Chaotic Modeling and Simulation (CMSIM) 2: 207–218, 2016

Permutation, Linear Combination and Complexity of Short Time Series

Zoran Rajilić

University of Banja Luka, M. Stojanovića 2, 51000 Banja Luka, Republic of Srpska, Bosnia and Herzegovina
(E-mail: `ZoranRajilic@netscape.net`)

Abstract. A new manner of estimating complexity, suitable for short time series, is proposed. Final part of a time series we represent as linear combination of previous subseries. Permutations inside subseries affect constants of the linear combination. Complexity Cmp is defined thorough the changes of these constants. In other to justify such approach, Cmp is related to the number of different frequencies in regular oscillations, Lyapunov exponent, level of noise, accuracy of the Monte Carlo integration and coefficient of nonlinearity in the acting force expression. Increasing of each of these five quantities is followed by increasing of Cmp. If the level of noise and Lyapunov exponent are low enough, we distinguish short time series from clean noise. Considering nonlinear damped oscillations of a particle, we mark an essential property of chaos: if coefficient of nonlinearity is large enough, complexity of chaotic motion is inside the interval of noise complexity, although there is not fluctuating force acting on the particle. Computing Cmp as a function of the time series final point, we can forecast this point if there is a sharp minimum. We foreacast and estimate the forecasting reliability without knowledge about the rules producing time series. Values of Cmp for some real time series are computed.
Keywords: Complexity, Time series, Chaos, Noise, Permutation, Linear combination, Lyapunov exponent, Nonlinearity, Forecasting.

1 Introduction

Bandt and Pompe proposed permutation entropy as a complexity measure for time series, based on comparison of neighboring points [1,3]. Permutation entropy is adapted for estimating complexity of short time series by changing the time delay [18]. Roots of complexity like dimension, nonlinearity and nonstationarity are related to the generating process, while roots of complexity like noise, aggregation and finite length are related to measurement [8].

Complexity of real time series is correlated with predictability. It is hard to forecast crisis from a short and noisy economic time series [6,9,13]. We can forecast some time series using, for example, wavelets [4,10], neural networks [2,7] or linear combination of logistic map [12]. It is of practical interest in

Received: 20 March 2015 / Accepted: 12 December 2015

ISSN 2241-0503

finance [5] and technical systems optimization, also for fundamental research, because the relation between prediction and learning the rule which has produced time series is not simple. Chaotic time series can be learned but not predicted, while quasiperiodic time series can be predicted but not learned [7]. Here we consider time series of length 110. We make permutations in ten subseries which linear combination is equal to the eleventh subseries. The effect of permutations is change of the linear combination constants. This change determines complexity *Cmp*. After investigating of our approach adequacy, computing *Cmp* of different regular, chaotic and stochastic time series, we show that our conception is useful in forecasting. We can approximate the 110th point and estimate the reliability of this forecasting. *Cmp* is computed for many real time series.
The basic difference between permutation entropy and *Cmp* is the following. One computes permutation entropy counting the existing permutations in the time series, while for *Cmp* we make a new time series by permutations inside the original time series.

2 Definition of *Cmp*

We divide time series A_1, A_2, ..., A_{110} into eleven subseries. First ten subseries are

$$\begin{aligned} F_{1,j} &= A_j \\ F_{2,j} &= A_{j+10} \\ F_{3,j} &= A_{j+20} \\ &\cdots \\ F_{10,j} &= A_{j+90} \end{aligned} \tag{1}$$

where $j = 1, 2, \ldots, 10$. The eleventh subseries we represent as linear combination of previous ones. For many nontrivial time series the equations

$$\begin{aligned} A_{101} &= \sum_{i=1}^{10} c_i F_{i,1} \\ A_{102} &= \sum_{i=1}^{10} c_i F_{i,2} \\ &\cdots \\ A_{110} &= \sum_{i=1}^{10} c_i F_{i,10} \end{aligned} \tag{2}$$

are independent and we can find out constants of the linear combination $< c_1, c_2, \ldots, c_{10} >$. We now make permutations inside first ten subseries and get new subseries

$$\begin{aligned} F'_{1,1} = A_{10},\, F'_{1,j} = A_{j-1} \\ F'_{2,1} = A_{20},\, F'_{2,j} = A_{j+9} \\ F'_{3,1} = A_{30},\, F'_{3,j} = A_{j+19} \\ \cdots \\ F'_{10,1} = A_{100},\, F'_{10,j} = A_{j+89} \end{aligned} \tag{3}$$

where $j = 2, 3, \ldots, 10$. The equations

$$\begin{aligned} A_{101} = \sum_{i=1}^{10} c'_i F'_{i,1} \\ A_{102} = \sum_{i=1}^{10} c'_i F'_{i,2} \\ \cdots \\ A_{110} = \sum_{i=1}^{10} c'_i F'_{i,10} \end{aligned} \tag{4}$$

yield new constants of the linear combination $< c'_1, c'_2, \ldots, c'_{10} >$. Using 2-norm of vectors, we define complexity.

$$Cmp = -ln \frac{|| < c'_1, c'_2, \ldots, c'_{10} > - < c_1, c_2, \ldots, c_{10} > ||}{|| < c_1, c_2, \ldots, c_{10} > ||} \tag{5}$$

Complexity of many regular, chaotic and stochastic time series is computed. Minimal value of Cmp increases in direction regularity–chaos–stochasticity (figure 1).

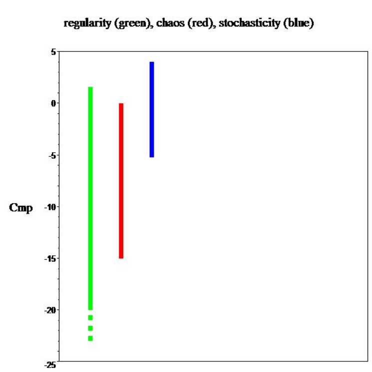

Fig. 1. Intervals where values of complexity are placed, for regular, chaotic and stochastic time series of length 110.

3 Regular Oscillations

Blue, red and green lines in figure 2 represent complexity of regular time series

$$\sum_{j=1}^{N} cos\big((2.9+0.6j)i-0.041j\big)$$
$$\sum_{j=1}^{N}(-0.9)^j cos\big((2.8+0.7j)i-0.03j\big) \tag{6}$$
$$\sum_{j=1}^{N}(-1.1)^j cos\big((2.1+0.8j)i-0.09j\big)$$

where i is time. High complexity corresponds to large number of different frequencies N.

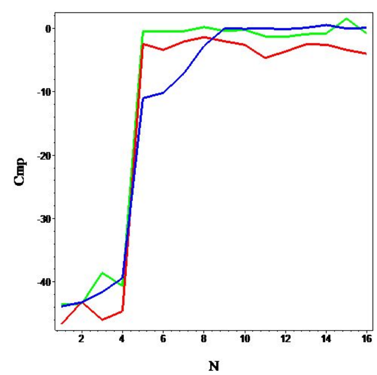

Fig. 2. Number of different frequencies is in high correlation with complexity.

4 Chaos and Noise

Computing complexity for 14000 of realizations of Gaussian noise we found

$$-5.221 \le Cmp \le 4.039 \tag{7}$$

For time series generated by Feigenbaum map (table 1) we find:
(1) $Cmp \le -6.46$, if $\lambda \le 0.18$ (we distinguish from clean noise) and
(2) $Cmp \ge -1.7$, if $\lambda \ge 0.19$ (we can not distinguish from clean noise).
If a chaotic time series contains noise (table 2), we can see it as chaotic, computing Cmp, if the level of noise and Lyapunov exponent are low enough.

5 Monte Carlo Integration

If we compute integrals using randomly distributed points (x_j, y_j) $(j = 1, 2, ..., 110)$, the integration is more accurate for more complex stochastic time series x_j and y_j (table 3).

q	λ	Cmp	q	λ	Cmp
1.3	-0.425	-41.59	1.46	0.19	-0.9
1.35	-0.097	-36.36	1.794	0.351	-0.507
1.402	0.028	-8.31	1.57	0.361	0.008
1.405	0.054	-6.46	1.68	0.403	-0.825
1.41	0.094	-8.85	1.83	0.481	-0.01
1.42	0.11	-8.05	1.89	0.548	-0.37
1.45	0.17	-6.96	1.94	0.585	-1.7
1.44	0.18	-7.84	1.99	0.654	-0.5

Table 1. Lyapunov exponent and complexity for time series generated by $z_n = 1 - qz_{n-1}^2$, with $z_0 = 0.7$.

q	λ (without noise)	Cmp (with 0.1% of noise)	Cmp (with 1% of noise)
1.405	0.054	-6.36	-3.55
1.402	0.028	-6.19	-3.7
1.406	0.069	-5.41	-3.6
1.44	0.18	-3.4	-2.95
1.94	0.585	-1.67	-1.44
1.99	0.654	-0.49	-0.43

Table 2. Lyapunov exponent and complexity for time series containing noise.

integral	Cmp_{x1}	Cmp_{y1}	$error_1(\%)$	Cmp_{x2}	Cmp_{y2}	$error_2(\%)$
$\int_0^{1.1} e^{-v} cos^2 5v dv$	-0.967	-0.812	5.9	-0.735	0.937	0.9
$\int_0^{2.53} \left[0.3v + \frac{2}{\sqrt{5+3sin4v}}\right] dv$	-1.42	-0.82	14.16	-0.26	-0.09	10.82
$\int_0^{5.41} e^{v-(v-2)^2} dv$	-0.992	-0.711	17.12	-0.472	0.178	5.04

Table 3. Complexity and errors of Monte Carlo integration.

6 Nonlinear Damped Oscillations

Here we analyze the coordinate of a particle $x(0.1j)$, with $j = 1, 2, ..., 110$ and acting force

$$F = -x - \beta x^3 - 0.005v \tag{8}$$

where β is coefficient of nonlinearity and v is velocity. Three types of motion are found (table 4):
(1) regular, where $Cmp < -11$ and $\lambda < 0$,
(2) chaotic with $Cmp < -5.221$ and $0 < \lambda \leq 0.063$, we distinguish from noise because it is outside of the interval of noise complexity and
(3) chaotic with large β, $-5.221 < Cmp < 4.039$ and $0.089 \leq \lambda$, we can not distinguish from noise computing complexity.
If cofficient of nonlinearity is large enough, complexity of chaotic motion is inside the interval of noise complexity, although there is not fluctuating force acting on the particle. It is an essential property of chaos.

β	Cmp	λ	β	Cmp	λ	β	Cmp	λ
1	-14.86	-0.032	7	-11.73	0.003	5941	-3.03	0.089
2	-17.80	-0.027	8	-11.82	0.007	6000	-1.43	0.098
3	-15.15	-0.015	9	-12.68	0.012	6300	-3.70	0.097
4	-13.92	-0.003	10	-14.58	0.010	6500	-3.66	0.095
5	-12.23	-0.002	20	-12.10	0.021	6501	-3.64	0.096
6	-11.14	-0.00009	500	-7.55	0.063	6541	-3.47	0.097

Table 4. Coefficient of nonlinearity, complexity and Lyapunov exponent of nonlinear damped oscillations. We can see three types of motion: regular, chaotic with $Cmp < -5.221$ and chaotic with $-5.221 < Cmp < 4.039$. Cmp is in correlation with β and λ. Here $x(0) = 0.1$ and $v(0) = -0.2$.

7 Forecasting of the 110th Point

Let us assume that we know $A_1, A_2, \ldots, A_{109}$, but do not know A_{110}. Computing Cmp for different possible values of A_{110}, not far from $A_1, A_2, \ldots, A_{109}$, and looking for minimum, we try to forecast A_{110}. We expect that if there are any rules in $A_1, A_2, \ldots, A_{109}$, complexity will be minimal for A_{110} not breaking these rules. In some cases forecasting is successful (figures 3 and 4). If complexity is high, we can not forecast. The forecasting reliability is connected with the minimum sharpness.

Fast computing with short time series and simple predictability and the forecasting reliability estimation are advantages of proposed forecasting method. Moreover we can forecast the 110th point without any knowledge about the rule producing $A_1, A_2, \ldots, A_{109}$.

Here we consider nonlinear damped oscillations (table 5) and time series generated by Feigenbaum map (figures 3 and 4). In section 8 we try to forecast real time series (figures 6 and 7). Red line marks true value of A_{110}.

β	true	forecasted	minimal Cmp	β	true	forecasted	minimal Cmp
5000	0.0976	0.0976	-7.81	8500	0.1004	0.1004	-7.15
5600	-0.0974	-0.0972	-5.23	8600	0.0863	0.0864	-6.07
5635	-0.0881	-0.0881	-5.60	8700	0.0519	0.0521	-6.92
6000	0.0803	0.0820	-3.15	8711	0.0476	0.0478	-6.98
7500	0.0256	0.0256	-4.66	8719	0.0445	0.0446	-6.93
8148	-0.0146	-0.0146	-8.43	8800	0.0121	0.0123	-6.29

Table 5. Forecasting of 110th point for nonlinear damped oscillations. We can compare true and forecasted value of $x(11)$. For minimal Cmp equals -3.15, the error is 2.1%. For significantly lower minimal Cmp the error is significantly lower.

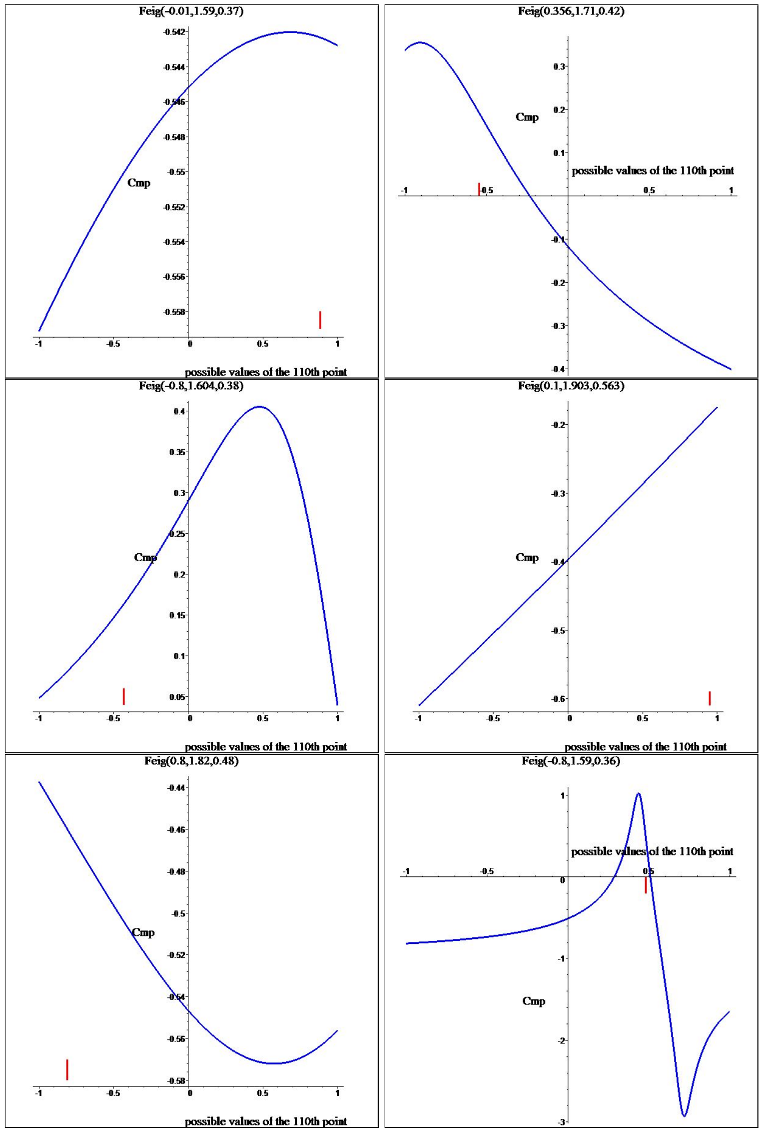

Fig. 3. Cmp as function of z_{110}. Feig(-0.01,1.59,0.37) means that time series is generated by $z_n = 1 - qz_{n-1}^2$, with $z_0 = -0.01$, $q = 1.59$ and Lyapunov exponent $\lambda = 0.37$. Red line marks true value of z_{110}. Here $\lambda \geq 0.36$ and $Cmp > -3$. There is not sharp minimum and forecasting is unsuccessful. Only in the case Feig(-0.8,1.59,0.36) we have roughly approximate forecasting.

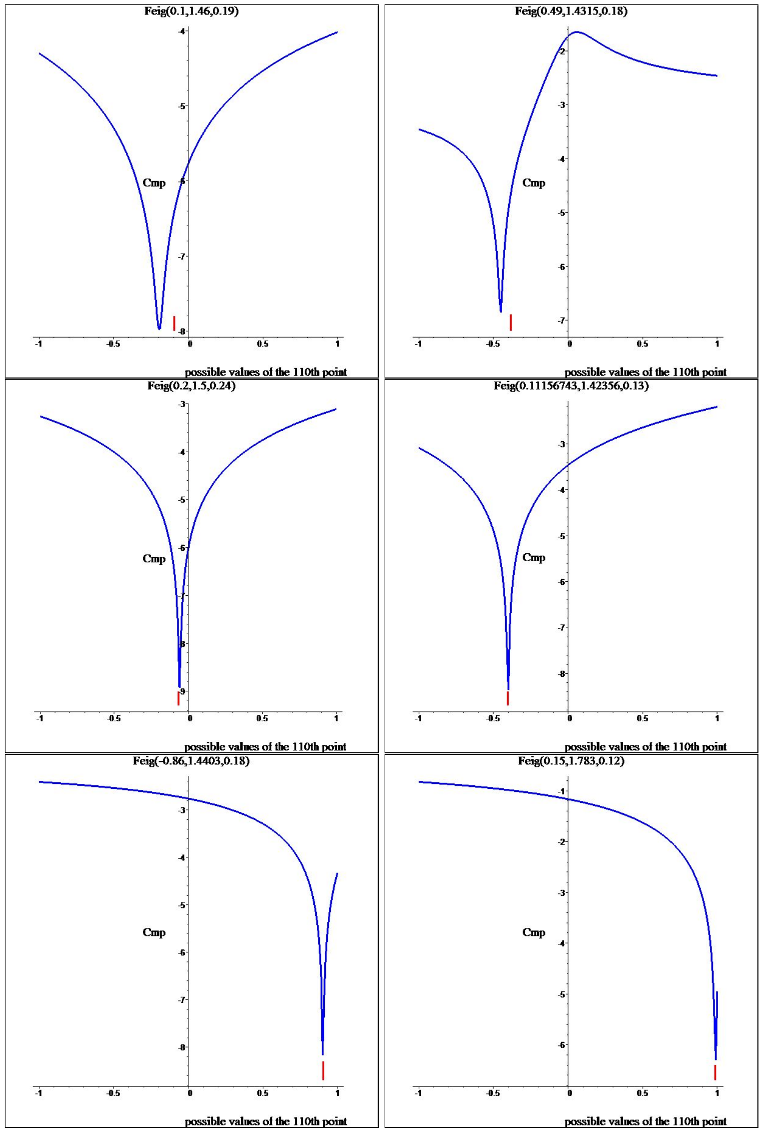

Fig. 4. Here $\lambda \leq 0.24$ and minimal Cmp is approximately from -9 to -6. In some cases forecasting is very successful.

8 Real Time series

8.1 Gross Domestic Product

Here we consider quarterly GDP in UK from 1955 until 2014 [15]. We compute $Cmp(p)$ using

$$A_i = GDP_{i+p}, \quad i = 1, \ldots, 110 \tag{9}$$

and find average values

$$\frac{1}{11} \sum_{p=0}^{10} Cmp(p) = -2.366 \tag{10}$$

for years 1955–1984, and

$$\frac{1}{10} \sum_{p=120}^{129} Cmp(p) = -0.827 \tag{11}$$

for years 1985–2014. This is significantly higher in comparison with the former period.

8.2 Stock market index S&P 500

For a year we compute average complexity of the stock market index S&P 500 time series [16] (figure 5)

$$< Cmp >= \frac{1}{141} \sum_{p=0}^{140} Cmp(p) \tag{12}$$

using

$$A_i = SP_{i+p}, \quad i = 1, 2, \ldots, 110; \quad p = 0, 1, \ldots, 140 \tag{13}$$

Approximate forecasting is possible (figure 6). The results are denoted by S&P500(year,p).

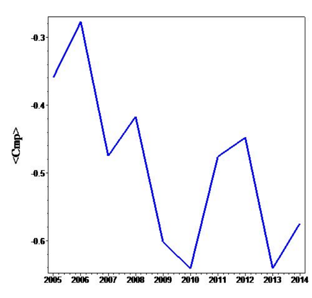

Fig. 5. Average complexity of the index S&P 500. $< Cmp >$ is very high, but after 2008 complexity is relatively low.

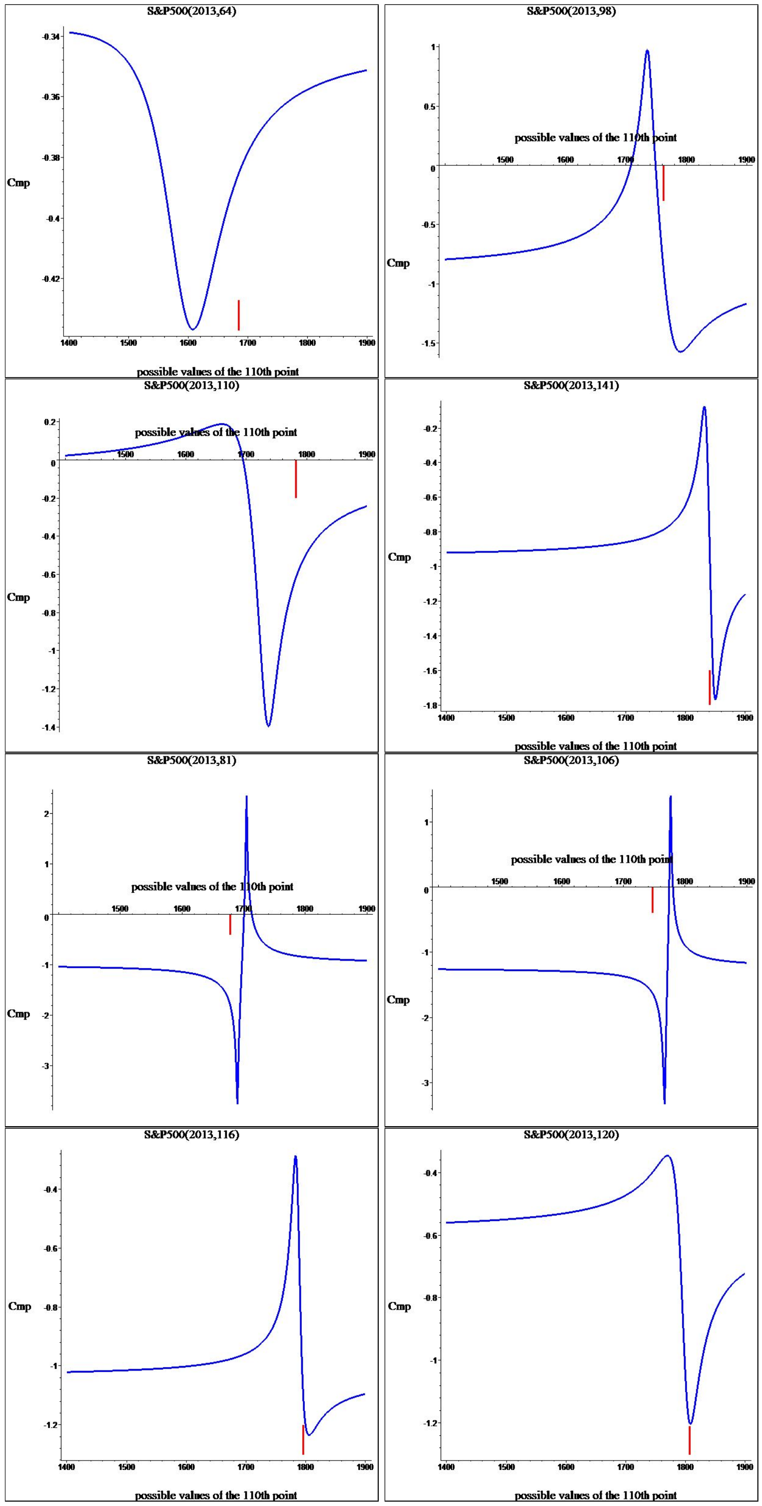

Fig. 6. Forecasting of the stock market index S&P 500 in 2013. It is better for deeper and sharper minimum.

8.3 RLC Circuit

The output voltage K_i is measured in the experiment with periodically driven RLC circuit [11,14]. The largest Lyapunov exponent is 0.33. We take

$$A_i = K_{i+p}, \quad i = 1, 2, \ldots, 109 \tag{14}$$

and try to forecast A_{110}. The results are denoted by Kodba(p) (figure 7).

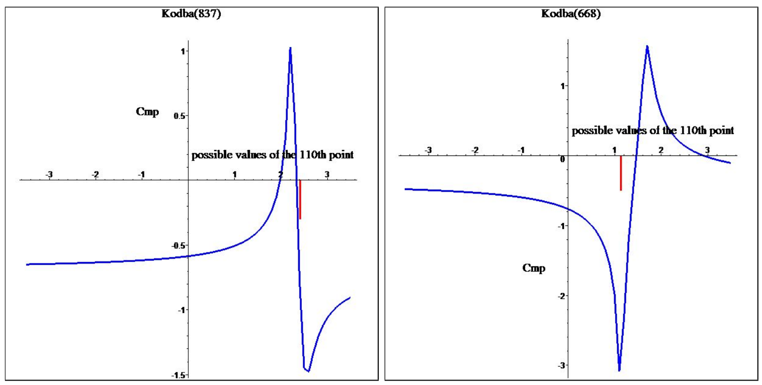

Fig. 7. Forecasting of the output voltage A_{110}, measured in the experiment with periodically driven RLC circuit. Forecasting is very accurate for deep and sharp minimum of complexity.

8.4 EEG Time Series

EEG time series EEG_j ($j = 1, 2, \ldots, 3595$) is recorded on a patient undergoing ECT therapy for clinical depression [17]. We take

$$A_i = EEG_{i+p}, \quad i = 1, 2, \ldots, 110 \tag{15}$$

and find out average Cmp in the interval $s \le p \le s + 90$ (figure 8).

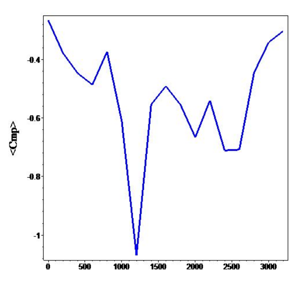

Fig. 8. Average complexity of the EEG time series, in the interval $s \le p \le s + 90$. We find high jumping $< Cmp >$.

9 Conclusion

Cmp, as a measure of complexity, is defined using permutation and linear combination. *Cmp* is related with (i) number of different frequencies in regular oscillations, (ii) Lyapunov exponent of the chaotic time series, (iii) level of noise, (iv) accuracy of the Monte Carlo integration, (v) coefficient of nonlinearity in the acting force expression. If complexity is low enough, it is possible to forecast using *Cmp* and also to estimate the forecasting reliability. Reliability of forecasting is large if mimimum of *Cmp* is sharp and deep. We can forecast without knowledge about the rules producing time series.

References

1. J. Amigó. *Permutation Complexity in Dynamical Systems*, Springer, Berlin, 2010.
2. G. Atsalakis and K. Tsakalaki. Simulating annealing and neural networks for chaotic time series forecasting. *CMSIM Journal*, 1, 81–90, 2012.
3. C. Bandt and B. Pompe. Permutation Entropy: A Natural Complexity Measure for Time Series. *Phys. Rev. Lett.*, 88, 17, 174102, 2002.
4. J. C. Berg. *Wavelets in Physics*, Cambridge University Press, Cambridge, 1998.
5. C. Chatfield. *Time–Series Forecasting*, Chapman & Hall/CRC, Boca Raton, 2000.
6. M. Faggini and A. Parziale. The Failure of Economic Theory. Lessons from Chaos Theory. *Modern Economy*, 3, 1–10, 2012.
7. A. Freking, W. Kinzel and I. Kanter. Learning and predicting time series by neural networks. *Phys. Rev. E*, 65, 5, 050903, 2002.
8. J. Garland, R. James and E. Bradley. Model-free quantification of time-series predictability. *Phys. Rev. E*, 90, 5, 052910, 2014.
9. A. Jakimowicz. Nonlinear Dynamical Systems Theory and Economic Complexity. *CMSIM Journal*, 4, 657–667, 2013.
10. T. W. Joo and S. B. Kim. Time series forecasting based on wavelet filtering. *Expert Systems with Applications*, 42, 8, 3868–3874, 2015.
11. S. Kodba, M. Perc and M. Marhl. Detecting chaos from a time series. *Eur. J. Phys.*, 26, 1, 205–215, 2005.
12. E. I. Malyutina and V. I. Shiryaev. Time Series Forecasting Using Nonlinear Dynamic Methods and Identification of Deterministic Chaos. *Procedia Computer Science*, 31, 1022–1031, 2014.
13. H. Niu and J. Wang. Quantifying complexity of financial short-term time series by composite multiscale entropy measure. *Communications in Nonlinear Science and Numerical Simulation*, 22, 1–3, 375–382, 2015.
14. M. Perc. Observed time series in the experiment with periodically driven RLC circuit. http://www.matjazperc.com/ejp/time.html
15. Quarterly National Accounts – Gross domestic product, YBHA. Office for National Statistics. http://www.ons.gov.uk/ons
16. Stock market index S&P 500. http://research.stlouisfed.org/fred2/series/SP500/downloaddata
17. M. West. EEG time series recorded at Duke University on a patient undergoing ECT therapy for clinical depression. http://www.isds.duke.edu/∼mw/ts_data_sets.html
18. X. Zhao, P. Shang and J. Huang. Permutation complexity and dependence measures of time series. *EPL*, 102, 4, 40005, 2013.