AAPT PHYSICS EDUCATION | AMERICAN JOURNAL of PHYSICS

Predicting the future from the past: An old problem from a modern perspective

F. Cecconi, M. Cencini, M. Falcioni, and A. Vulpiani

Citation: Am. J. Phys. 80, 1001 (2012); doi: 10.1119/1.4746070
View online: http://dx.doi.org/10.1119/1.4746070
View Table of Contents: http://ajp.aapt.org/resource/1/AJPIAS/v80/i11
Published by the American Association of Physics Teachers

Related Articles

Five popular misconceptions about osmosis
Am. J. Phys. 80, 694 (2012)

Resource Letter SS–1: The Spin-Statistics Connection
Am. J. Phys. 80, 561 (2012)

Introducing thermodynamics through energy and entropy
Am. J. Phys. 80, 627 (2012)

Analytic determination of the mean free path of sequential reactions
Am. J. Phys. 80, 316 (2012)

Quantum chaos: An introduction via chains of interacting spins 1/2
Am. J. Phys. 80, 246 (2012)

Additional information on Am. J. Phys.

Journal Homepage: http://ajp.aapt.org/
Journal Information: http://ajp.aapt.org/about/about_the_journal
Top downloads: http://ajp.aapt.org/most_downloaded
Information for Authors: http://ajp.dickinson.edu/Contributors/contGenInfo.html

ADVERTISEMENT

Predicting the future from the past: An old problem from a modern perspective

F. Cecconi and M. Cencini
Istituto dei Sistemi Complessi, Consiglio Nazionale delle Ricerche, Via dei Taurini 19, I-00185 Rome, Italy

M. Falcioni[a)]
Dipartimento di Fisica, Università "Sapienza," Piazzale A. Moro 2, 00185 Rome, Italy

A. Vulpiani
Dipartimento di Fisica, Università Sapienza, Piazzale A. Moro 2, 00185 Rome, Italy and Istituto dei Sistemi Complessi, Consiglio Nazionale delle Ricerche, Via dei Taurini 19, I-00185 Rome, Italy

(Received 16 December 2011; accepted 1 August 2012)

The idea of predicting the future from the knowledge of the past is quite natural, even when dealing with systems whose equations of motion are not known. This long-standing issue is revisited in the light of modern ergodic theory of dynamical systems and becomes particularly interesting from a pedagogical perspective due to its close link with Poincaré's recurrence. Using such a connection, a very general result of ergodic theory—Kac's lemma—can be used to establish the intrinsic limitations to the possibility of predicting the future from the past. In spite of a naive expectation, predictability is hindered more by the effective number of degrees of freedom of a system than by the presence of chaos. If the effective number of degrees of freedom becomes large enough, whether the system is chaotic or not, predictions turn out to be practically impossible. The discussion of these issues is illustrated with the help of the numerical study of simple models.

[http://dx.doi.org/10.1119/1.4746070]

I. INTRODUCTION

Predicting the future state of a system has always been a natural motivation for science, with applications such as weather forecasting and tidal prediction. Understanding the limitations to the predictability of a system evolution is often crucial.

In deterministic systems, where the future is uniquely determined by the present, two main approaches to the predictability problem can be taken. The first applies to systems whose evolution laws are known, in terms of either differential or difference equations. In this case, predictability is limited mainly by the presence of sensitivity to initial conditions (deterministic chaos). As taught in dynamical system courses, this sensitivity is characterized by the Lyapunov exponent. The second approach applies to phenomena whose governing laws are not known but whose evolution can be measured and recorded. In such a case, the best practical strategy is to use the past, as a full-scale model of the system, to make predictions of the future evolution.

The present paper discusses at an introductory level the latter method, which was developed in the framework of nonlinear time series analysis.[1–3] This topic is seldom included in basic courses and is closely related to an apparently distant classical theme, the Poincaré recurrences.[4] Surprisingly, although simple to establish,[5] such a connection has been overlooked even by specialists, as recently remarked by Altmann and Kantz.[6] Such a link also allows us to clarify the practical role of theoretical concepts such as the attractor dimension of a dynamical system. Indeed, as we shall see, when the evolution laws are unknown, the actual constraints to our prediction capabilities are rather set by the number of degrees of freedom (attractor dimension) than by the presence of chaos. This fact is often overlooked in favor of the widespread folklore of the so-called butterfly effect.[7] In this respect, it is important to stress that such limitations to predictability are a consequence of rather general results of ergodic and dynamical-system theory. Although the main ideas had been already put forward by Boltzmann,[8] many misguided applications of nonlinear time series analysis appeared in the literature after the rediscovery of chaos (see, e.g., Ref. 9).

One of the main reasons for excluding this topic from basic courses is the necessity to introduce advanced technical tools[3] such as the embedding technique.[3,10] Therefore, here, we present the problem in its simplest formulation. Often, when recording the evolution of a system with unknown dynamics, not all the variables necessary to identify the states or even their number are known. Moreover, if we are lucky enough to know them, we can access only one or a few scalar functions of them, typically affected by measurement errors. Throughout this paper, we will disregard all these technical difficulties (which can be to a large extent handled with specific techniques[3]) and assume that the necessary variables can be recorded with arbitrary precision. Even with such an ideal working hypothesis, the above-mentioned fundamental constraints to predictability are unavoidable.

The material is organized as follows. In Sec. II, after some historical notes, we introduce the method of analogs as the simplest procedure to predict the future from past time series. Section III introduces the model system used to clarify the main issues. In Sec. IV, we discuss the link between analogs and Poincaré recurrences, and show how the actual limitation to predictability from data stems from the effective number of degrees of freedom. Section V discusses two cases where the method works successfully; one is illustrated by a numerical example and the other refers to the important practical problem of tidal predictions. Finally, Sec. VI is devoted to conclusions.

II. THE METHOD OF ANALOGS

"If a system behaves in a certain way, it will do so again" seems a rather natural claim when referred, for instance, to

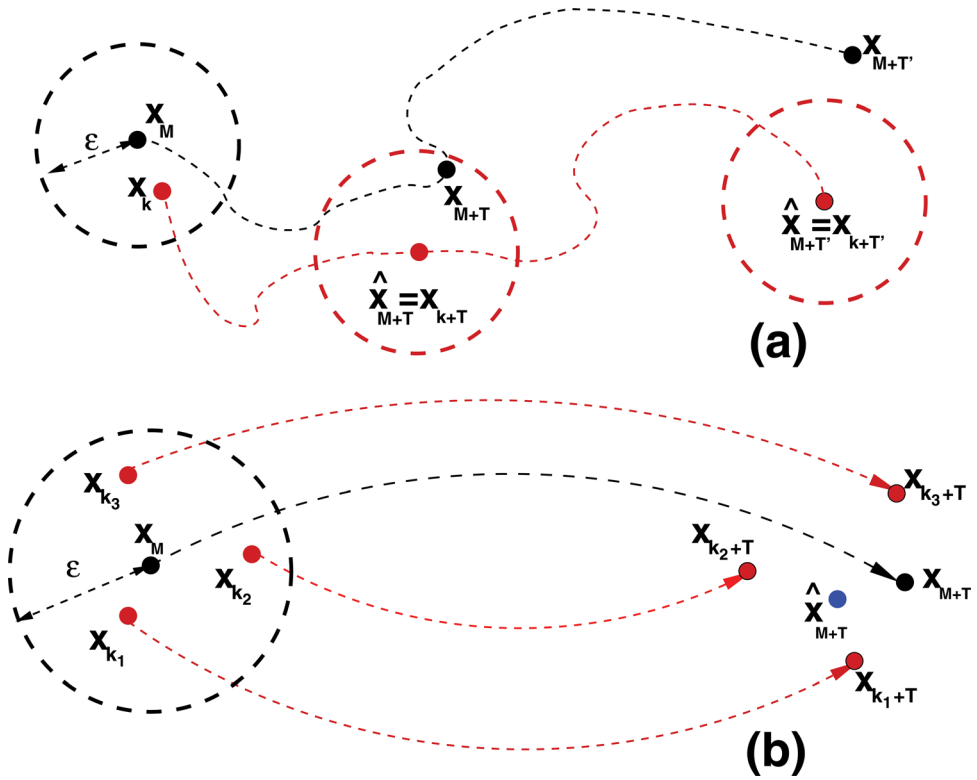

Fig. 1. (Color online) Sketch of the method of analogs: (a) illustration of Eq. (1) and of the error growth; (b) generalization of the method to more than one analog. In particular, if N_a analogs $\{\boldsymbol{x}_{k_n}\}_{n=1}^{N_a}$ are found, Eq. (1) can be replaced by $\widehat{\boldsymbol{x}}_{M+T} = \sum_{n=1}^{N_a} \mathbb{E}_n \boldsymbol{x}_{k_n+T}$ where the matrices $\mathbb{E}_n$ can be computed by suitable interpolations.

the diurnal and seasonal cycles. This claim is also supported by Biblical tradition:[11] "What has been will be again, what has been done will be done again; there is nothing new under the sun." This idea, together with the belief in determinism ("from the same antecedents follow the same consequents"), is at the basis of prediction methods. However, as Maxwell argued,[12] "It is a metaphysical doctrine that from the same antecedents follow the same consequents. … But it is not of much use in a world like this, in which the same antecedents never again concur, and nothing ever happens twice. … The physical axiom which has a somewhat similar aspect is 'That from like antecedents follow like consequents.'" These words no more surprise scientists, aware, by now, of the almost exceptional character of periodic behaviors and of the ubiquitous presence of irregular evolutions due to deterministic chaos; but at that time, they constituted a departure from tradition.

In spite of Maxwell's authoritative opinion, until World War I, weather forecasters substantially used empirical implementations of the naive idea, exploiting their experience and memory of past similar "patterns" (roughly surfaces of discontinuity between warm and cold air masses) to produce weather map predictions.[13] In the preface to his seminal book *Weather Prediction by Numerical Process*, Richardson criticizes the empirical approaches and, through an argument similar to Maxwell's,[14] contends that for weather forecasting, it is much more useful to integrate the partial differential (thermo-hydrodynamical) equations ruling the atmosphere. Although, as history witnessed, the successful approach to predictions is that foreseen by Richardson, it is interesting to discuss the range of applicability of predictions based on the past evolution of a deterministic system.

Lorenz introduced a mathematical formulation of the idea, called *method of analogs*,[15,16] which is the most straightforward approach to predictability in the absence of a detailed knowledge of the physical laws.

In its simplest form, the method works as follows. Assume that the known state $\boldsymbol{x}(t)$ of a process can be sampled at times $t_k = k\Delta t$ with arbitrary precision. The sampling interval Δt is also assumed to be arbitrary but not too short. We collect the sequence of states $\boldsymbol{x}_k = \boldsymbol{x}(t_k)$ with $k = 1 \ldots M$. If from the present state $\boldsymbol{x}_M$, we would like to forecast the future $\boldsymbol{x}_{M+T}$ at time t_{M+T} $(T \geq 1)$, the basic idea is to search in the past $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{M-1})$ for that state, say $\boldsymbol{x}_k$, most similar to $\boldsymbol{x}_M$, and to use its consequents as proxies for the future evolution of $\boldsymbol{x}_M$. Mathematically, we require that $|\boldsymbol{x}_k - \boldsymbol{x}_M| \leq \epsilon$, and we dub $\boldsymbol{x}_k$ an ϵ-*analog* to $\boldsymbol{x}_M$. If the analog were perfect $(\epsilon = 0)$, the system (being deterministic) would surely be periodic and the prediction trivial $\boldsymbol{x}_{M+T} \equiv \boldsymbol{x}_{k+T}$ for any T. If it is not perfect $(\epsilon > 0)$, we can use the forecasting recipe

$$\widehat{\boldsymbol{x}}_{M+T} = \boldsymbol{x}_{k+T}, \tag{1}$$

where the $\widehat{}$ symbol denotes an approximation. This prediction is depicted in Fig. 1(a) and simply expresses the principle that from similar antecedents follow similar consequents. For the prediction of Eq. (1) to be meaningful, the analog $\boldsymbol{x}_k$ must not be a near-in-time antecedent. When more than one analog can be found, the generalization of Eq. (1) is depicted in Fig. 1(b).

Once a "good" analog (meaning ϵ reasonably small) has been found, the next step is to determine the accuracy of the prediction (1), namely the difference between the forecast and the actual state, $|\widehat{\boldsymbol{x}}_{M+T} - \boldsymbol{x}_{M+T}|$. In practice, the ϵ-analog is the present state with an uncertainty, $\boldsymbol{x}_k = \boldsymbol{x}_M + \delta_0$ $(\delta_0 \leq \epsilon)$,

and the prediction (1) can be considered acceptable until the error $\delta_T = |\boldsymbol{x}_{M+T} - \boldsymbol{x}_{k+T}|$ remains below a tolerance Δ, dictated by the practical needs. The predictability time $\widehat{T} = \widehat{T}(\delta_0, \Delta)$ is then defined by requiring $\delta_T \lesssim \Delta$ for $T \leq \widehat{T}$.

Accuracy and predictability time are clearly related to (possible) sensitivity to initial conditions, as pioneered by Lorenz himself.[17] As taught in basic dynamical systems courses, chaotic evolutions exponentially amplify an infinitesimal error

$$\delta_T \simeq \delta_0 e^{\lambda_1 T}, \tag{2}$$

where λ_1 is the maximal Lyapunov exponent.[18] For a gentle introduction to Lyapunov exponents, the reader may refer to Ref. 19. Therefore, given a good analog, the prediction will be Δ-accurate up to a time

$$\widehat{T}(\delta_0, \Delta) \approx \frac{1}{\lambda_1} \ln \frac{\Delta}{\delta_0}. \tag{3}$$

Strictly speaking, for the above equation to be valid, both δ_0 and Δ must be very small.[18] It is worth remarking that the evaluation of the error growth rate (2) provides, at least in principle, a way to determine the Lyapunov exponent λ_1 from a long time series.

Conversely, deterministic non-chaotic systems are less sensitive to initial conditions: The error grows polynomially in time, and usually, $\widehat{T}(\delta_0, \Delta)$ turns out to be longer than that of chaotic systems, making long-term predictions possible.

For those familiar with chaotic systems, we have apparently reached the obvious conclusion that the main limit to predictions based on analogs is the sensitivity to initial conditions, typical of chaos. But, as realized by Lorenz himself, the main issue is actually to find good (small ϵ) analogs:[16] "In practice, this procedure may be expected to fail, because of the high probability that no truly good analogues will be found within the recorded history of the atmosphere." Lorenz also pointed out that the method is limited by the need for a very large data set,[15] independent of the presence of chaos.

It is worth concluding this historical presentation with a brief comment on the application of the method of analogs in Lorenz's original work.[15] Lorenz was strongly supporting weather forecasting based on solving the (approximate) equations of the atmosphere, as outlined by Richardson. He realized that the intrinsic limits to weather forecasting cannot be established by estimating the intrinsic error growth of these solutions. This work represents the first attempt to estimate the Lyapunov exponent from data, pioneering modern time series analysis.[20] He also realized that, unfortunately, the true Lyapunov exponent of the atmosphere cannot be estimated from data, as good analogs cannot be found and the difference between mediocre analogs may be expected to amplify more slowly than the difference between good analogs, because nonlinear effects play a greater role when the errors are large.

III. STUDY OF A SIMPLE MODEL

The difficulties in finding good analogs can be quantified by studying analog statistics. As an illustrative example, we compute numerically the probability of finding ϵ-analogs to a state in a simple model system introduced by Lorenz in 1996,[21] hence called the Lorenz-96 model. It consists of the following nonlinearly coupled ordinary differential equations:

$$\frac{dX_n}{dt} = X_{n-1}(X_{n+1} - X_{n-2}) - X_n + F, \tag{4}$$

where $n = 1, \ldots, N$ and periodic boundary conditions $(X_{N\pm n} = X_{\pm n})$ are assumed. The variables X_n may be thought of as the values of some atmospheric representative observable along the latitude circle, so that Eq. (4) can be regarded as a one-dimensional caricature of atmospheric motion.[21] The quadratic coupling conserves energy, $\sum_n X_n^2$. In the presence of forcing F and damping $-X_n$, the energy is only statistically conserved. The motion is thus confined to a bounded region of $\mathbb{R}^N$. Moreover, dissipation constrains the trajectories to evolve onto a subset of this region possibly with dimension $<N$, namely an attractor (fixed points, limit cycles, or a strange attractor if the dynamics is chaotic). The dynamical features are completely determined by the forcing strength F and by the system dimensionality N. In particular, for $F > 8/9$ and $N \geq 4$, the system displays chaos with exponential separation of nearby initial conditions.[22]

In principle, the statistics of the analogs of system (4) can be determined according to the following procedure. Given a state of the system $\boldsymbol{x}_M$ on the attractor, we have to consider its precursors $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{M-1})$ along the trajectory ending in $\boldsymbol{x}_M$ sampled at regular time intervals of duration Δt, $\boldsymbol{x}_i = \boldsymbol{x}(t_i = i\Delta t)$. Hence, the ϵ-analogs of $\boldsymbol{x}_M$ are those states $\boldsymbol{x}_j$ such that $|\boldsymbol{x}_j - \boldsymbol{x}_M| \leq \epsilon$. Finally, the fraction of ϵ-analogs,

$$C_M(\epsilon) = \frac{1}{M-1} \sum_{j=1}^{M-1} \Theta(\epsilon - |\boldsymbol{x}_j - \boldsymbol{x}_M|), \tag{5}$$

provides an estimate of the probability to find ϵ-analogs to $\boldsymbol{x}_M$ as a function of both the desired degree of similarity ϵ and the length of the history M we recorded. Being interested in typical behaviors and not just in the properties around a specific state $\boldsymbol{x}_M$, it is convenient to average $C_M(\epsilon)$ over r independent reference states. Therefore, instead of considering only the end point $\boldsymbol{x}_M$, we select r states $\{\boldsymbol{x}_\star^k\}_{k=1}^r$ along the trajectory, well spaced in time to be considered independent configurations on the attractor, and we replace Eq. (5) by the average fraction of ϵ-analogs,

$$C_{r,M}(\epsilon) = \frac{1}{Mr} \sum_{k=1}^{r} \sum_{j=1}^{M} \Theta(\epsilon - |\boldsymbol{x}_j - \boldsymbol{x}_\star^{(k)}|). \tag{6}$$

Because in our case we know the evolution laws (4), it is not really necessary to look at the backward time series of the reference states. In practice, we can select the $\{\boldsymbol{x}_\star^k\}_{k=1}^r$ and look at their forward ϵ-analogs.

The latter procedure is used to produce Fig. 2, where we show $C_{r,M}(\epsilon)$ obtained with $r = 10^3$ reference states and different lengths M of the time series, from 10^3 to 10^7. Of course, when the degree of similarity ϵ becomes larger than the attractor size, say $\epsilon_{\max}$, the fraction $C_{r,M}(\epsilon)$ saturates to 1. Therefore, it is meaningful to normalize the degree of similarity by $\epsilon_{\max}$. As for the dynamics (4), the forcing is fixed to $F = 5$ and we consider two system sizes, $N = 20$ and $N = 21$. In both cases, the system is chaotic. While for $N = 21$, analogs can be found with reasonable probability even for small values of ϵ $(\lesssim 10^{-4}\epsilon_{\max})$, for $N = 20$, analogs are found only for large values of ϵ $(\gtrsim 10^{-2}\epsilon_{\max})$, even for $M = 10^7$. The solid lines in Fig. 2 indicate that for $\epsilon \ll \epsilon_{\max}$, the probability to find an analog is fairly well approximated by a power law,

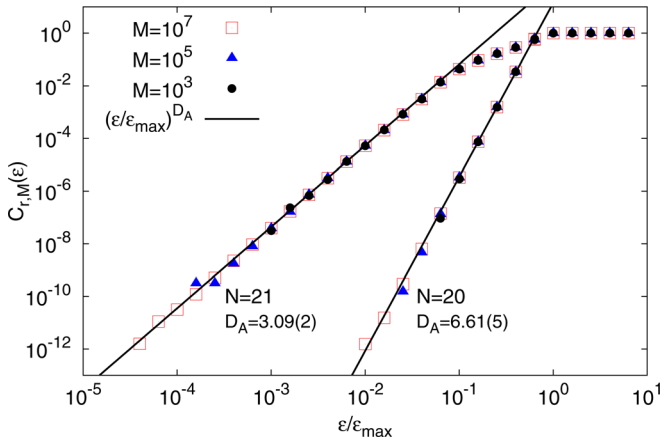

Fig. 2. (Color online) $C_{r,M}(\epsilon)$ vs $\epsilon/\epsilon_{\max}$ for $F=5$, $N=20$ and $N=21$; the reference states are $r=1000$ and different values of M ranging from 10^3 to 10^7 are considered. The solid lines are the fits of the data by means of relation (9).

$$C_{r,M}(\epsilon) \propto \epsilon^{D_A}. \tag{7}$$

In particular, we find $D_A \simeq 3.1$ and $D_A \simeq 6.6$ for $N=21$ and $N=20$, respectively. Therefore, the exponent D_A quantifies the difference between the two cases: upon lowering ϵ, the probability to find ϵ-analogs with $N=20$ becomes about $\epsilon^{3.5}$ times smaller than with $N=21$.

The probability to find ϵ-analogs is expected to decrease upon increasing the number of degrees of freedom N, as more constraints on the single components of the state vector should be satisfied. In this perspective, the above result seems at odds with intuition unless the exponent D_A in Eq. (7) is interpreted as the "effective" number of degrees of freedom.

We end this section by warning the reader that the counter-intuitive inequality $D_A(N=21) < D_A(N=20)$ is a peculiar consequence of the choice of the parameters F and N.[22] Generally, D_A is expected to increase with N.[18] Here, we made this choice to emphasize the importance of the effective number of degrees of freedom that, in general, is not trivially related to (and can be much smaller than) the number of variables N. As we shall see in Sec. IV, D_A is nothing but the attractor dimension, a measure of the effective number of degrees of freedom.

IV. DEGREES OF FREEDOM, RECURRENCE TIMES, AND ANALOGS

In this section, we recall some basic notions of ergodic dynamical systems and underline their connections with the analogs. In particular, we link the difficulty of finding analogs to the presence of long recurrence times.

A. The role of dimensions

The founding principle of ergodic theory is that the long-time statistical properties of a system can be equivalently described in terms of the invariant (time-independent) probability, μ, such that $\mu(\sigma)$ is the probability of finding the system in any specified region σ of its phase space. The phase space of a system described by N degrees of freedom is a region of $\mathbb{R}^N$, that is, an N-dimensional space.

If the evolution conserves phase-space volumes (as in the Hamiltonian motion of classical systems), then the probability $d\mu(\boldsymbol{x})$ of finding the state in a small region of volume dV, as defined in elementary geometry, around $\boldsymbol{x}$ is proportional to dV, i.e., to the Lebesgue measure of that region. In dissipative systems, phase-space volumes are contracted on average and the invariant probability $d\mu(\boldsymbol{x})$ is not proportional to dV, but concentrates on a set (the attractor) $A \subset \mathbb{R}^N$ of dimension $D_A < N$. Slightly more formally, the dimension D_A describes the small-scale ($\ell \ll 1$) behavior of the probability $\mu(B_{\boldsymbol{y}}^N(\ell))$ of finding points $\boldsymbol{x} \in A$, which are in the N-dimensional sphere of radius ℓ around $\boldsymbol{y}$

$$\mu(B_{\boldsymbol{y}}^N(\ell)) = \int_{B_{\boldsymbol{y}}^N(\ell)} d\mu(\boldsymbol{x}) \sim \ell^{D_A}. \tag{8}$$

Therefore, the trajectories of dissipative systems are effectively described by a number $D_A < N$ of degrees of freedom, though defined in an N-dimensional space.

For a noninteger D_A, the attractor and probability are said to be fractal. In general, attractors are nonhomogeneous with D_A, in Eq. (8), depending on $\boldsymbol{y}$, and an infinite set of dimensions is needed to fully characterize the invariant probability; we speak of multifractal objects.[23] For the sake of our discussion, these technical complications can be ignored, and the attractor can be assumed homogeneous and characterized by a single dimension D_A.

Upon reconsidering $C_M(\epsilon)$ defined in Eq. (5), we see that it is nothing but the fraction of time the trajectory spends in a sphere or radius ϵ centered in $\boldsymbol{x}_M$. For large M, as a consequence of ergodicity, $C_M(\epsilon)$ gives the probability of finding the system in that sphere, and the quantity (6) is an averaged probability. Therefore, for sufficiently large M and small ϵ, Eq. (8) implies

$$C_{r,M}(\epsilon) \approx \langle \mu(\epsilon) \rangle \sim \epsilon^{D_A}. \tag{9}$$

Strictly speaking, in Eq. (9), the right exponent should be the correlation dimension D_2, which controls the small-scale asymptotics of the probability to find two points on the attractor at distance $\leq \epsilon$.[18,23] Thanks to the homogeneity assumption; however, we have $D_2 \simeq D_A$.

Relation (9) links the observed behavior (7) in Fig. 2 to the attractor dimension, showing that the limiting factor to find good analogs is the attractor dimension, which quantifies the number of "active" degrees of freedom of the system. For those accustomed to chaotic systems, this result is rather obvious as $C_{r,M}(\epsilon)$ in Eq. (6) provides a standard approximation[3] to the correlation sum, $2/(M(M-1))\sum_{i,j>i}\Theta(\epsilon - |\boldsymbol{x}_i - \boldsymbol{x}_j|)$, at the basis of the Grassberger and Procaccia method to determine the correlation dimension D_2.[24] Indeed, the correlation sum is an unbiased estimator of the probability $P_2(\epsilon)$ to find two randomly chosen points on the attractor (using a long trajectory on it) at a distance $\leq \epsilon$. For small ϵ, $P_2(\epsilon) \sim \epsilon^{D_2}$ and thus the correlation dimension can be estimated.[18]

B. Poincaré recurrence theorem and Kac's lemma

The quantity $C_M(\epsilon)$, besides approaching (for large M) the probability to find the system state ϵ-close to $\boldsymbol{x}_M$, relates to the average time interval $\bar{\tau}_R$ between two consecutive ϵ-analogs of $\boldsymbol{x}_M$, which is given by

$$\bar{\tau}_R = \frac{(M-1)\Delta t}{\mathcal{M}(\epsilon)}, \tag{10}$$

$\mathcal{M}(\epsilon)$ being the number of ϵ-analogs in the interval $[t_1 : t_{M-1}]$. As by definition $C_M(\epsilon) = \mathcal{M}(\epsilon)/(M-1)$, we have

$$C_M(\epsilon) = \Delta t/\bar{\tau}_R. \tag{11}$$

This is actually a classical result of the ergodic theory, known as Kac's lemma.

To clarify this connection, it is worth recalling the Poincaré recurrence theorem[4] stating that, in Hamiltonian systems with a bounded phase space Ω, the trajectories exiting from a generic set $\sigma \subset \Omega$ will return back to σ infinitely many times. The theorem holds for almost all points in σ except for a possible subset of zero probability. In general, it applies to the class of systems with volume-preserving dynamics in phase-space, of which Hamiltonian ones are a particular sub-class. Actually, although often not stressed in elementary courses, it can be straightforwardly extended to dissipative ergodic systems provided one only considers initial conditions on the attractor, and "zero probability" is interpreted with respect to the invariant probability on the attractor.

Poincaré's theorem merely proves that a trajectory surely returns to the neighborhood of its starting point but does not provide information about the time between two consecutive recurrences—the Poincaré recurrence time. The latter is crucial to the method of analogs, because long recurrence times critically spoil its applicability [see Eq. (11)].

To estimate the average recurrence time, let us assume that an infinitely long sequence of states can be stored. Without loss of generality, we consider a discrete time sequence $\boldsymbol{x}_k = \boldsymbol{x}(k\Delta t)$ $(k = 0, \ldots, \infty)$ of states generated by a deterministic evolution from the initial condition $\boldsymbol{x}_0$. Given a set σ including $\boldsymbol{x}_0$, the recurrence time of $\boldsymbol{x}_0$ relative to σ, $\tau_\sigma(\boldsymbol{x}_0)$ can be defined as the minimum k such that $\boldsymbol{x}_k$ is again in σ

$$\tau_\sigma(\boldsymbol{x}_0) = \inf_k \{k \geq 1 | \boldsymbol{x}_0 \in \sigma \text{ and } \boldsymbol{x}_k \in \sigma\}. \tag{12}$$

Note that we are using dimensionless times, with $\Delta t = 1$. The mean recurrence time relative to σ, $\langle \tau_\sigma \rangle$ can then be computed as

$$\langle \tau_\sigma \rangle = \frac{1}{\mu(\sigma)} \int_\sigma d\mu(\boldsymbol{x}) \tau_\sigma(\boldsymbol{x}), \tag{13}$$

μ being the invariant probability with respect to the dynamics, defined in Sec. IV A. For ergodic dynamics, a classical result known as Kac's lemma states that[25]

$$\int_\sigma d\mu(\boldsymbol{x})\, \tau_\sigma(\boldsymbol{x}) = 1, \quad \text{so that} \quad \langle \tau_\sigma \rangle = 1/\mu(\sigma). \tag{14}$$

That is, the average recurrence time to a region σ is just the inverse of the probability of that region. We stress that Eq. (14) is a straightforward consequence of ergodicity.[26]

In a system with phase-space volume preservation (those for which the Poincaré theorem is typically invoked) with N degrees of freedom, if σ is a hypercube of linear size ϵ, one has

$$\mu(\sigma) \sim \left(\frac{\epsilon}{L}\right)^N \quad \text{and} \quad \langle \tau_\sigma \rangle \sim \left(\frac{L}{\epsilon}\right)^N, \tag{15}$$

where L is the typical excursion of each component of $\boldsymbol{x}$. Thus, the mean return time grows exponentially with N. Consequently in a macroscopic body ($N \gg 1$), $\langle \tau_\sigma \rangle$ is astronomically large, for any σ. The result (15) is surely positive for the validity of statistical mechanics, as recognized by Boltzmann himself who (without knowing Kac's lemma) replied to Zermelo's criticism of irreversibility, "Of course if one waits long enough, the initial state will eventually recur, but the recurrence time is so long that there is no possibility of ever observing it."[8] But the result is dramatically negative for the possibility to find analogs in high-dimensional systems.

In the case of ergodic dissipative systems, where the coarse-grained probabilities are ruled by the dimension D_A [compare Eq. (8)], Kac's result (15) applies with N replaced by D_A.

We conclude this digression on Poincaré recurrences by noting that the limitations to find the analogs set by relation (15) are unrelated to chaos. For instance, Eq. (15) also applies to a chain of n harmonic oscillators with incommensurable frequencies, a system with regular (quasiperiodic) behavior. Strictly speaking, such a system is not ergodic in the whole angle-action phase space, but in the space of angles only. Therefore in Eq. (15), instead of $N = 2n$, one has to set $N = n$.[27,28]

C. Consequences of Kac's lemma

The above results allow us to quantify Lorenz's pessimism with respect to the number of data necessary for finding good analogs in the atmosphere.[15] Clearly, we must require $M\Delta t \gtrsim \bar{\tau}_R$, which from Eq. (11) implies $M \gtrsim 1/C_M(\epsilon)$. Then using Eq. (9), we can see that the minimum length of the time series is

$$M \sim \left(\frac{L}{\epsilon}\right)^{D_A}, \tag{16}$$

where L is the typical excursion of each component of $\boldsymbol{x}$.

Equation (16) implies that, at least in principle, the method can work for deterministic systems having an attractor of finite dimension provided that the time series is suitably long. However, the exponential dependence on D_A in Eq. (16) imposes, upon putting in numbers, constraints that are too severe even if we content ourselves of poor precision, i.e., not too small ϵ-analogs. For instance, in Fig. 3, we show how the

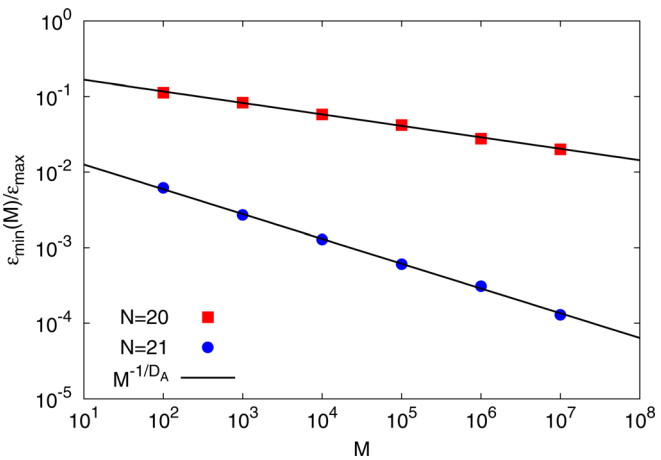

Fig. 3. (Color online) The ratio $\epsilon_{min}/\epsilon_{max}$ vs M. The parameters of the model are the same as in Fig. 2: $F = 5$, $N = 20$, and $N = 21$; the reference states are $r = 1000$. The solid lines are the fits of the data by means of relation (16).

distance between a reference point and its best analog ($\epsilon_{\min}$) scales with M. We see that for $\epsilon_{\min}/\epsilon_{\max} = 10^{-2}$, a sequence of 10^2 points is sufficiently long in the case $N=21$ ($D_A \approx 3.1$) while, on the contrary, even 10^7 points are not yet enough in the case $N=20$ ($D_A \approx 6.6$). Indeed, by inverting Eq. (16), we should expect $\epsilon_{\min} \propto M^{-1/D_A}$, as shown in Fig. 3. The differences between the cases $N=21$ and $N=20$ in Figs. 2 and 3 are thus a mere consequence of the different attractor dimensionality, namely $D_A(N=21) < D_A(N=20)$.

Relation (16) also lays the basis for understanding the limits of the Grassberger and Procaccia method[24] to compute the correlation dimension from the scaling behavior of the correlation sum or its approximation (6). In fact, it states that the larger the dimension of the attractor, the larger the number of points M necessary to sample it within a given accuracy ϵ. For example, a segment of size L will require $M \simeq L/\epsilon$ points, a square will require $M \simeq (L/\epsilon)^2$, and so on. Smith[29] proposed a minimum number of points of $M \sim 42^{D_A}$ (about a decade and a half of scaling region) to get reliable results. For $D_A = 5$ or 6, Smith's recipe requires from hundreds of millions to billions of data points, too large for standard experiments. The above considerations on the limits of applicability of the Grassberger and Procaccia[24] technique may sound trivial. However, in the 1980s, when nonlinear time series analysis started to be massively employed in experimental data analysis, the limitations due the length of the time series were overlooked by many researchers and a number of misleading papers appeared even in important journals (for a critical review see Ref. 9).

In conclusion, the possibility to predict the future from the past using analogs has its practical validity only for low-dimensional systems. More than one century after Maxwell, scientists working on prediction problems rediscovered his warning: "same antecedents never again concur, and nothing ever happens twice," whenever the system's dimension is moderately high.

D. Remarks on the case of unknown phase space

So far, we have assumed that the vector $\boldsymbol{x}$ determining the state of the system is known and can be measured with arbitrary precision. The real situation is less simple: Usually, we do not know the whole set of variables (not even their number) that define the state of a system. Moreover, even knowing them, in experimental measurements, we normally have access only to very few scalar observables u_t depending on the state of the system: $u_t = G[\boldsymbol{x}_t]$. In these cases, there exists a powerful technique (based on Takens' delay embedding theorem[10]) that can reconstruct the phase space, providing a rigorous foundation for the use of analogs.[30] Beyond the technical (often nontrivial) aspects, the main limit of the method, namely the exponential increase of M with D_A, still remains. Moreover, in practical implementations, the presence of unavoidable measurement errors introduces a further source of complications. Ways to deal with the general case of phase-space reconstruction and measurement errors have been developed, but their discussion is beyond the scope of this paper, so we refer the reader to specialized monographs.[2,3]

V. TWO EXAMPLES WHERE THE METHOD OF ANALOGS WORKS

Chaotic low dimensional attractors ($D_A \approx 2$–4) may occur in many physical systems such as electric circuits, lasers, fluid motion, etc.; see Ref. 3. Other natural phenomena, such as weather, are instead characterized by high dimensional attractors with D_A proportional to the total number of variables involved, which is huge in the case of the atmosphere. Thus, the conclusions of Sec. IV are very pessimistic: When D_A is that large, only mediocre analogs (rather large ϵ) can be found and those are, from the point of view of predictability, usually not so informative about the future evolution of the system.[15]

It is instructive, however, to consider here two exceptions to this rule. One is a variation on the theme of the Lorenz-96 model (4). We will then briefly discuss tidal predictions, which represent, to the best of our knowledge, one of the few instances in which the idea of using the past to predict the future works and has important practical applications.

A. Systems with multiscale structure

We consider here systems with a multiscale structure, where the vector state $\boldsymbol{x}$ can be decomposed into a slow component $\boldsymbol{X}$ which is also the "largest" one, and a fast component $\boldsymbol{y}$ that is "small" with respect to $\boldsymbol{X}$ (i.e., $y_{\rm rms} \ll X_{\rm rms}$). If the slow components can be described in terms of an "effective number" of degrees of freedom much smaller than those necessary to characterize the whole dynamics, mediocre (referred to the whole system) analogs can be used to forecast at least the slower evolving component.

As an illustration of such a system, we consider a variant of the model (4) introduced by Lorenz himself[21] to discuss the predictability problem in the atmosphere, where indeed a multiscale structure is present. The model reads

$$\frac{dX_n}{dt} = X_{n-1}(X_{n+1} - X_{n-2}) - X_n + F - \frac{hc}{b}\sum_{k=1}^{K} y_{k,n}, \qquad (17)$$

$$\frac{dy_{k,n}}{dt} = cby_{k+1,n}(y_{k-1,n} - y_{k+2,n}) - cy_{k,n} + \frac{hc}{b}X_n, \qquad (18)$$

where $n = 1, \ldots, N$ and $k = 1, \ldots, K$, with boundary conditions $X_{N+n} = X_{+n}$, $y_{K+1,n} = y_{1,n+1}$, and $y_{0,n} = y_{K,n-1}$. Equation (17) is essentially (4) except for the last term which couples $\boldsymbol{X}$ to $\boldsymbol{y}$. The variables $\boldsymbol{y}$ evolve with a similar dynamics but are c times faster and b times smaller in amplitude. The parameter h, set to 1, controls the coupling strength.

We repeat the computation to measure the probability of ϵ-analogs for the dynamics of this system, by assuming that the whole state of the system $\boldsymbol{x}(t) = (\boldsymbol{X}(t), \boldsymbol{y}(t))$ is accessible, and by ignoring which are the slow and fast variables, so that we must search for the analogs in the sequence of states $\boldsymbol{x}_k = (\boldsymbol{X}(t_k), \boldsymbol{y}(t_k))$, with $t_k = k\Delta t$. Figure 4 shows $C_{r,M}(\epsilon)$ as a function of $\epsilon/\epsilon_{\max}$ for a long sequence, $M = 10^7$, for fixed time scale separation $c=10$ and taking the fast component $\boldsymbol{y}$ to be respectively $b=20$, 50, and 100 times smaller than the slow one $\boldsymbol{X}$. The phase-space dimensionality is 50, with $N=5$ slow and $K=10$ fast degrees of freedom. The attractor dimension of the whole system D_A, given by the scaling $C(\epsilon) \sim \epsilon^{D_A}$ at very small ϵ, is rather large ($D_A \approx 10$). However, for $\epsilon/\epsilon_{\max} > O(1/b)$, we see a second power law $C(\epsilon) \sim \epsilon^{D_A^{\rm eff}}$ with $D_A^{\rm eff} \approx 3 < D_A$, which defines a sort of effective dimension at large scale.

Therefore, if we are interested in predicting the slow evolving component of the system, provided it is described by a relatively low number of effective degrees of freedom, as here, we can exploit the mediocre analogs (i.e., the

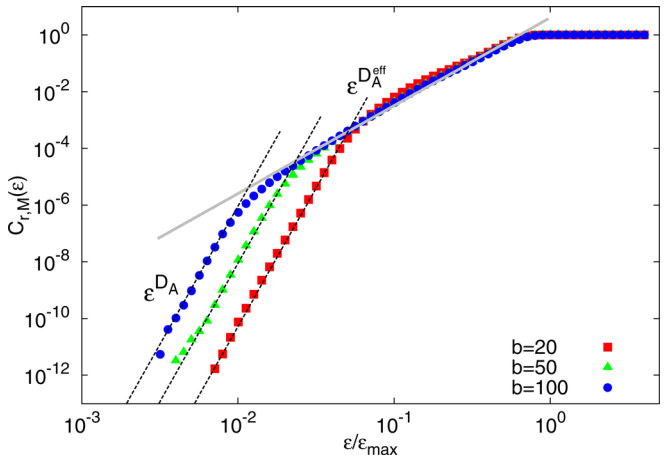

Fig. 4. (Color online) $C_{r,M}(\epsilon)$ vs $\epsilon/\epsilon_{\max}$ for the model of Eqs. (17) and (18), computed for three scale separations b (as labeled), holding the other parameters fixed at $h=1$, $c=10$, $F=10$, $N=5$, and $K=10$. The gray straight line has slope ≈ 3.1 while the dashed lines all have the same slope ≈ 9.8. Expression (6) has been computed with $r=10^3$ and $M=10^7$.

ϵ-analogs with $\epsilon/\epsilon_{\max} > O(1/b)$). Moreover, with reference to Eq. (2), it is reasonable to expect that the prediction error related to mediocre analogs grows as $\sim \epsilon e^{\lambda(\epsilon)T}$, where $\lambda(\epsilon)$ can be much smaller than the Lyapunov exponent λ_1 (indeed, as shown in Ref. 31, $\lambda(\epsilon) \approx \lambda_1/c$). This implies that slow variables can be predicted over a longer term than the whole state of the system, as already realized by Lorenz.[21] In general multiscale systems, increasing ϵ amounts to performing a coarse-graining on the system description, which implies the "elimination" of the fastest degrees of freedom, associated to the smallest scales. Consequently, coarse-graining reduces the number of effective degrees of freedom ($D_A^{\text{eff}}(\epsilon) < D_A$) and the error growth rate ($\lambda(\epsilon) < \lambda_1$).

The previous example is somehow the simplest multiscale system, i.e., $C(\epsilon)$ vs ϵ shows only two logarithmic slopes, D_A^{eff} and D_A. More generally one can have a logarithmic slope $D(\epsilon)$ with a series of plateaus: $D(\epsilon) \approx D_1^{\text{eff}}$ for $\epsilon \in [\epsilon_0, \epsilon_1]$, $D(\epsilon) \approx D_2^{\text{eff}} > D_1^{\text{eff}}$ for $\epsilon \in [\epsilon_1, \epsilon_2]$, and so on ($\epsilon_0 > \epsilon_1 > \epsilon_2 \ldots$). The interested reader may reproduce such a behavior by computing the correlation integral of the discrete-time system discussed in Ref. 32.

B. Tidal prediction from past history

Tidal prediction is a problem of obvious importance for navigation. The appropriate governing equations were established long ago by Laplace. It is necessary to study the water level, with suitable boundary conditions, under the gravitational forcing of the Moon, the Sun, and the Earth.[33] Due to the practical difficulties in the treatment of boundary conditions (only partially known and hard to solve numerically), even with powerful computers the fundamental equations cannot be directly used for tide forecasting.

However, and remarkably, already in the first half of the 19th century, there existed efficient empirical methods to compile numerical tables of tides in any location where a record of past tides was known.[34] As recognized by Laplace, a great simplification comes from the periodicity of the forcing (related to the motions of celestial bodies), which naturally suggests treating tides in terms of Fourier series, whose frequencies are known from celestial mechanics. Lord Kelvin and George Darwin (Charles' son) showed that water levels can be well predicted by a limited number of harmonics (say 10 or 20), determining the Fourier coefficients from the past time data at the location of interest. To make the numerical computations automatic, minimizing the possibility of error, Kelvin and Darwin built a tide-predicting machine: a special-purpose mechanical computer made of gears and pulleys. Tide-predicting machines were in use until half century ago, when they were replaced by digital computers to compute the Fourier series.[35]

Since tides are chaotic, it is natural to wonder why their prediction from past records is a relatively easy task. The reason is the low number of effective degrees of freedom involved. In a detailed description of tides, small-scale phenomena are also involved, with very short characteristic times, e.g., micro-turbulence; therefore the "true" D_A is surely very large, together with λ_1. Hence, the success of tidal prediction is mainly a consequence of the multiscale character of the system, which has a small D^{eff} (and also a small $\lambda(\epsilon)$) on the interesting not-too-small scales, in a way similar to the multiscale model of Sec, V A. Indeed, quite recently, investigations[36] of tidal time series by using the standard method of nonlinear time series analysis (such as embedding; see Sec. IV D) found effective attractor dimensions quite low (between 3 and 4) with effective Lyapunov exponents of the order of 5 days^{-1}. That explains *a posteriori* the success of the empirical method. Thanks to the low D^{eff}, analogs can be found. Moreover, to forecast tides a few hours in advance, the relatively low value of the Lyapunov exponent makes the predictability time long enough for practical purposes. Of course, quantitative details (the precise values of D^{eff} and of $\lambda(\epsilon)$) depend on the locations,[37] but for the method to work, the very important aspect is the limited value of the effective attractor dimension.

VI. CONCLUSIONS

It is a common belief that chaos is the main limiting factor to predictability in deterministic systems. This is correct as long as the evolution laws of the system under consideration are known. However, if the information on the system evolution is based only on observational data, the bottleneck lies in Poincaré recurrences which, in turn, depend on the number of effective degrees of freedom involved. Indeed, even in the most optimistic conditions, if the state vector of the system were known with arbitrary precision, the amount of data necessary to make the meaningful predictions would grow exponentially with the effective number of degrees of freedom, independently of the presence of chaos. However, when, as for tidal predictions, the number of degrees of freedom associated with the scales of interest is relatively small, the future can be successfully predicted from past history.

We stress that the necessity of an exponentially large (with D_A) amount of data constitutes a genuine intrinsic difficulty of every analysis based on time series without any guess on the underlying dynamics. Such a difficulty is not a peculiarity of the method of analogs but is inherent to all methods based on the occurrence frequency of sequences of states to estimate the average of observables. In other words, the problem arises whenever one needs to collect enough recurrences. This obstacle may be partially overcome by suitable information-theoretic techniques (see, e.g., Ref. 38) allowing for optimized reconstructions of the dynamics, whose dimensionality, however, increases with the required accuracy. These conclusions are further supported by a recent work by Cubitt and

coworkers,[39] showing that the reconstruction of dynamical equations from data is a computationally NP-hard problem, because the needed observation time scales exponentially with the number of degrees of freedom.

In general, the best strategy for meaningful prediction is that envisaged by Richardson, as a clever compromise between modeling and data analysis. In this regard, we would like to conclude by mentioning that, in the era of information technology, the enormous capacity of data storage, acquisition, and elaboration may encourage some people to believe that meaningful predictions can be extracted merely from data. For example, recently *Wired Magazine* provocatively titled an article "The end of theory: The data deluge makes the scientific method obsolete,"[40] asserting that nowadays, with the availability of massive data, the traditional way science progresses by hypothesizing, modeling, and testing is becoming obsolete. In this respect, we believe that, while it is undeniable that the enormous amount of data opens new opportunities, the role of modeling cannot be discounted. When the number of effective degrees of freedom underlying a dynamical process is even moderately large, predictions based solely on observational data soon become problematic, as in the case of weather forecasting.

ACKNOWLEDGMENTS

The authors thank U. Marini Bettolo Marconi for a careful reading of the manuscript. They are grateful to an anonymous referee for useful remarks and suggestions.

[a]Electronic mail: massimo.falcioni@roma1.infn.it

[1]*Time Series Prediction: Forecasting the Future and Understanding the Past*, edited by A. S. Weigend and N. A. Gershenfeld (Addison-Wesley, Reading, 1994).
[2]H. D. I. Abarbanel, *Analysis of Observed Chaotic Data* (Springer, New York, 1996).
[3]H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge U.P., Cambridge, 1997).
[4]H. Poincaré, "Sur le problème des trois corps et les équations de la dynamique," Acta Math. **13**, 1–270 (1890).
[5]C. Nicolis, "Atmospheric analogs and recurrence time statistics: Toward a dynamical formulation," J. Atmos. Sci. **55**, 465–475 (1998).
[6]E. G. Altmann and H. Kantz, "Recurrence time analysis, long-term correlations, and extreme events," Phy. Rev. E **71**, 056106 (2005).
[7]J. Gleick, *Chaos: Making a New Science* (Viking, New York, 1987).
[8]C. Cercignani, *Ludwig Boltzmann, The Man Who Trusted Atoms* (Oxford U.P., Oxford, 1988).
[9]D. Ruelle, "Deterministic chaos: The science and the fiction," Proc. R. Soc. London, Ser. A **427**, 241–248 (1989).
[10]F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence*, edited by D. A. Rand and L. S. Young (Springer, Berlin, 1981), pp. 366–381.
[11]The Qohelet's Book 1:9, NIV.
[12]L. Campbell and W. Garnett, *The Life of James Clerk Maxwell* (Macmillan, London, 1882).
[13]P. Lynch, *The Emergence of Numerical Weather Prediction: Richardson's Dream* (Cambridge U.P., Cambridge, 2006).
[14]L. F. Richardson, *Weather Prediction by Numerical Process* (Cambridge U.P., Cambridge, 1922).
[15]E. N. Lorenz, "Atmospheric predictability as revealed by naturally occurring analogues," J. Atmos. Sci. **26**, 636–646 (1969).
[16]E. N. Lorenz, "Three approaches to atmospheric predictability," Bull. Am. Meteorol. Soc. **50**, 345–349 (1969).
[17]E. N. Lorenz, "Deterministic nonperiodic flow," J. Atmos. Sci. **20**, 130–141 (1963).
[18]M. Cencini, F. Cecconi, and A. Vulpiani, *Chaos: From Simple Models to Complex Systems* (World Scientific Publishing, Singapore, 2009).
[19]S. De Souza-Machado, R. W. Rollins, D. T. Jacobs, and J. L. Hartman, "Studying chaotic systems using microcomputer simulations and Lyapunov exponents," Am. J. Phys. **58**, 321–329 (1990).
[20]A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining Lyapunov exponents from a time series," Physica D **16**, 285–317 (1985).
[21]E. N. Lorenz, "Predictability—A problem partly solved," in *Proc. Seminar on Predictability* (ECMWF, Reading, UK, 1996), pp. 1–18.
[22]A. Karimi and M. R. Paul, "Extensive chaos in the Lorenz-96 model," Chaos **20**, 043105 (2010).
[23]G. Paladin and A. Vulpiani, "Anomalous scaling laws in multifractal objects," Phys. Rep. **156**, 147–225 (1987).
[24]P. Grassberger and I. Procaccia, "Characterization of strange attractors," Phys. Rev. Lett. **50**, 346–349 (1983).
[25]M. Kac, "On the notion of recurrence in discrete stochastic processes," Bull. Am. Math. Soc. **53**, 1002–1010 (1947).
[26]We sketch the derivation in the simple case of a discrete-state system as obtained by partitioning the phase-space of the given system into a discrete number of regions $\{\rho_j\}$, each one defining a coarse-grained state j, such that the probability p_j of a state j is given by $p_j = \mu(\rho_j)$. We can focus on a state s. Assume that during the time interval $\mathcal{T}$, a system trajectory has visited $\mathcal{N}_s$ times the sth state (representing here the set σ). Denoting with $\tau_s(i)$ the time taken by the ith return to the state s, the mean recurrence time is $\sum_{i=1}^{\mathcal{N}_s} \tau_s(i)/\mathcal{N}_s = \mathcal{T}/\mathcal{N}_s$, that is, the inverse of the fraction of time spent by the system in the sth state, $f_s = \mathcal{N}_s/\mathcal{T}$. Invoking ergodicity, in the limit $\mathcal{T} \to \infty$, we have $\mathcal{T}/\mathcal{N}_s \to \langle \tau_s \rangle$ and $f_s \to p_s$, where p_s is the probability of the sth state (i.e., the probability $\mu(\sigma)$) so that $\langle \tau_s \rangle = 1/p_s$, which represents Eq. (14). Replacing $\mathcal{T}$ and $\mathcal{N}_s$ with $M-1$ and $\mathcal{M}(\epsilon)$, respectively, we obtain Eqs. (10) and (11), with $\Delta t = 1$, so that for large M, Eq. (11) gives again Kac's result.
[27]P. C. Hemmer, L. C. Maximon, and H. Wergeland, "Recurrence time of a dynamical system," Phys. Rev. **111**, 689–694 (1958).
[28]M. Kac, "Remark on recurrence times," Phys. Rev. **115**, 1 (1959).
[29]L. A. Smith, "Intrinsic limits on dimension calculations," Phys. Lett. A **133**, 283–288 (1988).
[30]E. J. Kostelich and D. P. Lathrop, "Time series prediction by using the method of analogues," in *Time Series Prediction: Forecasting the Future and Understanding the Past*, edited by A. S. Weigend and N. A. Gershenfeld (Addison-Wesley, Reading, 1994), pp. 283–295.
[31]G. Boffetta, M. Cencini, M. Falcioni, and A. Vulpiani, "Predictability: A way to characterize complexity," Phys. Rep. **356**, 367–474 (2002).
[32]E. Olbrich, R. Hegger, and H. Kantz, "Analysing local observations of weakly coupled maps," Phys. Lett. A **244**, 538–544 (1998).
[33]W. H. Munk and D. E. Cartwright, "Tidal spectroscopy and prediction," Philos. Trans. R. Soc. London, Ser. A **259**, 533–581 (1966).
[34]S. Ducheyne, "Whewell's tidal researches: Scientific practice and philosophical methodology," Stud. Hist. Philos. Sci., Part A **41**, 26–40 (2010).
[35]T. Ehret, "Old brass brains—Mechanical prediction of tides," ACSM Bull. **233**, 41–44 (2008).
[36]V. Khokhlov, A. Glushkov, N. Loboda, N. Serbov, and K. Zhurbenko, "Signatures of low-dimensional chaos in hourly water level measurements at coastal site of Mariupol, Ukraine," Stochastic Environ. Res. Risk Assess. **22**, 777–787 (2008).
[37]T. W. Frison, H. D. I. Abarbanel, M. D. Earle, J. R. Schultz, and W. D. Scherer, "Chaos and predictability in ocean water levels," J. Geophys. Res. **104**, 7935–7951, doi:10.1029/1998JC900104 (1999).
[38]F. Creutzig, A. Globerson, and N. Tishby, "Past-future information bottleneck in dynamical systems," Phys. Rev. E **79**, 041925 (2009).
[39]T. S. Cubitt, J. Eisert, and M. M. Wolf, "Extracting dynamical equations from experimental data is NP hard," Phys. Rev. Lett. **108**, 120503 (2012).
[40]C. Anderson, "The end of theory: The data deluge makes the scientific method obsolete," Wired Magazine 16.07, <http://www.wired.com/science/discoveries/magazine/16-07/pb_intr>.