Research Article

Statistical Complexity of Low- and High-Dimensional Systems

Vladimir Ryabov[1] and Dmitry Nerukh[2]

[1] *Department of Complex System, School of Systems Information Science, Future University Hakodate, 116-2 Kamedanakano-Cho, Hakodate-Shi, Hakodate, Hokkaido 041-8655, Japan*
[2] *Non-Linearity and Complexity Research Group, Aston University, Birmingham B4 7ET, UK*

Correspondence should be addressed to Dmitry Nerukh, d.nerukh@aston.ac.uk

We suggest a new method for the analysis of experimental time series that can distinguish high-dimensional dynamics from stochastic motion. It is based on the idea of statistical complexity, that is, the Shannon entropy of the so-called ϵ-machine (a Markov-type model of the observed time series). This approach has been recently demonstrated to be efficient for making a distinction between a molecular trajectory in water and noise. In this paper, we analyse the difference between chaos and noise using the Chirikov-Taylor standard map as an example in order to elucidate the basic mechanism that makes the value of complexity in deterministic systems high. In particular, we show that the value of statistical complexity is high for the case of chaos and attains zero value for the case of stochastic noise. We further study the Markov property of the data generated by the standard map to clarify the role of long-time memory in differentiating the cases of deterministic systems and stochastic motion.

1. Introduction

Statistical complexity is a measure that had been introduced by Crutchfield and Young in 1989 [1]. It has been proven useful for describing various complex systems, including those with hundreds of degrees of freedom [2]. According to our earlier paper [3], the statistical complexity of high-dimensional trajectories generated by the dynamics of an ensemble of water molecules grows up to the time scale of 1 microsecond, that is, an extremely long-time interval for a typical molecular dynamics simulation. Moreover, this property is much less pronounced for so-called surrogate time series that have exactly the same power spectrum and, hence, autocorrelation function as the original time series.

For example, in Figure 1 we plot the dependence of statistical complexity on the length of the time series for the symbolic data obtained from a Poincaré the section of 3D velocities describing the motion of a hydrogen atom in an ensemble of 392 water molecules [3]. The details of computing the atomic trajectories as well as the method used for partitioning the phase space and obtaining a symbolic string from the initially floating point data can be found in [4]. In the same figure, we draw the curves calculated for

so-called phase-shuffled surrogate time series [5], the data having identical autocorrelation function, and hence power spectrum as the original velocity trajectories. One can notice significant differences between the statistical complexity of the physical and the artificially generated data.

We then put forward a hypothesis that this property, that is a high value of statistical complexity, can be used for distinguishing between deterministic and stochastic systems (see also [6]). The phenomenon of the complexity growth with the length of time series that ensures the difference between the cases of deterministic and stochastic behaviour remains still unexplored. In order to elucidate the mechanism that makes the value of complexity high, we performed numerical experiments with the standard map (known also as the Chirikov-Taylor map) [7], one of the most studied paradigmatic models in nonlinear dynamics. We observed that statistical complexity was high indeed in the case of the standard map, and it had much lower value for the surrogate time series, being close to zero for the case of noncorrelated noise from a random number generator.

For the purpose of estimating statistical complexity for a symbolic time series, we utilize the CSSR algorithm [8] that
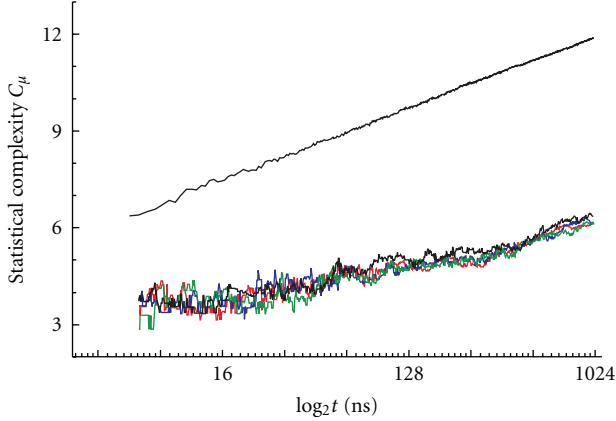
FIGURE 1: Statistical complexity versus the (log of) length of the analysis interval for the hydrogen velocity time series (top curve) and four surrogate time series (bottom curves): three independent realisations of the phase-shuffling algorithm (red, green, and blue), and single time series of a white noise passed through a low-pass linear filter (black) (from [3]). Note that the value of statistical complexity for the data obtained from simple random number generator is close to zero and does not depend on the length of time series for large enough value of the latter (not shown).

had been reported as an efficient, reliable, and easy to use software. The algorithm constructs an ϵ-machine, a Markov-chain with l-step memory, which constitutes a probabilistic model for the analysed data series. Statistical complexity measures an information content of the ϵ-machine via its Shannon entropy. Our analysis shows, however, that although the CSSR algorithm always converges well and produces a finite value of complexity, in some cases the approximation of data with a Markov-chain-type model is inadequate, making the complexity value dependent on the length of the analysed data. Finally, we came to the conclusion that, at least in the case of standard map, the main reason for the growth of complexity is the property of stickiness of periodic islands in the chaotic sea, a generic phenomenon in Hamiltonian systems [9]. It has been noted in [10] that due to the sticking property of the regular component in a subcritical domain ($K < 0.9716$), the dynamics of the standard map is subdiffusive that can be well approximated with a continuous time random walk model. Anomalous properties of the temporal behaviour of nonextensive entropy, a generalization of the usual Boltzmann-Gibbs entropy, have been also analysed in [11].

In the present work, we mainly study the domain of $K \gg 1$ where the area occupied by periodic islands is small, and the chaotic motion can be expected to be strongly mixing and ergodic. Nevertheless, as our results show, the presence of stickiness is still an important factor defining the long-term statistical measures. In terms of the CSSR algorithm, the property of stickiness breaks the independence of the data points separated by a history long-time interval, thus making the Markov-chain approximation invalid.

Finally, we discuss a conjecture that the property of the non-Markovianity of the ϵ-machine and growth of statistical complexity can be used in a constructive way for distinguishing deterministic and stochastic behaviours. The problem of detecting determinism in a noise looking chaotic time series is a long standing one. An extensive review of the issues related to the difference between chaos and noise and to inherent difficulties encountered in the high-dimensional cases can be found in [12]. We suggest that there is a significant difference between the statistical complexities of Hamiltonian chaos and coloured noise with identical power spectrum, the main reason for which consists in the presence of the long-time memory in time series obtained from Hamiltonian systems. This property originates from the stickiness of periodic islands that are abundant in the chaotic sea due to multiple resonances that occur in the phase space.

We would also like to note that since the phase space of Hamiltonian systems has a complicated structure of chaotic areas intermingled with periodic islands, this leads in some cases to the necessity of distinguishing between chaos and complex quasiperiodic motion. A measure called *orbital complexity* had been introduced for this purpose in the context of analysing the orbital motions of planets [13–15]. This measure, although being based on the calculation of the Shannon entropy (but in the spectral domain), has quite different meaning, purpose, and scope of applicability compared to statistical complexity.

2. Systems and Method of Analysis

The standard map is defined as

$$
\begin{aligned}
p_{n+1} &= p_n + K \sin \theta_n \bmod 2\pi, \\
\theta_{n+1} &= \theta_n + p_{n+1} \bmod 2\pi,
\end{aligned}
\tag{1}
$$

where K is a single parameter defining the dynamics of this system. In all the calculations below, the value of the parameter K has been chosen at $K = 6.908745$.

First, at the step called "symbolization," the original real-valued time series is transformed to a symbolic sequence by introducing a suitable partitioning of the phase space (Figure 2(a)).

At the next stage, the sequence of symbols is transformed to the sequence of histories, the l-symbol strings representing a refinement of the partitioning in the phase space [3]. ϵ-machine reconstruction requires a grouping of histories to "causal states," based on the analysis of the predictive properties of each history by one step forward in time. Finally, the statistical complexity is calculated from the ϵ-machine as Shannon entropy of the probability distribution of the causal states:

$$
C_\mu = - \sum_{i=1}^{N_c} p_i \log p_i,
\tag{2}
$$

here p_i are probabilities of the causal states in the ϵ-machine, and N_c is the total number of the causal states.

3. Numerical Experiments

We have calculated the statistical complexity using the algorithm called CSSR [8] for the standard map and plotted
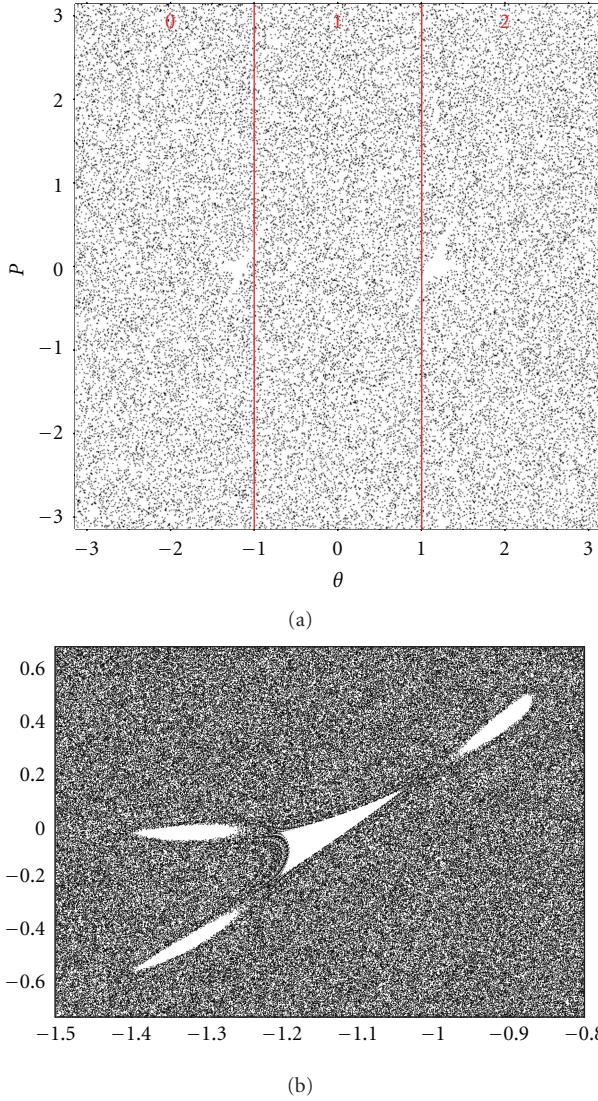
(a)

(b)

FIGURE 2: The standard map. Symbolization with a three-symbol alphabet (a). Two periodic islands are embedded into the chaotic sea. One of them zoomed is shown in (b).

(a)

(b)

FIGURE 3: The values of C_μ and the number of causal states for various history lengths (from bottom to top $l = 2 \cdots 9$) for the standard map trajectory as a function of the data length.

the graphs of complexity versus the amount of data (the length of symbolic sequence). We have also studied how the calculated values depend on the method of partitioning the map, initial conditions, and the parameter K of the system.

A typical plot of statistical complexity C_μ and the number of causal states for the history length $l = 2 \cdots 9$ are shown in Figure 3.

The results for the surrogate data generated using the same trajectory of the standard map are shown in Figure 4. Changing the initial conditions, the type of partitioning the phase space at the stage of symbolization, and/or value of the parameter K brings qualitatively the same results, that is, there is a significant difference between the complexity values calculated for the data obtained from the map and those for the surrogate time series. It should be noted though that the complexity value moderately increases with parameter K as shown in Figure 5. This behaviour is similar to that of other
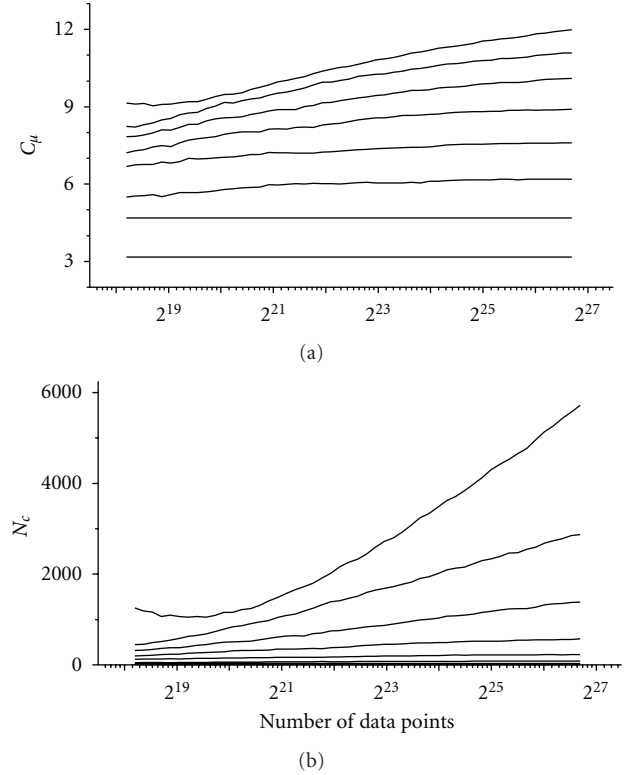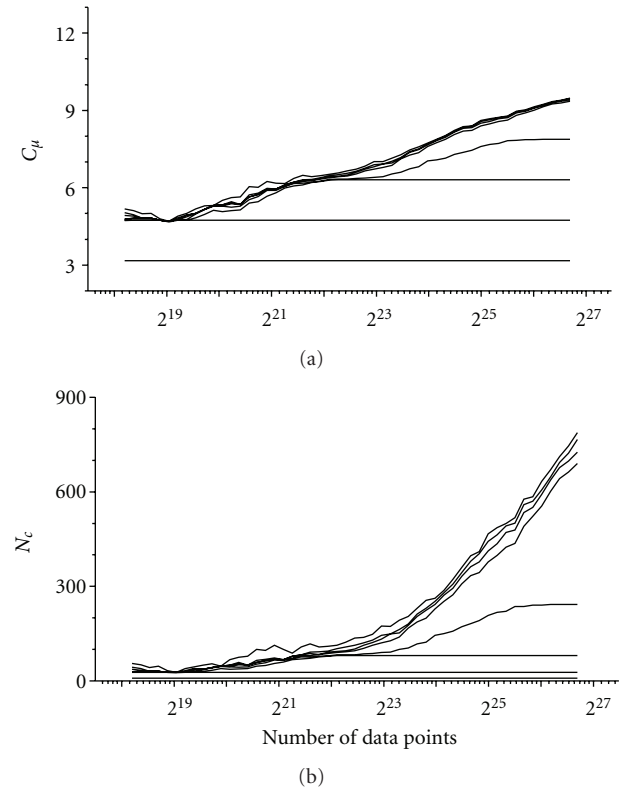

(a)

(b)

FIGURE 4: Same as in Figure 3 but for the random surrogate.
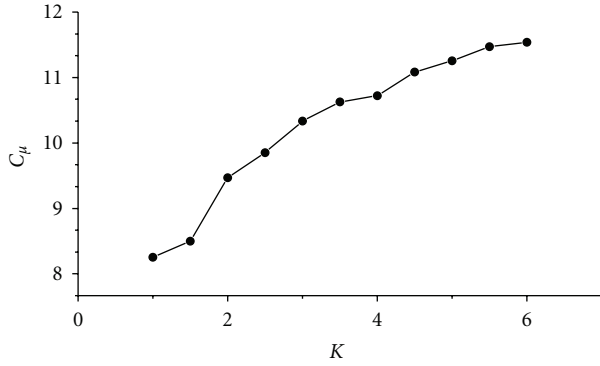
Figure 5: Dependence of statistical complexity on the parameter K (the number of data points is 10^8).
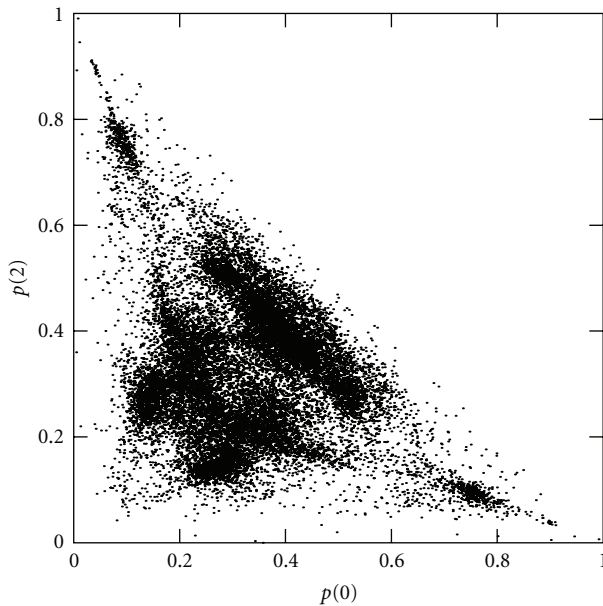

Figure 6: Conditional distributions of the next symbol for all histories at history length $l = 8$. Every point represents a history. The total number of points (histories) is 3^8.
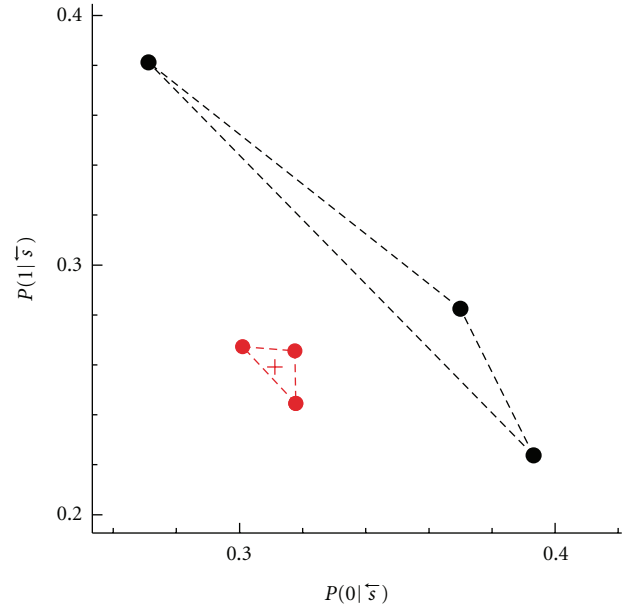

Figure 7: Conditional probability distribution for the history with the largest deviation from the Markov property (black triangle). Adding a symbol to the history changes drastically the position of the point in the diagram. Three circles correspond to adding "0," "1," or "2" to the history $l = 8$ (crosses). The red triangle represents a randomly chosen history with clear Markov property, that is, the distribution of probabilities does not depend on the added symbol.

characteristics used in nonlinear dynamics, like Lyapunov exponents or measure of the chaotic area reported in [16].

4. A Hypothesis on Markov Property

In this section, we would like to demonstrate that the large complexity values observed in the case of the standard map are caused by the presence of certain segments in the chaotic trajectory (which become histories after symbolization) that do not possess a property necessary for building a Markov-chain from the data. Consider the stage when the symbolic string has been converted to the sequence of histories, that is, symbolic words of length l. The Markov-chain (ϵ-machine) can be built from such a sequence, if the conditional probability distribution of the next symbol in the symbolic sequence depends only on the l-symbol string preceding the symbol, and it is independent on the previous symbol, that

is, the symbol that occurred $l + 1$ time steps before. In other words, if we consider the conditional probability distribution for a given history, it should not change (in statistical sense), if we increase the length of a history by one symbol to the past.

In Figure 6, we present a scatter diagram that demonstrates the distribution of the conditional probabilities for each history at the history length $l = 8$. Every point in the diagram corresponds to a single history. The large spread around the point with coordinates $(1/3, 1/3)$ evidences significant difference compared to the case of uniform distribution. The change from $l = 8$ to $l = 9$ does not change the overall pattern of point distribution shown in Figure 6. However, the analysis of the conditional probabilities for individual histories reveals huge changes in the position of points depending on the extra symbol added at the beginning of the history. In Figure 7, we depict the conditional distribution for the next symbol for two histories: one that shows strong deviation from Markov property and a "normal" one, that is, a randomly chosen history. The large deviation in the distribution of conditional probabilities can be concluded from a big distance between the vertexes of the upper triangle (distribution of the conditional probabilities at $l = 9$) and the cross corresponding to the conditional probability at $l = 8$. The probabilities at $l = 9$ are computed by adding one of the three symbols (012) at the beginning of the history of $l = 8$.

Finally, we would like to show that the segments of chaotic trajectories that correspond to the history with large deviation from the Markov property are located in the areas
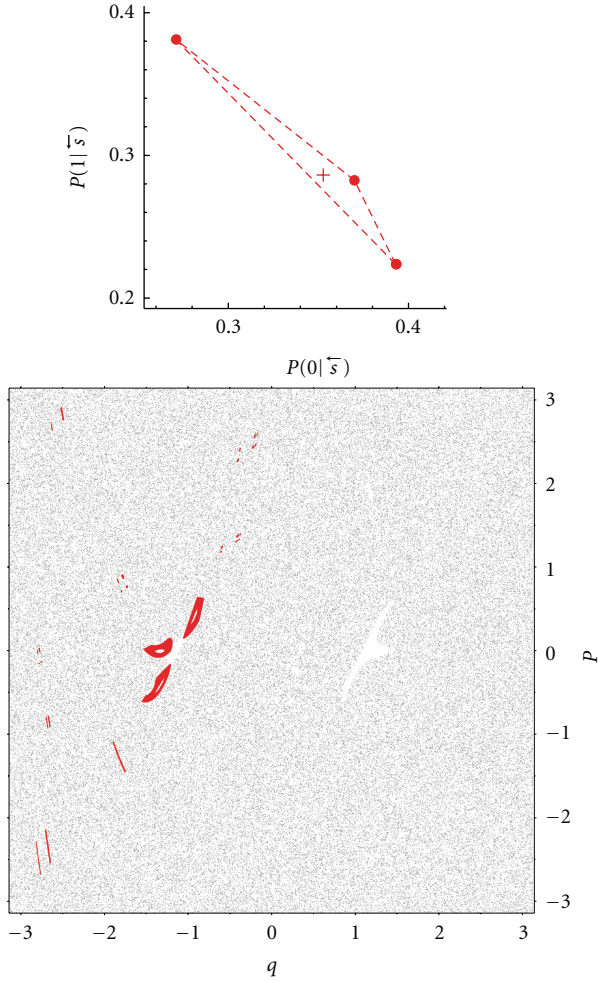
FIGURE 8: Parts of chaotic trajectory corresponding to the history with large deviation from the Markov-chain property. Only points corresponding to the central three symbols in the history are shown. Apparently, the history includes one of the two periodic islands shown in Figure 2.
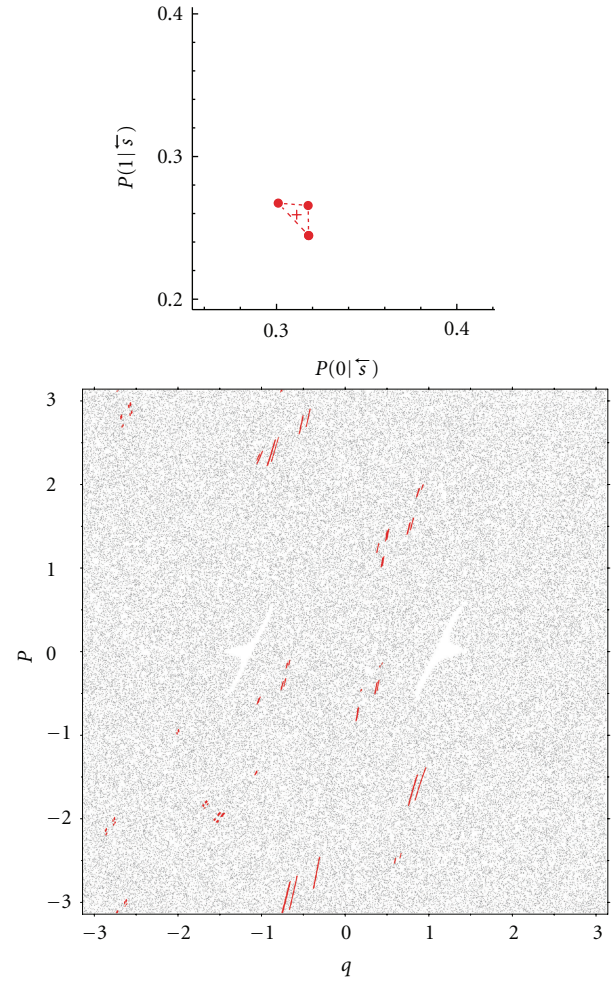

FIGURE 9: Parts of chaotic trajectory corresponding to the history with no deviation from the Markov-chain property. Only points corresponding to the central three symbols in the history are shown. Apparently, the history does not include any of the two periodic islands shown in Figure 2.

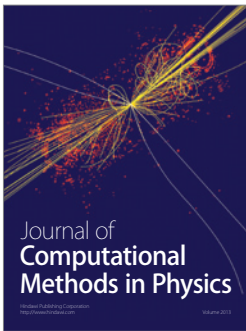of the phase space close to periodic islands. For this purpose, we plotted in Figure 8 only the points that correspond to the history with large deviation from the Markov-chain property. A comparison to Figure 2 suggests that the history with large deviation in the distributions is located close to the periodic islands. Therefore, we suppose that the breaking of Markovianity can be interpreted as a manifestation of the well-known phenomenon of "stickiness" [9] of trajectories in the areas close to periodic islands. Prolonged wandering of a trajectory around the island is equivalent to existing of long-time memory in the corresponding segments of the chaotic time series. Figure 8 should be also compared to Figure 9, which presents the segments of the chaotic trajectory corresponding to a history possessing the Markov property. Apparently, it has no relation to periodic islands. Such histories represent a vast majority in the ensemble of 3^l histories, the non-Markovian histories constituting only a fraction of percent.
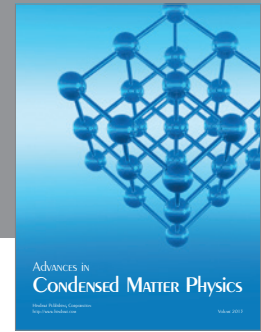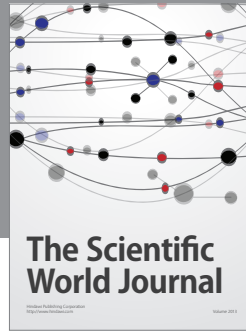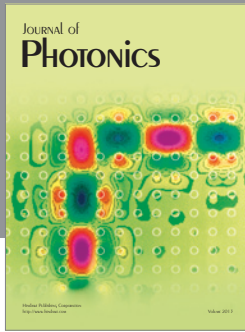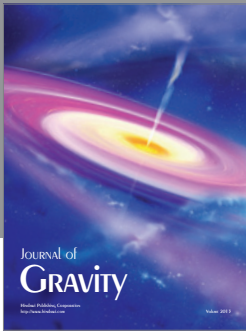
5. Discussion

It has been demonstrated in this paper that statistical complexity appears to be a useful measure for distinguishing Hamiltonian chaos in low- and high-dimensional systems from correlated noise with identical autocorrelation function. Its value for the symbolic time series calculated from the dynamics of Hamiltonian systems is substantially larger than that for a white noise time series or the time series obtained from the phase shuffling surrogate algorithm. Our explanation of the origin of this phenomenon in terms of Markov-chain theory consists in breaking down the Markov property by the symbolic sequences obtained from Hamiltonian systems.

We believe that the large value of complexity observed in our numerical experiments is defined by the presence of periodic islands with sticky borders in the phase space of Hamiltonian systems. The stickiness of certain areas in the phase space leads to long-time memory effects that are

responsible for breaking down the statistical independence of the future states from the past ones. This, in turn, makes the procedure of grouping the histories into causal states constituting the core of CSSR algorithm unstable. As a result, the algorithm finds more and more causal states necessary for building the ϵ-machine as a Markov chain, and the value of complexity grows with the number of causal states.

References

[1] J. P. Crutchfield and K. Young, "Inferring statistical complexity," *Physical Review Letters*, vol. 63, no. 2, pp. 105–108, 1989.

[2] D. P. Feldman, C. S. McTague, J. P. Crutch-field et al., "The organization of intrinsic computation: complexity-entropy diagrams and the diversity of natural information processing," *Chaos*, vol. 18, no. 4, Article ID 043106, 15 pages, 2008.

[3] D. Nerukh and V. Ryabov, "Computational mechanics of molecular systems," in *Computational Mechanics Research Trends, Computer Science, Technology and Applications*, H. P. Berger, Ed., Nova Science, 2010.

[4] D. Nerukh, V. Ryabov, and R. C. Glen, "Complex temporal patterns in molecular dynamics: a direct measure of the phase-space exploration by the trajectory at macroscopic time scales," *Physical Review E*, vol. 77, no. 3, Article ID 036225, 2008.

[5] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. Doyne Farmer, "Testing for nonlinearity in time series: the method of surrogate data," *Physica D*, vol. 58, no. 1-4, pp. 77–94, 1992.

[6] J. M. Amigó, S. Zambrano, and M. A. F. Sanjuán, "Combinatorial detection of determinism in noisy time series," *EPL*, vol. 83, no. 6, Article ID 60005, 2008.

[7] B. V. Chirikov, "A universal instability of many-dimensional oscillator systems," *Physics Reports*, vol. 52, no. 5, pp. 263–379, 1979.

[8] C. R. Shalizi and K. L. Shalizi, "Blind construction of optimal nonlinear recursive predictors for discrete sequences," in *Proceedings of the Uncertainty in Artificial Intelligence, 20th Conference*, M. Chickering and J. Halpern, Eds., pp. 504–511, AUAI Press, 2004.

[9] G. M. Zaslavsky, "Chaos, fractional kinetics, and anomalous transport," *Physics Report*, vol. 371, no. 6, pp. 461–580, 2002.

[10] J. H. Misguich, J.-D. Reuss, Y. Elskens, and R. Balescu, "Motion in a stochastic layer described by symbolic dynamics," *Chaos*, vol. 8, pp. 248–256, 1998.

[11] F. Baldovin, C. Tsallis, and B. Schulze, "Nonstandard entropy production in the standard map," *Physica A*, vol. 320, pp. 184–192, 2003.

[12] G. Boffetta, M. Cencini, M. Falcioni, and A. Vulpiani, "Predictability: a way to characterize complexity," *Physics Reports*, vol. 356, no. 6, pp. 367–474, 2002.

[13] N. T. Faber, C. M. Boily, and S. Portegies Zwart, "On time-dependent orbital complexity in gravitational N-body simulations," *Monthly Notices of the Royal Astronomical Society*, vol. 386, no. 1, pp. 425–439, 2008.

[14] H. E. Kandrup, B. L. Eckstein, and B. O. Bradley, "Chaos, complexity, and short time Lyapunov exponents: two alternative characterisations of chaotic orbit segments," *Astronomy and Astrophysics*, vol. 320, no. 1, pp. 65–73, 1997.

[15] I. V. Sideris and H. E. Kandrup, "Chaos and the continuum limit in the gravitational N-body problem. II. Nonintegrable potentials," *Physical Review E*, vol. 65, no. 6, Article ID 066203, 2002.

[16] I. I. Shevchenko, "Isentropic perturbations of a chaotic domain," *Physics Letters A*, vol. 333, no. 5-6, pp. 408–414, 2004.