2016 International Symposium on Nonlinear Theory and Its Applications,
NOLTA2016, Yugawara, Japan, November 27th-30th, 2016

# String Entropy as a Measure of Complexity in Chaotic Time Series

Takaya Miyano and Kenichiro Cho

Department of Mechanical Engineering, Ritsumeikan University
1–1–1 Noji-higashi, Kusatsu, Shiga 525-8577, Japan
Email: tmiyano@se,rutsumei.ac.jp, kcho@fc.ritsumei.ac.jp

**Abstract**—Chaotic time series numerically generated by deterministic nonlinear dynamics are often used as pseudorandom numbers for cryptography. Here, we show a method for time series analysis to characterize complexities in chaotic dynamics in terms of string entropy, as a class of information entropy, estimated from the relative frequencies of binary-coded characters transformed from a chaotic time series. We apply our method to various chaotic time series and discuss their performance as pseudorandom numbers for chaos-based cryptography.

## 1. Introduction

Chaotic dynamics have been often applied to cryptography, where the chaotic time series numerically generated by the dynamics were used as pseudorandom numbers to encrypt plaintexts and message signals [1]–[12]. Recently, we have derived the augmented Lorenz (AL) equations as a nondimensionalized dynamical model for turbulent thermal convection with a high Rayleigh number exceeding $10^6$ from the Newtonian equations of motion of a chaotic gas turbine [13]. The AL equations have been applied to a one-time pad chaotic cryptography, where the chaotic time series generated by the AL equations were used as pseudorandom numbers to mask a speech signal and a plaintext [14].

The randomness of pseudorandom numbers is of critical importance when using them in cryptosystems to encrypt messages. A low degree of randomness will facilitate code breaking by cryptanalysists. In fact, the statistical tests for the randomness of pseudorandom numbers to be used in cryptosystems, i.e., NIST 800-22, is published by the National Institute of Standards and Technology (NIST) [15], which is widely recognized as the standard protocol to assess the randomness.

We have recently proposed an information-theoretical method, referred to as the string entropy method [16], to evaluate the degree of complexity in a chaotic time series. Although the string entropy test is essentially the same as the entropy test included in NIST 800-22, it appears to also be effective to characterize the dynamical nature of chaotic dynamics. In this paper, we apply our method to the chaotic time series numerically generated by various chaotic dynamics and show how it is effective to characterize the chaotic nature.

## 2. String Entropy

Given a time series $\{x_i\}_{i=1}^N$ with a sampling time interval of $T$, the string entropy $S$ is defined as follows. First, $x_i$ is transformed into binary digits $b_i$ with the threshold crossing as

$$b_i = 0 \quad \text{if } x_i < x_c \ , \tag{1}$$
$$b_i = 1 \quad \text{otherwise} \ , \tag{2}$$

where $x_c$ is an appropriately chosen threshold around which $b_i$ is distributed with equal probability. We partition the binary series $\{b_i\}_{i=1}^N$ into a sequence of $Q$ segments consisting of $D$ binary digits, where $N = DQ$. Each segment is mapped to an "alphabet" binary-coded in $D$ bits, where each alphabet symbolizes the time evolution represented by $D$ successive realizations of the dynamical system. Thus, we obtain a string of $Q$ alphabets as random realizations of $2^D$ possible alphabets denoted as $\{a_n\}_{n=1}^{2^D}$. In the case of $D = 7$, we have $2^7 = 128$ alphabets {'0000000', '0000001', ... , '1111111'} corresponding to {'0', '1' ... , '127'} in the decimal expression. In this case, the total number of the alphabets is equal to that of the ASCII codes. With the frequency of appearance for each alphabet $a_n$ ($n = 1, \ \dots \ , \ 2^D$), we estimate a histogram, from which the probability density function $p(a_n)$ is calculated. The histogram represents the statistical distribution of the coarse-grained time evolution. The string entropy $S$ is defined and calculated as

$$S = -\sum_{n=1}^{2^D} p(a_n) \log_2 p(a_n) \ . \tag{3}$$

$S$ takes the maximum value $S_{max} = D$ [bits] if and only if $p(a_n) = 2^{-D}$ regardless of $n$. Hence, $S$ is normalized with respect to $S_{max}$, and the normalized string entropy $H$ is defined as $H = S/D$, where $0 \leq H \leq 1$ and $H = 1$ is obtained for completely random processes.

Let us consider the relationship between the string entropy and the Lyapunov exponent for a simple case,

i.e., a one-dimensional chaotic map $x_{n+1} = f(x_n)$. In this case, the Lyapunov exponent $\lambda$ is defined as

$$\lambda = \log \mid f'(x_0) \mid \ , \tag{4}$$

where $f'$ denotes the derivative of $f$ with respect to $x$ and $x_0$ denotes the initial point. The Lyapunov exponent estimated using Eq. (4) is usually averaged over many initial points in the chaotic attractor (denoted as $\Omega$) to obtain the global Lyapunov exponent as

$$\begin{aligned} \lambda &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \log \mid f'(x_n) \mid \ , \\ &= \int_{\Omega} \mu(x) \log \mid f'(x) \mid dx \ , \end{aligned} \tag{5}$$

where $\mu(x)$ denotes the probability density function of $x$. The string entropy is estimated using the probability density function $p(a)$ of the alphabets coded in $D$ bits. Hence, the global Lyapunov exponent is related via Eq. (5) with the string entropy, since $p$ is the $D$ successive products of $\mu$ of the coarse-grained $x$ with the binary expression along a trajectory in $\Omega$.

A similar relationship holds for multi-dimensional chaotic maps and chaotic flows by considering the Jacobian of the chaotic dynamics.

## 3. Numerical Analysis and Discussion

We conducted numerical experiments on estimating $H$ with the parameter settings of $D = 7$, $N = 120000$, $Q = 20000$, $T = 1$ for the logistic map, tent map, the Lorenz equations, and the AL equations, where all numerical calculations were performed in double precision on a 32-bit machine. No particular methods were used to reduce the accumulation of roundoff errors.

The logistic map is defined as $x_{n+1} = \alpha x_n(1 - x_n)$, where $0 \le x_n \le 1$. Figures 1(a), (b), (c), and (d) show a typical example of the estimated histograms under randomly chosen initial points ($x_0$) for $\alpha$ = 3.95, 3.98, 3.99, and 4, respectively. The threshold value was set to $x_c = 0.5$. The initial 5000 data points were discarded to eliminate the initial transient parts from the analysis. Estimates of $H$ are summarized in Table 1. At $\alpha = 3.95$, 3.98, and 3.99, there are missing alphabets regardless of the choice of $x_0$. These missing alphabets can be the clues to identify the value of $\alpha$. At $\alpha = 4$, all alphabets appear with approximately equal probability and $H$ approaches unity.

The results for the tent map are shown in Fig. 2 and Table 1. The tent map is defined as $x_{n+1} = 1 - 2 \mid x_n - 0.5 \mid$. The convergence of $x_n$ to the fixed points of $x = 0$ and $x = 1$ was circumvented by restricting the domain of $x_n$ to $\epsilon \le x_n \le 1 - \epsilon$ with $\epsilon = 10^{-6}$. The threshold value was set to $x_c = 0.5$, and the initial 5000 data points were eliminated from the analysis. In contrast to the logistic maps, there are many missing alphabets, whereas the alphabet '0000000' ('0' in the decimal expression) appears very frequently. These observations were independent of the choice of $x_0$. The estimated $H$ is considerably smaller than those for the logistic maps. Hence, the tent map cannot be used as a secure pseudorandom number generator.

Table 1: Estimates of the normalized string entropy $H$ for chaotic time series ($D = 7$ and $T = 1$).

| Dynamics | $H$ |
|---|---|
| Logistic map ($\alpha = 3.95$) | 0.8877 |
| Logistic map ($\alpha = 3.98$) | 0.9486 |
| Logistic map ($\alpha = 3.99$) | 0.9553 |
| Logistic map ($\alpha = 4$) | 0.9994 |
| Tent map | 0.6625 |
| Lorenz equations: $x$ | 0.9954 |
| Lorenz equations: $y$ | 0.9961 |
| AL equations: $x$ | 0.9991 |
| AL equations : $y_{100}$ | 0.9984 |

The results for the Lorenz equations [17] are shown in Figs. 3(a) and (b),and in Table 1. The Lorenz equations are defined as a three-dimensional system of ordinary differential equations: $\dot{x} = \sigma(y - x)$, $\dot{y} = rx - y - xz$, $\dot{z} = -\beta z + xy$, where $\sigma = 10$, $r = 28$, and $\beta = 8/3$. The equations were numerically integrated using the 4th-order Runge-Kutta method with a time width of $\Delta t = 0.01$. The initial conditions for $x$, $y$, and $z$ were given as pseudorandom numbers subject to the standard normal distribution. The initial 5000 numerical solutions were eliminated from the analysis. The threshold values were set to $x_c = 0$ and $y_c = 0$. For both $x$ and $y$, there are no missing alphabets and several alphabets appear more frequently than other alphabets. Estimates of $H$ for $x$ and $y$ are close to unity but slightly smaller than that for the logistic map with $\alpha = 4$.

Our final case study is concerned with the AL equations defined as

$$\dot{x} = \sigma\Big\{ \mathrm{tr}\left[(\mathbf{M}^{-1})^2 \mathbf{y}\right] - x \Big\} \ , \tag{6}$$

$$\dot{\mathbf{y}} = \mathbf{R}x - \mathbf{M}\mathbf{z}x - \mathbf{y} \ , \tag{7}$$

$$\dot{\mathbf{z}} = \mathbf{M}\mathbf{y}x - \mathbf{z} \ , \tag{8}$$

$$\mathbf{R} = R_0 \mathbf{M^2 \Phi W} \ , \tag{9}$$

where $x$ is a dimensionless scalar variable, $\mathbf{y} = \mathrm{diag}(y_1, \ \ldots \ , \ y_n, \ \ldots \ , \ y_K)$ and $\mathbf{z} = \mathrm{diag}(z_1, \ \ldots \ , \ z_n, \ \ldots \ , \ z_K)$ are dimensionless $K \times K$ diagonal matrices, $tr(\cdot)$ represents the diagonal sum of a matrix, $\sigma$ and $R_0$ are dimensionless parameters corresponding to the Prandtl and reduced Rayleigh numbers, respectively, and $\mathbf{M}$ denotes the diagonal matrix given by $\mathbf{M} = \mathrm{diag}(m_1, \ \ldots \ , \ m_n, \ \ldots \ , \ m_K)$ with

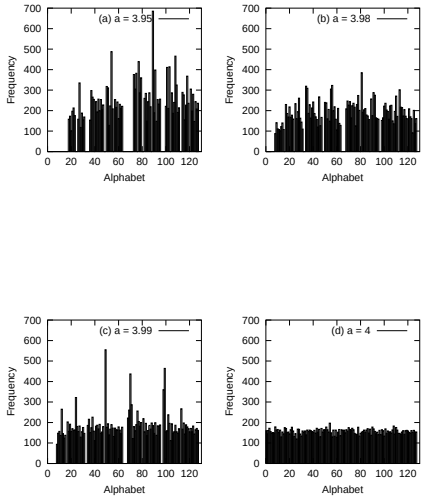


Figure 1: (Histograms of alphabets for the logistic maps with (a) $\alpha = 3.95$, (b) $\alpha = 3.98$, (c) $\alpha = 3.99$, and (d) $\alpha = 4$. $D = 7$ and $T = 1$ (one time step).

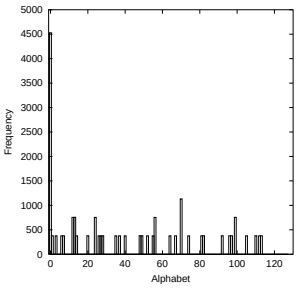


Figure 2: (Histograms of alphabets for the tent map. $D = 7$ and $T = 1$ (one time step).

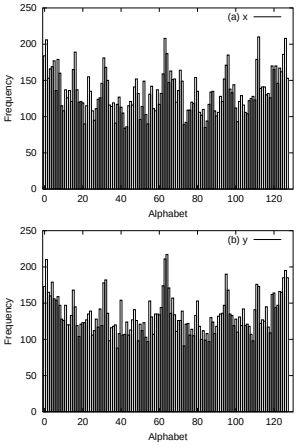


Figure 3: Histograms of alphabets for (a) $x$ and (b) $y$ of the Lorenz model. $D = 7$ and $T = 1$ (100 time steps).

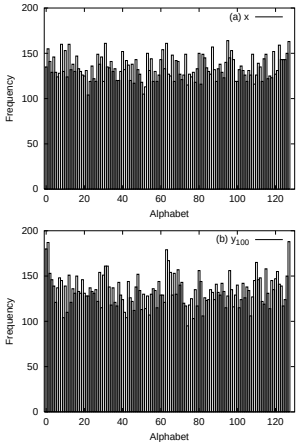


Figure 4: Histograms of alphabets for (a) $x$ and (b) $y_{100}$ of the AL model. $D = 7$ and $T = 1$ (2500 time steps).

$m_1 = 1$ and $m_2$ through $m_K$ randomly taking values of $m_n = n$ or $m_n = n + 0.5$. $\mathbf{M}$ can be used as a secret key for cryptography [14]. For the definitions of the diagonal coefficient matrices $\mathbf{\Phi}$ and $\mathbf{W}$, see [14]. The bifurcation parameters $\sigma$, $R_0$, and $\phi$ were set to $\sigma = 25$, $R_0 = 3185$, and $\phi = 0.36$ [rad].

We numerically integrated the AL equations in a similar way to the Lorenz equations, except with a time width of $\Delta t = 4 \times 10^{-4}$ and $K = 101$ under a random setting of $\mathbf{M}$ and the initial conditions of $x$, $\mathbf{y}$, and $\mathbf{z}$ given as pseudorandom numbers subject to the standard normal distribution. Time series $x$ and $y_{100}$ were obtained by discrete sampling of the numerical solutions with a sampling time of $T = 1$ (2500 time steps) and the threshold values $x_c = 0$ and $y_c = 0$.

Figures 4(a) and (b) show typical examples of the histograms for $x$ and $y_{100}$, respectively. Estimates of $H$ are summarized in Table 1. For both $x$ and $y_{100}$, there are no missing alphabets. Estimates of $H$ for $x$ and $y_{100}$ are close to unity.

We next estimated $H$ as a function of the bifurcation parameter $R_0$ for $y_{100}$ of the AL equations. Figure 5 shows the estimates of $H$ as a function of $R_0$, where the sampling time interval $T$ was set to 0.01 (25 time steps) to reduce the total computational time for the numerical integration of the AL equations and $R_0$ was increased from 1000 to 3190 with an increment width of $\Delta R_0 = 10$. $H$ increases as $R_0$ increases, and approaches unity at $R_0 > 3000$, where the AL equations appear to generate fully developed chaotic time series.

In conclusion, our numerical analysis indicates that the string entropy method is capable of characterizing chaotic dynamics. In particular, the missing alphabets in the estimated histograms provide important features of the chaotic trajectories governed by the

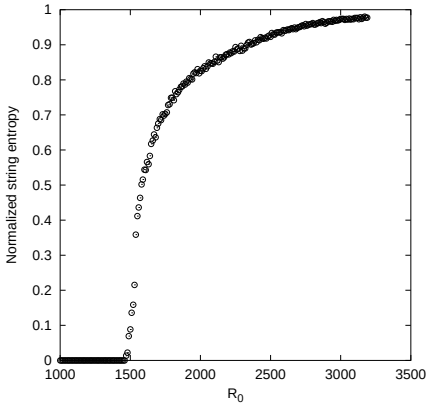


Figure 5: Estimates of the normalized string entropy as a function of the bifurcation parameter $R_0$ for $y_{100}$ of the AL equations. $\Delta R_0 = 10$, $D = 7$, and $T = 0.01$ (25 time steps).

dynamics. When applying the chaotic dynamics to cryptography, $H$ should be as close as possible to its maximum value of unity, and there should be no missing alphabets. In this sense, the AL equations is useful for a pseudorandom generator for a one-time pad cipher. Recently, we have verified that the pseudorandom numbers generated by the AL equations pass the statistical tests of NIST 800-22, which will be reported in a future paper.

**Acknowledgment**

This study was partly supported by JSPS KAKENHI Grant Number JP15K00353.

**References**

[1] K. M. Cuomo and A. V. Oppenheim, "Circuit implementation of synchronized chaos with applications to communications," *Phys. Rev. Lett.*, vol. 71, no. 1, pp. 65–68, 1993.

[2] A. Argyris, D. Syvridis, L. Larger, V. Annovazzi-Lodi, P. Colet, I. Fisher, J. Garcia-Ojarvo, C. R. Mirasso, L. Pesquera, and K. A. Shore, "Chaos-based communications at high bit rates using commercial fibre-optic links," *Nature*, vol. 437, pp. 343–346, 2005.

[3] H. Dedieu, M. Kennedy, and M. Hasler, "Chaos shift keying: Modulation and demodulation of a chaotic carrier using self-synchronizing Chua's circuits," *IEEE Trans. Circuits Syst.-II*, vol. 40, pp. 634–642, 1993.

[4] S. Hayes, C. Grebogi, E. Ott, and A. Mark, "Experimental control of chaos for communications," *Phys. Rev. Lett.*, vol. 73, pp. 1781–1784, 1994.

[5] Y. Lai, E. Bolt, and C. Grebogi, "Communicating with chaos using two-dimensional symbolic dynamics," *Phys. Lett.* A, vol. 255, pp. 75–81, 1999.

[6] T. Miyano, K. Nishimura, and Y. Yoshida, "Chaos-based communications using open-plus-closed-loop control," *IEICE Trans. Fundam.*, vol. E94-A, pp. 282–289, 2011.

[7] R. Tenny, L. S. Tsimring, L. Larson, and H. D. I. Abarbanel, "Using distributed nonlinear dynamics for public key encryption," *Phys. Rev. Lett.*, vol. 90, pp. 047903-1–047903-4, 2003.

[8] L. Kocarev, M. Sterjev, A. Fekete, and G. Vattay, "Public-key encryption with chaos," *Chaos*, vol. 14, no. 4, pp. 1078–1082, 2004.

[9] N. Masuda, G. Jakimoski, K. Aihara, and L. Kocarev, "Chaotic block ciphers: from theory to practical algorithms," *IEEE Trans. Circuits Syst.-I*, vol. 53, no. 6, pp. 1341–1352, 2006.

[10] F. Anstett, G. Millerioux, and G. Bloch, "Chaotic cryptosystems: Cryptanalysis and Indentifiability," *IEEE Trans. Circuits Syst.-I*, vol. 53, no. 12, pp. 2673–2680, 2006.

[11] W. Xu, L. Wang, and G. Chen, "Performance analysis of the CS-DCSK/BPSK communication system," *IEEE Trans. Circuits Syst.-I*, vol. 61, pp. 2624–2633, 2014.

[12] G. Alvarez and S. Li, "Some basic cryptographic requirements for chaos-based cryptosystems," *Int. J. Bifur. Chaos*, vol. 16, pp. 2129–2151, 2006.

[13] K. Cho, T. Miyano, and T. Toriyama, "Chaotic gas turbine subject to augmented Lorenz equations," *Phys. Rev.* E, vol. 86, pp. 036308-1–036308-12, 2012.

[14] K. Cho and T. Miyano, "Chaotic cryptography using augmented Lorenz equations aided by quantum key distribution," *IEEE Trans. Circuits Syst.-I*, vol. 62, no. 2, pp. 478–487, 2015.

[15] National Institute of Standards and Technology (U.S. Department of Commerce), Special publication 800-22 revision 1a: http://csrc.nist.gov/groups/ST/toolkit/rng/index.html.

[16] K. Cho and T. Miyano, "Entropy test for complexity in chaotic time series," *Nonlin. Theor. Appl. IEICE*, vol. 7, no. 1, pp. 21–29, 2016.

[17] E. N. Lorenz, "Deterministic non-periodic flow," *J. Atoms. Sci.*, vol. 20, pp. 130–141, 1963.