Neural Comput & Applic (2015) 26:1827–1838
DOI 10.1007/s00521-015-1845-8

ORIGINAL ARTICLE

Predictive clustering on non-successive observations for multi-step ahead chaotic time series prediction

V. A. Gromov · E. A. Borisenko

Received: 6 June 2014 / Accepted: 11 February 2015 / Published online: 22 February 2015

Abstract Predictive clustering algorithm based upon modified Wishart clustering technique is applied to predict chaotic time series. Concept of predictable and non-predictable observations is introduced in order to distinguish between reliable and unreliable predictions and, consequently, to enhance an ability to predict up to considerable number of positions ahead. Non-predictable observations are easily ascertained in the frameworks of predictive clustering, regardless used clustering technique. Clustering vectors are composed from observations according to set of patterns of non-successive positions in order to reveal characteristic observations sequences, useful for multi-step ahead predictions. The employed clustering method is featured with an ability to generate just enough clusters (submodels) to cope with inherent complexity of the series in question. The methods demonstrate good prediction quality for Lorenz system time series and satisfactory results for weather, energy market and financial time series.

Keywords Chaotic time series · Multi-step ahead prediction · Predictable and non-predictable observations · Predictive clustering

1 Introduction

Constant interest in chaotic system and models expressed by researchers of various sciences and lines of investigation [1–4] is associated with both fundamental importance of non-linear phenomena for natural and social processes description and inherent complexity of the processes as well. The problem of chaotic time series prediction is of considerable importance in the field.

The mathematical models and concepts traditionally employed for time series prediction are used to develop a single model; moreover, the number of parameters involved in the model is supposed to be minimal, if only it is consistent with data (see, for instance, [5] and references therein). It is based on the supposition that the larger the number of parameters, the less meaningful the model is.

Recently, a lot of computational intelligence models have been proposed to derive the structure of the strange attractor out of a series and to predict the series [6]. One can divide the models in three main groups in accordance with the artificial intelligence theories used to reconstruct the state space [7] and to analyse time series. The first group is neural networks that are able (being universal adaptive approximators) to reveal various local trends present in chaotic time series and to approximate them [8–11]. The second group comprises fuzzy and neuro-fuzzy approaches used to develop robust and logically transparent prediction models [12–15]. Finally, the third group is associated with distributed artificial intelligence methods such as genetic algorithms [10], swarm intelligence [7, 16], ant colony optimization [17, 18] and so on.

Meanwhile, real-world time series prediction leads to the conclusion that such a unified model may not exist or may be too complicated to deal with. As far as chaotic time series concerned, various parts of a strange attractor (of the system generating the series of interest) are associated with various lengths of the series and, respectively, it is reasonable to develop a separate model for each part.

On the other hand, medical statistics, electricity market [19], weather and earthquake forecasting often provide one

V. A. Gromov (✉) · E. A. Borisenko
Oles Honchar Dnepropetrovsk National University, Dnepropetrovsk 49050, Ukraine
e-mail: stroller@rambler.ru

E. A. Borisenko
e-mail: lena.borisenko@mail.ru

with bunches of related time series such that series belonging to a single bunch should be analysed together. These considerations, taken together, lead to predictive clustering problem statement [20]. The statement includes both clustering problem and prediction problem statements and implies that obtained clusters allow not only time series prediction but also knowledge extraction from each cluster.

The paper [21] proposes to apply decision trees to develop predictive clustering algorithms. In a series of papers [19, 22], pattern sequence-based forecasting (PSF) based on K-means clustering techniques is employed to predict daily sales of Spanish electricity market. As far as K-means clustering implies that the number of clusters is known in advance, authors performed wide-ranging simulation to find optimal number of clusters. Ant colony optimization is used to construct clusters and to predict chaotic time series in the paper [23].

The chunks of time series are obtained in order to calculate polynomial regression coefficients in [24]. Then, the regression coefficients are clusterized using the modified ECM algorithm. The algorithm is applied to Indonesian stock market indexes. Another study of the authors [25] employs similar approach to predict multiple (related) time series. A concept of “motif” is used to extract the most relevant information in order to clusterize it [26]. Authors modified K-means clustering technique by means of appropriate choosing of initial clusters’ centres.

In the frameworks of this approach, a large number of parameters involved are not considered as a method drawback. Ultimately, each prediction model is a tool to extract and to compress information (wherever the information is finally represented, say, as regression coefficients, neural networks weights, fuzzy membership function parameters, etc.) from data, and, as it is, is a result of a certain trade-off between demands of prediction efficiency and model compactness.

Actually, each predictive clustering algorithm tries to extract the most relevant (for future prediction) pieces of information from time series prior to the clustering (“regression coefficients” [24], “motifs” [26], etc.). The present paper uses patterns of non-successive observations. The modified Wishart algorithm is taken to be the clustering technique.

One should stress that the majority of the methods mentioned deals with single-step ahead prediction and, consequently, demonstrates from satisfactory to excellent prediction quality. For time series generated by Lorenz system and Mackey–Glass time series—typical benchmarks in the field—prediction error for single-step ahead prediction, to the best authors’ knowledge, is lesser that 0.5 % for now (if the best available method is used). The situation is reversed, if multi-step ahead prediction is concerned. The exponential prediction error growth (with a speed determined by the highest Lyapunov exponent) deteriorates prediction results rapidly, as the position to be predicted shifts from the last known observation position.

The rest of the paper is organized as follows. The next section states the problem and outlines the prediction model and the used clustering method as well. The third one provides the prediction results for time series generated by Lorenz system, for a weather time series for Dnepropetrovsk city, for Spanish energy market and for US gold prices time series as well. Finally, the last section presents conclusions.

2 Problem statement

Chaotic time series observations $y_0, y_1, \ldots, y_t$ are considered in order to predict subsequent observations $y_{t+1}, y_{t+2}, \ldots, y_{t+K}$ with prognosis absolute error lesser than a specified value $\left|\tilde{y}_{t+i} - y_{t+i}\right| < \alpha, \quad i = \overline{1, K}$.

We assume that all transient processes in the system that generate the time series in hand (the system is usually unknown) have been completed, and the time series reflects the trajectory movement in the neighbourhood of strange attractor, however complex it is. The second assumption is that the series meets Takens’ theorem conditions and, respectively, one can analyse the attractor structure using time series observations.

If one considers multi-step ahead prediction, one should take into account that for a chaotic time series, a prognosis can be performed in a proper way up to a certain limit of steps ahead (prediction horizon) [27]. The prediction horizon is explained by the fact that, for the chaotic series, data error, small at the initial time point, grows exponentially owing to divergence of initially close trajectories.

As trajectory of the system moves along the same area of the attractor frequently, one can meet analogous (similar) sequences in the time series associated with the area. If one reveals these areas, describes them and develops simplest prediction models for each area, one makes it possible to predict chaotic time series up to a considerable time limit [23]. The clustering method presented below (actually, Wishart algorithm) is employed to collect together sequences belonging to the same cluster.

Usually, to ensure that Takens’ theorem conditions are satisfied and, respectively, the dynamics analysed is equivalent to the true system dynamics, and vectors are composed from time series observations [28]. Point of interest here is that vectors composed of succeeding observations proved to be essentially less efficient with respect to prediction purposes than vectors composed of nonsucceeding observations according to a certain pattern. For the best prediction, one should run over all or, at least, a considerable portion of all reasonable patterns and single the most appropriate out. Different attractor areas may be associated with different clusters.

One should emphasize that each simplest model mentioned above is an averaged representation of the clustered time series sequences (or, alternatively, trajectories belonging to the respective attractor area). Consequently, it leads to deterioration of multi-step ahead prediction quality due to averaging (the predicted values are obtained with employment of the cluster centres) and, simultaneously, to its improvement in virtue of the fact that the "chaotic" exponential growth is alleviated. A clustering method used determines a compromise between these two tendencies.

The proposed prediction algorithm is subdivided into two parts. The first one analyses a time series at hand in order to clusterize sequences made of its observations according to predefined patterns and then to reveal typical observations sequences as cluster centres. The second part provides prognosis for the time series with employment of the revealed typical sequences. The series is considered to be normalized.

2.1 Clustering algorithm

First, we consider Wishart clustering method [28] modified by Lapko and Chentsov [29]. The method employs graph theory concepts and non-parametric probability density function estimator of k-nearest neighbours to clusterize vectors concatenated according to the pattern used (hereinafter, samples):

$$p(x) = \frac{k}{V_k(x)n}$$

Here, $V_k(x)$ and $d_{k(x)}(x)$ are, respectively, a volume and a radius of the minimal hypersphere with its centre at point x containing at least k samples. A similarity relation graph $G(Z_n, U_n)$ defined by its vertex set Z_n (all samples considered) and its edge set $U_n = \{(x_i, x_j) : d(x_i, x_j) \leq d_k(x_i), \quad i \neq j\}$ is the method primary concept. $G(Z_i, U_i)$ is generated subgraph of the graph $G(Z_n, U_n)$ with a vertex set $Z_i = \{x_j, \quad j = \overline{1, i}\}$ and an edge set containing all edges from U_n such that their final vertices belong to the set Zi.

The cluster c_l ($l > 0$) is defined to be a height significant one with respect to height value $h > 0$, if $\max\{|p(x_i) - p(x_j)| \; \forall x_i x_j \in c_l\} \geq h$.

Thus, the algorithm consists of the following steps:

1. To determine distances $d_k(x_i)$ for each sample between the sample and its k-nearest neighbour and to range the samples ascending in accordance with the distances calculated.
2. If $w(x_i)$ is a class number for ith sample, then to set $i = 1$.
3. For subgraph $G(Z_i, U_i)$, the following alternatives are possible.

   If x_i is an isolated vertex of $G(Z_i, U_i)$, then to start generation of the new cluster; after that to go to the step 4.
   If a vertex x_i is connected to vertices of lth class only and the class is completed, then to set $w(x_i) = 0$. Else (if the class is not completed yet) then to set $w(x_i) = 1$; after that to go to the step 4.
   If a vertex x_i is connected to vertices of classes $l_1, l_2, \ldots, l_t$, $t > 1$:
   If all t classes are completed, then to set $w(x_i) = 0$; after that to go to the step 4. $z(h) \leq t$ is a number of significant classes.
   If $z(h) > 1$, then to set $w(x_i) = 0$, to label significant classes as completed, and to delete insignificant ones setting $w(x_i) = 0$ for all samples belonging to them. Otherwise, to incorporate clusters $l_2, \ldots, l_t$ to l_1 by setting $w(x_j) = l_1$ for all samples belonging to them and to set $w(x_i) = l_1$ for the sample itself.

4. To set $i = i + 1$. If $i \leq n$ is lesser than the number of samples, than to go to the step 3.

To be specific, we used the following values $k = 11$, $h = 0.2$, $n = 4$. One should stress that the main advantage of the algorithm is its ability to ascertain the optimal number of clusters in the course of clustering.

2.2 Prediction algorithm

To predict time series values in the framework of the second part of algorithm, the centres of clusters (typical sequences) are calculated for all used patterns and obtained clusters. For a given position to be predicted (in the time series in question) and for a given cluster, one composes a vector from time series observations according to the pattern having been used to generate the cluster with the position associated with the last vector element (respectively, undefined), truncates the vector and the cluster centre (all elements but the last ones are included in the truncated vectors), and calculates the Euclidian distance between the truncated observation vector and the truncated cluster centre. One searches over all patterns and clusters in order to find the cluster with the minimal distance. The centre of this cluster is employed to predict the observation, namely its centre last element is used as a predicted value for the position in question.

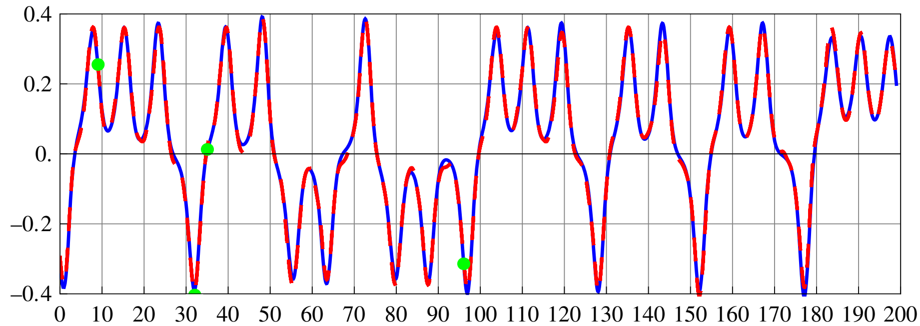


Fig. 1 Single-step ahead prediction for Lorenz time series

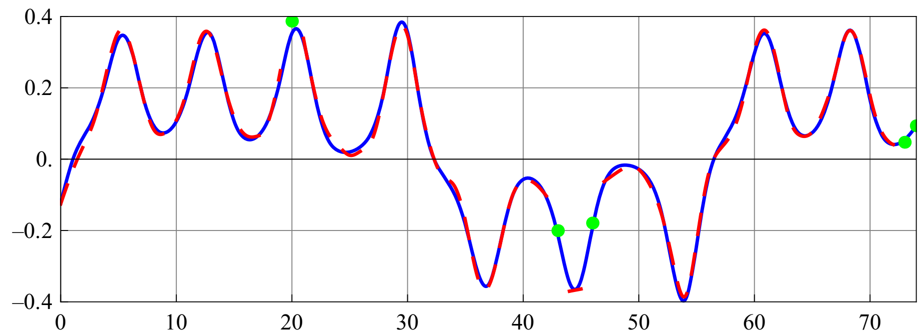


Fig. 2 Multi-step ahead prediction for Lorenz time series

Table 1 Prediction quality for Lorenz time series

Step count	MAE (σ)	MER (σ, %)	RMSE (%)	Non-predictable (%)
1	0.014 (0.01)	1.92 (1.92)	0.82	9.83
5	0.017 (0.011)	2.57 (2.21)	1.33	17.3
10	0.019 (0.012)	2.73 (2.37)	1.88	22.17

MAE, mean absolute error; MER, mean error relative; RMSE, root-mean-square error; and non-predictable, number of non-predictable observations related to the total number of testing set

Moreover, in virtue of the fact that exponential "chaotic" error growth is essentially absent, it is possible to use predicted values for further prediction and to obtain predicted values for positions far apart from the last known observations.

Admittedly, the methods using clustering techniques to reveal typical sequences and to predict with employment the sequences revealed possess the following limitation. Sometimes, for a given position to be predicted, it is impossible to find a respective cluster (there is no cluster centre matching observations from the time series length preceding the given position). Hereinafter, such observations are called non-predictable, and their number related to the total number of testing set observations (along with a prediction error averaged over all other [predictable] observations of the testing set) is taken to be a measure of method prediction quality.

In order to assess the relative importance of the clusters constructed, additional set, of the same size as the testing one, is introduced, and clusters' prediction qualities I_k is calculated over this set. For each cluster (for a given maximum prediction error value α), its prediction quality is computed as

$$I_k(\alpha) = \sum_{i \in S_k} \frac{\overline{\varepsilon_i}}{\varepsilon_{ki}} \frac{1}{|V_i| - 1}, \quad \overline{\varepsilon_i} = \frac{1}{|V_i|} \sum_{j \in V_k} \varepsilon_{ji}$$

where V_i—a set of clusters able to predict the ith observation with error lesser than the predefined value α; ε_{ji}—error obtained for ith observation prediction with employment of jth cluster centre, $j \in V_i$; $\overline{\varepsilon_i} = \frac{1}{|V_i|} \sum_{j \in V_k} \varepsilon_{ji}$—an average prediction error for ith observation over all clusters involved; S_k—a set of observations such that kth cluster centre predicts them with error lesser than the predefined value α.

3 Numerical results

The method discussed in the previous section is applied to a time series generated by Lorenz system [30], weather

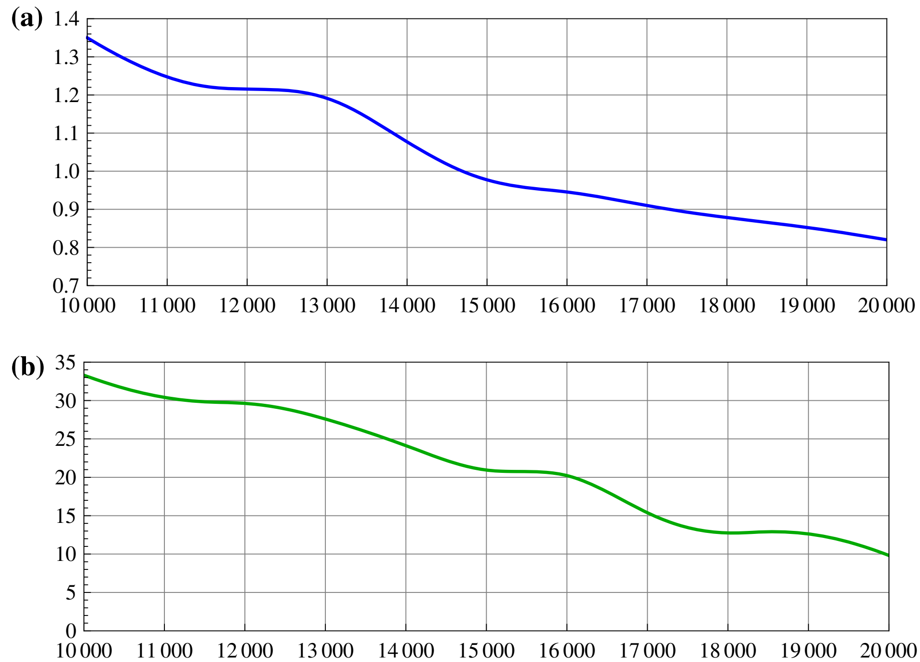


Fig. 3 Lorenz time series. Single-step ahead prediction. **a** RMSE (per cents) versus training set size. **b** Percentage of non-predictable points versus training set size

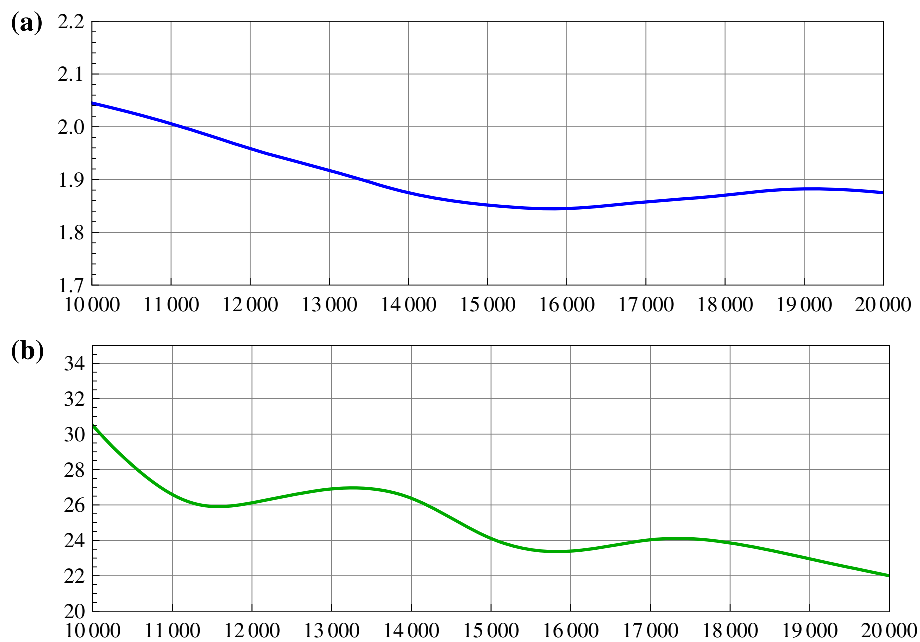


Fig. 4 Lorenz time series. Multi-step ahead prediction (10 steps ahead). **a** RMSE (per cents) versus training set size. **b** Percentage of non-predictable points versus training set size

time series for Dnepropetrovsk city, Australian electricity price market and US gold price time series as well.

The Lorenz system (with standard "chaotic" parameters $\sigma = 10$, $b = \frac{8}{3}$, $r = 28$) integration with employment of Runge–Kutta's fourth-order method (integration step is equal to 0.1) yields a time series hereinafter referred to as Lorenz series. Respectively, its data accuracy according to Runge–Kutta's integration error estimates is 10^{-5}. For Lorenz series, the first 800 observations are discarded in order to ensure that trajectory moves in the neighbourhood of the respective strange attractor. The testing set for the series consists of 3000 observations, while a training set size is varied and, actually, is a crucial parameter for the method considered.

The weather time series reflects temperature variations for Dnepropetrovsk city since 2008–2013 and comprises 15,000 observations. The testing set consists of 3000 observations, and the training sets of various sizes are considered as well.

Australian electricity price market time series presents price changes for the period since September, 2013 till

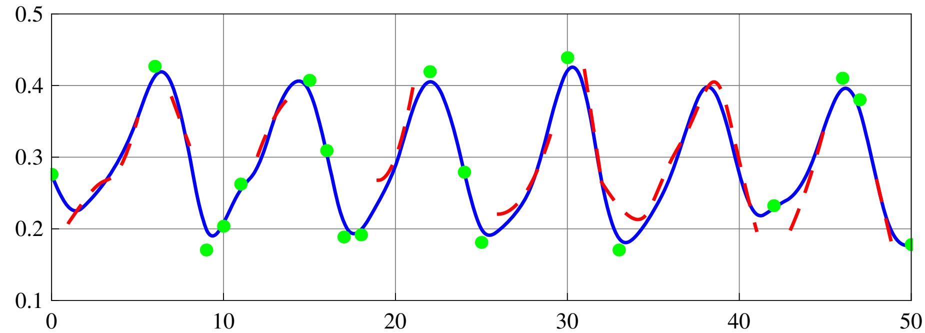


Fig. 5 Single-step ahead prediction for weather time series

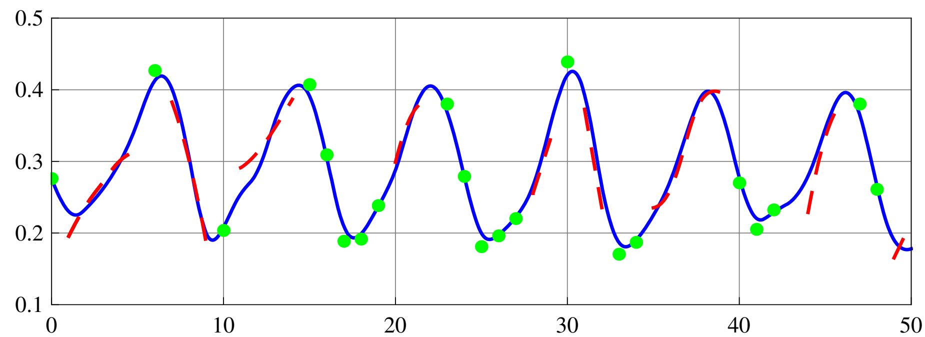


Fig. 6 Multi-step ahead prediction for weather time series

Table 2 Prediction quality for weather time series

Step count	MAE (σ)	MER (σ, %)	RMSE (%)	Non-predictable (%)
1	0.019 (0.013)	9.26 (0.086)	2.59	43.33
5	0.021 (0.013)	8.35 (0.089)	2.61	49.97
10	0.022 (0.014)	6.67 (0.092)	2.41	56.9

MAE, mean absolute error; MER, mean error relative, RMSE, root-mean-square error; and non-predictable, number of non-predictable observations related to the total number of testing set

September, 2014; training set consists of 20,000 observations.

Finally, time series of financial nature is exemplified by US gold prices taken from 1983 to 2007 years (6000 observations).

For all figures presented in this section, blue solid lines are associated with observed data, whereas red dashed lines are associated with predicted values. Green discs represent non-predictable points.

For all series, a set of four-point patterns with a maximal distance between adjacent nodes equal to 10 is considered, thus the maximal number of patterns is 1000. All presented results are obtained with employment of all patterns.

The maximal Lyapunov exponent was calculated for all studied time series with employment nearest neighbour method [27, 31]. All of them appear to be positive that is clear evidence for chaotic nature of the series under investigation; namely the values are 0.91, 2.18, 0.44 and 0.2 for Lorentz, weather, energy market and US Gold prices, respectively. The value for Lorentz time series is in a good agreement with results of [31] (see on p. 217).

Figure 1 presents single-step ahead prediction results for Lorentz time series. The size of training set is 20,000 observations. The percentage of non-predictable observations is about 10 %, while the average prediction error for predictable observations is equal to 0.82 %.

Figure 2 shows multi-step ahead prediction results for Lorentz time series; prediction with employment of predicted values is used. Similarly, the size of training set is 20,000. The time series length plotted in Fig. 2 demonstrates that it is possible to obtain reliable prediction values up to 79 steps ahead. Table 1 summarizes prediction results for single and multi-step prediction error. The first column presents information on how many steps ahead the data is predicted. The three next columns contain three various errors (MAE—mean absolute error, MER—mean error related to average value and RMSE—root-mean-square error), averaged over testing set. For the

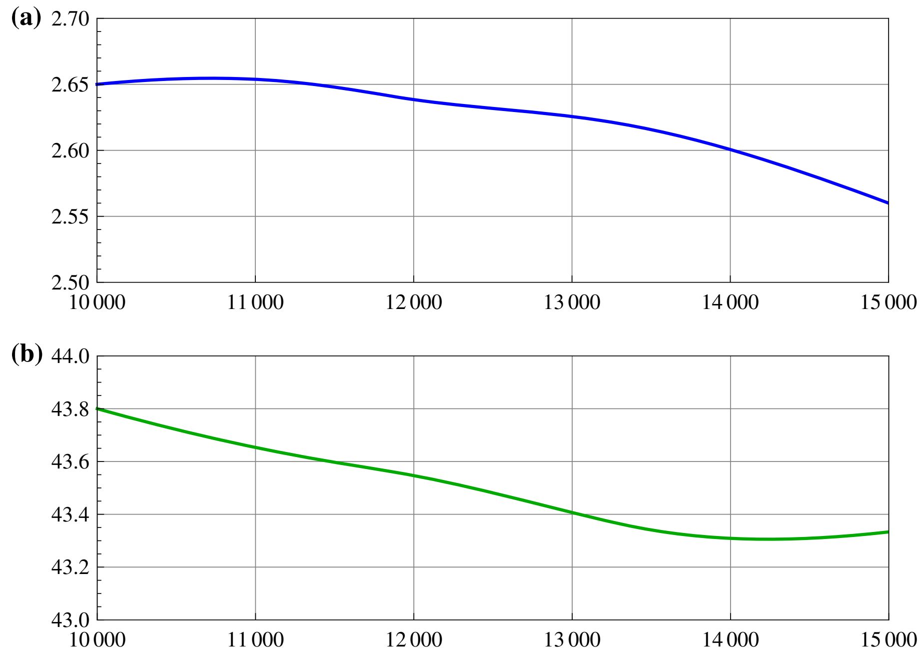


Fig. 7 Weather time series. Single-step ahead prediction. **a** RMSE (per cents) versus training set size. **b** Percentage of non-predictable points versus training set size

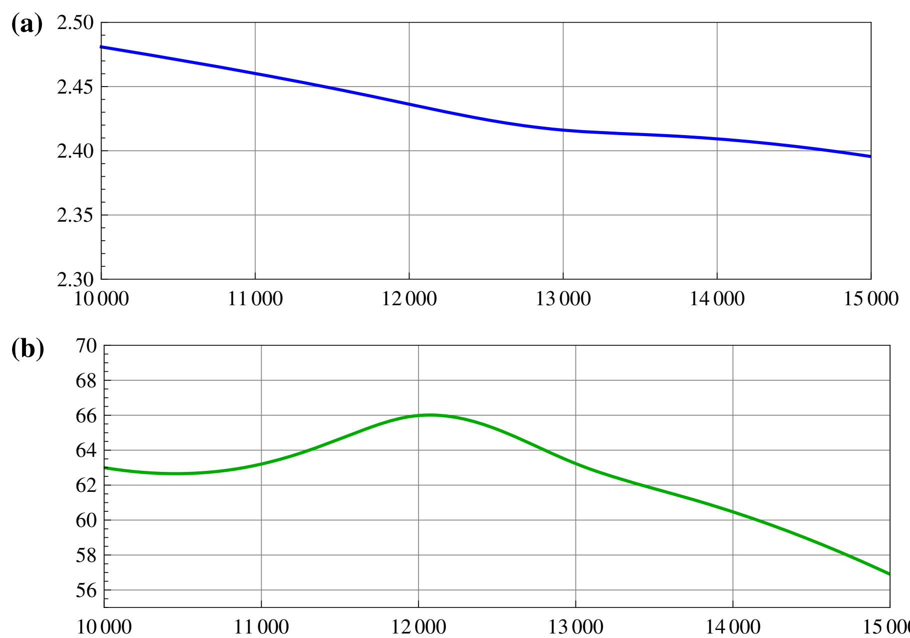


Fig. 8 Weather time series. Multi-step ahead prediction (10 steps ahead). **a** RMSE (per cents) versus training set size. **b** Percentage of non-predictable points versus training set size

Table 3 Prediction quality for Australian electricity price time series

Step count	MAE (σ)	MER (σ, %)	RMSE (%)	Non-predictable (%)
1	0.007 (0.006)	1.52 (0.014)	0.99	0
5	0.014 (0.012)	2.8 (0.025)	1.93	9.23
10	0.017 (0.013)	2.75 (0.028)	2.21	24.5

MAE, mean absolute error; MER, mean error relative; RMSE, root-mean-square error; and non-predictable, number of non-predictable observations related to the total number of testing set

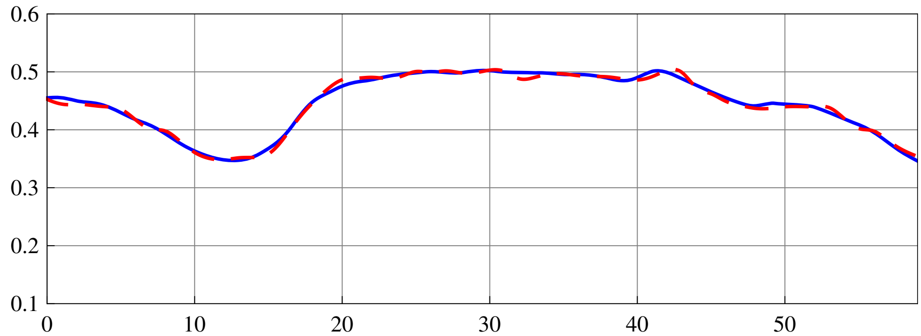


Fig. 9 Single-step ahead prediction for Australian electricity price series

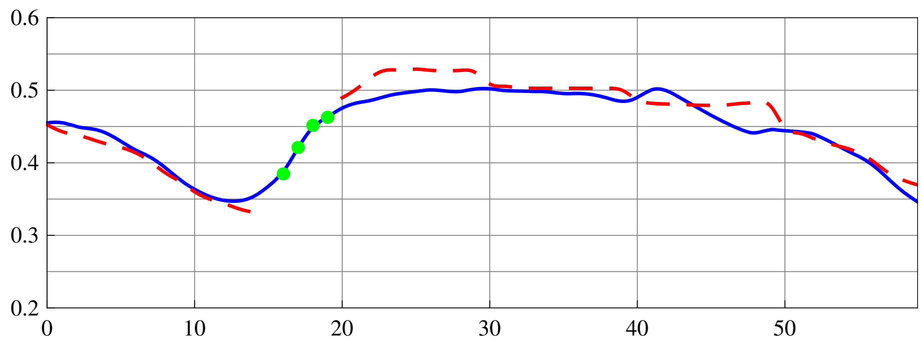


Fig. 10 Multi-step ahead prediction for Australian electricity price series

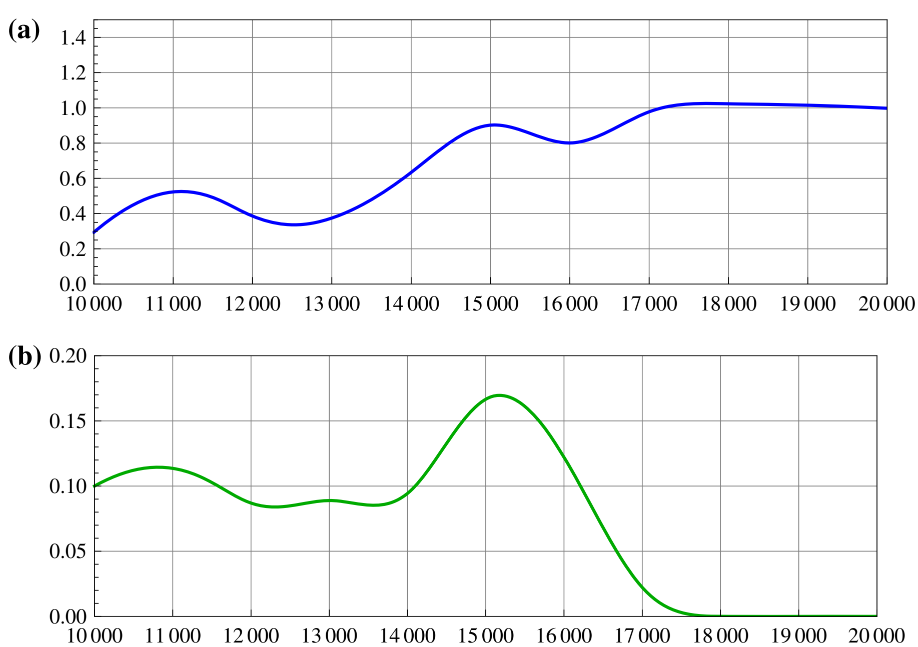


Fig. 11 Australian electricity price market. Single-step ahead prediction. **a** RMSE (per cents) versus training set size. **b** Percentage of non-predictable points versus training set size

two first errors, standard deviation is presented in the parenthesis. Finally, a percentage of non-predictable points is shown in the last column.

Two next figures (Figs. 3, 4) reflect the fact that the larger a training set is, the more areas (even with small values of ergodic probability measure) of the strange attractor is visited, and, consequently, the more accurate prognosis is obtained. Figure 3a displays a percentage of non-predictable observations versus training set size for single-step ahead prediction; Fig. 3b displays an average prediction error for predictable observations versus training set size. Figure 4a, b shows the same curves for multi-step

Fig. 12 Australian electricity price market. Multi-step ahead prediction (10 steps ahead). **a** RMSE (per cents) versus training set size. **b** Percentage of non-predictable points versus training set size

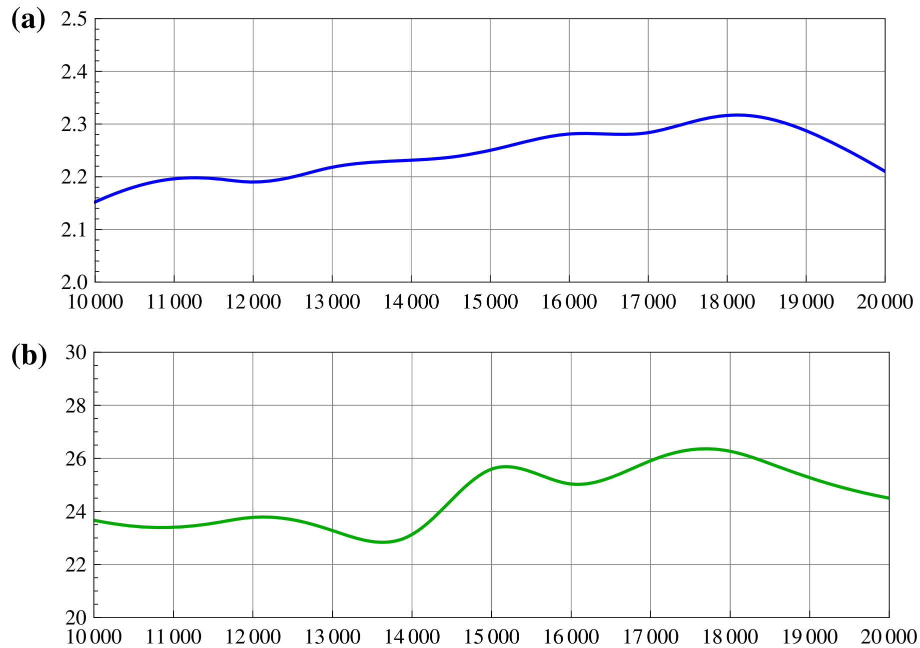
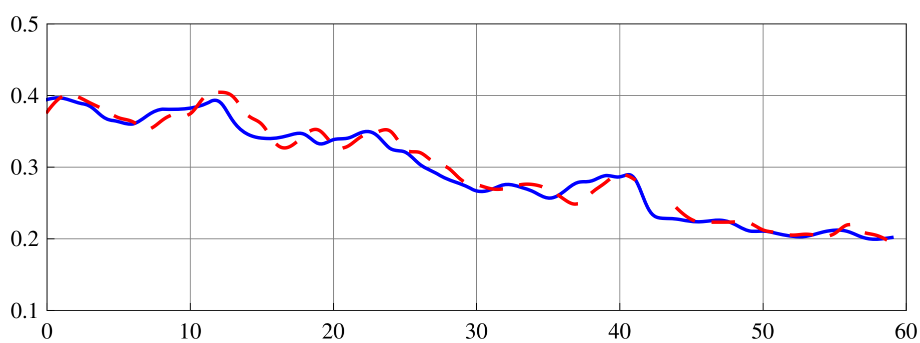


Table 4 Prediction quality for US gold price time series

Step count	MAE (σ)	MER (σ, %)	RMSE (%)	Non-predictable (%)
1	0.008 (0.007)	3.657 (0.032)	1.102	0.2
5	0.0111 (0.01)	4.847 (0.044)	1.499	2
10	0.012 (0.01)	5.13 (0.047)	1.619	2.2

MAE, mean absolute error; MER, mean error relative; RMSE, root-mean-square error; and non-predictable, number of non-predictable observations related to the total number of testing set

Fig. 13 Single-step ahead prediction for US gold price series



ahead prediction. For both cases, both characteristics decrease nearly monotonously and vanish as a size of training set limits to infinity.

Description of results for the weather time series traces the same route as the description for Lorenz series presented above. Thus, Figs. 5 and 6 present prediction results for single- and multi-step ahead prediction, respectively. Here, the size of training set is 15,000. A percentage of non-predictable observations for the single-step case is 43.3 %; an average prediction error for predictable observations is 2.59 %. The Table 2 summarizes prediction results for single- and multi-step prediction. The errors presented along similar lines with Table 1.

Figures 7 and 8 display dependences of a percentage of non-predictable observations and an average prediction error on training set size. The dependences (percentage of

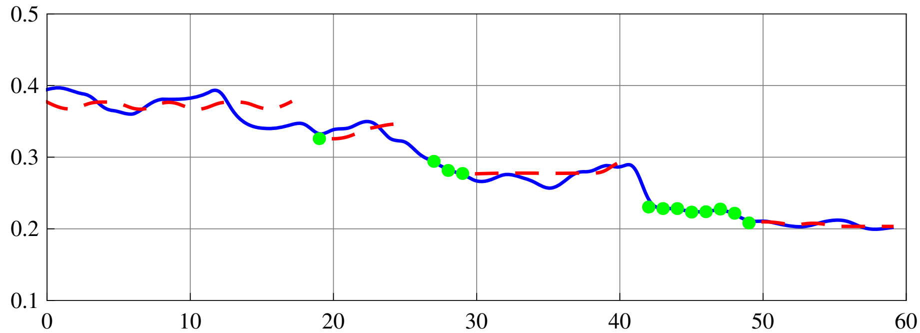


Fig. 14 Multi-step ahead prediction for US gold price series

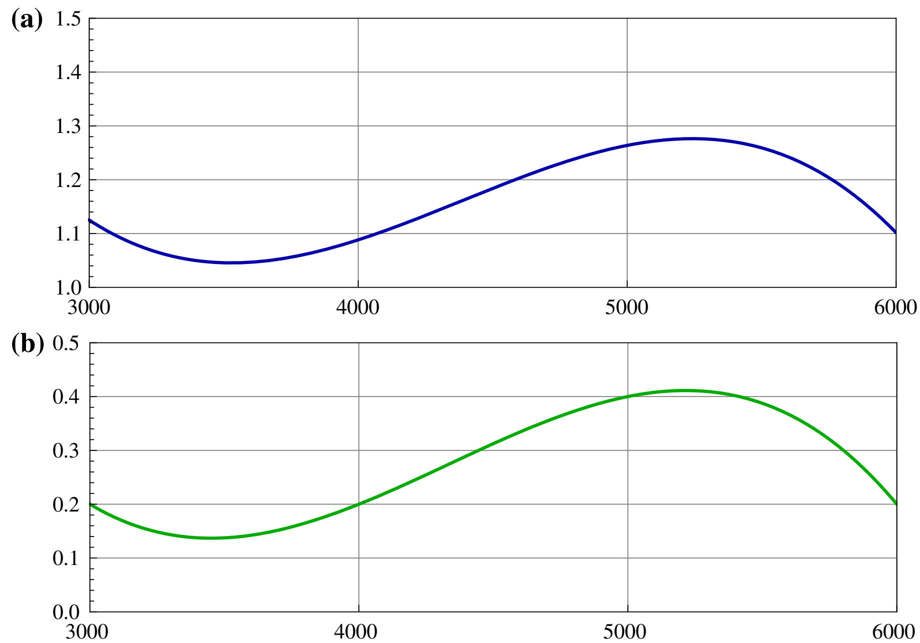


Fig. 15 US gold price. Single-step ahead prediction. **a** RMSE (per cents) versus training set size. **b** Percentage of non-predictable points versus training set size

non-predictable observations) decrease, however, sometimes non-monotonously, as similar dependences for Lorenz series do.

Table 3 and a series of Figs. 9, 10, 11 and 12 present information about prediction results for Australian electricity price time series. Information is presented analogously to the way it was presented for two previous series.

For that case, the error for predictable points does not exceed a predefined threshold of 3 %, while the number of non-predictable points decreases non-monotonously for single-step ahead prediction (Fig. 11b) and oscillates for multi-step ahead one.

Table 4 with Figs. 13, 14, 15 and 16 exhibits information about US gold prices time series; the prediction error and the number of non-predictable observations does not exceed the predefined threshold of 3 % for both single and multi-step ahead prediction (Tables 5, 6).

The results obtained with employment of the proposed prediction method are compared with those reported in the literature (see [19] and references therein). All columns of these tables except last ones are borrowed from the cited paper (PSF, DWT, MLP stands for pattern sequenced-based forecasting, discrete wavelet transformation and multilayer perceptron, respectively); the last columns display results obtained using the proposed method (PCW stands for predictive clustering with Wishart clustering algorithm).

4 Conclusions

Several conclusions may be reached.

Firstly, the concept of non-predictable points allows one to predict chaotic time series up to considerable number of positions ahead. Non-predictable points are easily

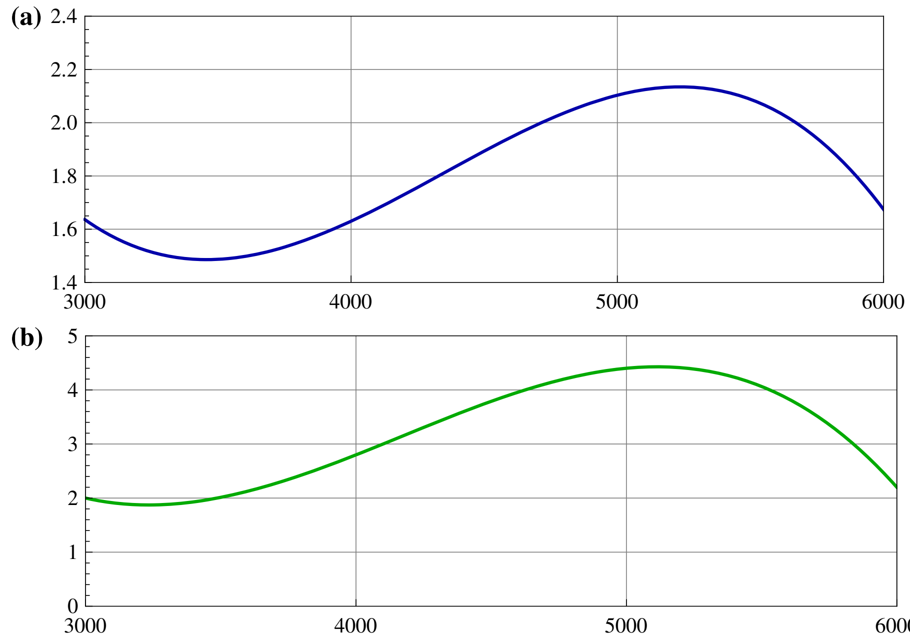


Fig. 16 US gold price. Multi-step ahead prediction (10 steps ahead). **a** RMSE (per cents) versus training set size. **b** Percentage of non-predictable points versus training set size

Table 5 MER for some days of the year 2004 (Australia's national electricity market—Price)

Day	ARIMA (%)	SVM (%)	PSF (%)	PCW (%)
5th June	32.31	18.09	16.72	1.94
17th June	29.09	13.31	8.31	1.72
20th June	33.73	17.11	14.23	1.32
21th June	24.18	19.20	18.93	1.94
Average	29.82	16.93	14.55	1.73

ARIMA, autoregressive integrated moving average; SVM, support vector machines; PSF, pattern sequenced-based forecasting; and PCW, predictive clustering with Wishart clustering algorithm

ascertained in the frameworks of predictive clustering, regardless the specific clustering technique used. Furthermore, the concept leads to distinguishing predictable and non-predictable points and provides much more information about considered method prediction ability than plain prediction error.

Secondly, the presented prediction methods based upon Wishart clustering algorithms demonstrate good prediction quality for Lorenz system time series and satisfactory results for weather time series.

Thirdly, for benchmark and real-world data, prediction quality, generally, increases as a training set size increases. Exceptions are associated with the presence of data generated by transient processes in time series. The clustering method considered is featured with an ability to generate just enough clusters (submodels) to cope with inherent complexity of the series in question. Ultimate clusters' structure stems from a certain trade-off between prediction efficiency and model (as a set of submodels) compactness.

Table 6 MER for some weeks of the year 2004 (Australia's national electricity market—Price)

Week	DWT (%)	MLP (%)	SVM (%)	PSF (%)	PCW (%)
Second of January	12.94	25.81	23.37	15.62	1.33
First of July	12.23	8.36	15.03	9.12	1.47
First of August	16.17	15.85	36.18	13.98	1.28
Third of December	10.01	47.41	33.74	10.23	1.11
Average	12.84	24.36	27.08	12.23	1.30

DWT, discrete wavelet transformation; MLP, multilayer perceptron; SVM, support vector machines; PSF, pattern sequenced-based forecasting; PCW, predictive clustering with Wishart clustering algorithm

References

1. Barkley JR (2000) From catastrophe to chaos: a general theory of economic discontinuities, 2nd edn. Kluwer, Boston
2. Barnett WA, Deissenberg C, Feichtinger G (eds) (2004) Economic complexity: non-linear dynamics, multi-agents economies and learning. Elsevier, Amsterdam
3. Murray JD (2002) Mathematical biology 1: an introduction. Springer, Berlin
4. Reichl LE (2004) The transition to chaos. Conservative classical systems and quantum manifestations. Springer, New York
5. Konishi S, Kitagava G (2008) Information criteria and statistical modeling. Springer, New York
6. Palit AK, Popovic D (2005) Computational intelligence in time series forecasting. Theory and engineering applications. Springer, London
7. Wang J, Chi D, Wu J, Lu H (2011) Chaotic time series method combined with particle swarm optimization and trend adjustment for electricity demand forecasting. Expert Syst Appl 38:8419–8429
8. Chen D, Wang J, Zou F, Yuan W, Hou W (2014) Time series prediction with improved neuro-endocrine model. Neural Comput Appl 24:1465–1475
9. Gan M, Peng H, Peng X, Chen X, Inoussa G (2010) A locally linear RBF network-based state-dependent AR model for nonlinear time series modeling. Inf Sci 180:4370–4383
10. Mirzaee H (2009) Linear combination rule in genetic algorithm for optimization of finite impulse response neural network to predict natural chaotic time series. Chaos, Solitons Fractals 41: 2681–2689
11. Ghazali R, Hussain AJ, Liatsis P, Tawfik H (2008) The application of ridge polynomial neural network to multi-step ahead financial time series prediction. Neural Comput Appl 17:311–323
12. Abiyev RH (2011) Fuzzy wavelet neural network based on fuzzy clustering and gradient techniques for time series prediction. Neural Comput Appl 20:249–259
13. Gholipour A, Lucas C, Mirmomeni AM, Shafiee M (2007) Extracting the main patterns of natural time series for long-term neurofuzzy prediction. Neural Comput Appl 16:383–393
14. Gu H, Wang H (2007) Fuzzy prediction of chaotic time series based on singular value decomposition. Appl Math Comput 185:1171–1185
15. Singh P, Borah B (2013) High-order fuzzy-neuro expert system for time series forecasting. Knowl Based Syst 46:12–21
16. Hong WC (2010) Application of chaotic ant swarm optimization in electric load forecasting. Energy Policy 38:5830–5839
17. Niu D, Wang Y, Wu DD (2010) Power load forecasting using support vector machine and ant colony optimization. Expert Syst Appl 37:2531–2539
18. Pan Y, Jiang JC, Wang R, Jiang JJ (2011) Predicting the net heat of combustion of organic compounds from molecular structures based on ant colony optimization. J Loss Prev Process Ind 24:85–89
19. Martınez-Alvarez F, Troncoso A, Riquelme JC, Riquelme JM (2011) Energy time series forecasting based on pattern sequence similarity. IEEE Trans Knowl Data Eng 23(8):1230–1243
20. Blockeel H, De Raedt L, Ramon J (1998) Top-down induction of clustering trees. In: 15th international conference on machine learning, pp 55–63
21. Dzeroski S, Gjorgjioski V, Slavkov I, Struyf J (2006) Analysis of time series data with predictive clustering trees. In: Proceedings of the 5th international conference on knowledge discovery in inductive databases, pp 63–80
22. Martınez-Alvarez F, Troncoso A, Riquelme JC, Riquelme JM (2007) Partitioning-Clustering Techniques Applied to the Electricity Price Time Series. In: Intelligent data engineering and automated learning—IDEAL-2007, Springer, Berlin
23. Gromov VA, Shulga AN (2012) Chaotic time series prediction with employment of ant colony optimization. Expert Syst Appl 39:8474–8478
24. Widiputra H, Kho H, Pears R, Kasabov N (2009) A novel evolving clustering algorithm with polynomial regression for chaotic time-series prediction. Neural Inf Process 5864:114–121
25. Widiputra H, Pears R, Kasabov N (2011) Multiple time-series prediction through multiple time-series relationships profiling and clustered recurring trends. Adv Knowl Discov Data Min 6635: 161–172
26. Phu L, Anh DT (2011) Motif-based method for initialization the K-means clustering for time series data. In: AI 2011: advances in artificial intelligence, vol 7106. Springer, Berlin, pp 11–20
27. Kantz H, Schreiber T (2004) Nonlinear time series analysis. Cambridge University Press, Cambridge
28. Bock HH (1974) Automatic classification. Vandenhoeck and Rupert, Göttingen
29. Lapko AV, Chentsov SV (2000) Non-parametric information processing systems. Nauka, Moscow (in Russian)
30. Jackson EA (1985) The Lorenz system: I. The global structure of its stable manifolds. Phys Scr 32:469–475
31. Malinetskii GG, Potapov AP (2002) Modern problems of nonlinear dynamics. Editorial URSS, Moscow (in Russian)