# Deterministic Chaos in Tropical Atmospheric Dynamics

H. WAELBROECK

*Instituto de Ciencias Nucleares, UNAM, Mexico City, Mexico*

(Manuscript received 12 April 1994, in final form 5 January 1995)

ABSTRACT

An 11-year dataset from the tropical weather station of Tlaxcala, Mexico, is examined. It is found that mutual information drops quickly with the delay, to a positive value that relaxes to zero with a timescale of 20 days. The mutual dependence of the observables is also examined and it is concluded that the dataset gives the equivalent of eight variables per day, known to a precision of 2%. It is determined that the effective dimension of the attractor is $D_{\text{eff}} \approx 11.7$ at the scale $3.5\% < R/R_{\max} < 8\%$. Evidence is found that the effective dimension increases as $R/R_{\max} \to 0$, supporting a conjecture by Lorenz that the climate system may consist of a large number of weakly coupled subsystems, some of which have low-dimensional attractors. A local reconstruction of the dynamics in phase space is performed; the short-term predictability is modest and agrees with theoretical estimates. Useful skill in predictions of 10-day rainfall accumulation anomalies reflects the persistence of weather patterns, which follow the 20-day decay rate of the mutual information.

## 1. Introduction

In temperate regions, the large-scale approach to meteorology based on the reduction of fluid equations to a finite grid has been effective in meeting the demands of agricultural planning and disaster prevention. This is not the case in tropical regions, where the atmospheric system is strongly dependent on local "details," such as the presence of a hill or a lake, which are too subtle to be accurately modeled on a grid system. Often it is not even known which details are important and which are not. One is tempted to turn to the empirical approach: to reconstruct the dynamical model not from physical equations, but from data drawn from observations of the system. This approach has been applied successfully to scalar time series analysis (Linsay 1991; Abarbanel et al. 1990; Farmer and Sidorowich 1987), leading to useful predictions for chaotic systems for which the underlying equations are not known a priori. Before applying this method to meteorology, one must gain a better understanding of tropical atmospheric dynamics from the point of view of chaos theory. That is the purpose of this article.

The study of chaos in extended systems, such as atmospheric dynamics, is attracting a growing interest as the next challenge in dynamical systems analysis. A key question that must be addressed is whether such systems can have low-dimensional attractors, so that one could hope to reconstruct the dynamics in phase space from observations of the system at regular intervals of time.

Most attempts at calculating an attractor dimension from weather data have encountered the difficulty that the datasets are too scarce to trust that the dimension found is really the correlation dimension of the attractor. Rather, the scaling graphs generally provide information only on the large-scale structure of the attractor, and there is no compelling reason to expect that this structure should be similar at finer scales. Instead, Lorenz has conjectured that the global climate may be well represented by a large number of weakly coupled subsystems, some of which may have low-dimensional attractors (Lorenz 1991). If the coupling is sufficiently weak, the external perturbations would not be detected in the large-scale reconstruction of the attractor, but would imply a much larger attractor dimension at smaller scales, reflecting the greater complexity of the coupled dynamical system. Lorenz based his conclusion on a study of a model system, which consisted of a number of identical copies of the set of three modes of the Oberbeck-Boussinesq equations for the convection of a fluid heated from below (Lorenz 1963), with coupling terms connecting the different copies.

In this article, we will consider Lorenz' conjecture in the practical setting of a tropical weather dataset, and attempt to decide to what extent a reconstruction of the large-scale structure of the attractor can be helpful for the purpose of designing prediction models for extended chaotic systems.

The dataset consists of 19 weather variables measured daily at a single weather station, located near the town of Tlaxcala, Mexico. The measurements were performed without interruption or change of equipment

*Corresponding author address:* Dr. Henri Waelbroeck, Instituto de Ciencias Nucleares, UNAM, Circuito Exterior, C.U., A. Postal 70-543, Mexico D. F. 04510, Mexico.

© 1995 American Meteorological Society

Unauthenticated | Downloaded 10/19/21 11:55 AM UTC

over a span of 4015 days (11 yr). Holes and flagrant errors in the dataset amounted to less than 0.5% of the data and were replaced by linear interpolation. The decision to use 19 variables at a single location, rather than fewer variables at several locations, was based on evidence that the data had been compiled most rigorously at the main weather station near Tlaxcala (hereafter referred to as Tlax). The 19 variables are not all independent; we shall discuss this in more detail in the next section where we show that the dataset carries as much information as eight mutually independent variables, or a single observable over a time span eight times as long.

The 19 variables are listed in Table 1. The range of each variable (= max − min) was chosen to equal $2^{b(i)}$ or a decimal fraction thereof, where $b(i)$ is the number of significant bits according to our estimate of the measurement accuracy. It was then normalized to the unit interval: $x_k^{(i)} \in [0, 1]$, $i = 1, \cdots, N$, $k = 1, \cdots, n$ ($N = 19$, $n = 4015$) and the annual cycle was subtracted. With this normalization, the greatest possible distance between two points is equal to one: we will denote the normalized distance by $R/R_{\max}$ to recall that it is less than or equal to one (see the precise definition of the distance function in section 3). The dataset is not available on any public-domain databank: please contact the author directly for electronic transfers.

In section 2, we examine the information content of the dataset with an adaptation of the mutual information method for short datasets. In section 3, we show that the data reflects a deterministic chaotic system down to the scale $R/R_{\max} = 3.5\%$, and determine the effective dimension of the attractor in the range $3.5\% < R/R_{\max} < 8\%$. A strong increase in the slope just below this range, but above the scale where the shortcomings of the dataset become noticeable, provides evidence in support of Lorenz' conjecture on the nature of the climate system. The theoretical predictability limit of the dataset is also estimated by two different methods. In section 4, we describe the results of a local reconstruction of the dynamics and conclude, based on those results and the previous data analysis, on how the large-scale structure of the attractor can be used in prediction tasks.

## 2. Mutual information analysis

The mutual information between two random variables $x^{(i)}, x^{(j)} (j \neq i)$, with statistics specified by a finite set of events with a determined measurement error, is defined as follows. One divides the range of the variables into a fixed number of intervals ("bins"), such that the size of each bin is greater than the error and each allowed bin is visited a statistically significant number of times ($\geqslant 40$). For two weather variables with a delay of $\tau$ days, the dataset gives us $n - \tau$ events $(x_t^{(i)}, x_{t+\tau}^{(j)}; t = 1, \cdots, n - \tau)$; a division into 64 bins was found to satisfy both criteria for all variables but numbers 9, 10, and 12 (cloud cover, average, and dominant wind directions).

The probability distribution of a variable is given by the fraction of the events that fall into each bin: the "probability of the bin $O_\alpha$" is

$$P(x^{(i)} \in O_\alpha) = \frac{\#\{x_t^{(i)}(t = 1, \cdots, n) : x_t^{(i)} \in O_\alpha\}}{n}, \tag{2.1}$$

TABLE 1. The 19 variables observed each day are given, as well as the number of relevant bits indicating the measurement accuracy, and the information calculated by partitioning the range of each variable into 64 subintervals. The maximum and minimum value of each variable prior to normalization is also given. Atmospheric pressure is corrected for elevation.

| | Variable $i$ | $b(i)$ | Max | Min | $n$ | $I(x)$ |
|---|---|---|---|---|---|---|
| 1 | max. vapor pressure | 8 | 51.2 hPa | 0 | 6 bits | 5.0 |
| 2 | min. vapor pressure | 8 | 51.2 hPa | 0 | 6 bits | 4.7 |
| 3 | max. dewpoint | 8 | 51.2°C | 0 | 6 bits | 4.9 |
| 4 | min. dewpoint | 8 | 25.6°C | −25.6°C | 6 bits | 5.2 |
| 5 | max. relative humidity | 7 | 100% | 0 | 6 bits | 2.6 |
| 6 | min. relative humidity | 7 | 100% | 0 | 6 bits | 4.8 |
| 7 | max. atmos. pressure | 8 | 790.6 mm | 765 mm | 6 bits | 4.0 |
| 8 | min. atmos. pressure | 8 | 780.6 mm | 755 mm | 6 bits | 4.1 |
| 9 | cloud cover (octants) | 4 | 8 | 0 | 4 bits | 3.1 |
| | wind | | | | | |
| 10 | dominant direction | 5 | 15 | 0 | 5 bits | 3.5 |
| 11 | average speed | 7 | 128 km $h^{-1}$ | 0 km $^{-1}$ | 6 bits | 4.1 |
| 12 | direction at max. | 5 | 15 | 0 | 5 bits | 3.6 |
| 13 | max. speed | 8 | 256 km $h^{-1}$ | 0 km $h^{-1}$ | 6 bits | 4.2 |
| 14 | precipitation | 11 | 204.8 mm | 0 mm | 6 bits | 1.6 |
| 15 | evaporation | 10 | 102.4 mm | 0 mm | 6 bits | 5.1 |
| 16 | max. temperature | 8 | 51.2°C | 0 | 6 bits | 4.9 |
| 17 | min. temperature | 8 | 25.6°C | −25.6°C | 6 bits | 5.2 |
| 18 | dry bulb temperature | 8 | 25.6°C | −25.6°C | 6 bits | 5.4 |
| 19 | hours of insolation | 11 | 20.48 h | 0 | 6 bits | 4.6 |

Unauthenticated | Downloaded 10/19/21 11:55 AM UTC

where the index $\alpha$ labels the bins and # is the cardinality, or number of elements in the set.

The average information needed to specify the variable $x^{(i)}$, knowing its probability distribution, is given by

$$I(x^{(i)}) = -\langle \log_2 P(x^{(i)}) \rangle, \tag{2.2}$$

where the average is a sum over $\alpha$ weighted by the probability $P(x^{(i)} \in O_\alpha)$. One easily checks that the information is maximal when $P(x^{(i)})$ is homogeneous and $I(x^{(i)})$ is equal to the number of bits needed to count the bins: for 64 bins, the maximum information is equal to 6 bits (Table 1). In general, the more $P[x^{(i)}]$ is peaked around a particular value of $x^{(i)}$, the less information one needs, on average, to specify the value of $x^{(i)}$: $I[x^{(i)}]$ measures how surprised one feels when, knowing the distribution $P[x^{(i)}]$, one is told a typical value of the variable (Fraser and Swinney 1983; Fraser 1988).

The two-point probability $P[x^{(i)} \in O_\alpha, x^{(j)} \in O_\beta]$ is given by the fraction of events where $x^{(i)}$ falls into the bin $O_\alpha$ and $x^{(j)}$ falls into the bin $O_\beta$. The information needed to specify two variables is given by

$$I[x^{(i)}, x^{(j)}] = -\langle \log_2 P[x^{(i)}, x^{(j)}] \rangle, \tag{2.3}$$

where the average is now a sum over $\alpha$ and $\beta$ weighted by $P[x^{(i)}, x^{(j)}]$.

The information needed to specify two correlated variables jointly is less than that which is needed to specify each of the variables independently; the difference measures the degree of mutual dependence of the two variables and is called "mutual information." With the notation $I^{(i)} \equiv I(x^{(i)})$, etc., the mutual information is given by

$$I_m^{(i,j)} = I^{(i)} + I^{(j)} - I^{(i,j)} \geq 0. \tag{2.4}$$

If one is concerned with the mutual information for two measurements of the same variable with a delay $\tau$, the relevant events are $\{x_t^{(i)}, x_{t+\tau}^{(i)}\}$. Since $I[x_0^{(i)}] = I[x_\tau^{(i)}] = I^{(i)}$, the mutual information is given by

$$I_m^{(i)}(\tau) = 2I^{(i)} - I^{[i(0),i(\tau)]}. \tag{2.5}$$

Likewise, the mutual information between the variable $x_t^{(i)}$ and the delayed variable $x_{t+\tau}^{(j)}$ is given by

$$I_m^{(i,j)}(\tau) = I^{(i)} + I^{(j)} - I^{[i(0),j(\tau)]}. \tag{2.6}$$

For a chaotic system, $I_m^{(i,j)}(\tau) \to 0$ and $I_m^{(i,j)}(\tau) \to 0$ as $\tau \to \infty$, $\forall i, j$.

The mutual information is considered to be the best measure of mutual dependence of variables, for nonlinear systems (Fraser 1989).

Unfortunately, the implementation of this method for short datasets encounters the difficulty that the two-point distribution $P[x^{(i)}, x^{(j)}]$ is not accurately defined, since the number of events which fall into each allowed bin is not statistically significant (insufficient sampling). The probability distribution comes out rough, which implies that probability is "localized" at statistically insignificant peaks in $P[x^{(i)}, x^{(j)}]$. Therefore, the two-point information is underestimated and the mutual information is overestimated. This problem can be corrected by smoothing the distribution $P[x^{(i)}, x^{(j)}]$. We used the discretized diffusion equation

$$P_{\alpha\beta}^{k+1} = (1 - \epsilon) P_{\alpha\beta}^{k} + \epsilon \left( \frac{P_{\alpha\beta+1}^{k} + P_{\alpha\beta-1}^{k} + P_{\alpha+1\beta}^{k} + P_{\alpha-1\beta}^{k}}{4} \right), \tag{2.7}$$

which we iterate four times ($k = 1, \cdots, 4$), so that the final value of $P_{\alpha\beta}$ is a weighted average of $P$ over a square of $8 \times 8 = 64$ neighboring bins, with the strongest weights assigned to the bins closest to $(\alpha\beta)$. The probability distribution gradient is set to zero on the boundary, where $\alpha$ or $\beta = 0$ or 65. The smoothing parameter $\epsilon$ is chosen independently for each pair $(i, j)$ so that $I_m^{(i,j)}(\tau) \to 0$ as $\tau \to \infty$, and for each $i$, so that $I_m^{(i)}(\tau) \to 0$. To implement the $\tau \to \infty$ limit, we required that the average mutual information over the range $\tau = 150, \cdots, 250$ be equal to zero; the rms of $I_m$ over this set of 100 trials gives one a measure of the degree of precision of the mutual information estimates: an upper bound $\epsilon < 0.7$ was set on the smoothing parameter, and the cases where the rms was greater than 0.05 bits were considered not trustworthy due to the short dataset.

The mutual information among the 19 variables is given in Table 2. One notes that in most cases mutual informations are low, between 0 and 0.2 bits, with a few notable exceptions: for example, the minimum dewpoint is strongly dependent on the minimum registered value of the vapor pressure, and likewise for the corresponding maximum values. This is not surprising, given that these variables are closely related and reach their extrema more or less simultaneously. Also the dry bulb temperature is dependent on the minimum temperature, cloud cover, minimum vapor pressure, and minimum dewpoint. The total information at one time step is given by the sum of the informations of the 19 variables (diagonal terms in Table 2), minus the sum of the mutual informations $I_m^{(i,j)}$ for $i < j$, or $70.4 - 23.6 = 46.8$ bits, roughly equivalent to eight variables of 6 bits each. In this counting, one is neglecting the $n - point$ mutual information, $n \geq 3$.

The mutual information of each variable as a function of the delay is given in Table 3. The mutual information drops to about 10% after a delay of one day: this is sufficiently small that one can take observations for consecutive days as distinct coordinates and obtain a faithful representation of the phase space (i.e., one such that coordinates $x_i, x_j$ of data points do not cluster near the hyperplane $x_i = x_j$). The total amount of independent data available for the reconstruction is then $8 \times 4015 = 32\,120$.

Some variables have a strong persistence, which is revealed by a relatively large mutual information at de-

Unauthenticated | Downloaded 10/19/21 11:55 AM UTC

TABLE 2. The mutual information of each pair of observables at the same time step is given; again the interval (0,1) is divided into 64 subintervals, so that the maximum possible information is equal to 6 bits. Note that the diagonal terms represent the information of each variable individually. The sum of all the terms on the diagonal gives the information needed to specify the 19 variables independently, while the sum of all terms below the diagonal gives the total amount of mutual information among them. The difference between these two quantities is equal to the information needed to specify the 19 variables together.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 5.0 | | | | | | | | | | | | | | | | | | |
| 2 | 0.2 | 4.7 | | | | | | | | | | | | | | | | | |
| 3 | 2.0 | 0.2 | 4.9 | | | | | | | | | | | | | | | | |
| 4 | 0.3 | 1.0 | 0.2 | 5.2 | | | | | | | | | | | | | | | |
| 5 | 0.1 | 0.1 | 0.1 | 0.1 | 2.6 | | | | | | | | | | | | | | |
| 6 | 0.1 | 0.2 | 0.1 | 0.2 | 0.2 | 4.8 | | | | | | | | | | | | | |
| 7 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 4.0 | | | | | | | | | | | | |
| 8 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 | 4.1 | | | | | | | | | | | |
| 9 | 0.3 | 0.4 | 0.2 | 0.6 | 0.0 | 0.5 | 0.0 | 0.3 | 3.1 | | | | | | | | | | |
| 10 | 0.1 | 0.2 | 0.1 | 0.0 | 0.0 | 0.0 | 0.2 | 0.1 | ? | 3.5 | | | | | | | | | |
| 11 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 4.1 | | | | | | | | |
| 12 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.2 | 0.0 | 0.3 | ? | 0.2 | 3.6 | | | | | | | |
| 13 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.3 | ? | 4.2 | | | | | | |
| 14 | 0.1 | 0.2 | 0.1 | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 1.6 | | | | | |
| 15 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.2 | 0.0 | 0.0 | 0.8 | 0.2 | 0.0 | ? | 0.1 | 0.1 | 5.1 | | | | |
| 16 | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 | 0.2 | 0.0 | 0.1 | 0.0 | ? | 0.0 | ? | 0.0 | 0.0 | 0.3 | 4.9 | | | |
| 17 | 0.2 | 1.2 | 0.2 | 1.3 | 0.0 | 0.1 | 0.0 | 0.0 | 0.8 | 0.3 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 5.2 | | |
| 18 | 0.1 | 0.9 | 0.1 | 1.0 | 0.1 | 0.2 | 0.0 | 0.0 | 0.5 | ? | 0.0 | ? | 0.0 | 0.1 | 0.0 | 0.0 | 1.1 | 5.4 | |
| 19 | 0.0 | 0.2 | 0.0 | 0.2 | 0.0 | 0.1 | 0.0 | 0.0 | 0.5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.2 | 0.1 | 0.2 | 0.2 | 4.6 |

lays $\tau \sim 15$, particularly the dewpoint and the vapor pressure, and the dry bulb temperature, and the minimum temperature: this indicates a moderate potential for the application of statistical prediction models such as the linear autorregressive models or the low-threshold linear reconstruction model described in sections 3 and 4. The slow decline of mutual information with the delay appears to indicate a strong persistence of some features of weather patterns, which would be well exploited in such a model. The rainfall shows very little persistence but has significant mutual dependence with other more persistent variables. This indicates some potential for medium-range predictions that would exploit the predictability of these other variables to give trend predictions for precipitation anomalies (section 4). This type of predictability through the persistence of another variable could be quantified as mutual information by computing the three-point mutual informations.

The total mutual information as a function of the delay gives one a measure of how much additional information one gets by adding one day to a set of $T$ consecutive days of data. A complete description of the state of the weather system should be such that no new

TABLE 3. The mutual information of each variable is given as a function of the delay, as well as the total over all variables. Some variables have a relatively large persistence even for delays of 14 days, while for other variables the mutual information drops very quickly with the delay. The total mutual information at a delay $t = 1$ day is less than 10 total information in one day of data, so one can consider that consecutive days provide reasonably independent coordinates for a phase space reconstruction.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 11 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 5.0 | 4.7 | 4.9 | 5.2 | 2.6 | 4.8 | 4.0 | 4.1 | 4.1 | 4.2 | 1.6 | 5.1 | 4.9 | 5.2 | 5.4 | 4.6 | 70.4 |
| 1 | .76 | .56 | .74 | .75 | .22 | .60 | .60 | .55 | .10 | .15 | .10 | .17 | .62 | .59 | .42 | .28 | 6.61 |
| 2 | .58 | .36 | .57 | .51 | .10 | .39 | .19 | .19 | .06 | .09 | .06 | .13 | .39 | .43 | .37 | .11 | 4.53 |
| 3 | .51 | .29 | .51 | .41 | .08 | .33 | .10 | .12 | .03 | .08 | .04 | .10 | .29 | .39 | .35 | .08 | 3.71 |
| 4 | .47 | .23 | .47 | .39 | .06 | .27 | .08 | .11 | .03 | .08 | .04 | .08 | .23 | .35 | .31 | .05 | 3.25 |
| 5 | .44 | .23 | .44 | .34 | .04 | .26 | .07 | .10 | .03 | .06 | .04 | .08 | .20 | .28 | .27 | .04 | 2.92 |
| 6 | .42 | .19 | .42 | .35 | .03 | .25 | .05 | .08 | .03 | .06 | .04 | .08 | .18 | .29 | .28 | .03 | 2.78 |
| 7 | .40 | .17 | .41 | .31 | .03 | .25 | .04 | .07 | .03 | .06 | .03 | .07 | .16 | .26 | .28 | .03 | 2.60 |
| 8 | .39 | .18 | .40 | .31 | .03 | .24 | .03 | .07 | .03 | .05 | .03 | .07 | .15 | .31 | .28 | .04 | 2.61 |
| 9 | .39 | .18 | .40 | .35 | .03 | .23 | .02 | .06 | .02 | .06 | .02 | .06 | .14 | .30 | .27 | .04 | 2.57 |
| 10 | .38 | .18 | .39 | .31 | .03 | .22 | .02 | .05 | .02 | .05 | .02 | .07 | .14 | .27 | .27 | .02 | 2.44 |
| 11 | .37 | .18 | .38 | .33 | .03 | .21 | .02 | .05 | .02 | .04 | .02 | .06 | .13 | .23 | .25 | .03 | 2.35 |
| 12 | .36 | .17 | .37 | .31 | .03 | .20 | .02 | .04 | .02 | .05 | .01 | .06 | .13 | .25 | .22 | .02 | 2.26 |
| 13 | .36 | .17 | .37 | .28 | .03 | .20 | .02 | .05 | .02 | .04 | .02 | .06 | .12 | .28 | .25 | .01 | 2.28 |
| 14 | .37 | .15 | .38 | .33 | .03 | .20 | .02 | .04 | .03 | .04 | .02 | .07 | .11 | .22 | .25 | .02 | 2.28 |

Unauthenticated | Downloaded 10/19/21 11:55 AM UTC

information is gained by adding more data. Recalling that a single day of data consists of 46.8 bits of information, the information added by introducing a second day is 46.8 − 6.6 = 40.2 bits; the third day adds 46.8 − 6.6 − 4.5 = 35.7 bits, etc.; the information added per day is represented in Fig. 1. The fact that no additional information is gained after 14 days of data indicates that the state of the system is well described if one gives the value of these variables for 14 consecutive days. This figure is an overestimate: if one could compute the $n$ − *point* mutual informations, the total mutual informations would be larger and the number of days beyond which no new net information is gained would be less than 14. We will see below that the embedding dimension, computed from the scaling graphs, corresponds to seven consecutive days of data, and that the local reconstruction of the dynamics gives optimal predictions for $T_e = 10 - 14$. The mutual consistency of these results is encouraging.

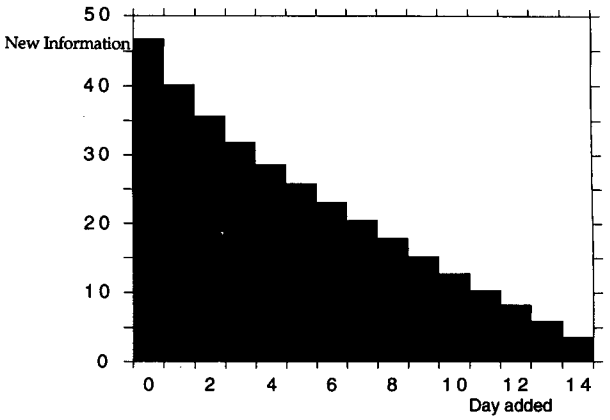


FIG. 1. The additional information needed to specify the variables at day $T + 1$, knowing these variables at days number 1, 2, . . . , $T$, is given as a function of $T$. No new information is gained by extending the phase space vector beyond 14 consecutive days of data. This indicates that the dynamics is deterministic with an embedding dimension corresponding to $T_e = 14$.

## 3. Phase space reconstruction

The phase space reconstruction is based on the following idea. One assumes that the system can be modeled by $d$ first-order differential equations; we will see below how one verifies the validity of this assumption and determines $d$. Any set of $d$ independent measurements of the state of the system provides a complete set of initial data, which determines the trajectory uniquely. In the phase space reconstruction method, one begins by choosing $d$ observables to provide a convenient representation of the state of the system in $\mathbb{R}^d$. These can be measurements of a single physical variable at regular intervals of time, as in the method of delays, or measurements of $d$ different variables when another variable equals a predetermined value (Poincaré section). Here we will take the 19 available variables for $T$ consecutive days of data ($d = 19 \times T$). Some of them are minimum or maximum values, others are averages over a 24-h period; all are well-defined functions of the continuous-time physical variables over non-overlapping intervals of time. Following the assumption that the dynamics is determined by first-order differential equations, these induce recursion relations for the evolution of the observables from one 24-h period to the next. The phase space reconstruction proposes 1) to examine the geometry of the reconstructed attractor (fractal dimension, etc.) and 2) to find an approximation of these recursion relations.

We define a point in phase space as a set of $19 \times T$ variables corresponding to $T$ consecutive days of observations. The dataset then gives one $4015 - T$ points with which we shall attempt to reconstruct the attractor in the embedding phase space: $\vec{x}_k = \{x_{k-t}^{(i)}: i = 1, \cdots, N; t = 0, \cdots, T-1\}$. We will use the arrowed vector notation for phase space points, and bold-face letters for a single day of observations: $\mathbf{x}_k \in \mathbb{R}^N$. The distance between two points in phase space is taken to be the normalized Manhattan distance

$$d(\vec{x}_k, \vec{x}_l) = \frac{1}{N} \sum_{i=1}^{N} \sum_{t=0}^{T-1} \alpha(t) |x_{k-t}^{(i)} - x_{l-t}^{(i)}|, \quad (3.1)$$

where $\alpha(t)$ is a partition of the identity for $t = 0, 1, \cdots, T-1$, which in this section will be taken to be uniform: $\alpha(t) = 1/T, \forall t$. Since the range of each variable $x^{(i)}$ was normalized, one easily checks that $d(\vec{x}_k, \vec{x}_l) \leqslant 1, \forall k, l$.

To determine the correlation dimension, one computes the number of pairs of points with distance $d(\vec{x}_k, \vec{x}_l) < \rho$: $N(\rho)$ (Grassberger and Procaccia 1983). If $T = T_e$ is chosen so that $\mathbb{R}^{N \times T_e}$ embeds the attractor of dimension $D_a$, then $N(\rho) \sim \rho^{D_a}$. If $T$ is too small for a proper embedding, $T < T_e$, then $N(\rho) \sim \rho^{D(T)}$, where $D(T)$ is the dimension of the attractor projected onto the first $T \times N$ coordinates. The dimensions $D(T)$ are determined from the slopes of logarithmic plots of $N(\rho)$, by linear regression (Fig. 2). If there is a deterministic dynamical rule, the graph of $D(T)$ reaches an upper bound $D = D_a$ when $T = T_e$. There are several obstacles that can inhibit a linear "scaling region" in the graph of $\log(N)$ versus $\log(\rho)$.

### a. Noise

A stochastic signal superimposed on the deterministic signal can disperse the points along directions orthogonal to the $D_a$-dimensional attractor; this is evidenced by a steeper slope of the logarithmic graph. In the case of our dataset, measurement error can be estimated to be less than 2% of the range of each variable; if we assume that the signed error is equally distributed on both sides of the origin, the measurement error in the Manhattan distance between points in $\mathbb{R}^{N \times T}$ is of the order of (Feller 1968)

Unauthenticated | Downloaded 10/19/21 11:55 AM UTC

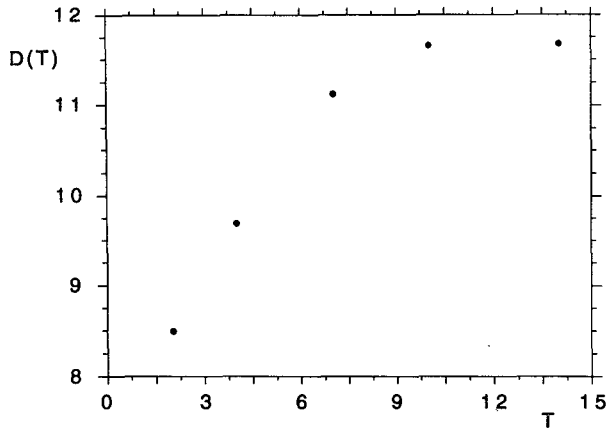


FIG. 2. The dimension $D(T)$ of the attractor's projection on a $19T$-dimensional subspace of the embedding space is given as a function of $T$. A plateau is clearly reached at $T = 7$, indicating an effective dimension of the attractor $D_a = 11.7$.

$$\epsilon_m \sim \frac{2}{(T \times N)^{1/2}} \%. \tag{3.2}$$

For systematic measurement errors the random-sign argument is not valid and one must consider the upper bound on the error, $\epsilon_m < 2\%$.

### b. Lack of data

For a short dataset, at small $\rho$ the total number of pairs of points with distance less than $\rho$ becomes too small to be considered statistically significant and one should require that

$$N(\rho) \geqslant 50. \tag{3.3}$$

In our case, the graphs of $N(\rho)$ reveal that this places a lower bound $\rho \approx 3\%-4\%$, depending on $T$, below which the points of the graphs scatter randomly due to the bad statistics. Since this lower bound is relatively large, one cannot safely assume that, for an infinitely long dataset, the scaling region would continue with the same slope all the way down to $\rho \to 0$. This problem is frequently encountered in climate and weather datasets (see Grassberger 1986; Lorenz 1991, and references therein) and implies that the correlation dimension, defined as the effective dimension in the limit $\rho \to 0$, cannot be computed. One should then consider the dimension estimates as effective dimensions at a certain scale. As Lorenz has suggested, it is possible that this effective dimension be a good approximation of the correlation dimension of a simpler local subsystem which is more or less decoupled from the global climate system: at lower values of $\rho$, the coupling with external variables would become relevant, leading to a much higher correlation dimension. We will describe evidence in support of Lorenz' conjecture below.

### c. Edge effect

For large values of $\rho$, comparable to the size of the attractor, the number of pairs saturates when a sphere of radius $\rho$ centered at a typical point on the attractor extends beyond the edge of the attractor, in an empty region of phase space: the maximum possible number of pairs taken from $n$ data points is $n^2$. The scaling region therefore ends below this saturation point, and

$$N(\rho) \ll n^2. \tag{3.4}$$

To see exactly where saturation becomes significant, one can draw only the complete graph of $N(\rho)$ and observe at which point saturation occurs. From Figs. 2–4, one sees that saturation becomes significant for $N(\rho) \geqslant n^2/100$.

### d. Scale-dependent dimension

Some attractors have different apparent dimensions at different scales; in some cases even the small-scale limit of the effective dimension is not well defined and the correlation dimension does not exist, as, for example, with Zaslavskii's map, where $D_a$ alternates between $D_a \approx 1$ and $D_a \approx 1.6$ in the limit $\rho \to 0$ (Grassberger and Procaccia 1983). If the effective dimension varies with $\rho$ in the region between the lower bounds (subsections a and b) and the upper bound (subsection c), it may be impossible to find a physically significant scaling region.

In our case, a clear scaling region is found up to $T = 10$ (Fig. 3) and to a lesser extent at $T = 14$ (Fig. 4).

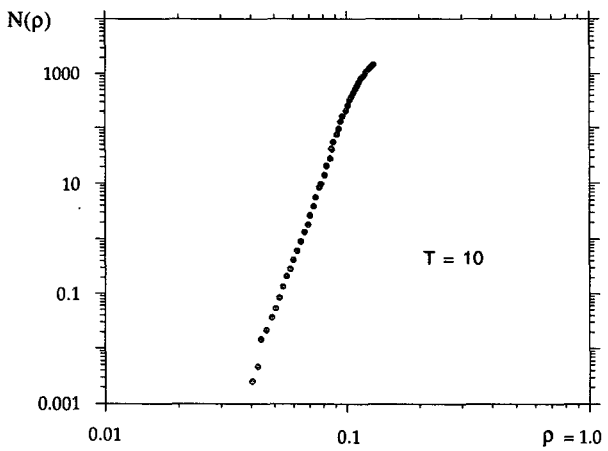


FIG. 3. The scaling graph gives the number of pairs of data points in phase space with distance less than $r$, as a function of $r$, if we use a slightly redundant embedding with 10 consecutive days of data. One distinguishes four regions at increasing values of $r$: at the lowest values there appears to be a tendency for the slope to increase, but this occurs at values of $N$ too low to trust the statistics. Next comes the scaling region, with a slope equal to $D_a = 11.67$. One then enters a nonlinear region where the slope increases at first, to about 14, before it begins to decrease as a consequence of the saturation, when $r$ becomes comparable to the size of the attractor.

Unauthenticated | Downloaded 10/19/21 11:55 AM UTC

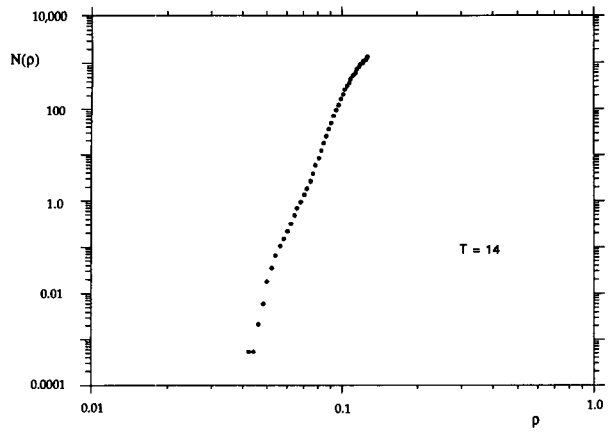


FIG. 4. The scaling graph for $T = 14$ consecutive days of data displays the same features as for $T = 10$, although the scaling region with a slope equal to $D_a = 11.68$ is narrower. The increase in the slope at low values of $r$ is more significant. This may reflect a higher effective dimension of the system coupled to external variables, as suggested by Lorenz' work; a time of 14 days would then appear to be sufficient to affect the local system up to the statistically significant scale $r = 0.055$.

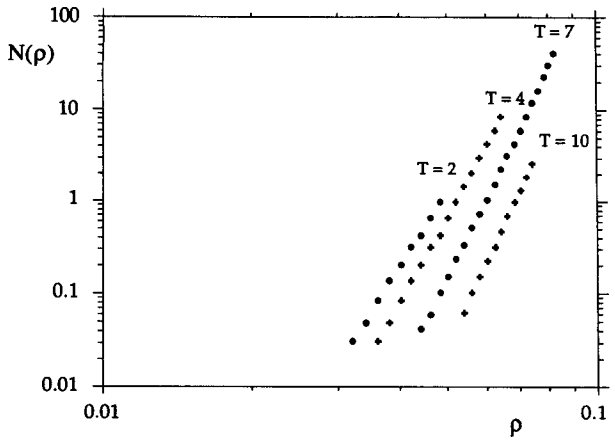


FIG. 5. The number of pairs within a ball of radius $r$, $N(r)$, scales like a power of $r$. The bilogarithmic plot gives a straight line that can be determined by linear regression. The scaling graphs are represented for four different values of the embedding dimension, corresponding to $T = 2$, 4, 7, and 10. The projected dimension $D(T)$ of the attractor on the chosen subspace of the true embedding space is given by the slope of the graph. The slope does not increase from $T = 7$ to $T = 10$, indicating that 7 consecutive days of data suffice to define the state of the system (a point in phase space).

The scaling regions are taken to begin above $N(\rho)/n = 0.01(N(\rho) \approx 40)$ and extend to larger $\rho$ with the limitation that the linear regression coefficient be greater than or equal to 0.999. The slopes of the scaling regions (Fig. 5) give $D(2) = 8.5$, $D(4) = 9.7$, $D(7) = 11.12$, $D(10) = 11.67$, and $D(14) = 11.68$, indicating an effective attractor dimension $D_{\text{eff}} = 11.7$ at the scale $0.04 < \rho < 0.08$.

At larger scales ($0.08 < \rho < 0.12$), one notes a slight *increase* in the effective dimension, to $D_{\text{eff}} \sim 14$, although the scaling region is not sufficient to give an accurate estimate of the slope. This contrasts with the decrease in the slope with saturation, at $\rho > 0.12$.

At the other end of the correlation graph, the first few points for which we can consider that the statistics are satisfactory suggest a steeper slope, particularly for larger values of $T$ (Figs. 4, 6); this hints that the effective dimension of the dynamics at small scales may be greater than in the scaling region, perhaps due to a weak coupling to external variables as suggested by Lorenz' work (Lorenz 1991). Let us assume that this external coupling can be approximated by a random variable with variance $\sigma^2$ and mean zero, added on to each phase space variable. The average squared distance between two points in phase space would then be $\rho^2 = \rho'^2 + 19T\sigma^2$, where $\rho'$ is the average without noise. For example, one can determine from Figs. 2–4 the average distance $\rho_{200}$ such that there are $N(\rho_{200}) = 200$ pairs of data points with distance less than $\rho_{200}$. From $N(\rho_{200}) \approx 0.04$ at $T = 7$ and $\rho_{200} \approx 0.05$ at $T = 14$ one finds $19\sigma^2 \approx 1.3 \times 10^{-4}$. This allows one to compute $\rho_{200} \approx 0.065$ at $T = 29$, in agreement with Figure 6. Although this is by no means a formal argument, the accuracy of the prediction for $T = 29$ indicates that one may be well justified to interpret the correlation graphs as follows: the correlation graphs indicate the existence of a deterministic system with attractor dimension $D_a = 11.7$ coupled to external perturbations, which can be reasonably well modeled as a random term added onto each variable, with variance $\sigma^2 \approx 1.3 \times 10^{-4}$ and mean zero.

If the deterministic signal is a representation of the local atmospheric dynamics and the random part is a representation of external perturbations, then one might

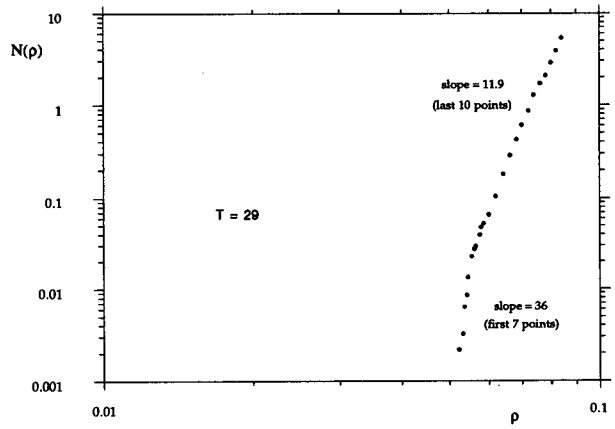


FIG. 6. The scaling graph for $T = 29$ consecutive days of data shows very clearly an increase in the slope at low values of $r$, up to $r = 0.065$ where there is no question that the effect observed is statistically significant, since the total number of pairs observed is then equal to $4000 \times 0.05$. The higher slope $D_{\text{eff}} = 36$ shows that the correlation dimension of the coupled system is at least this large.

Unauthenticated | Downloaded 10/19/21 11:55 AM UTC

think that a fully deterministic model could be obtained if one included in the phase space the so-called external variables, perhaps by including measurements of other variables besides the $N = 19$ variables of the dataset. A lower bound on the dimension of the attractor of the coupled system can be derived from the slope of the correlation graph at low values of $\rho$; this would give an effective dimension in excess of $D_{\text{eff}} \approx 35$ for $\rho \leq 0.04$ ($T = 7$) or $\rho \leq 0.05$ ($T = 14$). Such a large attractor dimension indicates that a reconstruction of the phase space would require a very large number of data points, corresponding to daily observations over at least $10^{35/11.7}$ yr just to reconstruct the attractor at the same very low resolution with which we reconstructed the 11.7-dimensional attractor. The approach suggested in the previous paragraph, to treat the perturbations as random noise, stands out as more realistic.

Whether or not the speculations of the previous two paragraphs stand up to future scrutiny, one can surely not assume that the effective dimension $D_{\text{eff}} \approx 11.7$ is valid at all scales, or that it is a valid estimate of the correlation dimension of the attractor, defined as the effective dimension in the limit $R \to 0$.

At the scale $0.04 < \rho < 0.08$, the effective dimension $D_{\text{eff}} \approx 11.7$ appears to be trustworthy: each point in the scaling region is based on a large number of pairs, so that $N(\rho)/n^2$ is insensitive to the size of the dataset (this has been checked explicitly by further reducing the dataset), and is well below the saturation, as is evidenced by the fact that the slope does not drop off until much larger values of $N(\rho)$. The following numerical experiments lend further confidence to the effective dimension result:

1. The sensitivity to the choice of delay was checked by averaging the data by sets of three days and thereby reducing the length of the dataset to $n/3$ days; the same attractor dimension was recovered.

2. The sensitivity to the choice of variables was examined by considering only the first six variables for each day instead of 19. Again, the same attractor dimension was found. A further reduction, to the first 2 variables, led to an underestimation of the effective dimension, $D_{\text{eff}} \approx 11.1$, which indicates that these two variables were not perfectly coupled to the system as a whole. Although this result may surprise strong believers in the Takens rule, following which any one variable should reflect the dynamics of the entire system through its interactions with all others, in practice if the interaction is too weak then they are detected only in the fine structure of the attractor, below the scale $\rho = 0.04$. This implies that if one is going to use a single variable in the correlation method, then it is important to choose one that is strongly coupled to the others, as noted by Lorenz (Lorenz 1991).

3. Finally, we performed the standard check of generating an artificial dataset with the same statistics but no deterministic dynamics, and found that the upper bound $D(T) \leq D_{\text{eff}}$ disappears. We generated such a dataset by randomly mixing the $n = 4015$ patterns, and found that the graph $D(T)$ increases monotonously, passing $D = 34$ at $T = 14$.

The largest Liapunov exponents can be estimated by considering the average rate of divergence of nearby trajectories (Briggs 1990; Gade and Amritkar 1990; Ellner et al. 1991; Galbraith 1992; Wales 1991; Zeng et al. 1991). For each available data point in the phase space, $\vec{x}_k \in \mathbb{R}^{T_e \times N}$, the nearest neighbor is identified and its evolution is compared to that of the original pattern, step by step. Here we excluded near neighbors that are also neighbors in time (measurements within 20 days of each other) since their trajectories do not diverge. The error at each time step is averaged over the patterns. The error grows rapidly with a timescale roughly equal to 2 days, then reaches a slowly rising upper bound that is significantly lower than the average distance between randomly chosen points (Fig. 7). The upper bound indicates a potential for longer-term predictability based on the persistence of weather patterns. The doubling of the error in two days would indicate a Liapunov exponent $\Lambda_1 \sim 1/2 \ln(2)$, although the dataset is clearly insufficient to consider this a serious estimate.

A quantitative estimate of the theoretical predictability limit can be obtained as a byproduct of the correlation method (Fraedrich 1987; Fraedrich and Leslie 1989; Nese 1989):

$$T_{\text{pre}} = \frac{\ln(N(\rho; T)) - \ln(N(\rho; T + m))}{m}. \quad (3.5)$$

From the slope of the graph of $N(\rho; T)$ as a function of $T$, with $\rho = 0.05$, one finds $T_{\text{pre}} = 1.83$ days (Fig. 8). Another estimate is given by the slope of the graph

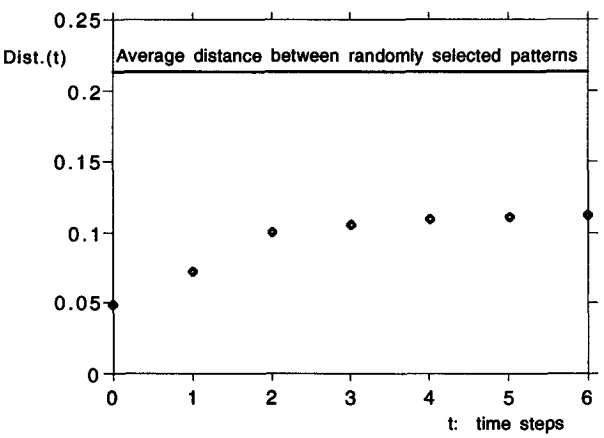


FIG. 7. The distance between two initially nearby points as they evolve in the pattern set is given as a function of the evolution time. After a rapid rise that reflects the short-term chaotic dynamics, this distance follows a slowly rising slope toward the upper limit, which corresponds to the distance between two randomly chosen patterns.

Unauthenticated | Downloaded 10/19/21 11:55 AM UTC

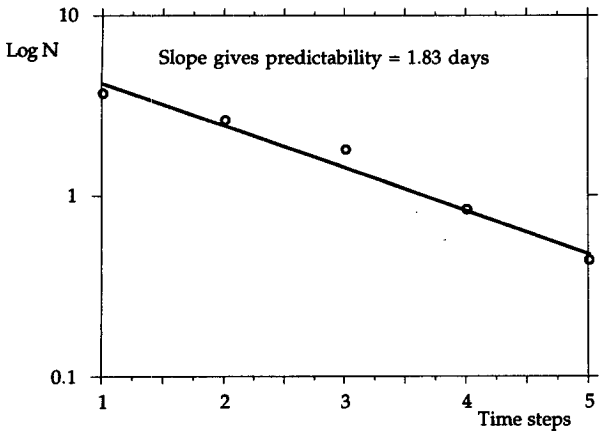


FIG. 8. An estimate of the short-term (chaotic) predictability is given by the graph of the number of pairs found in a ball of fixed radius as a function of the projected embedding dimension, $d = 19T$. This number decreases exponentially with the dimension $d$, and the rate of exponential decline is a measure of predictability.

$D(T)$ in the linear region previous to the saturation at $D(T_e) = D_{\text{eff}}$. A linear regression (Fig. 9) gives $T_{\text{pre}} = 2.1$ by this criterion. The similarity between these estimates and the consistence with the divergence rates discussed in the previous paragraph hints that they may be valid in spite of the very short dataset. The prediction error at time $T$ would then be $\epsilon(T) \approx \epsilon_0 e^{T/T_{\text{pre}}}$ for small $T$, where $\epsilon_0$ is the scale at which the reconstruction of the attractor is meaningful, $\epsilon_0 \sim 4\%$. Note that the predictability limit falls well short of the 2–3 week limit of large-scale weather patterns studied in General Circulation Models (GCMs). This confirms the claim made in the introduction that the high sensitivity of local atmospheric dynamics in the Tropics invalidates the grid approach of the GCMs for aspects of the weather that are sensitive to the local variables, such as tropical precipitation. Grid reductions of the fluid equations make sense only if the inevitable simplifications of the equations and the grid reduction itself do not betray the physics of the system: if the key factor which actually triggers rainstorms is the presence of a mountain, any physical model that does not take it into account, either directly or through a renormalization of the parameters, will necessarily give inaccurate predictions. For highly unstable systems, the attractor reconstruction method is an obvious candidate to avoid oversimplifying the dynamics. Unfortunately, the high dimensionality of the attractor severely limits the accuracy of the phase space reconstruction method in this particular example, as we shall see shortly.

## 4. Local reconstruction of the dynamical map

In the previous section, we saw that the state of the weather system can be described by a set of $T_e \approx 10$ consecutive days of measurements, corresponding to one point in the phase space $\mathbb{R}^{T_e \times N}$. Here, we will be concerned with the reconstruction of the dynamics (Farmer and Sidorowich 1987, 1988; Abarbanel et al. 1990; Giona et al. 1991; Gouesbet 1991; Linsay 1991; Tsonis 1992; Elsner and Tsonis 1992). Given an initial point $\vec{x}_0$ in this phase space, we wish to compute its orbit. The dynamical evolution integrated over one time step is a map $\mathbf{f}: \mathbb{R}^{T_e \times N} \to \mathbb{R}^{T_e \times N}$; we will denote by $\mathbf{f}_L$ the local reconstruction of this map in the vicinity of $\vec{x}_0$. A generalization of the method of analogs will be used (Lorenz 1963, 1969; van den Dool 1989; Toth 1989): the evolution of the initial point $\vec{x}_0$ will be based on the data points $\vec{x}_k$ closest to $\vec{x}_0$ and their known evolution one day ahead, $\mathbf{x}_{k+1} = \mathbf{f}(\vec{x}_k)$.

We will use the distance (3.1), with weights $\alpha(t)$ decreasing linearly with the delay, so that $\alpha(T_e - 1) = \alpha(0)/2$,

$$\alpha(t) = \frac{4(T_e - 1) - 2t}{3T_e(T_e - 1)}. \tag{4.1}$$

This choice of weighting function gives slightly better results than a flat weighting, $\alpha(t) = 1/T_e$. The difference, however, does not appear to be statistically significant.

We will consider the data points $\vec{x}_k$ such that $d(\vec{x}_k, \vec{x}_0) < \eta$, where $\eta$ is the radius of the ball around $\vec{x}_0$ where we wish to determine the local map $\mathbf{f}_L$. The reconstructed map is given by

$$\mathbf{f}_L(\vec{x}_0) = \frac{\sum_{k=1}^{n} \lambda[\eta - d(\vec{x}_k, \vec{x}_0)]\mathbf{x}_{k+1}}{\sum_{k=1}^{n} \lambda[\eta - d(\vec{x}_k, \vec{x}_0)]}, \tag{4.2}$$

where $\lambda$ is a threshold-linear weight function,

$$\lambda(x) = x\theta(x) \tag{4.3}$$

for a local linear reconstruction ($\theta$ is the step function).

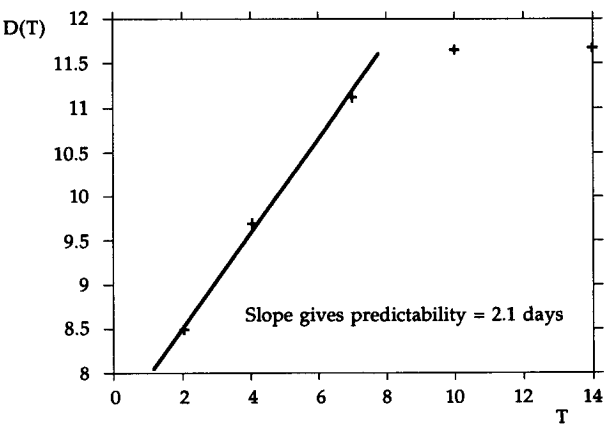


FIG. 9. Another estimate of the short-term predictability is given by the graph of the projected attractor dimension versus the projected embedding space dimension, $D(T)$, in the linear region. This measure gives the result $T_{\text{pre}}$ 2.1, comparable to that of Fig. 8.

Unauthenticated | Downloaded 10/19/21 11:55 AM UTC

Thus, the dynamical model (4.2) is a weighted average of $\mathbf{x}_{k+1} = \mathbf{f}(\vec{x}_k)$ over the points $\vec{x}_k$ in an $\eta$ ball centered at $\vec{x}_0$, weighted by $\eta - d(\vec{x}_k, \vec{x}_0)$.

For low values of $\eta$, this model functions as an associative memory to recall a temporal sequence of patterns, with a storage capacity that can be computed analytically using statistical methods developed for neural networks (Zertuche et al. 1994a,b). Here, we are interested in supercritical values of $\eta$, where the learned sequence is unstable and the model acquires positive Liapunov exponents, like the physical system being modeled.

The prediction model has two parameters that must be determined by optimization: the radius $\eta$, and the number of time steps that determine a point in phase space: $T_e$.

Based on the previous results on mutual information estimates and the application of Grassberger and Procaccia's algorithm, one would expect that $T_e$ should be sufficient to give a proper embedding of the attractor, $T_e \geqslant 7$, and $\eta$ should be such that a typical initial point has a sufficient number of neighbors with distance less than $\eta$ for a meaningful linear interpolation. For the given embedding dimension, the number of points should be $N(\eta) \approx 500$; which would require that $\eta \approx 0.1$. Unfortunately, this is not far below the saturation point ($N \leqslant 4000$), so one cannot expect a very fine reconstruction of the dynamics: the true dynamical map over such a large $\eta$-ball is not likely to be approximately linear; therefore, $\mathbf{f}_L(\vec{x}_0) \neq \mathbf{f}(\vec{x}_0)$.

The optimization was performed by removing the last year of data and using it to extract a statistical sampling of initial test points $\vec{x}_0$. The skill for one-day lead forecasts was found to be optimal for $T = 14$, $\eta = 0.1$, where

$$\text{skill} = 1 - \langle\langle d_1[\mathbf{f}_L(\vec{x}_0), \mathbf{x}_{k+1}]\rangle\rangle, \qquad (4.4)$$

$$d_1(\mathbf{x}_k, \mathbf{x}_l) = \frac{1}{N}\sum_{i=1}^{N} |x_k^{(i)} - x_l^{(i)}|. \qquad (4.5)$$

The skill of the optimal model ($T = 14$, $\eta = 0.1$) is

$$\text{Max(skill)} = 0.919, \qquad (4.6)$$

which compares to 0.904 for persistence, 0.905 for the seasonal average, and 0.865 for random pattern selection. More details on the results of the prediction code are being published separately (Waelbroeck et al. 1994).

At optimum, we also measured the skill in the prediction of the anomaly of rainfall accumulations over a 10-day period, a variable chosen for its relevance in agriculture: the skill was equal to 64% of the variance of the observed anomaly over the 11-yr span of measurements, where the anomaly of a variable at a given date is defined as the difference between the value of the variable and its seasonal average, evaluated over all available data within 10 days of this date, for each of the 11 yr of data (an average over 231 samples).

The 10-day rainfall predictions for all of 1992 (the last year in the dataset) are compared to the seasonal average and to the observed precipitations in Fig. 10. Although the predictions clearly do not follow the observations very closely, the unusual early beginning of the rain season was predicted by the model.

## 5. Summary

In summary, we have shown that the effective dynamics of a tropical weather system, at a scale corresponding to 3.5%–8% of the range of the variables, reflects an underlying deterministic dynamics with an attractor of dimension $D_a \approx 11.7$. The local linear structure of the attractor indicates a theoretical predictability of about 2 days for linear reconstructions of the dynamics.

A sharp increase in the effective dimension at smaller scales, $R/R_{\max} < 4\%$, may indicate that external perturbations of the local system become significant, in accordance with Lorenz's conjecture that the climate system consists of a number of weakly coupled subsystems whose attractors can be detected with the Grassberger–Procaccia algorithm.

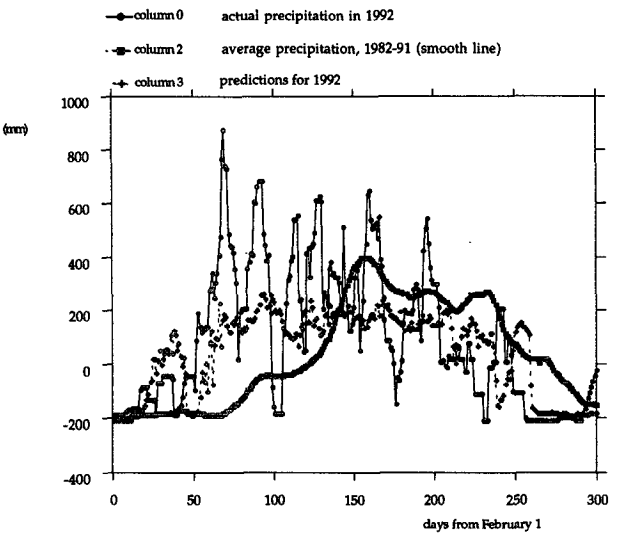


FIG. 10. The actual precipitation curve for 1992, smoothed by 10-day averaging (solid line) is compared to the normal precipitation curve, defined as an average over 10 yr of data for dates within 10 days of the current date (thick dotted line) and the predictions of the model for 10-day cumulative precipitation (fine dotted line). The shift in the $y$ axis reflects the annual average that was removed. Note that 1992 was an unusual year in that the rain season began almost two months ahead of schedule. This sort of phenomenon is precisely what a model should be able to predict if it is to be useful in practical applications, such as agriculture. The predictions slightly exaggerate the out-of-season winter rains (first bump at the left), but predict accurately the beginning of the rain season at day number 60. The other unusual phenomenon, a dry spell followed by a recovery of rains late in the rain season, is predicted, but with a delay of about 10 days.

Unauthenticated | Downloaded 10/19/21 11:55 AM UTC

A simple reconstruction of the dynamics was proposed, which showed some skill in 10-day predictions: this long-term predictability appears to reflect the persistence of weather patterns, which would also cause the slow decline of the mutual information as a function of the delay.

Altogether, three data analysis methods were applied and gave reasonably coherent results for the following characteristics of the system.

1. Embedding dimension: The mutual information analysis gave $T_e = 14$, the scaling graphs led to $T_e = 7$, and the optimization of the local reconstruction model gave $T_e = 14$.
2. Short-term predictability, limited by $\Lambda_1$: The mutual information analysis indicates that $T_{\text{pre}} \leqslant 1$ day, the scaling graphs gave $T_{\text{pre}} \leqslant 2$ days, and the local reconstruction model, $T_{\text{pre}} \leqslant 1$ day.
3. Long-term predictability, from the persistence of weather patterns: After a quick initial drop, the mutual information fades out more slowly with a time scale of 20 days. Also, the distance of nearby trajectories quickly reaches a maximum distance $\rho \approx 0.1$ after which it increases more slowly, with a time scale of about 20 days. Finally, the prediction model shows that useful skill can be extracted for 10-day predictions.

The general coherence of these results lends some confidence that they reflect genuine physical properties of the dynamical system. There appears to be three effective dynamics at three different scales.

1. At very fine scales, one has a very complex system involving the local system and its couplings to the external variables, with an effective dimension of the attractor $D_a \geqslant 35$.
2. At the scale $0.04 < \rho < 0.08$, the local weather system dominates, with an effective attractor dimension $D_a \approx 11.7$, a Liapunov exponent $\Lambda_1 \geqslant 1/2 \ln(2)$, and a predictability limited to a couple of days.
3. At the scale $\rho \approx 0.1$, one has general features of weather patterns that follow a much softer dynamical evolution, characterized by long predictability times, of the order of 20 days.

The situation is similar to that encountered in linear regression forecasts, where a short-term predictability due to a first-order Markov process is followed by a longer "tail of autocorrelation," which allows prediction of monthly means well beyond what can be expected from the first-order process (Gutzler and Mo 1983; Horel 1985; Trenberth 1985; Van den Dool et al. 1986; Roads and Barnett 1984; Strauss and Halem 1981).[1] Of course in this article the short-term predictability is modeled by a nonlinear rule based on 14 consecutive days of data, but a similar "tail of mutual information" is likely to be responsible for the quality of the predictions at medium-long range, as anticipated in section 2.

The external perturbations affect only the system progressively, reaching the amplitude $\rho = 0.05$ after 14 days; this indicates that the short-term predictability could be improved and extended several days, if one could improve on the resolution with which the local dynamics is reconstructed: this would require improving the use of the available local data to better define the structure of the attractor near the scale $R/R_{\max} \approx 4\%$.

Beyond that point, any further improvement in prediction skill would require introducing the large-scale external perturbations from general circulation models. This last task requires a methodology for determining what aspects of large-scale weather patterns are most relevant to the local system. We hope that the preliminary work that we have described in this article has laid the first stones toward a new approach to small-scale meteorology, which is custom designed to predict local, chaotic phenomena.

*Acknowledgments.* This work was supported by CONACyT Grant 400349-5-1714E and by the Association Générale pour la Coopération et le Développement (Belgium). We would like to express our gratitude to Tomás Morales of the Center for Atmospheric Sciences and to members of the Tlaxcala regional center of the Comisión Nacional del Agua for access to the data, and thank S. Orozco, J. Jiménez, R. Orea, M. Vázquez, and A. Meneses for the preparation of the dataset. We also gladly express our gratitude to the Comité de Supercómputo of the UNAM for access to the Cray YM-P4. Finally, we are indebted to C. Duqué, director of the Association pour le Développement par la Recherche et l'Action Integrées and the Belgian Ambassador to Mexico, for their support of our project.

## REFERENCES

Abarbanel, H. D. I., R. Brown, and J. B. Kadtke, 1990: Prediction in chaotic nonlinear systems: Methods for time series with broadband Fourier spectra. *Phys. Rev.*, **41A,** 1782.

Anthos, R. A., 1991: Predictability of mesoscale meteorological phenomena. *Predictability of Fluid Motions,* G. Holloway and B. J. West, Eds., Amer. Inst. Phys., 247–270.

Briggs, K., 1990: An improved method for estimating Liapunov exponents of chaotic time series. *Phys. Lett. A,* **151,** 27.

Ellner, S., A. R. Gallant, D. McCaffrey, and D. Nychka, 1991: Convergence rates and data requirements for Jacobian-based estimates of Lyapunov exponents from data. *Phys. Lett. A,* **153,** 357.

Elsner, J. B., and A. A. Tsonis, 1991: Do bidecadal oscillations exist in the global temperature record. *Nature,* **353,** 551.

——, and ——, 1992: Nonlinear prediction, chaos, and noise. *Bull. Amer. Meteor. Soc.,* **73,** 49–60.

Essex, C., T. Lookman, and N. A. H. Nerenberg, 1991: Climate attractor over very short time scales. *Nature,* **326,** 64–66.

Farmer, J. D., and J. J. Sidorowich, 1987: Predicting chaotic time series. *Phys. Rev. Lett.,* **59,** 845.

——, and ——, 1988: Exploiting chaos to predict the future and reduce noise. Los Alamos preprint LA-UR-88-901.

[1] We thank the referee for this comment.

Unauthenticated | Downloaded 10/19/21 11:55 AM UTC

Feller, W., 1968: *An Introduction to Probability Theory and its Applications.* Wiley, 244 pp.
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.
——, and L. M. Leslie, 1989: Estimates of cyclone track predictability. Part I: Tropical cyclones in the Australian region. *Quart. J. Roy. Meteor. Soc.,* **115,** 79–92.
Fraser, A. M., 1988: Phase space reconstructions from time series. *Physica,* **D34,** 391.
——, 1989: Information and entropy in strange attractors. *I.E.E.E. Trans. Info. Theory,* **35,** 245.
——, and H. L. Swinney, 1983: Independent coordinates for strange attractors from mutual information. *Phys. Rev. A,* **33,** 1134.
Frison, T., 1990: Predicting nonlinear and chaotic systems behavior using neural networks. *J. Neural Net. Comput.,* **2,** 45.
Gade, P. M., and R. E. Amritkar, 1990: Characterizing loss of memory in a dynamical system. *Phys. Rev. Lett.,* **65,** 389.
Galbraith, J. W., 1992: Inference about trends in global temperature data. *Clim. Change,* **22,** 209–221.
Ghil, M., and R. Vautard, 1991: Interdecadal oscillations and the warming trend in global temperature time series. *Nature,* **350,** 324.
Giona, M., F. Lentini, and V. Cimagalli, 1991: Functional reconstruction and local prediction of chaotic time series. *Phys. Rev. A,* **44,** 3496.
Gouesbet, G., 1991: Reconstruction of the vector fields of continuous dynamical systems from numerical scalar time series. *Phys. Rev. A,* **43,** 5321.
Grassberger, P., 1986: Do climatic attractors exist? *Nature,* **323,** 609–612.
——, and I. Procaccia, 1983: Characterization of strange attractors. *Phys. Rev. Lett.,* **50,** 346.
Gutzler, D. S., and K. C. Mo, 1983: Autocorrelation of Northern Hemisphere geopotential heights. *Mon. Wea. Rev.,* **111,** 155–164.
Hammel, S. M., 1990: A noise reduction method for chaotic systems. *Phys. Lett. A,* **148,** 421.
Henderson, H. W., and R. Wells, 1988: Obtaining attractor dimensions from meteorological time series. *Advances in Geophysics,* Vol. 30, Academic Press, 205.
Hense, A., 1987: On the possible existence of a strange attractor for the Southern Oscillation. *Beitr. Phys. Atmosph.,* **60,** 34–47.
Horel, J. D., 1985: Persistence of the 500-mb height field during Northern Hemisphere winter. *Mon. Wea. Rev.,* **113,** 2030–2042.
Kaplan, D. T., and L. Glass, 1992: Direct test for determinism in a time series. *Phys. Rev. Lett.,* **68,** 427.
Katz, R. W., 1982: Statistical evaluation of climate experiments with general circulation models: A parametric time series modeling approach. *J. Atmos. Sci.,* **39,** 1446–1455.
Linsay, P. S., 1991: An efficient method of forecasting chaotic time-series using linear interpolation. *Phys. Lett. A,* **153,** 353.
Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.,* **20,** 130–141.
——, 1969: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.,* **26,** 636–646.
——, 1991: Dimension of weather and climate attractors. *Nature,* **353,** 241.
Nese, J. M., 1989: Quantifying local predictability in phase space. *Physica D,* **35,** 237.
Nicolis, C., and G. Nicolis, 1984: Is there a climate attractor? *Nature,* **311,** 529.
Pandit, S. M., and S. M. Yu, 1983: *Time Series and System Analysis With Applications.* Wiley, 272 pp.
Roads, J. O., and T. P. Barnett, 1984: Forecasts of the 500-mb height using a dynamically oriented statistical model. *Mon. Wea. Rev.,* **112,** 1354–1369.
Sharifi, M. B., K. P. Georgekakos, and I. Rodriguez-Iturbe, 1990: Evidence of deterministic chaos in the pulse of storm rainfall. *J. Atmos. Sci.,* **47,** 888–893.
Strauss, D. M., and M. Halem, 1981: A stochastic-dynamical approach to the study of natural variability of the climate. *Mon. Wea. Rev.,* **109,** 407–421.
Toth, Z., 1989: Long-range weather forecasting using an analog approach. *J. Climate,* **2,** 594–607.
Tsonis, A. A., 1992: *Chaos. From Theory to Applications.* Plenum Press, 274 pp.
——, and J. B. Elsner, 1988: Weather attractor over very short time scales. *Nature,* **333,** 545–547.
——, and ——, 1989: Chaos, strange attractors and weather. *Bull. Amer. Meteor. Soc.,* **70,** 14–23.
Van den Dool, H. M., 1989: A new look at weather forecasting through analogues. *Mon. Wea. Rev.,* **117,** 2230–2247.
——, W. H. Klein, and J. E. Walsh, 1986: The geographical distribution and seasonality of persistence in monthly mean air temperatures over the United States. *Mon. Wea. Rev.,* **114,** 546–560.
Vautard, R., and M. Ghil, 1989: Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D,* **35,** 395.
Waelbroeck, H., and Coauthors, 1994: Prediction of tropical rainfall by local phase space reconstruction. *J. Atmos. Sci.,* **51,** 3360–3365.
Wales, D. J., 1991: Calculating the rate of loss of information from chaotic time series by forecasting. *Nature,* **350,** 485.
Zeng, X., R. Eykholt, and R. A. Pielke, 1991: Estimating the Lyapunov exponent spectrum from short time series of low precision. *Phys. Rev. Lett.,* **66,** 3229.
Zertuche, F., and Coauthors, 1994a: Storage capacity of a neural network with state-dependent synapses. *J. Phys. A,* **27,** 1575–1583.
——, 1994b: Recognition of temporal sequences of patterns using state-dependent synapses. *J. Physics A,* **27,** 5879–5887.

Unauthenticated | Downloaded 10/19/21 11:55 AM UTC