Resampling Chaotic Time Series

Silvia Golia*

Dipartimento di Scienze Statistiche, Università di Perugia, Via Pascoli, C.P. 1315/succ.1, 06100 Perugia, Italy

Marco Sandri[†]

Centro di Informatica e Calcolo Automatico, Università di Verona, Via dell' Artigliere 19, 37129 Verona, Italy

A method for resampling time series generated by a deterministic chaotic data generating process (DGP) is proposed. Given an observed time series, this method potentially allows one to obtain an arbitrary number of time series of arbitrary length which can be considered as a product of the same unknown DGP. The notion of shadowing and brittleness of the pseudo-orbit proves to be particularly useful in characterizing the conditions for a correct resampling. A simple practical application of the method is shown. [S0031-9007(97)03296-1]

PACS numbers: 05.45.+b

One of the main tasks of time series analysis is to infer from a given time series the properties of the underlying data generating process, as, for example, nonlinearity, complexity, chaoticity, etc. There are many works on this subject. Among these, we must cite Theiler *et al.* [1] who developed a statistical method to test if the underlying dynamics is of a particular type, for example, white noise or linearly correlated noise. Brock *et al.* [2] suggest a method to test the hypothesis of independence and identical distribution, and to detect the presence of nonlinearities in the observed data. Wolf *et al.* [3] and many other authors [4] proposed algorithms for the estimation of the largest Lyapunov exponent (the main test for chaoticity) or the whole spectrum from a time series. The algorithm for the estimation of the correlation dimension [5] enables one to obtain a lower bound estimation of the Hausdorff dimension of the attractor of the unknown data generating process (DGP).

In many of these works, however, there is a lack of a distributional theory for the proposed tests, which would provide the necessary framework for the statistical hypothesis testing. From an analytical point of view, few results have been obtained so far; one example is Lai and Chen [6]. A promising direction of research is the one based on resampling techniques. Following Simon and Bruce [7], by resampling procedures we mean *"... a collection of computer intensive methods to use experimental data to obtain different estimates of a statistics or underlying distribution."* Among the resampling methods for time series, we mention Künsch [8], Politis and Romano [9], and Lall and Sharma [10]. However, these are particularly ill suited for resampling time series generated by a purely deterministic DGP, as a chaotic DGP, or with a strong deterministic component; as a matter of fact, they tend to introduce a spurious stochastic component in the resampled time series. Gencay's work [11] is the most interesting because it circumvents the obstacle by boot-strapping the points in the phase space rather than in the time domain. However, this method can only provide the empirical distribution of Lyapunov exponent estimators.

The method analyzed here, which was inspired by an idea contained in the work of Farmer and Sidorowich [12], is substantially different from the ones proposed so far and, in some sense, is more general because it potentially allows us to obtain the empirical distribution of a wider range of estimators such as, for example, correlation dimension, entropies, and Lyapunov exponents. The method is essentially a parametric resampling technique based on local (neural) approximations of the flow (or the map) which generated the observed data set. The only random factor introduced here is the choice of a starting point on the attractor. The resulting resampled series mimic (in a sense which will be specified later) the deterministic behavior of the underlying DGP. First, we present some notions which will be used later.

An m-dimensional differentiable dynamical system is a time evolution defined by an evolution equation (continuous-time case) $\dot{\mathbf{x}} = F(\mathbf{x})$, $\mathbf{x} \in \mathbf{R}^m$, which generates a smooth function $f^t(\mathbf{x})$, the *flow*, such that $\frac{d}{dt}(f^t(\mathbf{x}))_{t=\tau} = F(f^\tau(\mathbf{x}))$ for all τ, or a *map* (discrete-time case) $\mathbf{x}_{t+1} = f(\mathbf{x}_t)$, $\mathbf{x} \in \mathbf{R}^m$, where f or F are differentiable functions.

The quantities most frequently used to characterize the dynamics of the system are the correlation dimension and the Lyapunov exponents. They measure, respectively, the degree of geometrical complexity of the attractor and the degree of dynamical instability of the trajectories.

The *correlation dimension*, d_{corr}, is defined by

$$d_{\text{corr}} = \lim_{\varepsilon \to 0} \frac{\log C(\varepsilon)}{\log \varepsilon}, \qquad (1)$$

where $C(\varepsilon) = \lim_{N \to \infty} C_N(\varepsilon)/N^2$, N is the number of available points, and $C_N(\varepsilon)$ is the number of pairs of points on the attractor whose distance from one another is less that ε. It gives information on the minimum number of

variables present in the evolution of the corresponding dynamical system [5].

The Lyapunov exponents λ_i are measures of the average rate of divergence (or convergence) of typical trajectories in time. The largest Lyapunov exponent is given by

$$\lambda(\mathbf{x}, \mathbf{w}) = \lim_{t \to \infty} \frac{1}{t} \ln \|D_{\mathbf{x}} f^t(\mathbf{x}) \cdot \mathbf{w}\| \qquad (2)$$

for almost any vector $\mathbf{w}$, where $D_{\mathbf{x}} f^t(\mathbf{x})$ is the derivative of f^t in $\mathbf{x}$. The multiplicative ergodic theorem of Oseledec [13] ensures that (2) exists. The other exponents can in principle be obtained by diagonalizing the positive matrices $D_{\mathbf{x}}^* f^t D_{\mathbf{x}} f^t$, where $D_{\mathbf{x}}^* f^t$ is the adjoint of $D_{\mathbf{x}} f^t$, and using the fact that their eigenvalues behave the same as $\exp^{2t\lambda_1}, \exp^{2t\lambda_2}, \ldots$. If the largest Lyapunov exponent is positive, then this is a strong indication that the system is chaotic.

Finally, we introduce the notion of a *feed-forward neural network* [14]. Neural networks are a class of nonlinear models which are capable of approximating any continuous function f uniformly on compacta, that is, they are universal approximators. This property makes neural models very appealing in nonlinear regression.

The input values $\mathbf{x} \in \mathbf{R}^d$ (d is the embedding dimension) are received by the d input units, which pass them to the hidden units using a linear transformation determined by their connection strength γ_{ij}. Each hidden unit performs a nonlinear transformation on its total input, producing a total output using a sigmoid function $K(x)$, called activation function, which is the same for all the hidden units. The network output O can be represented as

$$O = \sum_{i=1}^{q} \beta_i K\left(\sum_{j=1}^{d} \gamma_{ij} x_j + b_i \right), \qquad (3)$$

where β_i, with $\beta_i \in \mathbf{R}^q$, are the weights of the output produced by every hidden unit, q is the number of hidden units, and b_i is the bias of the hidden units.

Now, let $\{x_t\}_{t=1}^n$ be a scalar time series arising from a deterministic dynamical system (for example, chaotic) and let $\{\mathbf{x}_t\}_{t=1}^N$ be the corresponding series of reconstructed vectors of $\mathbf{R}^d$, where $N = n - d + 1$. The aim of the present Letter is to develop an algorithm which is able to produce, using the information contained in the data set, an arbitrary number M of new series which can be considered realizations of the same DGP.

In the case where the flow (or the map) f is known, the natural way for resampling the series is to randomly select M initial conditions $\mathbf{x}_{0j}$, $j = 1, \ldots, M$, belonging to the basin of attraction of the system, compute the N iterations $X_j = \{f^i(\mathbf{x}_{0j})\}_{i=1}^N$, where $f^i = f \circ \cdots \circ f$ (i times), and discard the first n_T values which represent the transient. Therefore, the only stochastic component of this algorithm is the choice of the initial condition from which the orbit starts.

In the real case, the flow (or the map) f is obviously unknown. It is therefore necessary to (locally or globally) estimate f, fitting linear [12] or nonlinear models (such as

feed-forward neural networks, radial basis functions, recurrent neural networks, etc. [15]) to the $\{\mathbf{x}_t\}_{t=1}^N$ data set. Let $\hat{f}_{\mathbf{x}}$ be the estimate of f at the point $\mathbf{x}$. Of course, if $\hat{f}$ is a global estimate, $\hat{f}_{\mathbf{x}}(\cdot) = \hat{f}_{\mathbf{y}}(\cdot)$ for all $\mathbf{x}, \mathbf{y} \in \{\mathbf{x}_t\}_{t=1}^N$. The jth resampled orbit starting from (the randomly chosen) $\mathbf{x}_{0j}$ is $\hat{X}_j = \{\hat{\mathbf{x}}_t(\mathbf{x}_{0j})\}_{t=1}^N$, where $\hat{\mathbf{x}}_t(\mathbf{x}_{0j}) \equiv \hat{f}_{\hat{\mathbf{x}}_{t-1}}(\hat{\mathbf{x}}_{t-1}) = \hat{f}_{\hat{\mathbf{x}}_{t-1}} \circ \hat{f}_{\hat{\mathbf{x}}_{t-2}} \circ \cdots \circ \hat{f}_{\mathbf{x}_{0j}}(\mathbf{x}_{0j})$, and the first n_T (transient) iterations have been discarded.

Here, we propose using a local estimator $\hat{f}_{\mathbf{x}}$ based on the nearest neighbors technique [12] and feed-forward neural networks. The proposed resampling methodology can be summarized as follows:

(1) Starting from the observed time series, use the time delay method to reconstruct the points in the phase space, obtaining the points $\mathbf{x}_t \in \mathbf{R}^d$ defined as $\mathbf{x}_t = \{x_t, x_{t+j}, x_{t+2j}, \ldots, x_{t+(d-1)j}\}$, where d is the embedding dimension for the dynamical system under study and j is the time delay.

(2) Randomly choose an initial condition $\mathbf{x}_0$, keeping, for example, one of the N points of the orbit and adding a small random quantity η.

(3) Look for the k nearest points to $\mathbf{x}_0$, $\{\mathbf{x}_{n_i}\}_{i=1}^k$, i.e., find the k points which minimize the distance (here we consider the Euclidean distance) from the reference point $\mathbf{x}_0$; this set of points represents the total input of the neural network.

(4) Choose a time interval T and select the points $\{\mathbf{x}_{n_i+T}\}_{i=1}^k$, i.e., find the k points in the phase space which correspond to the evolution, after T steps, of the dynamical system starting from the set of points $\{\mathbf{x}_{n_i}\}_{i=1}^k$ selected in step (3); this set of points is the total output of the neural network.

(5) Making use of these two sets of points in $\mathbf{R}^d$, estimate the flow f fitting a neural model $\hat{f}_{\mathbf{x}_0}$, as explained before.

(6) Use $\hat{f}_{\mathbf{x}_0}$ to calculate $\hat{\mathbf{x}}_1 \equiv \hat{\mathbf{x}}_{0+T} = \hat{f}_{\mathbf{x}_0}(\mathbf{x}_0)$; this is the first point of the new orbit which is going to be constructed.

(7) Go back to (3), look for the k nearest points of $\mathbf{x}_1$, estimate $\hat{f}_{\hat{\mathbf{x}}_1}$, and then calculate $\hat{\mathbf{x}}_2 \equiv \hat{\mathbf{x}}_{0+2T} = \hat{f}_{\hat{\mathbf{x}}_1}(\hat{\mathbf{x}}_1)$.

The parameters which appear in the algorithm are the embedding dimension d, the time delay j, and the number of nearest neighbors k. Regarding the choice of the embedding dimension d, one can use, for example, the method of false nearest neighbors [16]; one has to remember that choosing a wrong embedding dimension causes a wrong reconstruction of the phase space. In fact, if d is too small, there are intersection problems for the orbits, so points which seem near are in reality very far; if d is too big, this reduces the density of the usable points. Also, the choice of j, the time delay, strongly influences the reconstruction of the phase space [17]. Finally, one has to determine the best k: The choice of too small a k involves difficulties in the estimation of f with the neural network; in fact, few points of input and output are used, so the estimation becomes more approximate and rough. However, too big a k introduces an error in the estimation of

f caused by having to consider points which are actually very far from the reference one. For the best choice of k, based on cross-validation criteria, see Casdagli [18].

In order to legitimately regard the trajectories $\hat{X}_j$ as coming from the same process which generates the original data set, two fundamental questions must be considered. First, it is important to point out that it is erroneous to state (see Farmer and Sidorowich [12], p. 316) that each resampled series starting from a given initial point $\mathbf{x}_{0j}$ must necessarily be a "good" forecast of the true orbit starting from the same point. This is a particularly strong condition, especially for chaotic systems where good forecasts are only possible for short time periods. What must be guaranteed instead is the existence of a trajectory $X = \{\mathbf{x}_t(\mathbf{x}_{0h})\}$ of the true unknown system ε close to the resampled one, i.e., $|\mathbf{x}_t(\mathbf{x}_{0h}) - \hat{\mathbf{x}}_t(\mathbf{x}_{0j})| < \varepsilon$ for any $t = 1, 2, \ldots, N$. In other words, we must prove that each trajectory $\hat{X}_j$ is ε *shadowed* by a (unknown) true orbit [19]. In the lucky cases where f has the shadowing property, the existence of a true orbit ε close to the resampled one, is ensured. In fact, f has the *shadowing property* if, for every $\varepsilon > 0$, there is a $\delta > 0$ such that every δ pseudo-orbit can be ε shadowed by an actual orbit, i.e., if $\{\mathbf{x}_0, \mathbf{x}_1, \dot{s}\}$ satisfies $d(f(\mathbf{x}_t), \mathbf{x}_{t+1}) \leq \delta$ for every $t \geq 0$, then there is an initial condition $\mathbf{x}$ such that $d(f^t(\mathbf{x}), \mathbf{x}_t) \leq \varepsilon$ for all $t \geq 0$. Actually, the existence of the shadowing property has been proved only for hyperbolic dynamical systems (the shadow lemma by Anosov and Bowen) and for a few simple maps such as tent maps [20].

In the cases where it is not possible to ascertain if f has the shadowing property (particularly when f is unknown), one could apply the shadowability test proposed by Dawson *et al.* [21]. This test is based on the notion of the *brittleness* of a pseudotrajectory, that is, the constant of proportionality between the shadowing distance (the distance in the phase space that the pseudo-orbit has to cover in order for it to be deformed into a true trajectory) and the original magnitude δ of the one-step error (in our case, the approximation error). A necessary condition for shadowability is that the brittleness times the error magnitude is smaller than the extent of the attractor.

Supposing that the problem of shadowability has been solved, a second fundamental question must be taken into account: What is the amount of noise present in the resampled series? In fact, for our purposes it is not sufficient to verify the existence of an ε-shadowing trajectory. It is also necessary to guarantee that ε is sufficiently small. If not, the resampled series cannot be considered as generated from the same true unknown DGP, and their statistical properties will differ from the original ones. In this case, the resampling algorithm will have no practical utility. Obviously, the magnitude of the error of the pseudotrajectory depends on the level of the error δ (which, in turn, depends on the approximation method used for estimating f and on the sample size n)
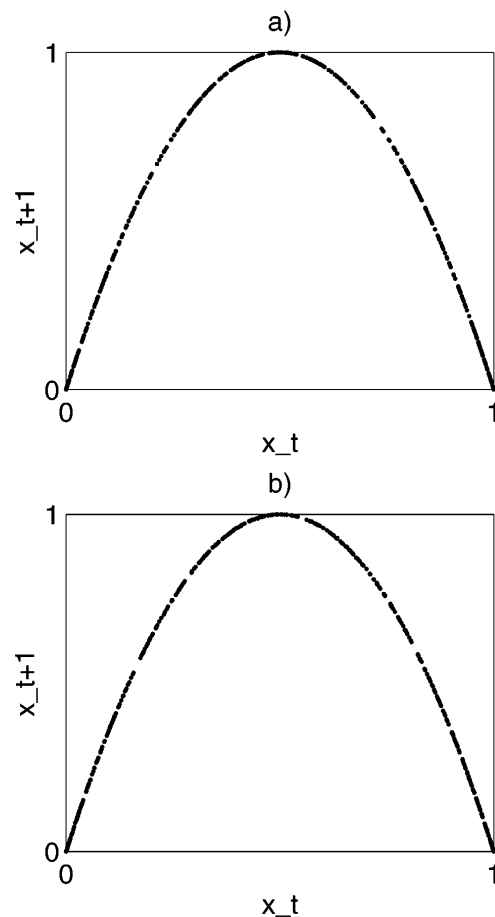

FIG. 1. Logistic map: (a) a 2D reconstruction of a true time series and (b) of a resampled one.

and on the "degree" of brittleness. The product of δ and the brittleness gives an estimate of the noise ε which is present in the resampled series [21].

We apply our resampling method to a series (500 points) generated by the logistic map $x_{t+1} = 4x_t(1 - x_t)$. The parameter values involved in our algorithm, in this specific case, are $d = 2$, $j = 1$, and $k = 20$. d and k are chosen so that the forecasting error (calculated on a testing set) of the local neural model is minimized. We show that the resampled series have the same dynamical and geometrical properties of the original DGP.

For the logistic map, the Lyapunov exponent is $\lambda_1 = \ln 2 = 0.69315$ and the correlation dimension is $d_{\mathrm{corr}} = 1$. First of all, we present the attractor reconstructed using two time series, the first coming from the "true" DGP and the second obtained from resampling the first, using the method proposed in this Letter. We can see that there are

TABLE I. Descriptive statistics of the 100 resampled time series.

	Median	Mean	Std. dev.
Lyapunov exponent	0.6942	0.6924	0.01844
Correlation dimension	0.9704	0.9621	0.05248

TABLE II. p^- values for the tests of equal mean, variance, and distribution.

	Mean	Variance	Distrib.
Lyapunov exponent	0.42	0.18	0.28
Correlation dimension	0.44	0.57	0.47

no remarkable discrepancies between the two figures (Fig. 1).

Then, we produced 100 true time series from the logistic map and 100 "resampled" time series following our algorithm. Using the method proposed by Rosenstein *et al.* [22], we calculated the Lyapunov exponent and the correlation dimension for all the time series and then we used these results for making tests on the null hypothesis of equal distribution of the two samples of estimates. Each time series contains 500 observations.

The sample median, mean and standard deviation of the 100 Lyapunov exponent and correlation dimension estimates of the resampled time series are shown in Table I. The estimates are very close to the true values of λ_1 and d_{corr}.

We conducted three types of nonparametric tests: First we tested the null hypothesis of equal mean between true and resampled estimates of λ_1 and d_{corr}, respectively, then the hypothesis of equal variance and finally the hypothesis of equal distribution (Kolmogorov-Smirnov test) [23]. On the basis of the p values reported in Table II, we can accept the whole set of hypotheses, within a significance level of 5%. (Given a hypothesis test, let T be a suitably chosen test statistic; we define p value as the probability that the statistic T "exceeds" the observed value t_{obs} when the null hypothesis is, in fact, true [24]. An empirical decision rule is to accept the null hypothesis if the p value is bigger than 0.10.) These results give us confirmation that in this application the method is working correctly: The dynamical and geometrical properties of the resampled time series and of the series directly generated by the DGP under study are the same. Further applications of the method will be reported in [25].

For the sake of simplicity, here we have not checked the degree of brittleness and shadowability of the resampled series. We will consider this crucial feature in future works. Moreover, preliminary explorations prove that the method can be applied, with proper modification, to noisy time series.

*Electronic address: sgol@stat.unipg.it
†Electronic address: msandri@chiostro.univr.it

[1] J. Theiler, B. Galdrikian, A. Longtin, S. Eubank, and J. D. Farmer, in *Nonlinear Modeling and Forecasting,* edited by M. Casdagli and S. Eubank, SFI Studies in the Sciences of Complexity Proc. XII (Addison-Wesley, Reading, MA, 1992).

[2] W. A. Brock, W. D. Dechert and J. A. Scheinkman, Social Systems Research Institute, University of Wisconsin, 1987 (unpublished).

[3] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, Physica (Amsterdam) **16D**, 285 (1985).

[4] R. Brown, P. Bryant, and H. D. I. Abarbanel, Phys. Rev. A **43**, 2787 (1991); J. P. Eckmann, S. O. Kamphorst, D. Ruelle, and S. Ciliberto, Phys. Rev. A **34**, 4971 (1986); R. Gencay and W. D. Dechert, Physica (Amsterdam) **59D**, 142 (1992); D. Mc Caffrey, S. Ellner, R. Gallant, and D. Nychka, J. Am. Stat. Assoc. **87**, 682 (1992).

[5] P. Grassberger and I. Procaccia, Physica (Amsterdam) **9D**, 189 (1983).

[6] D. Lai and G. Chen, Int. J. Bifurcation Chaos Appl. Sci. Eng. **6** (5), 1721 (1995).

[7] J. L. Simon and P. Bruce, Math. J. **3** (1), 48 (1993).

[8] H. R. Künsch, Ann. Stat. **17**, 1217 (1989).

[9] D. N. Politis and J. P. Romano, J. Am. Stat. Assoc. **89**, 1303 (1994).

[10] U. Lall and A. Sharma, Water Resour. Res. **32** (3), 679 (1996).

[11] R. Gencay, Physica (Amsterdam) **89D**, 261 (1996).

[12] J. D. Farmer and J. J. Sidorowich, in *Evolution, Learning and Cognition,* edited by Y. C. Lin (World Scientific, Singapore, 1988).

[13] V. I. Oseledec, Trans. Moscow Math. Soc. **19**, 197 (1968).

[14] H. White, *Artificial Neural Networks. Approximation and Learning Theory* (Blackwell, Oxford, 1992).

[15] A. S. Lapedes and R. Farber, Los Alamos National Laboratory Technical Report No. LA-UR-87-2662, 1987; J. Moody and C. J. Darken, Neural Comput. **1**, 281 (1989); M. C. Mozer, in *Predicting the Future and Understanding the Past,* edited by A. S. Weigend and N. A. Gershenfeld, SFI Studies in the Science of Complexity Proc. XV (Addison-Wesley, Reading, MA, 1994).

[16] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, Phys. Rev. A **45**, 3403 (1992).

[17] D. S. Broomhead and G. P. King, Physica (Amsterdam) **20D**, 217 (1986).

[18] M. Casdagli, J. R. Stat. Soc. B **54**, 303 (1992).

[19] S. M. Hammel, J. A. Yorke, and C. Grebogi, J. Complex. **3**, 136 (1987).

[20] E. M. Coven, I. Kan, and J. A. Yorke, Trans. Am. Math. Soc. **308**, 227 (1988).

[21] S. Dawson, C. Grebogi, T. Sauer, and J. A. Yorke, Phys. Rev. Lett. **73**, 1927 (1994).

[22] M. T. Rosenstein, J. J. Collins, and C. J. De Luca, Physica (Amsterdam) **65D**, 117 (1993).

[23] W. J. Conover, *Practical Nonparametric Statistics* (Wiley, New York, 1980), 2nd ed.

[24] D. R. Cox and D. V. Hinkley, *Theoretical Statistics* (Chapman and Hall, London, 1979).

[25] S. Golia and M. Sandri (to be published).