

Exploiting Ergodicity for the Analysis of Short Time Series

Matjaž Perc

*Department of Physics, Faculty of Natural Sciences and Mathematics,
University of Maribor, Koroška cesta 160, SI-2000 Maribor, SLOVENIA*

Qing Yun Wang* and Zhi Sheng Duan

*State Key Laboratory for Turbulence and Complex Systems,
Department of Mechanics and Aerospace Engineering,
College of Engineering, Peking University, Beijing 100871, CHINA*

(Received 13 May, 2008)

A new generally applicable method for analyzing short time series that exploits the ergodic nature of chaotic systems is presented and tested on a simple example. The method can be integrated into several existing algorithms and should prove valuable for characterizing short experimental signals.

PACS numbers: 05.45.Tp, 05.45.Ac

Keywords: chaotic systems, Lyapunov exponents

The accurate characterization of dynamics in an experimental time series is fundamental for understanding the observed behavior of any natural or artificial system. While numerous methods are available for this task [1], their accuracy more or less depends on the availability of long data sets. This requirement often hinders productive research from experimental data, especially in fields where acquisition of long data sets is impossible due to intrinsic system properties or lack of non-invasive recording techniques that destroy the system's dynamics. Such instances are frequent in biology, especially at the cellular level [2], and can also be encountered in economics [3], astronomy [4], as well as man-made systems in general.

Here we propose a new approach for analyzing short time series based on ergodicity of deterministic chaotic systems [5]. We exploit the fact that an arbitrarily small neighborhood of every point that forms the chaotic attractor is repeatedly visited by the trajectory during the temporal evolution of the system. Consider the reconstructed phase space vectors in m dimensions $\vec{y}_i = [x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau}]$ that

were obtained from a short scalar time series x_i [6]. If x_i is sufficiently short, the trajectory may occupy only a small portion of the whole phase space, as shown in Fig. 1(a). In this case, it is usually impossible to carry out further analyses. However, ergodicity assures that the last point of the trajectory, denoted as \vec{y}_n , has an arbitrary close neighbor \vec{y}_j , from which the trajectory continues, possibly further occupying the available phase space, as shown in Fig. 1(b). It is thereby crucial to note that j isn't necessarily the $n + 1$ point in time. Based on this observation we conclude that several short time series segments, although not obtained continuous in time, *i.e.* from a single experiment, may still yield a long enough continuous trajectory in the phase space to allow further analysis. Since initial conditions in real-life settings are never known perfectly it is certain that repetitions of the same experiment will always yield at least slightly different time series, thus providing further information about the underlying system. Obviously, parameter settings of the system must be kept constant during repetitions so that stationarity criteria are not violated [7].

Following the above reasoning let us summarize the algorithm. Suppose we possess r short time series each occupying n data points that were obtained from r repetitions of an experiment. Altogether then $s = rn$

*Also at School of Statistics and Mathematics, Inner Mongolia Finance and Economics College, Huhhot 010051, CHINA

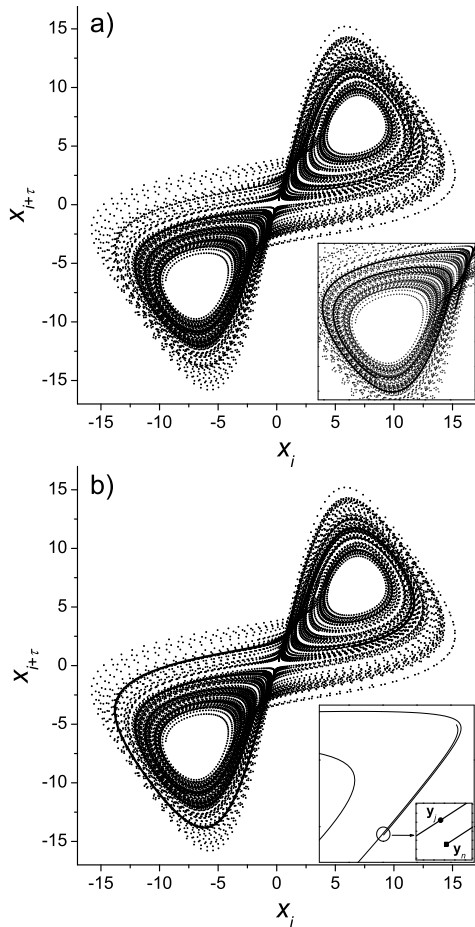


FIG. 1. Reconstructed phase space ($d = 3, \tau = 17$) from 50 randomly selected data segments of the Lorenz system [8], each occupying $n = 200$ points. The line connects points from the 3rd [panels (a) and (b)] and the 18th [panel (b)] segment.

data points are available for further analysis, which we can simply merge together, one data segment after another, to form a new time series x_i , where $i = 1, 2, \dots, s$ (provided points in all data segments were evenly sampled). This new time series is discontinuous whenever $(i \bmod n) = 0$. To correct this, each time the traditional embedding procedure [6] encounters a point x_i where $[n - (i \bmod n)] < (m - 1)\tau$ [\vec{y}_i cannot be formed without $x_{i+(m-1)\tau}$ going through a discontinuity] a search for a close phase space neighbor, denoted as $\vec{z}_j = [x_j, x_{j-\tau}, \dots, x_{j-(m-1)\tau}]$, begins that is less than ϵ apart from $\vec{z}_{i+t} = [x_{i+t}, x_{i+t-\tau}, \dots, x_{i+t-(m-1)\tau}]$,

where $t = [n - (i \bmod n)]$. Note that a different letter is used for these delay vectors since they are formed backwards in time as opposed to \vec{y}_i , whereas t is introduced in order to start the search algorithm at the last point of each data segment to minimize data wastage. Moreover, \vec{z}_j is considered as a suitable close neighbor only if $[n - (j \bmod n)] \geq [(m - 1)\tau - t]$ so that the point x_j has enough successors within its own data segment to completely form the unfinished delay vector \vec{y}_i . The ‘missing’ coordinate $x_{i+(m-1)\tau}$ of the delay vector \vec{y}_i is, upon finding a suitable \vec{z}_j , replaced by $x_{j-t+(m-1)\tau}$. Successive delay vectors then have to be formed from x_j onwards, whereby if $x_{j-t+(m-1)\tau}$ is introduced as the m -th delay coordinate of \vec{y}_i , then the first coordinate of \vec{y}_{i+1} (note that delay vectors are numbered consecutively) must be x_{j-t+1} . This procedure has to be repeated until as many points of the whole data set ($i = 1, 2, \dots, s$) as possible are used while keeping ϵ small. In general, a trade off between ϵ (continuity of the final ‘merged’ time series) and the percent of all occupied data points has to be made. The whole set of data points can be viewed as a labyrinth. However, the catch lies not simply in coming from one end to the other, but rather to do so in an as continuous manner as possible, whereby none of the points should be used more than m times, *i.e.* each time as a different delay coordinate. At the end, one obtains a continuous long set of delay vectors \vec{y}_i , *i.e.* a continuous trajectory in the phase space, where the time course of a delay coordinate represents a new continuous time series that has exactly the same properties as each individual data segment, and is due to its extended length suitable for further analysis.

Note that although the above procedure assumes that proper embedding parameters m and τ are known, this must not be the case in general. The algorithm can be integrated directly into existing methods for determining proper embedding parameters [9], whereby those m and τ that are being considered as suitable are used for calculations. In fact, the algorithm can be integrated into any existing methods that incorporate time operations of the form $x_a \rightarrow x_{a+b}$.

An excellent tool for optimizing the outcome of the presented algorithm is the determinism test introduced by Kaplan and Glass [10]. The test measures average directional vectors in a coarse-grained m -dimensional embedding space. If $x_{j-t+(m-1)\tau}$ is not a suitable replacement for $x_{i+(m-1)\tau}$, either because $\|\vec{z}_j - \vec{z}_{i+t}\|$ is too large, or the trajectory doesn't evolve further in the same direction, the norm of the average directional vector of all passes through the box that occupies the emergent discontinuity will be smaller than unity (if each pass is treated as a unit vector). Thus, we may introduce the weighted average of all lengths of average directional vectors $\bar{\Lambda}$ (defined as in [10]) as the crucial quantity that, for an optimal execution of the algorithm, has to be maximal. Of course $\bar{\Lambda}$ need not be exactly 1, since the time series itself may be burdened with measurement error or may even originate from a stochastic system. The coarse grained flow of the phase space that was merged together from 50 data segments of the Lorenz system, after the application of the algorithm with $\epsilon_{max} = 0.7$ (approximately 1/10 of the standard deviation of data), is presented in Fig. 2. It can be observed that basically all vectors are of unit length. Accordingly $\bar{\Lambda} = 0.99$. Since we demonstrate our results on a numerical example (the input data segments are noise-free) this high $\bar{\Lambda}$ is not surprising and confirms the successfulness of our algorithm.

As a final stringent test, we calculate Lyapunov exponents of the newly obtained phase space by approximating the flow with radial basis function as advocated in [11]. The obtained results presented in Fig. 3 show excellent agreement with accurate values obtained with the help of differential equations. Noteworthy, similar results were also obtained by using polynomial basis functions [12].

The method introduced in this Letter is set out to enable time series analysis of very short data segments by exploiting phase space ergodicity of chaotic systems. Thereby, the fact that an arbitrarily small neighborhood of every point of the chaotic attractor is repeatedly visited by the trajectory during the temporal evolution of the system is exploited. Although throughout

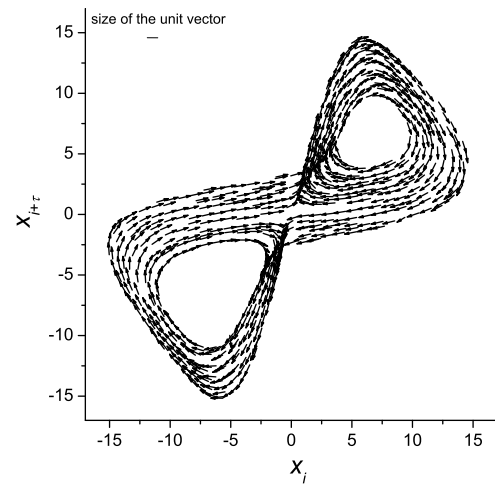


FIG. 2. Directional flow of the merged phase space that was coarse grained into a $30 \times 30 \times 30$ grid. The length of arrows is equal to the vector norm in 3D, whereas directions correspond to projections onto the $x_i - x_{i+\tau}$ plane.

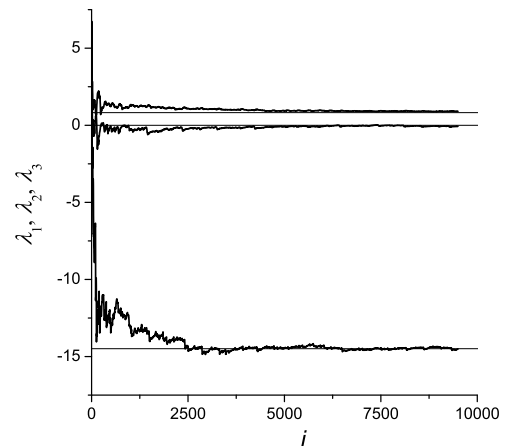


FIG. 3. Lyapunov exponents of the merged phase space shown in Fig. 2. Thin straight lines indicate accurate values. Thick lines depict the convergence of the maximal, middle and the lowest Lyapunov exponent from top to bottom.

the Letter stress is put on deterministic chaotic systems, which should definitely not be assumed ad hoc when attempting analysis of experimental time series, the method can be implemented on any data sets. As always, the outcome of the method has to be verified by known determinism [11, 13] and stationarity [7] tests

so the time series can be deemed suitable for further analysis. An obvious weak point of the method is that it assumes multiple short time series being at hand prior to implementation. Although conclusions drawn from an experiment are valid only if the experiment can be repeated and conclusions verified, this is a rather poor argument if the experiment is costly or otherwise difficult to repeat, as are for example certain medical conditions like rarely occurring heart arrhythmia or epileptic seizures. Nevertheless, there exist several experimental situations that are currently being studied predominantly with mathematical modeling and numerical simulations, as for example effects of different

hormones on intracellular calcium oscillations [14], for which the approach advocated in this paper may be especially useful.

Acknowledgments

This work was supported by the National Science Foundation of China (Funds No. 10702023 and 60674093) and China Post-Doctoral Science Foundation (Fund No. 20070410022). Matjaž Perc additionally acknowledges support from the Slovenian Research Agency (Project No. Z1-9629).

References

- [1] H. D. I. Abarbanel, *Analysis of Observed Chaotic Data* (Springer, New York, 1996); H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, 1997); J. C. Sprott, *Chaos and Time-Series Analysis* (Oxford University Press, Oxford, 2003).
- [2] A. Goldbeter, *Biochemical Oscillations and Cellular Rhythms* (Cambridge University Press, Cambridge, 1996).
- [3] R. S. Tsay, *Analysis of Financial Time Series* (John Wiley and Sons, New York, 2002).
- [4] A. S. Weigend and N. A. Gershenfeld (eds.), *Time Series Prediction: Forecasting the Future and Understanding the Past* (Addison-Wesley, Reading, 1993).
- [5] J.-P. Eckmann and D. Ruelle, *Rev. Mod. Phys.* **57**, 617 (1985).
- [6] F. Takens, in: D. Rand and L. S. Young (eds.), *Dynamical Systems and Turbulence* (Springer-Verlag, Berlin, 1981); T. Sauer, J. Yorke, and M. A. Casdagli, *J. Stat. Phys.* **65**, 579 (1991).
- [7] T. Schreiber, *Phys. Rev. Lett.* **78**, 843 (1997).
- [8] E. N. Lorenz, *Atmos. Sci.* **20**, 130 (1963); Parameters $r = 25$, $b = 8/3$, $c = 10$ and time step for numerical integration $dt = 0.01$.
- [9] A. M. Fraser and H. L. Swinney, *Phys. Rev. A* **33**, 1134 (1986); M. B. Kennel, R. Brown, and H. D. I. Abarbanel, *Phys. Rev. A* **45**, 3403 (1992).
- [10] D. T. Kaplan and L. Glass, *Phys. Rev. Lett.* **68**, 427 (1992).
- [11] J. Holzfuss and U. Parlitz, in: L. Arnold, H. Crauel, and J.-P. Eckmann (eds.), *Proceedings of the Conference on Lyapunov Exponents* (Springer-Verlag, Berlin, 1991); U. Parlitz, *Int. J. Bifurcat. Chaos*, **2**, 155 (1992).
- [12] K. Briggs, *Phys. Lett. A* **151**, 27 (1990); P. Bryant, R. Brown, and H. D. I. Abarbanel, *Phys. Rev. Lett.* **65**, 1523 (1990).
- [13] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, *Physica D* **58**, 77 (1992); T. Schreiber and A. Schmitz, *Physica D* **142**, 346 (2000).
- [14] K. S. R. Cuthbertson and P. H. Cobbold, *Nature* **316**, 541 (1985); N. M. Woods, K. S. R. Cuthbertson, and P. H. Cobbold, *Nature* **319**, 600 (1986).