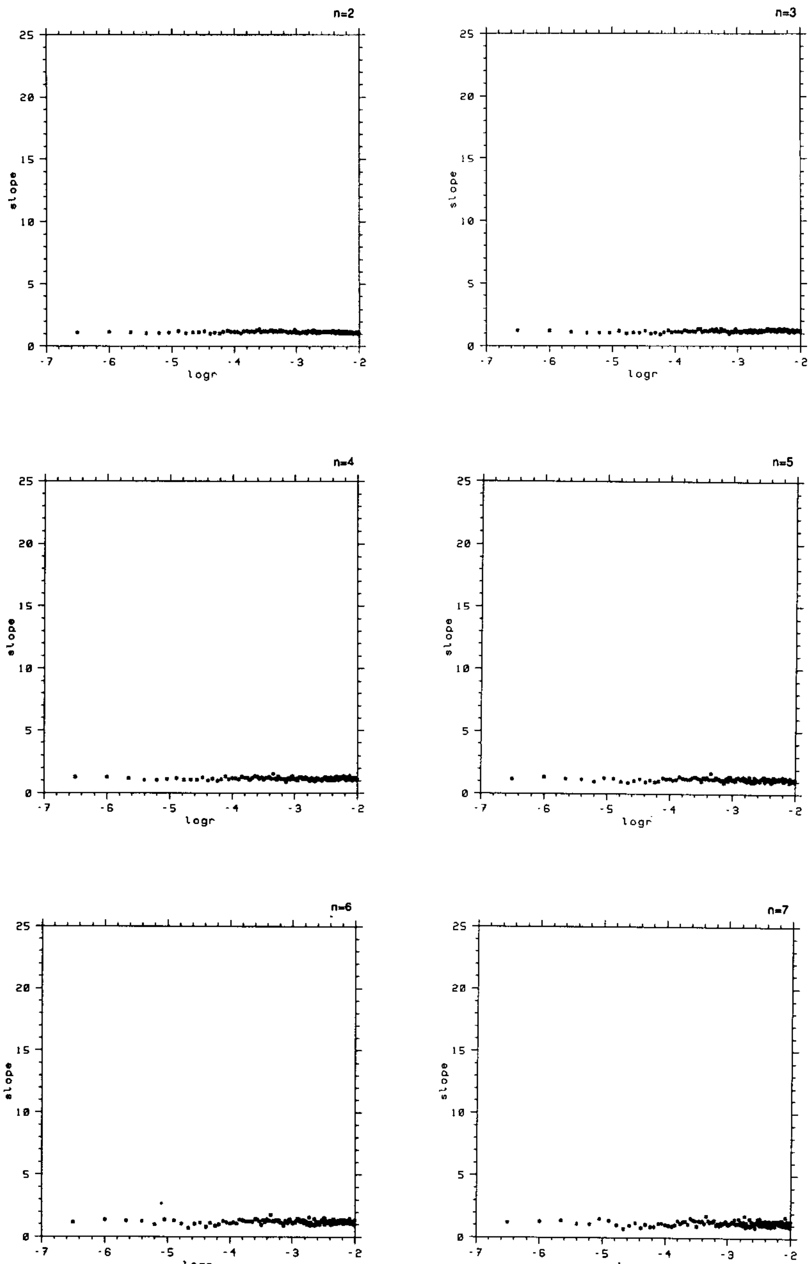
The aforementioned discussion unavoidably brings us to the next very important point. Because of the underlying assumptions all the theoretical calculations and derivations of the necessary number of points as a function of embedding dimension presented up to this point are valid only as long as the embedding dimension n is less than correlation dimension D_2. In addition, it is quite possible that estimates of the number of points would depend on the type of the attractor (nonuniform, fractal); an issue that has not yet been addressed in those calculations. Recently, Lorenz (1991) considered a three-variable chaotic system and produced a model by taking seven linearly coupled copies of that system. The model is described by 21 equations and for a choice of the coupling coefficient its dimension is equal to 17. Lorenz applied the Grassberger–Procaccia algorithm using 4000 values of a selected variable from the model. He found that if N is not too large then the algorithm underestimates the dimension. However, 1) different variables can yield different estimates, and 2) suitably selected variables can some times yield a fairly good estimate. Such a variable is one that is strongly coupled with the rest of the variables of the model. If this result holds for all chaotic systems then it is not the sample size we should worry about but the choice of the observable!

Over and above these important issues, it is not known at this point whether or not (and especially for fractal sets) the need for data increases at the same rate with embedding dimension when $n > D_2$. Experimentation with known dynamical systems indicates that in this case even though the need for data may increase it may not be as severe as it is predicted by the aforementioned formulas! For example, Fig. 2 shows slope versus log$r$ plots of a sample of size 2000 for an observable from the Hénon map for embedding dimensions 2, 3, 4, 5, 6, 7, and 8. We observe that for $n = 2$, $d_2 = 1.25$ (which the Hausdorff–Besicovitch dimension of the Hénon map). Subsequently, we observe that up to embedding dimension 8, fairly accurate estimates of the correlation dimension can be obtained. In fact, we do not observe too much fluctuation or an underestimation of d_2 with increasing embedding dimension. Similar results are obtained for other dynamical systems like the logistic map and the Lorenz system. The data used to produce Fig. 1 tell a similar story, but the slope versus log$r$ curves exhibit more scatter due the presence of noise in real world data. Of course, eventually (i.e., for $n \gg D_2$) fluctuations may mask any scaling region. These results provide some evidence that for fractal sets the need for data may increase at a much slower rate for embedding dimensions higher than the correlation dimension of the attractor. We should point out, however, that such a conclusion may not be valid for every dynamical system or dataset. Do we, therefore, in cases where saturation is observed at an embedding dimension $n_s > D_2$, need $N \sim 10^{(2+0.4n_s)}$ points or just $N \sim 10^{(2+0.4D_2)}$ points (again provided that n_s is not much higher than D_2)? If we assume that $N \sim 10^{(2+0.4D_2)}$ data points are required (see also Nerenberg and Essex 1990), we find from Fig. 3 that many of the studies reporting on low-dimensional attractors in weather and climate come very close to satisfying the requirements (and may come even closer if we fix our accuracy to $0.10n$ instead of $0.05n$).

4. What does a finite dimension mean?

Coming back to Fig. 1, we may now present the results obtained if a random sample of the size 2200 is used in the calculations. These results are indicated by crosses and clearly show that there is a great difference between the random sample and the rainfall data. Which brings us to the final point we would like to discuss. The results in Fig. 1 simply indicate that the rainfall data are different than a sequence of random numbers. Random samples are simulated so that they have similar statistical properties (such as probability distribution, spectra, etc.) with the actual data and are often referred to as surrogate data. Do the results in Fig. 1 alone provide strong evidence that there is an underlying attractor in the rainfall data?

The importance of this question lies in the fact that given a strange attractor, a finite correlation dimension follows, whereas the opposite statement may not be



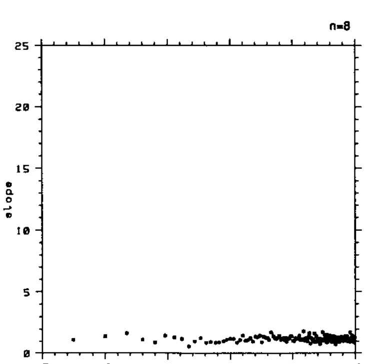


FIG. 2. Slope vs log$r$ for an observable from the Hénon map ($x_{t+1} = 1 - ax_t^2 + y_t$, $y_{t+1} = bx_t$, $a = 1.4$, $b = 0.3$).

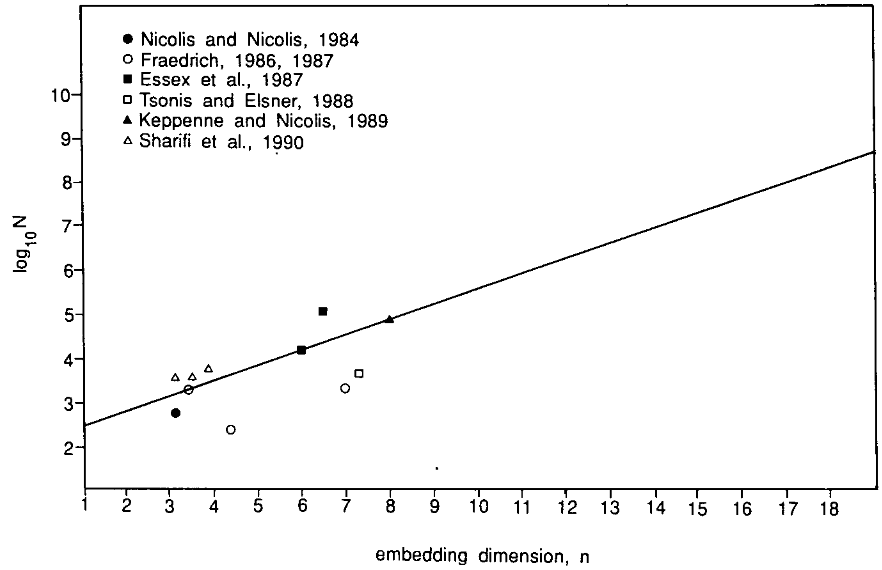


FIG. 3. The solid curve is the function expressed by Eq. (3). The other symbols depict number of points used, and attractor dimensions reported by various investigators. Everything close or above the solid line can be considered as convincing evidence regarding the existence of low-dimensional attractors in weather and climate. Note that the solid line represents the needed number of points assuming that $n = D_2$.

true. As many will argue there is no theorem that says given a finite correlation dimension an associated system has a strange attractor! To complicate things, counter examples of certain Hamiltonian, or random systems yielding finite correlation dimensions, or both have been presented.

Benettin et al. (1986), Osborne and Caponio (1990), Chernikov et al. (1989), Bishop and Lomdahl (1986), Bishop et al. (1986), and others have presented examples of Hamiltonian systems having a small finite correlation dimension. As we know, Hamiltonian systems cannot have an attractor.

In addition, Osborne et al. (1986) and Osborne and Provenzale (1989) have argued that a certain class of *random* sequences would exhibit a finite correlation dimension. This class of random sequences includes self-affine sequences that exhibit a power law spectra of the form $P(f_k) = Cf_k^{-a}$ for $1 < a < 3$ and are commonly referred to as fractional Brownian motions (fBms) or colored noise. In this case the *trail* of n-independent realizations (each one representing one phase space dimension) is self-similar with a theoretically predicted fractal dimension equal to $2/(a-1)$. When the Grassberger–Procaccia algorithm is applied to trails or to trajectories reconstructed from a single sequence via the method of delays, a finite correlation dimension (close to the fractal dimension of the trail) is obtained. Thus they suggested that a finite value for the correlation dimension cannot be indicative of a dynamical system with a finite number of degrees of freedom. It can only indicate a lower bound for the actual number of degrees of freedom, which might be infinite! For the first time it was shown that the algorithm could not differentiate between chaos and this type of noise (colored noise).

As it turns out, however, the observation of Osborne and Provenzale may have no relevance to the practical estimation of dimensions from time series. We must note that the concept of fractal dimension can be applied to time series in two distinct ways. The first is to indicate the *number of degrees of freedom* in the underlying dynamical system. The second is to quantify the *self-similarity* of the trajectory in phase space. The Grassberger–Procaccia algorithm yields the first one. It does not really provide an estimate of the self-similarity of the trajectory. For example, applying the algorithm to an observable from the Lorenz system results in a dimension of about 2.07. This value has nothing to do with the self-similarity of the Lorenz trajectory (which by the way is not self-similar). Theiler (1991) attacked this issue in an analytical way and proved that Osborne and Provenzale's anomalous scaling $C(r, n) \propto r^{2/(a-1)}$ would not have been observed had they used the required number of points (which in the case of fBms is very large due to their very long correlations) or if they had evaluated the correlation integral for smaller r's. Instead, the scaling $C(r, n) \propto r^n$ would have been observed.

The aforementioned developments raise several questions as far as the application of the algorithm is concerned. Can the algorithm really be "fooled" or is it simply not properly applied? Most importantly, what are Hamiltonian systems, or fBms, or both good for in the interpretation of data like those sampled from physical systems? Hamiltonian systems in nature are at best rare and fBms are nonstationary processes. For fBms the autocorrelation function $C(\zeta) \propto \zeta^{1-a}$ for $N \to \infty$. Therefore, $C(\zeta)$ never reaches the value of zero. Thus, no pairs are really independent. In addition for finite N the decorrelation time is a function of N. Because of their properties, Mandelbrot (1983) remarked that fBms are not effective candidates for modeling natural processes. If any natural process were an fBm, it would have by now grown enough to destroy nature. Therefore, by definition natural processes are not fBms and cannot exhibit $1/f^a$ spectra. The argument that some data might, during a finite time, mimic an fBm can be dealt with by simply testing the data for stationarity or by looking at the autocorrelation function for various lengths of the record or by making sure that the dataset is longer than the length of a dominant period or by nonlinear prediction (Tsonis and Elsner 1992).

5. Conclusions

This work was designed to address some of the current issues in estimating correlation dimensions. We have explained some ways according to which the algorithm might be working and at the same time we have demonstrated the existence of certain weaknesses, which make it (and other similar methods) somewhat more qualitative thus often requiring subjective judgment about where an attractor of a given dimension exists. In fact, both Grassberger (1986) and Ruelle (1990) in their critique of work reporting on low-dimensional attractors were not fully aware of issues discussed here, such as the critical embedding dimension (see also Essex and Nerenberg 1991). Consequently, we showed that several studies reporting on attractors in weather and climate might have used enough points.

There are several possibilities here. In weather and climate we deal with coarse data in which small-scale processes are absent to begin with. These large-scale coarse data are likely to obey to a closed dynamics, which need not appeal to the small-scale processes. Many of the attractors deduced in geophysics are likely to be of this kind. For one thing, owing to the sampling time, they discard fully developed turbulence, which certainly corresponds to a highly dimensional attractor. In his latest work Lorenz (1991) showed that many of the reported dimensions might be significantly lower than the actual dimensions if not enough points were used. On the other hand, he showed that the number of points need not be very large if the right variable is considered. Lorenz also demonstrated that if N is small, the estimated dimension using a weakly coupled variable from the 21-equation model ought to resemble the estimate using a variable of a original three-variable system. Consequently, real-data studies reporting on low-dimensional attractors may not be altogether meaningless; they just need to be reinterpreted. Lorenz states that, as Tsonis and Elsner (1989) suggested, the atmosphere might be viewed as loosely coupled subsystems. In this case these studies attempt to measure the dimension of a subsystem.

During our evaluation we have established that due to existing algorithm weaknesses all results present just evidence rather than proof of existence. Recall that the first work reporting on a low-dimensional climatic attractor (Nicolis and Nicolis 1984) had a question mark on its title!

We conclude that evidence for low-dimensional attractors should be "fortified" by additional evidence such as Lyapunov exponents, model simulations, and nonlinear forecasting. The latter especially may provide the best test of underlying determinism (Tsonis and Elsner 1992). As stated in Sugihara and May (1990) "prediction, is after all the sine qua non of determinism." Even though research in this area is in an early stage, preliminary results (Elsner and Tsonis 1992) have reported that nonlinear prediction using climatic records similar to those used in Nicolis and Nicolis (1984) supports previous conclusions about the existence of attractors in climate.

Acknowledgments. We thank Drs. C. Nicolis and C. Essex for valuable comments in an earlier version of the paper.

REFERENCES

Benettin, G., D. Casati, L. Galgani, A. Giolgilli, and L. Sironi, 1986: Apparent fractal dimension in conservative dynamical systems. *Phys. Lett.*, **118A,** 325–330.

Bishop, A. R., and P. S. Lomdahl, 1986: Nonlinear dynamics in damped sine-Gordon systems. *Physica,* **18D,** 54–66.

——, M. G. Forest, D. W. McLaughlin, and E. A. Overman II, 1986: A quasi-periodic route to chaos in near-integrable PDE. *Physica,* **23D,** 293–328.

Chernicov, A. A., R. Z. Sagdeev, D. A. Usikov, and G. M. Zavlasky, 1989: *Sov. Sci. Rev. C. Math. Phys.*, **8,** 83–172.

Elsner, J. B., and A. A. Tsonis, 1992: Nonlinear prediction, chaos and noise. *Bull. Amer. Meteor. Soc.*, **73,** 49–60.

Essex, C., and M. A. H. Nerenberg, 1991: Comments on "Deterministic chaos: The science and the fiction." *Proc. R. Soc. London A.*, **453,** 287–292.

——, T. Lookman, and M. A. H. Nerenberg, 1987: The climate attractor over short time scales. *Nature,* **326,** 64–66.

Fraedrich, K., 1986: Estimating the dimensions of weather and climate attractors. *J. Atmos. Sci.*, **43,** 419–432.

——, 1987: Estimating weather and climate predictability on attractors. *J. Atmos. Sci.*, **44,** 722–728.

Frazer, A. M., and H. L. Swinney, 1986: Independent coordinates for strange attractors from mutual information. *Phys. Rev. A,* **33,** 1134–1140.

Grassberger, P., 1986: Do climatic attractors exist? *Nature,* **323,** 609–612.

——, and I. Procaccia, 1983a: Characterization of strange attractors. *Phys. Rev. Lett.*, **50,** 346–349.

——, and ——, 1983b: Measuring the strangeness of strange attractors. *Physica,* **9D,** 189–208.

Keppenne, C. L., and C. Nicolis, 1989: Global properties and local structure of the weather attractor over Western Europe. *J. Atmos. Sci.*, **46,** 2356–2370.

Lorenz, E. N., 1991: Dimension of weather and climate attractors. *Nature,* **353,** 241–244.

Mandelbrot, B. B., 1983: *The Fractal Geometry of Nature.* Freeman, 461 pp.

Nerenberg, M. A. H., and C. Essex, 1990: Correlation dimension and systematic geometric effects. *Phys. Rev. A,* **42,** 7065–7074.

Nicolis, C., and G. Nicolis, 1984: Is there a climatic attractor? *Nature,* **331,** 529–532.

Osborne, A. R., and A. Provenzale, 1989: Finite correlation dimension for stochastic systems with power-law spectra. *Physica D,* **35,** 357–381.

——, and R. Caponio, 1990: Fractal trajectories and anomalous diffusion for chaotic motions in 2D turbulence. *Phys. Rev. Lett.,* **64,** 1733–1736.

——, A. D. Kirwan, Jr., A. Provenzale, and L. Bergamasco, 1986: A search for chaotic behavior in large and mesoscale motions in the Pacific Ocean. *Physica,* **23D,** 75–83.

Packard, N. H., J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, 1980: Geometry from a time series. *Phys. Rev. Lett.,* **45,** 712–716.

Ruelle, D., 1981: Chemical kinetics and differentiable dynamical systems. *Nonlinear Phenomena in Chemical Dynamics,* A. Pacault and C. Vidal, Eds., Springer-Verlag.

——, 1990: Deterministic Chaos: The science and the fiction. *Proc. Roy. Soc. London,* **A427,** 241–248.

Schuster, H. G., 1988: *Deterministic Chaos.* VCH Weinheim, 122.

Sharifi, M. B., K. P. Georgakakos, and I. Rodriguez-Iturbe, 1990: Evidence of deterministic chaos in the pulse of storm rainfall. *J. Atmos. Sci.,* **47,** 888–893.

Smith, L. A., 1988: Intrinsic limits on dimension calculations. *Physics Lett. A.,* **133,** 283–288.

Sugihara, G., and R. M. May, 1990: Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature,* **344,** 734–741.

Takens, F., 1981: *Dynamical Systems and Turbulence. Lecture Notes in Mathematics 898.* Springer-Verlag.

Theiler, J., 1991: Some comments on the correlation dimension of $1/f^{a}$ noise. *Phys. Lett. A,* **155,** 488–492.

Tsonis, A. A., and J. B. Elsner, 1988: The weather attractor over very short time scales. *Nature,* **33,** 545–547.

——, and ——, 1989: Chaos strange attractor and weather. *Bull. Amer. Meteor. Soc.,* **70,** 16–23.

——, and ——, 1990: Comments on "Dimension analysis of climatic data." *J. Climate,* **3,** 1502–1505.

——, and ——, 1992: Nonlinear prediction as a way of distinguishing chaos from random fractal sequences. *Nature,* **358,** 217–220.