



Jul 12th, 9:50 AM - 10:10 AM

Number of Patterns in Lempel-Ziv Algorithm Applied to Iterative Maps and Measured Time Series

G. Mimić

University of Novi Sad, gordan.mimic@df.uns.ac.rs

I. Arsenić

University of Novi Sad, guto@polj.uns.ac.rs

D. T. Mihailović

University of Novi Sad, ilija@polj.uns.ac.rs

Follow this and additional works at: <https://scholarsarchive.byu.edu/iemssconference>



Part of the [Civil Engineering Commons](#), [Data Storage Systems Commons](#), [Environmental Engineering Commons](#), [Hydraulic Engineering Commons](#), and the [Other Civil and Environmental Engineering Commons](#)

Mimić, G.; Arsenić, I.; and Mihailović, D. T., "Number of Patterns in Lempel-Ziv Algorithm Applied to Iterative Maps and Measured Time Series" (2016). *International Congress on Environmental Modelling and Software*. 63.

<https://scholarsarchive.byu.edu/iemssconference/2016/Stream-A/63>

This Event is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Number of Patterns in Lempel-Ziv Algorithm Applied to Iterative Maps and Measured Time Series

G.Mimić^a, I. Arsenić^b and D. T. Mihailović^b

^aDepartment of Physics, Faculty of Sciences, University of Novi Sad, Dositej Obradovic Sq. 4, 21000 Novi Sad, Serbia; gordan.mimic@df.uns.ac.rs

^bDepartment of Field Crops and Vegetable, Faculty of Agriculture, University of Novi Sad, Dositej Obradovic Sq. 8, 21000 Novi Sad, Serbia; guto@polj.uns.ac.rs, ilija@polj.uns.ac.rs

Abstract: Some tests with information measure based on Kolmogorov complexity, using Lempel-Ziv algorithm, were performed. Information measure, which indicates randomness, was applied on the modeled and measured time series. The logistic map iterations were used as the modeled time series and the measurements of velocity data from the turbulent flow collected in a laboratory channel, were used as the measured one. The complexity counter, indicating the number of distinct patterns in the time series, was computed. An analysis has shown that the number of patterns in the measured time series is increasing with the time series length, but having a saturation value in the modeled time series. Presented results indicate on rather more complex dynamic of the velocity of turbulent flow time series than in the deterministically created chaotic time series.

Keywords: turbulent flow, logistic map, time series analysis, randomness, patterns, Kolmogorov complexity

1 INTRODUCTION

Various methods of time series analysis are used in science today, aiming to investigate and understand characteristics of time series (Tang et al., 2015). Time series are consisted of data presenting the evolution of some system, sequentially passing from one state to the other, i.e. the dynamics of the system. A new discipline named nonlinear dynamical analysis is established to examine the behavior of nonlinear systems, and in particular complex systems, using information measures (Stam, 2005). Complex systems exhibit irregular behavior that is hardly predictable while the dynamic of the systems with low level of complexity is periodic, more regular and predictable. One of the main features of complexity is randomness (Grassberger, 2012), which is in the focus of this paper. Kolmogorov complexity has already been used in the investigation of the randomness in time series that come from measurements of different quantities in natural systems (Mihailović et al., 2015a). Many natural systems exhibit irregular behavior, especially fluid flows, among others. Information measures have been applied to examine geophysical fluid dynamics (Wijesekera and Dillion, 1997; Wesson et al., 2003) and velocity of the turbulent flow time series (Ichimiya and Nakamura, 2013; Mihailović et al., 2014; Mihailović et al., 2015b). Main goal of applying these measures is a quantification of the level of randomness in time series, searching for the hidden rules in the dynamics of the systems. In this paper beside the level of randomness, the complexity counter, presenting the number of different patterns in time series, was analyzed. For the comparison, two types of time series were used, the modeled and the measured one. Modeled time series was obtained by the iterations of logistic map, as a map showing chaotic behavior. Measured time series was obtained by measurement of the velocity of the turbulent water flow in a laboratory channel. The paper is structured in the following way. In Section 2 we have described origin of the data, as well as the Lempel-Ziv algorithm. In Section 3 the results of calculations are discussed, while the Section 4 is left for the conclusions.

2 MATERIALS AND METHODS

The logistic map is a typical example of a simple dynamical system exhibiting very complex behavior. Many processes in biology, ecology and economics can be described by logistic map (May, 1976). In particular, this equation has many applications in physics (Mimić et al., 2013, among others). It is given in the form of the iterative map

$$x_{n+1} = rx_n(1 - x_n) \quad (1)$$

where x is ranged in the interval between 0 and 1, while r is the control parameter ranging in the interval between 0 and 4. The most interesting behavior logistic map shows after the parameter r reaches 3, when the bifurcations start, leading to the chaos (Sprott, 2003). In all of the numerical experiments $x_0=0.2$ was selected for the initial condition. Parameter r was ranging between 3.8 and 3.9 since the highly developed chaos, with some windows of stability.

The experiment of turbulent flow was performed in a laboratory channel at the University of Federico II, Naples (Italy).^{*} The channel was 8 m long and 0.4 m wide (Figure 1) with the following experimental conditions: flow rate $Q=33$ l/s, slope of the channel $i=1\%$ and the depth of the flow $h_f=6.3$ cm. The velocity measurements were carried out in twenty-five vertical locations using a laser Doppler velocimeter (LDV) equipped with a frequency shifter and a frequency tracker. The sampling rate was 2000 Hz for 135 s. The turbulent flow velocity data were post-processed using the LabView software to derive the distribution of time-averaged velocity and standard deviation related to turbulence intensity. The number of the samples was $N=270000$.



Figure 1. Laboratory channel

The Kolmogorov complexity $K(x)$ of an object x is the length, in bits, of the smallest program that when run on an Universal Turing Machine (U) prints object x , which is maximized for random strings (Evans, et al., 2002). Detailed explanation of the Kolmogorov complexity can be found in Li and Vitanyi (1997). Main problem with the Kolmogorov complexity is that it is not computable exactly. For that reason Lempel and Ziv created algorithm which presents its computable interpretation (Lempel and Ziv, 1976). The idea is to code any time series with length N into the string consisted of 0 and 1, following the rule: if the value is less than the threshold then it will be written 0 on the corresponding place in time series, otherwise it will be written 1. Commonly, the mean value of the time series is used as the threshold (Zhang et al., 2001). Now the algorithm is counting minimum number of the distinct patterns in the given sequence. The complexity counter c will be described following the example given in Kaspar and Schuster (1987). Thus, for the sequence 0010, it will be

- 1) First digit is always the first pattern, which implies $\rightarrow 0\cdot$
- 2) $S=0, Q=0, SQ=00, SQ\pi=0, Q \in v(SQ\pi) \rightarrow 0\cdot0$
- 3) $S=0, Q=01, SQ=001, SQ\pi=00, Q \notin v(SQ\pi) \rightarrow 0\cdot01\cdot$
- 4) $S=001, Q=0, SQ=0010, SQ\pi=001, Q \in v(SQ\pi) \rightarrow 0\cdot01\cdot0$.

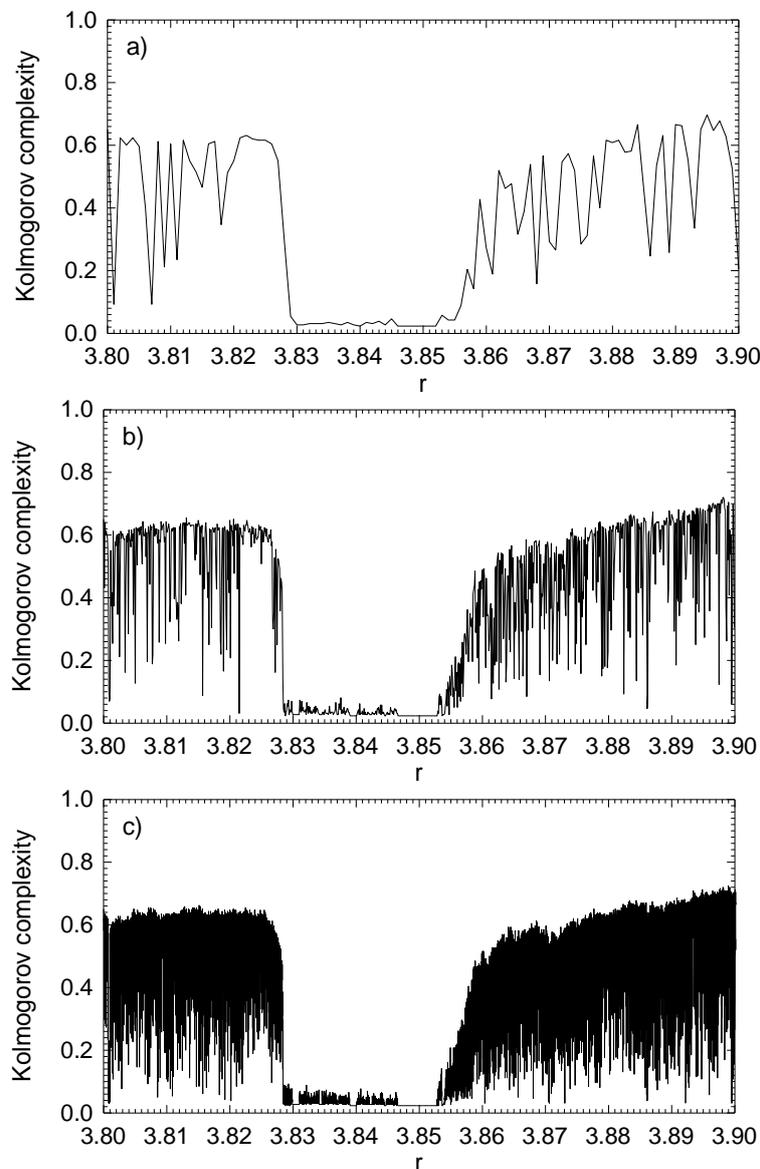
The complexity counter c is the number of distinct patterns separated by dots, while the last 0 in the sequence presents the rest. In this case $c=3$. The complexity counter c is a function of the length of the sequence N , so further it will be noted as $c(N)$. The ultimate value of the complexity counter is

^{*}Data provided by Prof. Carlo Gualtieri and Prof. Paola Gualtieri

$b(N) = N/\log_2 N$, which stands for the maximal possible entropy of the time series with length N . Normalized information measure $C_k(N)$ is defined as $C_k = c(N) / b(N)$. For a nonlinear time series $C_k(N)$ varies between 0 and 1.

3 RESULTS AND DISCUSSION

Chaos and turbulence are both related to irregular behavior, even though the details of their underlying rules are quite different. Chaos has very accurate mathematical definition while the turbulence is physical regime of fluid flow, without an accurate mathematical definition. Nevertheless, looking at the specific pictures, one can see the same kind of phenomenon. For example, there is a cellular automaton able to create the same patterns in the flow as ones observed in the experiments with turbulent flow (Wolfram, 2002). However, the origin of randomness is different. For the chaotic systems it lies in sensitive dependence on initial condition, while for the turbulent fluid it must be intrinsically generated inside the system itself. The notions of order and chaos in fluid flows have already been analyzed (Narasimha, 1987). Here, we wanted to compare the level of randomness in turbulent flow and chaotic system. For that reason we chose the logistic map as the simple example of chaotic behavior. Firstly, Kolmogorov complexity of the logistic map was tested depending on the different number of digits presented in the parameter r . The results are depicted in Figure 2.



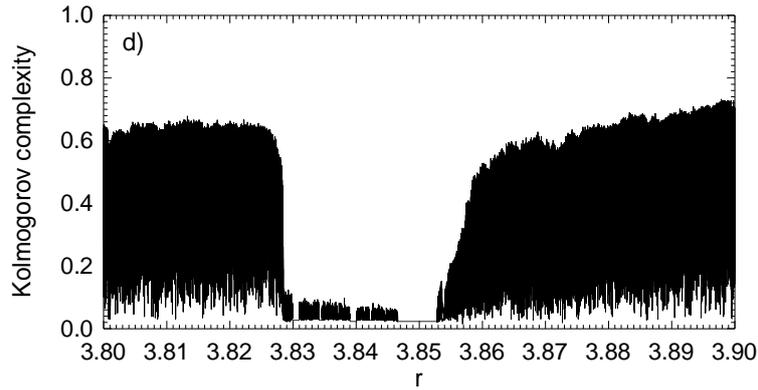


Figure 2. Kolmogorov complexity of the time series consisted of 3000 values, generated with the logistic map using $x_0=0.2$ and $r \in [3.8, 3.9]$ with the step: a) 10^{-3} , b) 10^{-4} , c) 10^{-5} and d) 10^{-6}

Logistic map exhibits highly developed chaotic behavior when r is ranged between 3.8 and 3.9. Kolmogorov complexity has values greater than 0.6 for the specific r . Yet, small changes in r results in drop of the Kolmogorov complexity. With decreasing the step for r for one order of value, the oscillations of the randomness of time series are more pronounced. Also, the loss of information about randomness is present if the step for r is larger. When r is in the interval between 3.83 and 3.85 there is the decrease of the randomness due to the window of stability. Now, the dependence of Kolmogorov complexity on the length of time series was tested. The idea was to increase the length of time series with steps of 1000, up to 270000, since the measured time series has the exact number of data. The initial condition for the logistic map was $x_0=0.2$ and parameter $r=3.8$. The velocity of the turbulent flow time series was obtained by measurement at the first level of the channel, 0.1 cm from the bottom. The results are depicted in Figure 3.

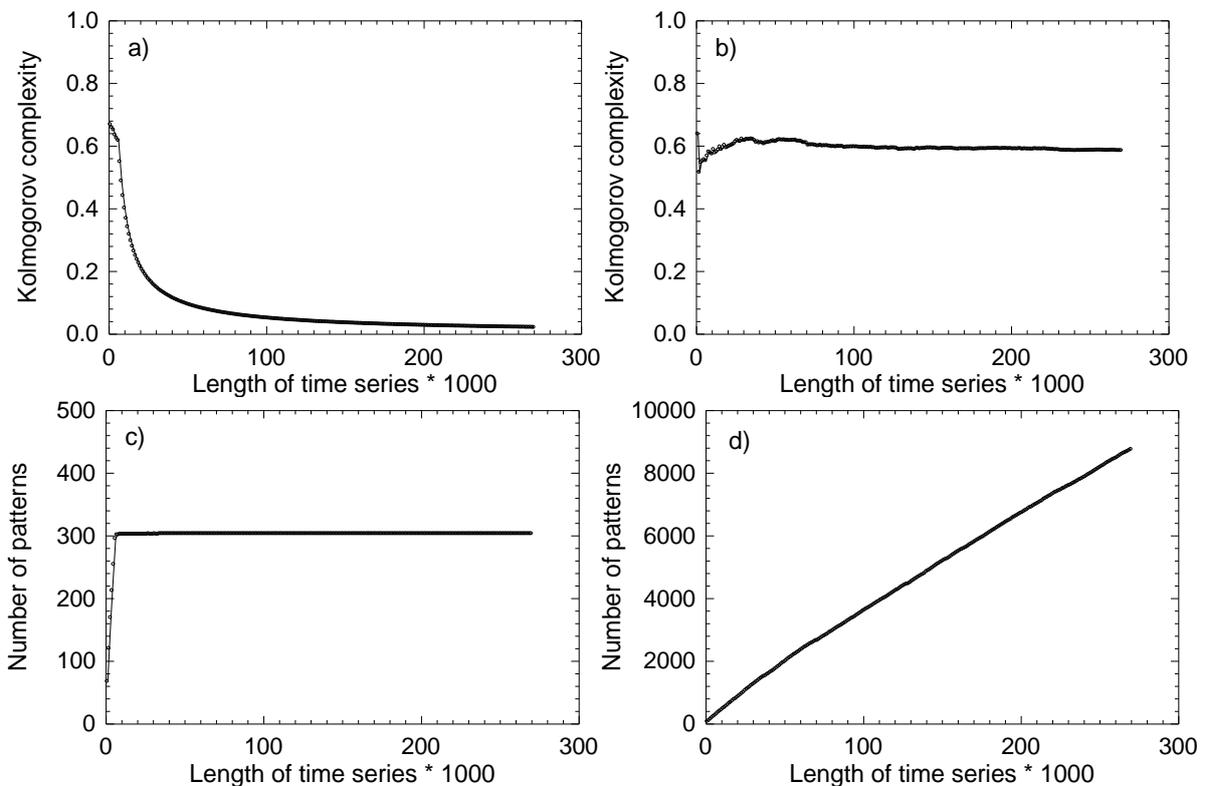


Figure 3. Kolmogorov complexity of the logistic map (a), and the velocity of the turbulent flow time series (b), and the number of patterns of the logistic map (c) and the velocity of the turbulent flow time series (d), in the time series of different length

Surprisingly, increasing number of iterations of the logistic map produces time series with the decreasing level of randomness (Figure 3a). One would expect that for highly developed chaos the level of randomness is uniform as for the velocity of the turbulent flow (Figure 3b) and it does not depend on the length of time series. Explanation for this could be found in Figure 3c. The number of distinct patterns, i.e. the complexity counter $c(N)$, has the saturation value. Kolmogorov complexity is decreasing with N because denominator $b(N)$ is increasing and $c(N)$ has a constant value. In the case of turbulent flow, the number of distinct patterns is increasing with N , which was expected (Figure 3d). It means that the turbulent flow exhibit more randomness than the chaotic behavior of the logistic map. This is very important for the reason that any numerical test with the logistic map demands the drop of a few hundreds or thousands iterations due to the transition behavior of the system (Mihailović, 2015a). Now, it is clear that one cannot drop arbitrary number of iterations in numerical experiments since there is the exact saturation value of the number of patterns in the logistic map time series, for a given initial condition and parameter r . The complexity counter has detected 301 patterns in 6112 iterations and in the time series with 270000 iterations there are only two more new patterns. Presented results point out to the fact that long term behavior described with the logistic map is not completely random and that the dynamics of logistic map is governed by some “hidden” rule. Note that the simulations were obtained in FORTRAN with variable x defined as REAL. If the variable is defined as DOUBLE PRECISION different number of patterns would be contained, but it would also reach some saturation value.

Calculating the complexity counter for the turbulent flow time series, measured at the different distance from the bottom shows the vertical profile of randomness in the fluid (Figure 4). Number of distinct patterns in time series varies between 5000 and 9000. Randomness is decreasing up to 1 cm from the bottom, it remains similar up to 3 cm from the bottom and then it increases up to the surface of the fluid. This could be explained by the existence of coherent structures in turbulent flow, organized in the layer few centimeters above the bottom (Hussain, 1983). Kolmogorov complexity has exactly the same vertical profile as the one for the complexity counter, with the values ranging in the interval 0.384–0.602, where the difference between the values origins in the nature of turbulent flow, stressing the randomness featured with the turbulence.

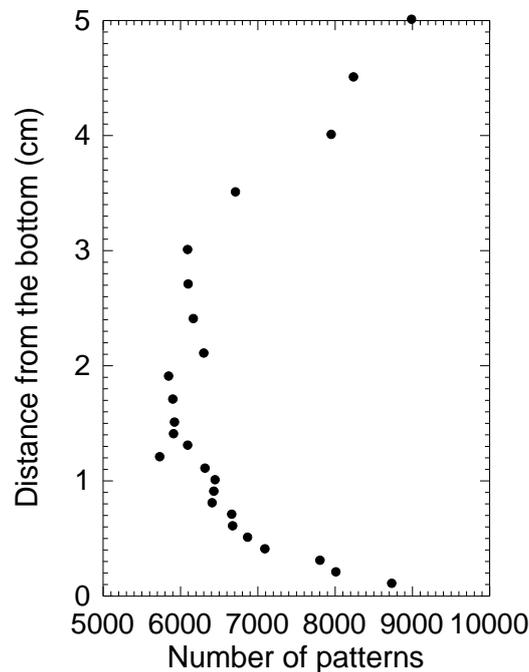


Figure 4. Number of distinct patterns in time series measured at different distance from the bottom in turbulent flow

4 CONCLUSIONS

In this paper, some tests with Kolmogorov complexity, as the information measure of randomness, were performed. Two types of time series, the modeled and measured one, were used. Iterations of the logistic map were used as the modeled time series while the measured time series was created from the measurements of the velocity data in turbulent flow from the laboratory channel. The logistic map exhibits chaotic behavior after the control parameter r reaches 3. Varying parameter r with different steps between 3.8 and 3.9 Kolmogorov complexity varies a lot. For the fixed value of r and the initial condition, randomness is decreasing with the length of time series. Reason for that lies in the fact that there is the saturation value of the number of distinct patterns in time series obtained by iterating the logistic map. This is important for the numerical experiments with logistic map in which it is common to drop few hundreds or thousands of iterations to pass the transition behavior. It is shown that the drop cannot be taken arbitrary. On the other hand, the number of distinct patterns in turbulent flow time series increases almost linearly with the length of time series. Presented results indicate on a more complex dynamic of the turbulent flow time series than the chaotic time series.

ACKNOWLEDGMENTS

This paper was realized as a part of the project No. 43007 financed by the Ministry of Education and Science of the Republic of Serbia for the period 2011-2015. The authors are very grateful to the Prof. Carlo Gualteri and Prof. Paola Gualtieri, from the University of Federico II, Naples (Italy) for providing us with the data of turbulent flow from the laboratory channel.

REFERENCES

- Evans, S.C., Hershey, J.E., Saulnier, G., 2002. Kolmogorov complexity estimation and analysis. 6th World Conference on Systemics, Cybernetics and Informatics, July 14-18, Orlando, Florida, USA.
- Grassberger, P., 2012. Randomness, Information and Complexity, arXiv: 1208.3459 [physics.data-an].
- Hussain, A. K. M. F., 1983. Coherent Structures - Reality and Myth. *Phys. Fluids* 26, 2816-2850.
- Ichimiya, M., Nakamura, I., 2013. Randomness representation in turbulent flows with Kolmogorov complexity (In laminar-turbulent transition due to periodic injection in an inlet boundary layer in a circular pipe). *J. Fluid Sci. Techn.* 8, 407-422.
- Kaspar, F. and Schuster, H.G., 1987. Easily calculable measure for the complexity of spatiotemporal patterns. *Phys. Rev. A* 36, 842-848.
- Lempel, A., Ziv, J., 1976. On a complexity of finite sequence. *IEEE Trans. Inform. Theory* 22, 75-81.
- Li, M., Vitanyi, P., 1997. *An Introduction to Kolmogorov Complexity and its Applications*, 2nd ed., Springer Verlag, Berlin, Germany.
- May, R.M., 1976. Simple mathematical models with very complicated dynamics. *Nature* 261 (5560), 459-467.
- Mihailović, D.T., Nikolić-Đorić, E., Drešković, N., Mimić, G., 2014. Complexity analysis of the turbulent environmental fluid flow time series. *Physica A* 395, 96-104.
- Mihailović, D. T., Mimić, G., Nikolić-Djorić, E., Arsenić, I., 2015a. Novel measures based on the Kolmogorov complexity for use in complex system behavior studies and time series analysis. *Open Phys.* 13, 1-14.
- Mihailović, D.T., Mimić, G., Drešković, N., Arsenić, I., 2015b. Kolmogorov Complexity Based Information Measures Applied to the Analysis of Different River Flow Regimes. *Entropy* 17(5), 2973-2987.
- Mimić, G., Mihailović, D.T., Budinčević, M., 2013. Chaos in computing the environmental interface temperature: nonlinear dynamic and complexity analysis of solutions. *Mod. Phys. Lett. B* 27(26), 1350190.
- Narasimha, R., 1987. Order and chaos in fluid flows. *Curr. Sci. India* 56(13), 629-645.
- Sprott, J.C., 2003. *Chaos and time-series analysis*, Oxford University Press, Oxford, UK.
- Stam, C.J., 2005. Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field. *Clin. Neurophysiol.* 116 (10), 2266-2301.
- Tang, L., Lv, H., Yang, F., Yu, L., 2015. Complexity testing techniques for time series data: A comprehensive literature review. *Chaos Soliton. Fract.* 81, 117-135.

- Wesson, H. K., Katul, G. G., Siqueira, M., 2003. Quantifying organization of atmospheric turbulent eddy motion using nonlinear time series analysis. *Bound-Lay. Meteorol.* 106, 507–525.
- Wijesekera, H. W., Dillion, T. M., 1997. Shannon entropy as an indicator of age for turbulent overturns in the ocean thermoclines. *J. Geophys. Res.* 102, 3279–3291.
- Wolfram, S., 2002. *A new kind of science*, Wolfram Media Inc., Champaign, IL, USA.
- Zhang, X. S., Roy, R. J., Jensen, E. W., 2001. EEG complexity as a measure of depth of anesthesia for patients. *IEEE Trans. Biomed. Eng.* 48, 1424-1433.