A Possible Explanation for Low Correlation Dimension Estimates for the Atmosphere

SHAFIQUL ISLAM*

Department of Civil and Environmental Engineering, University of Cincinnati, Cincinnati, Ohio

RAFAEL L. BRAS AND IGNACIO RODRIGUEZ-ITURBE†

Ralph M. Parsons Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Manuscript received 1 July 1991, in final form 4 April 1992)

ABSTRACT

There have been numerous attempts to detect the presence of deterministic chaos by estimating the correlation dimension. The values of reported correlation dimension for various geophysical time series vary between 1.3 and virtually infinity (i.e., no saturation). It is pointed out that analyzing variables that depend on physical constraints and thresholds, like precipitation, may lead to underestimation of the correlation dimension of the underlying dynamical system.

1. Introduction

Recent developments in the areas of nonlinear dynamics and chaos have highlighted the possibilities of distinguishing deterministic chaos from stochastic noise by analyzing a single variable time series from a complex physical system. A procedure introduced by Grassberger and Procaccia (1983) attempts to do so by estimating the correlation dimension of the attractor from a time series. The fundamental assumption of the procedure suggested by Grassberger and Procaccia (1983) is that if the exact mathematical description of a dynamical system is unknown, the state space can be reconstructed by a single variable time series (Packard et al. 1980). Once we accept this assumption, the procedure becomes a very powerful tool for analyzing natural systems to understand their behavior. If the correlation dimension is found to be small, then the reasonable conclusion is that the system is describable by a simple model with few [for example, less than or equal to 4 in the case of climate according to Nicolis and Nicolis (1984)] degrees of freedom. The significance of this is enormous, as Theiler (1990) nicely puts it: "by merely observing a single component of a potentially complex physical system, one can actually count the active degrees of freedom in the system." It is interesting to note, though, that Grassberger (1986) recognizes that just having an estimate of correlation dimension does not tell anything about the structure of the underlying dynamics.

The atmospheric sciences community has not missed the power of the above arguments. To what extent are the dynamics of the atmosphere captured by a system with only a few variables? Can we analyze a single variable, like pressure, temperature, winds, or precipitation, and deduce the number of degrees of freedom of the atmosphere? If the correlation dimension estimated from the analysis of an atmospheric variable is small, can we confidently search for a correspondingly small number of differential equations that describe the system?

There have been numerous attempts to estimate the correlation dimension of several observable variables at different space–time scales of the atmosphere and the ocean. The estimated values of correlation dimension vary between 1.3 and virtually infinity (i.e., no saturation). One is naturally puzzled: how can we explain such a wide range of values for the "same" atmospheric system? Can the atmosphere truly be a small dimensional system?

In the next section, we will quickly review the dimensional characterization of attractors. Particularly, we will describe a commonly used estimate of the correlation dimension. In section 3, we review some of the available results for the correlation dimension as obtained from the analysis of several types of atmospheric and oceanographic variables. Several existing arguments that question the validity of some of the low-dimensional results will be pointed out.

In section 4, we propose that low correlation dimension may be a result of the dependence of some atmospheric variables on thresholds and other physical

* Formerly at Ralph M. Parsons Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts.

† Instituto Interacional de Estudios Avanzados, Caracas, Venezuela.

Corresponding author address: Rafael L. Bras, Ralph M. Parsons Laboratory, MIT, Cambridge, MA 02139.

constraints. The correlation dimension resulting from the analysis of this type of variable should not be construed as representative of the underlying dynamical system that produces them. This paper does not question the concept of correlation dimension nor the procedures normally used to estimate it. It suggests limits to the interpretation that may be given to some results.

2. Dimensional characterization of attractors

Let $X_0(t)$ be the time series of a dynamical variable presumably from a potentially complex natural system. As the n variables $\{X_k(t)\}$ describing the system satisfy a set of first-order differential equations, successive differentiation in time reduces the problem to a single highly nonlinear differential equation of nth order for one of these variables. Thus, instead of $X_k(t)$, $k = 0, 1, \cdots, n-1$, we may use $X_0(t)$, the variable of the time series data, and its $(n-1)$ successive derivatives $X_0^{(k)}(t)$, $k = 1, \cdots, n-1$, to be the n variables of the problem spanning the phase space of the system (Packard et al. 1980). Therefore, in principle, sufficient information is given in a one-dimensional time series to construct a multidimensional phase space for studying the system's dynamics.

A simple procedure, suggested originally by Ruelle (1981), avoids the problem of calculating $X_0^{(k)}(t)$ from a time series of $X_0(t)$ and uses multiple time delays as a surrogate for successive derivatives. A point in an n-dimensional phase space is now defined by the variables:

$$X_0(t), X_0(t+\tau), \ldots, X_0[t+(n-1)\tau]. \quad (1)$$

To construct a well-behaved phase space by time delay, a careful choice of τ is critical. A procedure to do so will be discussed later.

The procedure to calculate the correlation dimension of the strange attractor from a time series (Grassberger and Procaccia 1983) will now be briefly discussed. Consider a set of N points on an attractor embedded in a phase space of n dimensions, constructed from a single time series:

$$\begin{gathered} X_0(t_1), X_0(t_2), \ldots, X_0(t_N) \\ X_0(t_1+\tau), X_0(t_2+\tau), \ldots, X_0(t_N+\tau) \\ \vdots \\ X_0[t_1+(n-1)\tau], X_0[t_2+(n-1)\tau], \\ \ldots, X_0[t_N+(n-1)\tau]. \quad (2) \end{gathered}$$

For notational convenience, let vector $\mathbf{X}_i$ denote a point in phase space whose coordinates are $\{X_0(t_i), \ldots, X_0[t_i + (n-1)\tau]\}$. Due to the exponential divergence of trajectories, most pairs $(\mathbf{X}_i, \mathbf{X}_j)$, except for $i = j$, will be dynamically uncorrelated; however, all the points must lie on the attractor. Therefore, they should be spatially correlated. This is the reasoning behind the use of correlation function $C(r)$ defined as

$$C(r) = \lim_{N\to\infty} \frac{1}{N^2 - N} \sum_{\substack{i,j=1 \\ i\neq j}}^{N} H(r - |\mathbf{X}_i - \mathbf{X}_j|), \quad (3)$$

where H is the Heaviside function, $H(u) = 0$ if $u \leq 0$, and $H(u) = 1$ if $u > 0$.

Grassberger and Procaccia (1983) showed that for $r \to 0$, there exists a power-law behavior between $C(r)$ and r, that is,

$$C(r) \sim r^{\nu}. \quad (4)$$

In other words, correlation dimension ν of the attractor is given by the slope of $\log C(r)$ versus $\log r$ for small r. They argue that correlation dimension makes an unambiguous distinction between deterministic chaos and random noise by its characteristic dependence on embedding dimension n. The slope attains a saturation limit beyond some relatively small n for a system with deterministic chaos, whereas it grows indefinitely with increasing n for random noise.

3. Review of weather and climate attractors

Now let us look at the reported values of correlation dimension for some geophysical time series (see Table 1). Nicolis and Nicolis (1984) analyzed 500 values of oxygen isotope data and identified the existence of climatic attractor with low dimensionality. Contrary to this claim, Grassberger (1986) found no evidence of a finite-dimensional attractor using the same dataset and he argued that the apparent low-dimensional attractors of Nicolis and Nicolis (1984) resulted from using too few, too highly smoothed data.

Fraedrich (1986, 1987) looked at several variables and also reported the presence of weather and climatic attractors. A 182-point climatic time series showed a saturation dimension of 4.4, which, in our judgment, is a drastically inadequate number of points to reconstruct the attractor even if it exists. For the weather attractor, no saturation was found until, perhaps somewhat artificially, winter and summer data were separated.

Essex et al. (1987) and Keppenne and Nicolis (1989) analyzed daily 500-mb geopotential heights at several meteorological stations across the globe and reported a very high value of correlation dimension (>6–8.4). Tsonis and Elsner (1988) investigated the presence of a weather attractor over very short time scales. They analyzed 10-s averages of the vertical wind velocity over an 11-h period (total number of points, $N = 3960$) and reported a correlation dimension of 7.3. Questions may be raised about the reliability of the estimate at such a high value of saturation with such a few points.

Rodriguez-Iturbe et al. (1989) and Sharifi et al. (1990) analyzed rainfall intensity data for several storms from Boston and reported correlation dimension values of less than 4. This is surprisingly low given

TABLE 1. Correlation dimension estimates for various geophysical time series.

Single variable time series	Correlation dimension	Reference
Oxygen isotope records of the deep-sea cores spanning the past one million years	3.1	Nicolis and Nicolis (1984)
Same record, different smoothing	No evidence of finite dimensionality	Grassberger (1986)
Daily local geopotential values at 500 mb taken at 1200 UTC extending over a span of 40 years	~6.0	Essex et al. (1987)
15-yr daily surface pressure	No saturation	Fraedrich (1986)
14 winter daily surface pressure	~3.2	
15 summer daily surface pressure	~3.9	
30-yr relative sunshine duration	No saturation	
29 winter relative sunshine duration	~3.1	
30 summer relative sunshine duration	~4.3	
10-yr zonal wave amplitude	No saturation	
10 winter zonal wave amplitude	~3.1	
10 summer zonal wave amplitude	~3.6	
14 winter daily surface pressure	≥6.8–7.1	Fraedrich (1987)
Lagrangian trajectories of three satellite-tracked buoys	1.4	Osborne et al. (1986)
Relative Lagrangian motion between pairs of drifters	1.3	Sanderson and Goulding (1990)
10-s averages of the vertical wind velocity recorded 10 m above the ground over an 11-h period	7.3	Tsonis and Elsner (1988)
15-s rainfall intensity data for a storm from Boston	3.8	Rodriguez-Iturbe et al. (1989)
Time between raingage signals each corresponding to an accumulation of 0.01 mm of rainfall; analyzed for three storms.	<4	Sharifi et al. (1990)
Daily 500-mb geopotential heights at nine different meteorological stations from western Europe for 24 yr—1 January 1961	6.8–8.4	Keppenne and Nicolis (1989)

the nature of the atmospheric system and the results of Tsonis and Elsner (1988) at similar time scales.

Ruelle (1990) showed that one should be wary of dimension estimates that are not well below 2 $\log_{10}N$, where N is the number of data points. Recently, this bound has been questioned by Essex and Nerenberg (1991); however, the results of Ruelle (1990) emphasize the importance of the length of time series for the adequate estimation of correlation dimension. The issue of data adequacy for practical estimation of dimension from time series has been discussed by various other authors (Takens 1985; Grassberger 1986; Essex et al. 1987; Smith 1988; among others).

Theiler (1990) pointed out two sources of error in the estimation of correlation dimension: statistical imprecision and systematic bias. The first type of error essentially results from the finite data and is reasonably tractable. The second type of error comes from a wide variety of sources. Theiler (1990) looked at a few sources such as the effect of noise, discretization, and autocorrelation in the time series. Most of these errors can be remedied or at least compensated for.

Lorenz (1992) argues that the procedure of Grassberger and Procaccia (1983) systematically underestimates the dimension if the variable chosen for analysis is only weakly coupled to much of the underlying dynamical system and if the dataset is only moderately large.

There may be another possible explanation of the low correlation dimension obtained by some authors. Following a thought similar to that of Lorenz (1992), atmospheric variables that are bounded or result from threshold-type phenomena are found to yield correlation dimensions (for moderately large datasets) that underestimate the true dimensions of the system attractor. At this point, it is important to note that in Table 1, variables like pressure and vertical velocity yield high correlation dimension, while derived variables like sunshine duration and rainfall result in low-dimension estimates.

4. The impact of thresholds on correlation dimension estimates

Thresholds and other constraints of thermodynamic or physical nature play significant roles in controlling the dynamics of the atmosphere. For example, the dynamical equation for water vapor mixing ratio q_v is not constrained, and it can essentially take any value. Physically, however, there is a definite functional dependence between the temperature and the saturation water vapor mixing ratio q_{vs}, known as the Clausius–Clapeyron relationship. Hence, the evolution of water vapor must be constrained [i.e., $q_{vs}(T) \geqslant q_v(T) \geqslant 0$] by this relationship although this is *not* a dynamical equation. Furthermore, there are several threshold-dependent behavior laws: for example, conversion of cloud water to rainwater. It is proposed that if the single variable time series chosen for analysis depends on physical constraints and thresholds, then its correlation dimension will be significantly less than that of the underlying dynamical system, explaining some of the inconsistent results of Table 1.

The point is argued by comparing the behavior of rainfall to that of the related vertical wind velocity. Rainfall R at the ground surface is given by $(\rho_a q_r \omega^*)$, where ρ_a is the air density, q_r is the rainwater mixing ratio, and ω^* is the effective fall velocity of the rain, whereas ω is the actual vertical velocity. In theory, if we can construct the phase space for ω (or any other variables) correctly, then the estimate of correlation dimension should correspond to the correlation dimension of the underlying dynamical system. The correlation dimension for ρ_a, ω, q_r, and R (and for that matter all other variables that describe the atmospheric processes) should be exactly the same. Rainfall is a special kind of variable, however, that depends on different threshold values of associated dynamical variables. Zero rainfall may occur due to zero q_r or $\omega < \omega^*$. The threshold in the vertical velocity appears because the upward component of the vertical velocity cannot produce rainfall at the surface and because drops with small vertical velocity (small drops) will evaporate before they can reach the surface. Rainfall is also affected by physical constraints like the Clausius–Clapeyron equation and the thresholds similar to the conversion of cloud water to rainwater.

In estimating the correlation dimension ν_R for R, a subset of $\omega > \omega^*$ is studied, whereas when the ν_ω is analyzed for ω, all possible values of ω are included. Therefore, ν_R ought to be significantly smaller than ν_ω. To prove this conjecture, rainfall intensity and the vertical velocity field generated from a three-dimensional cloud model have been analyzed (Clark 1977). The details of the simulation can be found in Islam et al. (1993).

An atmosphere in equilibrium is simulated by assuming that the surface heat fluxes are in balance with radiative cooling. This equilibrium condition is crucial for any analysis of strange attractors; it is important that the points used are actually on the attractor and are not on the transient part of the trajectories. Most published work ignores this condition when looking for deterministic chaos in the atmospheric phenomena. In our simulations, it took 50 h to reach the equilibrium and we have chosen the period between 60 and 80 h for the estimation of correlation dimension. The dataset has 7200 points at 10-s time steps.

Figure 1 shows the time series of rainfall intensity and vertical velocity at a point. From mere visualizations, it is clear that the details of the vertical velocity time series is richer in its complexities and fluctuations. In addition, since the vertical velocity also includes terminal velocity of raindrops, the rainfall intensity is zero whenever positive vertical velocity is given.

As previously stated, the choice of time delay to the construction of the phase-space time series is important. If unlimited infinite phase data were available,

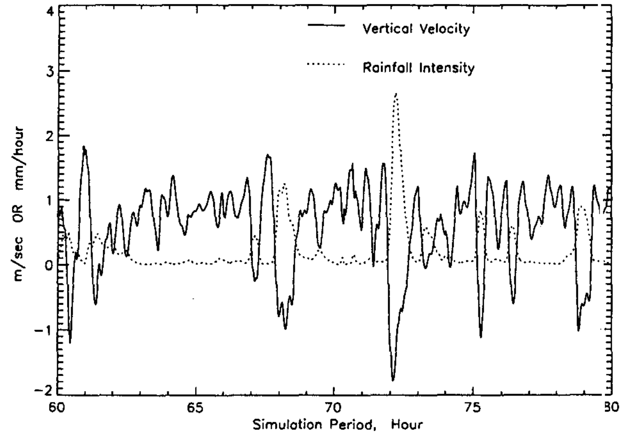
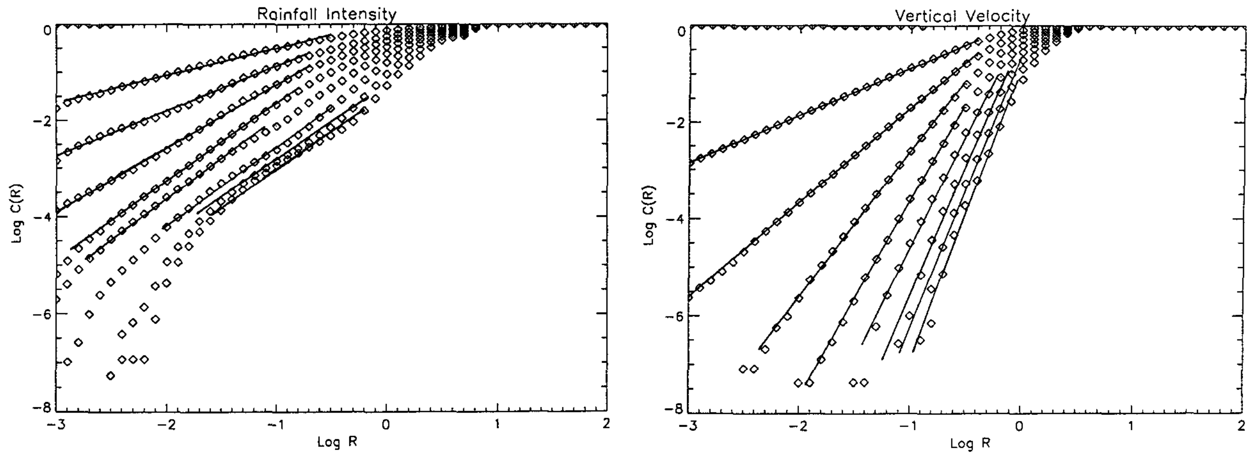


FIG. 1. The 10-s-average rainfall intensity and vertical velocity for a 20-h period in the quasi-equilibrium regime.

almost any delay time τ and embedding dimension $n > D$ (D being the fractal dimension) will reproduce the attractor. Choosing adequate values of n and τ for limited and possibly noisy experimental data, however, is not straightforward. For example, if the product $(n - 1)\tau$ is too large, then the components $X(t)$ and $X[t + (n - 1)\tau]$ of the reconstructed vector $\mathbf{X}$ will be effectively decorrelated, which will, in turn, inflate the estimated dimension (Theiler 1990). On the other hand, if $(n - 1)\tau$ is too small, the components $X(t) \ldots X[t + (n - 1)\tau]$ will be highly correlated and the reconstructed attractor will resemble a diagonal line. Generally, τ is chosen as a characteristic time scale such that two components of the vector $\mathbf{X}$ are not linearly dependent and $(n - 1)\tau$ is not too much greater than the chosen time scale. A popular choice for this characteristic time is based on the autocorrelation function of the original time series. The time delay τ is chosen such that autocorrelation drops to $1/e$. This is called the *decorrelation time*. Recently Fraser and Swinney (1986) proposed a different choice of characteristic time based on mutual information. The decorrelation time was used to define τ. For the simulated rainfall time series, this was 17.5 min. For vertical velocity, it was 12.0 min.

A correlation dimension analysis gives very low values for the rainfall-intensity field (Fig. 2a), whereas for the vertical velocity field the exponent, ν_ω (Fig. 2b), keeps increasing with the embedding dimension implying infinite correlation dimension. Figure 2c clearly shows saturation at a low correlation dimension for the rainfall-intensity data, whereas for the vertical velocity no saturation occurs with increasing embedding dimension. Other correlation dimension experiments (Islam 1991) not reported here, with much longer time series and independent embedding instead of time delay embedding, yield similar results. It seems that even if we can construct an attractor for rainfall and estimate its dimension ν_R, it does not tell us anything about the



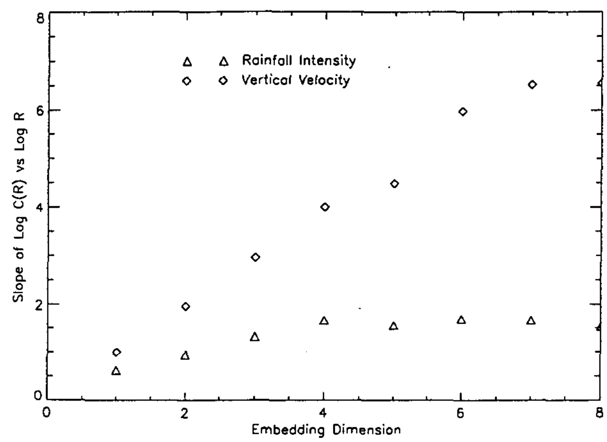


FIG. 2. (a) Plot of $\log_{10}C(R)$ against $\log_{10}R$ for the rainfall intensity $N = 7200$. Different curves represent different embedding dimension m, $m = 1, 2, \cdots, 8$. (b) Similar to (a) but for the vertical velocity. (c) The scaling exponent ν as a function of embedding dimension m for (a) and (b).

active degrees of freedom of the underlying dynamical system.

5. Concluding remarks

Most of the correlation dimension estimates for various geophysical time series seem surprisingly low. We have shown how the presence of thresholds and physical constraints on variables may be an explanation (along with others) for the seemingly low estimate of correlation dimension; the implication being that we cannot just choose any time series and estimate its correlation dimension to conclude that the underlying dynamical system is describable by ν variables. It does not mean that the low estimates of correlation dimension are meaningless. They do, however, need to be reinterpreted. A good and low estimate may simply mean that a low number of variables may capture certain dynamical aspects of the analyzed time series rather than the entire underlying dynamical system.

At this point, we can only speculate that thresholds may violate the conditions that make time embedding a good procedure to generate a phase space from a single time series.

Acknowledgments. We would like to thank Professor Edward Lorenz for many fruitful discussions concerning this work. This work is funded by the National Science Foundation under Grant ATM-9020832.

REFERENCES

Clark, T., 1977: A small-scale dynamic model using a terrain-following coordinate transformation. *J. Comput. Phys.*, **24**, 186–215.

Essex, C., and M. A. H. Nerenberg, 1991: Comments on "Deterministic chaos: The science and the fiction" by D. Ruelle. *Proc. Roy. Soc., Lond. A*, **435**, 287–292.

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

Fraedrich, K., 1986: Estimating the dimensions of weather and climate attractors. *J. Atmos. Sci.*, **43**, 331–344.

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

Fraser, A. M., and H. L. Swinney, 1986: Independent coordinates for strange attractors from mutual information. *Phys. Rev.*, **A33**, 1134.

Grassberger, P., 1986: Do climatic attractors exist? *Nature*, **323**, 609–42.

——, and I. Procaccia, 1983: Measuring the strangeness attractors. *Physica*, **D9**, 189–208.

Islam, S., 1991: Predictability of mesoscale precipitation. Sc.D. Thesis, Department of Civil Engineering, Massachusetts Institute of Technology, 240 pp.

——, R. L. Bras, and K. Emanuel, 1993: Predictability of mesoscale rainfall in the tropics. *J. Appl. Meteor.*, **32,** 297–310.

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

Lorenz, E. N., 1992: Dimension of weather and climate attractors. *Nature,* **353,** 241–244.

Nicolis, C., and G. Nicolis, 1984: Is there a climatic attractor? *Nature,* **311,** 529–532.

Osborne, A. R., A. D. Kirwan, 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.

Rodriguez-Iturbe, I., B. F. dePower, M. B. Sharifi, and K. P. Georgakakos, 1989: Chaos in rainfall. *Wat. Resources Res.*, **25,** 1667–1675.

Ruelle, D., 1981: Sensitive dependence on initial conditions and turbulent behavior. *Bifurcation Theory and Its Applications in Scientific Disciplines,* American Academic of Sciences, 136–229.

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

Sanderson, B. G., and A. Goulding, 1990: The fractal dimension of relative Lagrangian motion. *Tellus,* **42A,** 550–556.

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. *Phys. Lett.* A, **133,** 283–288.

Takens, F., 1985: On the numerical determination of the dimension of an attractor. *Lecture Notes in Mathematics.* Springer-Verlag,

Theiler, J., 1990: Estimating the fractal dimension of chaotic time series. *J. Opt. Soc. Amer.*, **A7,** 1055–1073.

Tsonis, A. A., and J. B. Elsner, 1988: The weather attractor over very short timescales. *Nature,* **333,** 545–547.