Patterns and pseudo-randomness using complex systems

Jeaneth Machicao[1]*, **Odemir M. Bruno**[1]

[1]Scientific Computing Group. São Carlos Institute of Physics,
University of São Paulo, São Carlos - SP, PO Box 369, 13560-970, Brazil.

machicao@usp.br, bruno@ifsc.usp.br

Thesis available at: **http://www.teses.usp.br/teses/disponiveis/76/76132/tde-28022018-144846/pt-br.php**

Abstract. *In this thesis, we developed a method that exploits the random-like properties of chaotic systems as a pseudo-random number generator (PRNG). We explored the k-digits to the right of the decimal separator (less significant digits) of an original orbit of a chaotic map. This approach called as "deep-zoom" demonstrated the relationship between the parameter k and the quality of the pseudo-random sequences, since it showed a rapid transition from "weak to strong" randomness as k tends to infinity, thus allowing to manipulate pseudo-randomness in a parametrically manner.*

1. Motivations and objectives

Pseudo-random number generators (PRNGs) are the backbone of the most diverse fields of application, ranging from statistics and probability theory, decision theory, numerical calculus, simulation, and systems modeling, in the gaming and entertainment industry, programming languages, and even in more critical scenarios such as cryptography. Classically, the construction of PRNGs is based on deterministic algorithms using linear recurrences, bitwise operations, and algebraic concepts, among other artifacts, often without mathematical foundations, e.g. the linear congruence generator (LCG) or the Mersenne Twister [Matsumoto and Nishimura 1998]. Unlike classical algorithms, chaos-based PRNGs are implemented based on chaotic systems that produce values with random-like properties. In fact, significant progress has been reported, for example, with the CB-PRNGs based on differential equations and recurrence maps [Radwan et al. 2016, WANG and YANG 2012, Öztürk and Kılıç 2015, FranÇois et al. 2014, Min et al. 2013, Hu et al. 2013], as well as using chaotic cellular automata [Tomassini et al. 2000, Spencer 2015, Hortensius et al. 1989].

One of the most important properties of the chaos theory is its sensitivity to the initial conditions since the smallest variation on the initial conditions (seed) can disturb completely the system over time (butterfly effect). Additionally, because of its random-like behavior, and its unpredictability, that is, the difficulty of being able to predict over a long period what the system's behavior will be in the future. All these features are of great applicability in different branches including cryptography and PRNGs, where there is a need to obtain sources of pseudo-randomness in chaotic systems with a high positive Lyapunov exponent. For this reason, the close relationship between chaos and pseudo-randomness has aroused great interest in the academic community, which is also commercially and militarily exploited, as demonstrated by the state-of-the-art over the past 32 years.

Notwithstanding, some researchers have expressed certain doubts about the chaotic properties of well-known chaotic maps, such as the logistic map in the context of cryptography, which presents non-uniform probability distribution (pattern U), where a plateau distribution is expected; dependence of the control parameter which may lead to periodic windows; large enough ciphertext samples that can estimate the parameter [Álvarez et al. 2003]; short cycle length orbits where long periodicity is expected depending on machine limitations [Persohn and Povinelli 2012]; degradation of digital chaotic system problems [Hu et al. 2014].

However, we have observed that the true potential of chaotic systems relies on the infinitesimal depth of the precision digits of their orbit points. For example, considering the well-known Mandelbrot set, when displayed on a computer screen, it reveals interesting but rather limited patterns. However, when this pattern is repeatedly extended, a large number of complex patterns can be distinguished and, certainly, is in these magnifications where legitimate chaos occurs. Thus, higher computational precision is required to exploit the deep-zoom of a chaotic system and hence to investigate the pseudo-random properties of such chaotic systems.

In this thesis, we presented patterns and pseudo-randomness as an approach that relates both concepts, which traditionally are seen as opposites. This approach uses the mathematical basis of complex systems for two purposes: on the one hand, to explore the spectrum of pseudo-randomness of chaotic systems in a quest to achieve true randomness and, on the other hand, the development of methods based on artificial life and complex networks such as method of pattern recognition that finally intertwined the search for patterns in pseudo-random sequences. In this thesis two important questions are developed: is it possible to generate pseudo-random numbers as close to true randomness? Is it possible to create a method that generates such random numbers that make it difficult to search for patterns? The answer to both questions is affirmative.

In this work, we developed a method that explores the deep-zoom properties of the chaotic systems, specifically in the logistic map and tent map, as sources of pseudo-randomness. We observe that the patterns disappear and the pseudo-randomness is increased by removing k-digits to the right of the decimal separator of each of the points of an original orbit of a chaotic map. Thus, a rapid transition from "weak" randomness to "strong" was evidenced as k tends to infinity, which allows assessing to a spectrum of pseudo-randomness manipulated parametrically. This conjecture becomes more evident since the different strategies analyzed (statistical tests, dynamics analysis of chaotic systems, analysis of the Lyapunov exponent and spectral analysis) corroborate that k is related to the pseudo-random qualities. Finally, this same approach was used to analyze the sequences of pseudo-random numbers generated by the gold standard of the k-logistic map in the context of pattern recognition. Our main results are summarized as follows:

- Regarding chaos theory, it has been corroborated that, despite the simplicity of some chaotic systems, it is possible to obtain good sources of pseudo-randomness.
- We developed two chaos-based PRNGs based on the k-logistic map and the k-tent map, respectively, that yielded an article published in the Chaos journal [Machicao and Bruno 2017].
- The main product of this thesis is that we obtained a parameterized gold-standard PRNG, which is the first of its kind into the literature. The gold standard repre-

sents an important research tool that can aid the development of three different areas: pattern recognition, cryptography, and cryptanalysis. Since it can generate virtually infinite sets of random numbers with known theoretical basis. Consequently, the proposed approach has brought significant advances to a wide range of fields including cryptography and cryptoanalysis.

2. Main results

In this work, it was observed that the pseudo-random properties of a chaotic map can be improved as k increases. In fact, by means of all the visualization tools (bifurcation diagram, Poincaré diagram, frequency histogram), analysis by the Lyapunov exponent, of the randomness tests, including spectral analysis; suggest that the quality performance of the proposed PRNG with $k \geq 4$-logistic map overpass the pseudo-randomness properties of classical PRNG such as LCG and Mersenne Twistter.

Hereafter, we summarized all of the main results obtained. All subsequent experiments were focused on one of the more chaotic regions of the k-logistic map provided by the parameter $\mu = 4$ which corresponds to the largest Lyapunov exponent and the most chaotic region as well. In Fig. 1a-b, we can observe the Poincaré diagram, which relates the sequences x_t^k, x_{t+1}^k and x_{t+2}^k, respectively. The original orbit k_0 shows the classic inverted parabola of the logistic map, whose pattern remains in the 3D plot. Then, from top to bottom, we present the phase diagrams from k_1 to k_4, we can observe the transformation of the parabola pattern into zig-zag patterns, which are progressively disappeared until they become visually random, as can be seen from k_2 onwards. In fact, it is observed that the phase space is being filled as k increases, that is, that k-logistic map produces almost all possible values between $[0, 1]$, as expected of a good pseudo-random number generator and also becomes a non-invertible map. So future and past numbers are becoming more uncorrelated. In Fig. 1c is observed the Fourier power spectrum displayed in two dimensions. From top to bottom, this spectral analysis is shown for different values of the parameter k, for k_0, k_1, k_2, k_3 and k_4. For the original logistic map (k_0) the parallel traces indicate the presence of patterns, however, as the value k is increased, the spectral density is approaching to a constant in the center, which is a clear indicator of the absence of patterns.

For the purposes of good construction of a PRNG, a uniform distribution is expected as much as possible [Álvarez and Li 2006, Arroyo 2009]. The U pattern of this distribution is evident by looking at the first graph of Fig. 1d, which represents the frequency curve of the original orbit k_0 using the parameter $\mu = 4$. This U pattern is well known in the theory of dynamic systems since the logistic map follows an invariant probability density function. However, from top to bottom, it can be seen that this distribution becomes more uniform as k increases, that is, the pseudo-random properties of the k-logistic map changes becoming, in spite of redundancy, more random as $k \gg 1$. In addition, this uniformization process is illustrated in Fig. 1e, which shows a comparison of the former curves [Machicao and Bruno 2017].

Besides the former plots, the results of the 18 statistical tests are reported in Table 1 and Table 2, for the DIEHARD suite and the NIST suite, respectively. Each column corresponds to the number of files that passed the sub-tests. In both tables, the tests that failed in at least 50 files were highlighted in gray, which can be observed in the case of k_0,

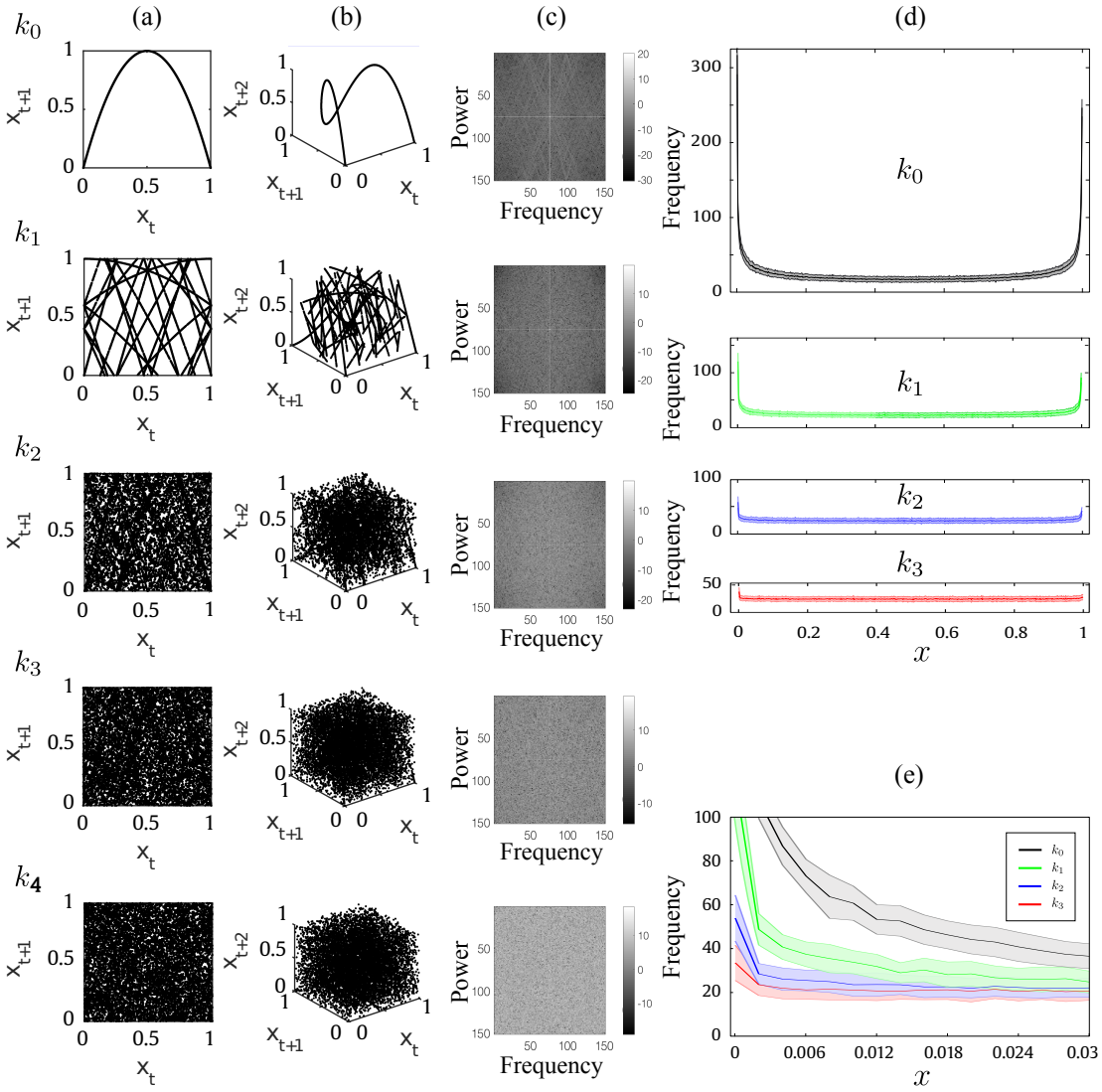

Figura 1. Different visual results for the k-logistic map for k_0, k_1, k_2, k_3 and k_4 (top to bottom) using $\mu = 4$. (a-b) Two- and three-dimensional diagrams are shown on the left and right column, respectively. The horizontal and vertical axes show the phase space of x_t^k against x_{t+1}^k. Each orbit contains 10^4 points started from random initial conditions, where the first 200 iterations were discarded (transient time). (c) 2D Fourier power spectrum for 150^3 numbers generated by PRNG k-logistic map. (d) Frequency distribution curves. Horizontal axis shows the $x \in [0, 1]$ (500 bins) and vertical axis shows the frequency of the 10^4 values discarding firsts 10^3 transient values. The curves represent the mean and standard deviation (shaded error bar) for sequences generated over 100 random initial conditions. (e) The inset plot depicts a zoom on the windows $x \in [0, 0.03]$. Adapted from [Machicao and Bruno 2017].

k_1, k_2 and k_3-logistic map. We observed that $k = 0$ fails to both Diehard and NIST tests, however the panorama changes as the parameter k increases. The k-logistic map passes on all of the tests from Diehard and NIST when $k \geq 4$.

Tabela 1. Average number of files that passed Diehard tests using the k-logistic map PRNG from 100 file samples. Severely failed tests are shown in gray. All tests passed using the interval $0.0001 <$ p-value < 0.9999. Source: [Machicao and Bruno 2017].

Diehard tests	k_0	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9
BirthdaySpacings [KS]	100	100	100	100	100	100	100	100	100	100
OverlappingPermutations	99	97	98	95	98	96	98	98	99	100
Ranks31x31 matrices	100	100	100	100	100	100	100	100	100	100
Ranks32x32 matrices	100	100	100	100	100	100	100	100	100	100
Ranks6x8 matrices [KS]	0	0	25	99	100	100	100	100	100	100
Monkey20bitsWords [KS]	0	99	100	100	100	100	100	100	100	100
OPSO [KS]	98	99	100	100	100	100	100	100	100	100
OQSO [KS]	98	100	100	100	100	100	100	100	100	100
DNA [KS]	100	100	100	100	100	100	100	100	100	100
Count1sStream	0	0	0	98	100	100	100	100	100	100
Count1sSpecific [KS]	0	0	0	0	94	100	100	100	100	100
ParkingLot [KS]	100	100	100	100	100	100	100	100	100	100
MinimumDistance [KS]	96	100	100	100	100	100	100	100	99	100
RandomSpheres [KS]	100	100	100	100	100	100	100	100	100	100
Squeeze [KS]	100	100	100	100	100	100	100	100	100	100
OverlappingSums [KS]	100	100	100	100	100	100	100	100	100	100
Runs (up)	100	100	100	100	100	100	100	100	100	100
Runs (down)	100	100	100	100	100	100	100	100	100	100
Craps (wins)	100	100	100	100	100	100	100	100	100	100
Craps (throws/game)	100	100	100	100	100	100	100	100	100	100

Tabela 2. Number of files that passed the NIST test suites [Rukhin et al. 2001] for the k-logistic map. Failed tests are shown in gray. All the tests passed to the $\alpha = 0.01$ significance level. Source: [Machicao and Bruno 2017].

NIST tests	k_0	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9
Frequency	98	99	99	99	99	99	99	99	99	99
BlockFrequency ($m = 128$)	0	1	66	95	98	98	99	100	99	99
CumulativeSums										
Forward sums	97	98	99	99	98	99	99	99	99	99
Reverse sums	97	99	99	99	99	99	99	99	99	99
Runs	0	0	14	91	98	99	99	99	99	100
LongestRun	0	0	15	89	98	99	98	100	99	99
Rank	99	100	99	99	99	99	99	99	99	99
FFT	77	98	99	99	99	99	99	99	99	99
Non-overlappingTemplate										
000000001	0	0	48	97	99	99	99	100	99	99
000000011	0	3	87	98	99	99	99	99	99	99
000000101	0	41	94	98	98	99	99	99	98	99
OverlappingTemplate	0	0	11	93	98	99	98	99	99	99
Universal	0	68	97	98	99	99	99	99	99	99
ApproxEntropy ($m = 10$)	0	0	64	98	99	99	99	100	99	99
RandomExcursions										
$x = -4$	90	98	99	99	98	99	99	99	99	100
$x = -3$	91	97	99	99	99	99	99	99	99	99
$x = -2$	94	99	98	99	99	98	99	99	99	99
$x = -1$	95	99	99	98	99	99	99	99	99	100
RandExcursVar										
$x = -9$	99	100	99	99	100	99	99	99	99	100
$x = -8$	99	99	99	99	99	99	99	99	99	100
$x = -7$	100	99	99	99	99	99	99	99	99	100
$x = -6$	100	99	99	99	99	99	98	99	99	99
$x = -5$	100	99	99	99	99	99	99	99	99	99
$x = -4$	99	100	99	99	99	99	99	99	99	99
$x = -3$	99	100	99	98	99	99	99	99	99	99
$x = -2$	99	99	99	98	99	99	99	100	99	99
$x = -1$	99	99	99	99	99	99	99	99	99	99
Serial ($m = 16$)										
Serial 1	0	1	82	96	98	99	99	98	99	99
Serial 2	10	81	95	98	98	99	99	99	98	99
LinearComplexity ($M = 500$)	99	98	99	99	99	99	99	98	99	99

In all of these analyzes, we have found that the pseudo-random properties of the logistic map can be noticeably improved when k is increased. In fact, patterns become increasingly widespread until they become visually indistinguishable ($k \geq 4$). With respect

to the pseudo-randomness tests, it was also corroborated that the sequences generated by using the map $k \geq 4$-logistic map passed successfully for the randomness tests of DIEHARD and NIST. Thus, we formulated the following conjecture: as the k parameter increases, the pseudo-randomness is improved, going from the regular pattern (k_0) to the most random (k_∞). Obviously, in computational terms, k_∞ would be impossible to prove, but we are not intending to exploit this side of the conjecture, but to exploit the fact the parameter k increases the pseudo-randomness and thereby creating a useful tool: **a PRNG gold standard**. With this gold standard, it is possible to generate datasets produced with different parameters k that provide the distinct classes, which allows the study and development of methods aimed at the recognition of patterns and non-linear time series.

3. Scientific production

3.1. Publications

- FILHO, H. A.; MACHICAO, J.; BRUNO, O. M. A hierarchical model of metabolic machinery based on the kcore decomposition of plant metabolic networks. Plos One, v. 13(5), p. e0195843, 2018.
- MACHICAO, J.; CORRÊA E. Jr.; MIRANDA, G.H.B.; AMANCIO, D.; BRUNO, O. M. Authorship attribution based on Life-Like network automata. Plos One, v. 13 (3), p. e0193703, 2018.
- MACHICAO, J.; RIBAS, L.; SCABINI, L; BRUNO, O. M. Cellular automata rule characterization and classification using texture descriptors. Physica A v. 497, p. 109—117, 2018.
- MACHICAO, J.; BRUNO, O. M. Improving the pseudo-randomness properties of chaotic maps using deep-zoom. Chaos: an interdisciplinary journal of nonlinear science v. 27 p. 053116, 2017.
- MACHICAO, J.; BRUNO, O. M. A cryptographic hash function based on chaotic network automata. Journal of Physics: conference series v. 936, p. 012058, 2017.
- FILHO, H. A.; MACHICAO, J.; BRUNO, O. M. Geometry from stomata networks at leaves of the Ctenanthe oppenheimiana. Journal of Physics: conference series v. 936, p. 012085, 2017.
- FILHO, H. A.; MACHICAO, J.; BRUNO, O. M. Geometric plasticity at leaves from Ctenanthe oppenheimiana probed by measure of distances between stomata. Journal of Physics: conference series v. 936, p. 012094, 2017.
- MIRANDA, G. H. B; MACHICAO, J.; BRUNO, O. M. Exploring spatio-temporal dynamics of cellular automata for pattern recognition in networks. Scientific Reports v. 6 n. 37329, 2016.
- MIRANDA, G. H. B; MACHICAO, J.; BRUNO, O. M. Network Analysis Using Spatio- Temporal Patterns. Journal of Physics: conference series v. 738 p. 012011, 2016.
- MACHICAO, J.; BAETENS, J. M.; MARCO, A. G.; DE BAETS, B.; BRUNO, O. M. A dynamical system approach to the discrimination of the modes of operation of cryptographic systems. Communications in Nonlinear Science and Numerical Simulation v. 29 n. 1–3, p. 102–115, 2015.
- MACHICAO, J.; BAETENS, J. M.; MARCO, A. G.; DE BAETS, B.; BRUNO, O. M. A Dynamical Systems Approach to the Discrimination of Cryptographic

Modes of Operation. In: 8th International Congress on Industrial and Applied Mathematics, 2015, Proceedings... Beijing, China 2015, p. 61.

3.2. In press

- MIRANDA, G. H. B; MACHICAO, J.; BRUNO, O. M. An optimized shape descriptor based on structural properties of networks. Digital Signal Processing.
- MIRANDA, G. H. B; MACHICAO, J.; BAETENS, J. M.; DE BAETS, B. BRUNO, O. M. A family of network automata based on neighborhood density. Automata 2018.

3.3. Submitted

- MACHICAO, J.; ALVES, M.; BAPTISTA, M.; BRUNO, O. M. Exploiting ergodicity of the k-logistic map to improve security in cryptographic systems. Chaos: an interdisciplinary journal of nonlinear science.
- MACHICAO, J.; ALMEIDA, H. A.; LAHR, D. J. G.; BUCKERIDGE, M.; BRUNO, O. M. Topological assessment of metabolic networks reveals evolutionary information. Scientific Reports.

3.4. In drafting stage

- MACHICAO, J.; ALMEIDA, H. A.; BRUNO, O. M. Analyses of the stomatic phenotypic plasticity by using Life-Like network automata (LLNA).
- MACHICAO, J.; RIBAS, L.; BRUNO, O. M. Binary Pattern on Life-Like Network Automata to Network Classification.
- MACHICAO, J.; BRUNO, O. M. Dynamical analysis of a deformation on unimodal maps.
- MACHICAO, J.; BRUNO, O. M. The logistic map and the number of the beast.
- MACHICAO, J.; ALVES, M.; BRUNO, O. M. Analyzing the Life-Like network automata rule space based on parametric pseudo-randomness.
- MACHICAO, J.; SCABINI, L; RIBAS, L.; BRUNO, O. M. Life-Like cellular automata rule space clustering using texture descriptors.
- MIRANDA, G.; MACHICAO, J.; BAETENS, J. M.; DE BAETS; BRUNO, O. M. Family of like-like network automata for pattern recognition.
- LARSEN, B. MACHICAO J.; BRUNO, O.M. A randomness test based on the probability intermittency distribution.

Referências

Álvarez, G. and Li, S. (2006). Some basic cryptographic requirements for chaos-based cryptosystems. *International Journal of Bifurcation and Chaos*, 16(8):2129–2151.

Álvarez, G., Montoya, F., Romera, M., and Pastor, G. (2003). Cryptanalysis of an ergodic chaotic cipher. *Physics Letters A*, 311(2–3):172–179.

Arroyo, D. (2009). *Framework for the analysis and design of encryption strategies based on discrete-time chaotic dynamical systems*. PhD thesis, Universidad Politécnica de Madrid, Madrid.

FranÇois, M., Grosges, T., Barchiesi, D., and Erra, R. (2014). Pseudo-random number generator based on mixing of three chaotic maps. *Communications in Nonlinear Science and Numerical Simulation*, 19(4):887–895.

Hortensius, P. D., McLeod, R. D., Pries, W., Miller, D. M., and Card, H. C. (1989). Cellular automata-based pseudorandom number generators for built-in self-test. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 8(8):842–859.

Hu, H., Deng, Y., and Liu, L. (2014). Counteracting the dynamical degradation of digital chaos via hybrid control. *Communications in Nonlinear Science and Numerical Simulation*, 19(6):1970 – 1984.

Hu, H., Liu, L., and Ding, N. (2013). Pseudorandom sequence generator based on the chen chaotic system. *Computer Physics Communications*, 184(3):765–768.

Machicao, J. and Bruno, O. (2017). Improving the pseudo-randomness properties of chaotic maps using deep-zoom. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27:053116.

Matsumoto, M. and Nishimura, T. (1998). Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, 8:3–30.

Min, L., Chen, T., and Zang, H. (2013). Analysis of fips 140-2 test and chaos-based pseudorandom number generator. *Chaotic Modeling and Simulation*, 2(1):273–280.

Öztürk, İ. and Kılıç, R. (2015). A novel method for producing pseudo random numbers from differential equation-based chaotic systems. *Nonlinear Dynamics*, 80(3):1147–1157.

Persohn, K. and Povinelli, R. (2012). Analyzing logistic map pseudorandom number generators for periodicity induced by finite precision floating-point representation. *Chaos, Solitons & Fractals*, 45:238–245.

Radwan, A. G., AbdElHaleem, S. H., and Abd-El-Hafiz, S. K. (2016). Symmetric encryption algorithms using chaotic and non-chaotic generators: A review. *Journal of Advanced Research*, 7(2):193 – 208.

Rukhin, A., Soto, J., Nechvatal, J., Smid, M., Barker, E., Leigh, S., Levenson, M., Vangel, M., Banks, D., and Heckert, A. (2001). A statistical test suite for random number generator for criptographic applications. Technical report, NIST, Gaithersburg, MD, USA: NIST. special publication 800-22.

Spencer, J. (2015). Pseudorandom bit generators from enhanced cellular automata. *Cellular Automata*, 10(3–4):295–317.

Tomassini, M., Sipper, M., and Perrenoud, M. (2000). On the generation of high-quality random numbers by two-dimensional cellular automata. *IEEE Transactions on Computers*, 49(10):1146–1151.

WANG, X.-Y. and YANG, L. (2012). Design of pseudo-random bit generator based on chaotic maps. *International Journal of Modern Physics B*, 26(32):1250208.