Volume 237, No. 3, March 2008

ISSN 0167-2789

ELSEVIER

PHYSICA D
NONLINEAR PHENOMENA

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Physica D 237 (2008) 283–291
PHYSICA D
www.elsevier.com/locate/physd

Adaptive metrics in the nearest neighbours method

M. Kulesh*, M. Holschneider, K. Kurennaya

Institute for Mathematics, University of Potsdam, Am Neuen Palais 10, 14469 Potsdam, Germany

Received 8 November 2006; received in revised form 12 July 2007; accepted 20 August 2007
Available online 12 September 2007

Communicated by S. Kai

Abstract

We present a modification of the nearest neighbours method in order to improve accuracy of forecast of a time series. This modification is characterized by an adaptive metrics, which is able to adjust its parameters to each segment of a time series. In order to assess the accuracy of the presented method synthetic and real time series were predicted. We show, that this modified metrics is advantageous in prediction in comparison with standard Euclidean metrics or weighted metrics.

Keywords: Time series; Prediction; Nearest neighbours method; Metrics; Optimization

1. Introduction

The nearest neighbours method is one of the most popular techniques in nonlinear time series analysis. This method is often used for prediction of seasonal time series in such fields of research, as nonlinear and chaotic time series investigations [1–3], meteorology [4,5], economics and finance [6,7] and industrial processes modelling [8].

In essence, the nearest neighbours method goes as follows: each time a forecast needs to be made the data series is searched for the situation, similar to the current one. The final fixed number of points of the data series is compared with all other parts of this time series, which have the same length. The comparison is quantified with the help of a chosen metrics. Then, prediction is based on what happened after the closest (in the chosen metrics) chunk of signal in the past.

Often the nearest neighbours method gives us one of the best fit for nearly periodic time series forecasting, but in spite of this fact, several points of the method are still open to questions. It is a problem how to determine the parameters of method as discussed in [5,9–11]. Comparison of different kinds of metrics is given in the papers [8,12] and overview of appropriate measure of nearness is presented in [9]. In [2,13,14], analysis is carried out not only for most popular metrics, but also for more general measure with several parameters. In [8] is proposed a distance, called shape distance, which is insensitive to scaling and translation.

However using the metrics, mentioned above, it is almost not possible to trace a tendency of global increasing or decreasing of time series amplitude. In case there is no special amplitude factor, forecast exhibits a tendency to amplitude attenuation. One possibility of dealing with this situation could be to remove all trends in a preprocessing step. This however might produce additional prediction mistakes, connected with trend estimation and prolongation.

In this paper, we present a modification of the nearest neighbours method and propose a new flexible metrics, which reduces problems of amplitude changing and trend determination.

This paper is organized as follows. In the next two sections, we give an overview of classical nearest neighbours method and propose our extension of this method. Then we cite an instance of synthetic time series prediction, and afterwards we apply our modification for forecasting of banking data.

2. *k*-nearest neighbours algorithm

Among the reasons that deterministic nonlinear modelling techniques of complex data series have received a great deal

* Corresponding author. Tel.: +49 (0)331 977 1823; fax: +49 (0)331 977 1578.
E-mail address: mkulesh@math.uni-potsdam.de (M. Kulesh).

0167-2789/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.physd.2007.08.019

of attention is that they can be used to forecast, at least in the short term, the evolution of a chaotic system whose underlying dynamics is unknown [14].

Most of these techniques can be grouped into two major classes: global and local ones. In global methods, the whole past information is used for predictions about the evolution of the system under study. Clearly, this has the disadvantage that if new information is taken into account, all the parameters of the model may change and then a long parameter estimation time may be required. The local methods overcome this drawback by utilizing only part of the history.

Let us consider a time series

$$\vec{Y} = (y_1, y_2, \dots, y_N), \quad y_i \in \mathbb{R}^+, \quad y_i \geq 0. \tag{1}$$

In fact, the basic idea, which supports local techniques, is that, if a deterministic mechanism governs the evolution of this data series, then, for sufficiently high values of m, any value y_{N+1} will be given by

$$y_{N+1} = F(y_N, y_{N-\tau}, \dots, y_{N-(m-1)\tau}) = F(\vec{Y}_N), \tag{2}$$

where F is a continuous rule, and we have introduced the *fitting set* $\vec{Y}_N$. A window $m\tau$ consists of the past values of time series, where m is the embedding dimension, and τ is the embedding delay. Often the length of such a window can be chosen by sight. In most cases, it is a length of seasonality, which can be determined by means of autocorrelogram.

To compare the fitting set $\vec{Y}_N$ with all other parts of this time series, which have the same length, we define a set of $(N - 2m\tau + 1)$ finite windows, called the *time delay embeddings*. Each vector $\vec{Y}_t$ of this set is defined as

$$\vec{Y}_t = (y_t, y_{t-\tau}, \dots, y_{t-(m-1)\tau}), \quad t \in [m\tau; N - m\tau].$$

In the k-nearest neighbours algorithm, a future value y_{N+1} (*target value*) of data series can be approximately predicted from the evolution of k vectors $\vec{Y}_t^{(j)}$, called *nearest neighbours*, which are chosen according to some metrics, to be specified later:

$$y_{N+1} = \Phi\left(\vec{Y}_t^{(1)}, \vec{Y}_t^{(2)}, \dots, \vec{Y}_t^{(k)}\right).$$

A particular choice might be to take the average value of the k target values of the corresponding k-nearest neighbours:

$$y_{N+1} = \frac{1}{k}\sum_{j=1}^{k} F\left(\vec{Y}_t^{(j)}\right) = \frac{1}{k}\sum_{j=1}^{k} y_{t+1}^{(j)}. \tag{3}$$

The criterion for selecting a vector to be a nearest neighbour is the measure of closeness of the concerned vector $\vec{Y}_t$ to the fitting set $\vec{Y}_N$. The determination of the measure of closeness is the major factor responsible for the prediction success. Closeness is usually defined in terms of a metrics distance on the Euclidean space. The most common choices are the *Minkowski* metrics:

$$L_M(\vec{Y}_N, \vec{Y}_t) = (|y_N - y_t|^d + |y_{N-\tau} - y_{t-\tau}|^d + \dots + |y_{N-(m-1)\tau} - y_{t-(m-1)\tau}|^d)^{\frac{1}{d}}. \tag{4}$$

This metrics has some particular cases:

(1) in case $d = 1$ we have the *cityblock* metrics

$$L_1(\vec{Y}_N, \vec{Y}_t) = |y_N - y_t| + |y_{N-\tau} - y_{t-\tau}| + \dots + |y_{N-(m-1)\tau} - y_{t-(m-1)\tau}|; \tag{5}$$

(2) case $d = 2$ gives us the widely used *Euclidean* metrics

$$L_2(\vec{Y}_N, \vec{Y}_t) = (|y_N - y_t|^2 + |y_{N-\tau} - y_{t-\tau}|^2 + \dots + |y_{N-(m-1)\tau} - y_{t-(m-1)\tau}|^2)^{\frac{1}{2}}; \tag{6}$$

(3) and while $d = \infty$ we have the *Chebychev* metrics

$$L_\infty(\vec{Y}_N, \vec{Y}_t) = \max(|y_N - y_t|, |y_{N-\tau} - y_{t-\tau}|, \dots, |y_{N-(m-1)\tau} - y_{t-(m-1)\tau}|).$$

3. Adaptive *k*-nearest neighbours method

In this paper, we are concerned with a modification of the k-nearest neighbours method in the case, where the dynamics F given by Eq. (2) has some invariance properties. Suppose that the dynamics is translation invariant:

$$F(y_N + \mu, y_{N-\tau} + \mu, \dots, y_{N-(m-1)\tau} + \mu) = F(\vec{Y}_N) + \mu \tag{7}$$

and that it is homogeneous with respect to scaling:

$$F(\lambda y_N, \lambda y_{N-\tau}, \dots, \lambda y_{N-(m-1)\tau}) = \lambda F(\vec{Y}_N). \tag{8}$$

Such an assumption seems natural in many cases, where the dynamic does not depend on the absolute value of the past, but rather on the shape of the fluctuations around some average value and their overall size. In this case, the dynamic does not depend on m variables, but rather on the reduced variables $(\vec{Y}_N - \langle\vec{Y}_N\rangle)\alpha$, where $\langle\vec{Y}_N\rangle$ denotes the average value, and α is a scaling parameter that fixes the variance to 1.

Note that other invariance of that type can be treated by an obvious modification of the method we are going to describe. However, for the sake of definiteness, we limit ourselves to this specific invariance. One possible way to take into account this invariance of the dynamics in the estimation of the nearest neighbours would be to translate to zero average value and scale to variance 1 each time delay embedding before applying a Minkowski distance metrics (4).

Here however we propose an alternative method to factorize out the invariance of the dynamics in the estimation process. We therefore introduce the following metrics where we factorize out the symmetry group of the dynamics:

$$L_A(\vec{Y}_N, \vec{Y}_t) = \min_{\lambda_t, \mu_t} \mathcal{F}_t(\lambda_t, \mu_t),$$
$$\text{where} \quad \mathcal{F}_t(\lambda_t, \mu_t) = (|y_N - \lambda_t y_t - \mu_t|^d + |y_{N-\tau} - \lambda_t y_{t-\tau} - \mu_t|^d + \dots + |y_{N-(m-1)\tau} - \lambda_t y_{t-(m-1)\tau} - \mu_t|^d)^{\frac{1}{d}}. \tag{9}$$

The parameter of minimization λ_t in (9) equilibrates the amplitude difference between $\vec{Y}_N$ and $\vec{Y}_t$, and parameter μ_t is responsible for the trend of time series. These parameters lie in the range of

$$\lambda_t \in [1; R_t/r_t], \quad \mu_t \in [0; R_t - r_t], \tag{10}$$

where R_t and r_t are the largest and the smallest elements of vector $\vec{Y}_t$ correspondingly.

Now based on this quotient metrics, we choose the k-nearest neighbours. From (7)–(8), we obtain that the associated predicted target value $\tilde{y}_{t+1}$ for each vector $\vec{Y}_t$ is then the back-translated and rescaled value:

$$\tilde{y}_{t+1} = \lambda_t F(\vec{Y}_t) + \mu_t = \lambda_t y_{t+1} + \mu_t.$$

Finally, the target value y_{N+1} (3) is predicted as an average of local target values $\tilde{y}^{(j)}_{t+1}$, $j = 1 \dots k$:

$$y_{N+1} = \frac{1}{k}\sum_{j=1}^{k}\left(\lambda_t^{(j)} y_{t+1}^{(j)} + \mu_t^{(j)}\right).$$

For given vectors $\vec{Y}_N$ and $\vec{Y}_t$, the optimization problem (9) can be solved in general for every d using, for example, the Simulated Annealing approach [15]. In the case $d > 1$, the derivatives of the function $\mathcal{F}_t(\lambda_t, \mu_t)$ exist and, therefore, we can obtain λ_t and μ_t using the algorithm of Levenberg-Marquardt [16] optimization or other gradient method.

However, this optimization problem has two interesting particular cases. In the first case of *Adaptive Euclidean* metrics by $d = 2$, we define

$$\mathcal{F}_t(\lambda_t, \mu_t) = ((y_N - \lambda_t y_t - \mu_t)^2 + (y_{N-\tau} - \lambda_t y_{t-\tau} - \mu_t)^2 + \cdots + (y_{N-(m-1)\tau} - \lambda_t y_{t-(m-1)\tau} - \mu_t)^2)^{\frac{1}{2}}.$$

If we then consider two conditions $\partial\mathcal{F}_t(\lambda_t, \mu_t)/\partial\lambda_t = 0$ and $\partial\mathcal{F}_t(\lambda_t, \mu_t)/\partial\mu_t = 0$ and solve the corresponding linear system, we can obtain the solution of the minimization problem analytically:

$$\mu_t = \frac{S_N S_{tt} - S_t S_{Nt}}{m S_{tt} - S_t^2}, \quad \lambda_t = \frac{m S_{Nt} - S_t S_N}{m S_{tt} - S_t^2},$$

where $S_t = \sum_{i=1}^{m} y_{t-(i-1)\tau}$, $S_{tt} = \sum_{i=1}^{m} y^2_{t-(i-1)\tau}$,

$$S_N = \sum_{i=1}^{m} y_{N-(i-1)\tau}, \quad S_{Nt} = \sum_{i=1}^{m} y_{N-(i-1)\tau} y_{t-(i-1)\tau}.$$

The second particular case of the metrics (9) with $d = 1$ is the *adaptive cityblock* metrics, where the function $\mathcal{F}_t(\lambda_t, \mu_t)$ is defined as

$$\mathcal{F}_t(\lambda_t, \mu_t) = |y_N - \lambda_t y_t - \mu_t| + |y_{N-\tau} - \lambda_t y_{t-\tau} - \mu_t| + \cdots + |y_{N-(m-1)\tau} - \lambda_t y_{t-(m-1)\tau} - \mu_t|. \quad (11)$$

In this case, the continuous optimization problem is not smooth; therefore, we cannot use a gradient optimization method. As an alternative to the simulated annealing algorithm, we propose a discrete global two-dimensional search, which can be effectively implemented using a fast median calculation:

(1) First, we perform the minimum search over discrete set of λ_t, where we divide the segment of λ_t (10) into p parts as

$$\lambda_t^{(i)} = 1 + i\frac{R_t/r_t - 1}{p}, \quad i \in [0, p] \in \mathbb{N}.$$

(2) Next, we use the property that the median Med $=$ argmin $\sum |x_i - m|$ is the central point, which minimizes the average of the absolute deviations of a set of points $\{x_i\}$. Using this property, for every fixed $\lambda_t^{(i)}$ we can find the value of the parameter $\mu_t^{(i)}$, which minimizes the function (11):

$$\mu_t^{(i)} = \mathrm{Med}(|y_N - \lambda_t^{(i)} y_t|, |y_{N-\tau} - \lambda_t^{(i)} y_{t-\tau}|, \dots, |y_{N-(m-1)\tau} - \lambda_t^{(i)} y_{t-(m-1)\tau}|).$$

To improve the calculation of $\mu_t^{(i)}$, an effective method for fast median search [17] can be used.

(3) Using the sets $\lambda_t^{(i)}$ and $\mu_t^{(i)}$, we can approximate the expression (9) as

$$L_A(\vec{Y}_N, \vec{Y}_t) = \min_i(|y_N - \lambda_t^{(i)} y_t - \mu_t^{(i)}| + |y_{N-\tau} - \lambda_t^{(i)} y_{t-\tau} - \mu_t^{(i)}| + \cdots + |y_{N-(m-1)\tau} - \lambda_t^{(i)} y_{t-(m-1)\tau} - \mu_t^{(i)}|).$$

At this step, for every vector $\vec{Y}_t$ we obtain the corresponding values i_0, which minimizes the above-mentioned equation. Therefore, the metrics (9)–(11) has its minimum for $\lambda_t = \lambda_t^{(i_0)}$ and $\mu_t = \mu_t^{(i_0)}$.

The parameter p determines here the fragmentation of λ_t and, therefore, has an influence on the precision of the optimization. However, an improvement in our ability to find the minimum of the metrics with larger p more precisely comes at the expense of computation speed.

Let us consider one simple example. If we assume the embedding dimension $m = 2$ and the embedding delay $\tau = 1$, then we obtain a fitting set and a time delay embeddings as follows:

$$\vec{Y}_N = (y_N, y_{N-1}), \quad \vec{Y}_t = (y_t, y_{t-1}).$$

We can then do a linear regression between these two vectors, i.e. find λ and μ such that

$$y_N = \lambda y_t + \mu + \varepsilon_t, \quad y_{N-1} = \lambda y_{t-1} + \mu + \varepsilon_{t-1},$$

where ε_t and ε_{t-1} are two error terms. As we have two data points and two parameters, we can actually find λ^* and μ^* such that both error terms will be equal to zero, that is,

$$y_N = \lambda^* y_t + \mu^*, \quad y_{N-1} = \lambda^* y_{t-1} + \mu^*.$$

Hence the distance (9) defined above, $L_A(\vec{Y}_N, \vec{Y}_t) = 0$ for any t, and we cannot predict a future value y_{N+1} by the value $\lambda^* y_{t+1} + \mu^*$. From this example we can see, that in comparison with the standard nearest neighbours method, the adaptive approach proposed in this paper has some restrictions on the choice of the embedding dimension: $m > 2$.

4. Numerical simulations

To illustrate the accuracy of the method, we generate and predict in this section several deterministic and chaotic time series as well as consider two real time series related to the bank

Table 1
Prediction summary for deterministic synthetic examples

Time series	k	m	MAPE(L_2)	MAPE(L_A)
Time series with seasonal dependence	2	100	0.5%	0.1%
Time series with multiplicative seasonality	2	15	20.8%	6.8%
High-frequency time series	3	70	26.4%	16.7

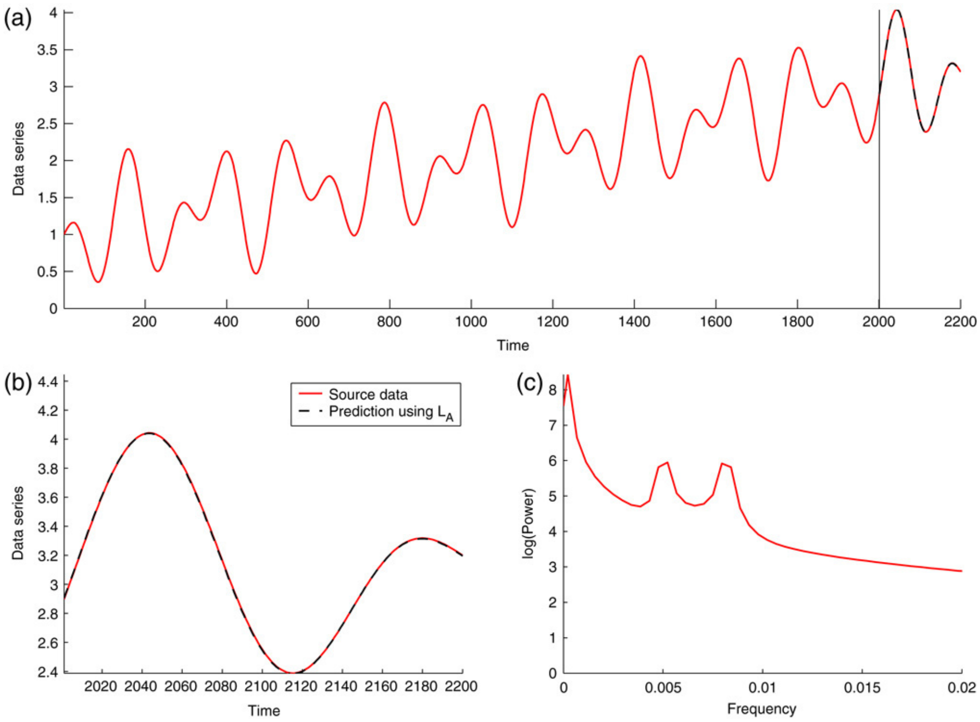

Fig. 1. (a) Time series with seasonal dependence, vertical line shows the prediction start. (b) Zoom of predicted values and (c) Fourier spectrum of source data.

data processing. For the accuracy control, we use the Mean Absolute Percentage Error, calculated as

$$\text{MAPE} = \frac{1}{M}\sum_{i=1}^{M}\left|\frac{y_i - \tilde{y}_i}{y_i}\right| \cdot 100\%,$$

where y_i is the source points, $\tilde{y}_i$ is the predicted points and M is the number of predicted points.

4.1. Deterministic synthetic examples

In this section, we test the proposed method on three deterministic time series with pronounced seasonal dependence, trend and amplitude change. Using adaptive cityblock metrics (11), we calculate the prediction and the MAPE(L_A) value for each time series. For this calculation, we select the embedding dimension similar to the seasonal period using an autocorrelation technique.

To compare the prediction with those obtained using standard Euclidean metrics (6), we also apply k-nearest neighbours algorithm (3) to the corresponded detrended time series and evaluate the MAPE(L_2) value. For all considered time series, we collected the calculation parameters and corresponding MAPE values in the Table 1.

(1) The first synthetic time series with seasonal dependence and linear increasing is showed in Fig. 1(a):

$$S(t) = \cos(t/25)\sin(t/100) + t/1000 + 1,$$
$$t \in [0, 2200],$$

where t denotes time. In Fig. 1(b), solid line corresponds to the original time series and dashed line corresponds to a forecast, made for 10% of series length using new adaptive approach. For this time series, we do not see any difference between source and predicted data. The Fourier spectrum of the series shown in Fig. 1(c) indicates clearly a modulation phenomenon (multiplication of two sinusoidal signals) in the data [18]. Amplitude of seasonal component does not change—it means that optimal value of parameter λ_t for all embedding dimension vectors is equal to zero. During the prediction, only the parameter μ_t, responsible for a changing trend, should be determined. Prediction applied to the detrended data using standard Euclidean metrics (6) also predict this time series perfect, as there is no amplitude changing in this data series. Note what other methods like described in [18] and based on a modulation model can predict this time series perfect with error close to zero. Obviously, this is assuming that the periods of the sinusoidal signals are known. In more general applications,

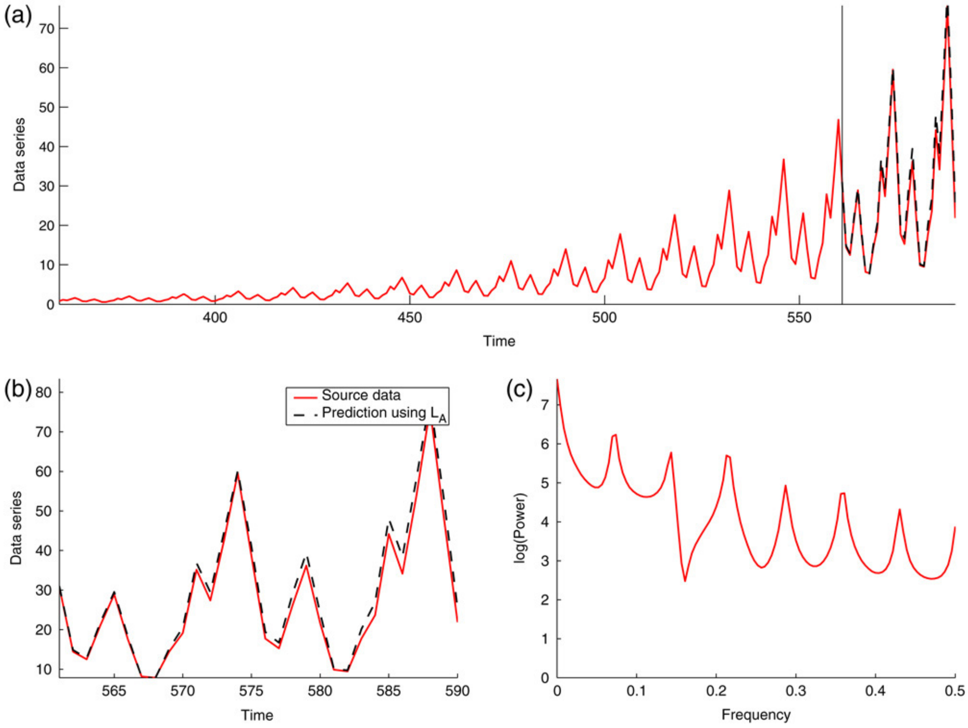

Fig. 2. (a) Time series with multiplicative seasonality, vertical line shows the prediction start. (b) Zoom of predicted values and (c) Fourier spectrum of source data.

where the periods have to be estimated, then the errors and forecasting performance would depend on how well the periods are estimated. Our propose method does not require to estimate these periods.

(2) The next nonlinear simulation shown Fig. 2(a) is a time series with multiplicative seasonality. This time series has a nonlinear trend, and the amplitude of seasonal oscillations increases with time. We model this time series as

$$S(t) = \begin{cases} R(t), & t \in [0; 79] \\ \dfrac{S(t-t_0)^2}{S(t-2t_0)}, & t \in [80; 590], \quad t_0 = 14, \end{cases}$$

where $R(t) = \dfrac{t}{70\,000}(\sin(t/350)\cos(9t/7) + 10)$.

Prediction is done for 15% of source time series length; from Fig. 2(b) we see very good agreement between the source (solid line) and predicted (dashed line) data. This time series has a peculiarity, that the amplitude of every following periodic segment is a fixed time as much as the previous segment. Evidently, the parameter λ_t has caught this dependency. This signal may be also modelled, for example, as a Dynamic Harmonic Regression [19] that includes a trend and a seasonal component.

(3) High-frequency time series with multiplicative seasonality and smoothly increasing amplitude is presented in Fig. 3(a). To produce this time series, we used the following explicit expression:

$$S(t) = \frac{t}{100}|\sin(t/2)| + |\cos(t/20)|, \quad t \in [0; 550].$$

It models a high-frequency series with seasonal periodicity. In Fig. 3(b), prediction is done for 10% of time series length. The Fourier spectrum (Fig. 3(c)) is relatively complicated, therefore this synthetic series is not usual to consider the interaction between trends and seasonality, and it is difficult to handle it with other standard methods.

4.2. *Chaotic time series*

To make some sensible comparisons to the 'conventional' nearest neighbours method, in this section we perform some tests on a typical and complicated chaotic time series such as a time series generated from the Duffing equation, from Mackey–Glass delay-differential equation and chaotic Ikeda map. These time series do not have any global amplitude changing or trend, therefore we can perform the direct comparison between three metrics: L_A, L_1, and L_2. As for deterministic synthetic examples shown in the previous section, we collected the calculation parameters and corresponding MAPE values in the Table 2.

(1) The first example is a solution of Duffing equation, shown in Fig. 4(a). For prediction, we use only the horizontal component of this chaotic two-component series. In the definition (1) we assume that the time series consists of the positive values $y_i \geq 0$, therefore we added the value x_0 to the source horizontal component (Fig. 4(b)). The prediction is done for 5% of time series length. Comparison of prediction using different metrics is shown in Fig. 4(c), where solid line corresponds to the source time series, dashed line depicts the prediction using the adaptive metrics (9), dot–dashed and dotted lines represent cityblock (5) and Euclidean (6) metrics correspondingly.

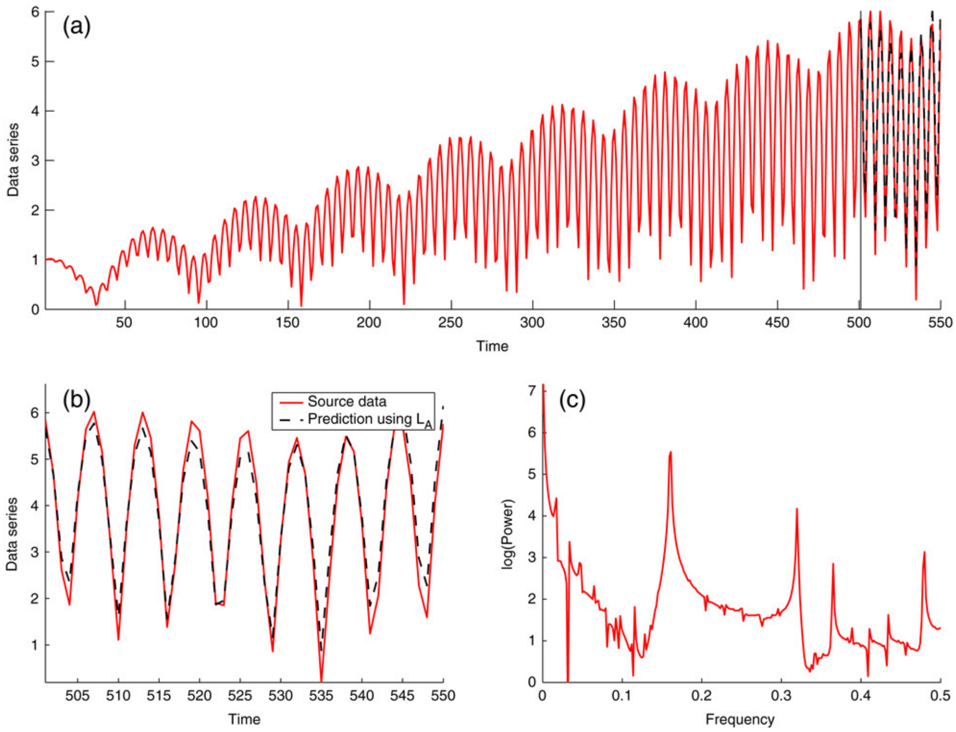

Fig. 3. (a) High-frequency time series with multiplicative seasonality, vertical line shows the prediction start. (b) Zoom of predicted values and (c) Fourier spectrum of source data.

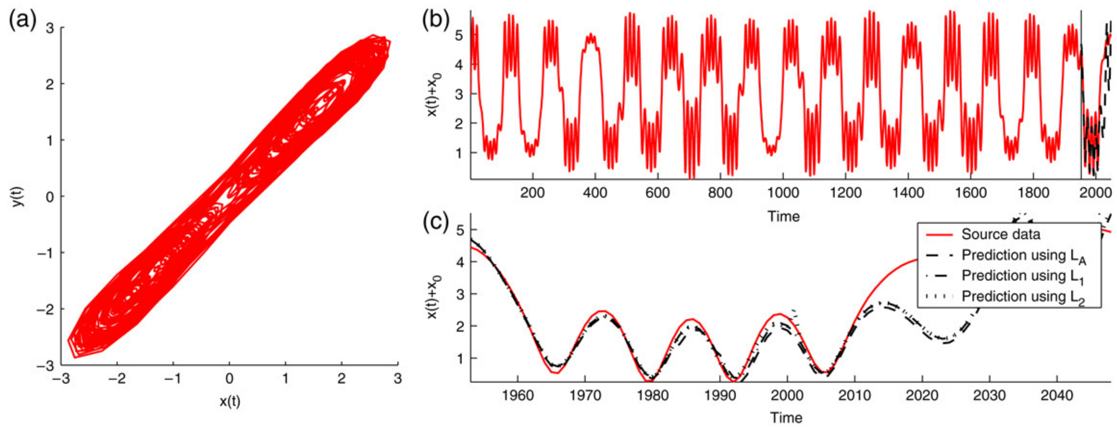

Fig. 4. (a) Chaotic time series based on the solution of Duffing equation. (b) Horizontal component of the time series, vertical line shows the prediction start. (c) Zoom of predicted values for three different metrics.

Table 2
Prediction summary for chaotic time series

Time series	k	m	MAPE(L_1)	MAPE(L_2)	MAPE(L_A)
Duffing equation	2	48	20.4%	18.7%	21.6%
Mackey–Glass delay differential equation	3	7	1.6%	4.6%	3.5%
Chaotic Ikeda map	3	7	33.4%	10.8%	26.9%

(2) The Mackey–Glass delay differential equation is defined in terms of a scalar function of time, $x(t)$ [2,20,21]:

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = -bx(t) + \frac{ax(t-\tau_0)}{1+x^c(t-\tau_0)},$$
$$a = 0.2, b = 0.1, c = 10, \tau_0 = 17.$$

The system is infinite dimensional because it is a time delay equation, and has an infinite number of Lyapunov exponents. The solution of this system is shown in Fig. 5(a). The prediction is done only for the horizontal component for 5% of time series length (Fig. 5(b)). According to the Takens'

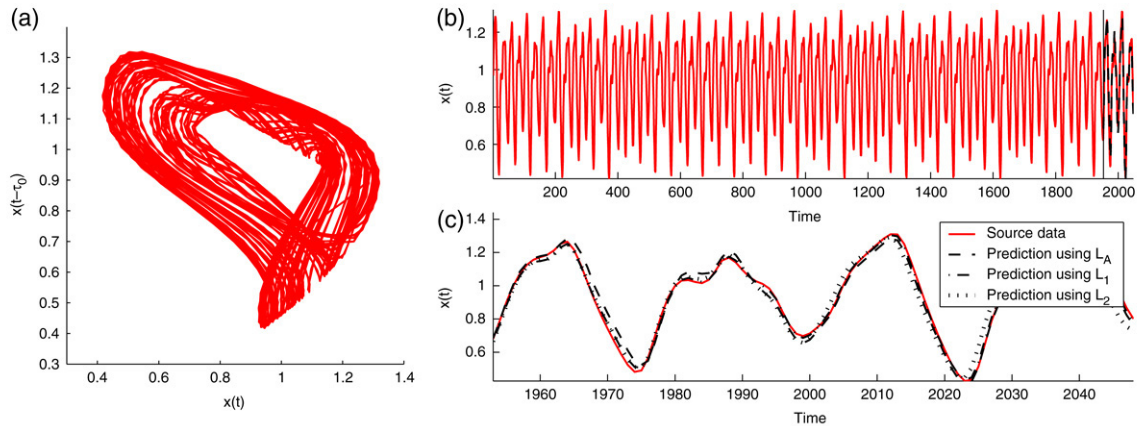

Fig. 5. (a) Chaotic time series based on the solution of the Mackey–Glass delay differential equation. (b) Horizontal component of the time series, vertical line shows the prediction start. (c) Zoom of predicted values for three different metrics.

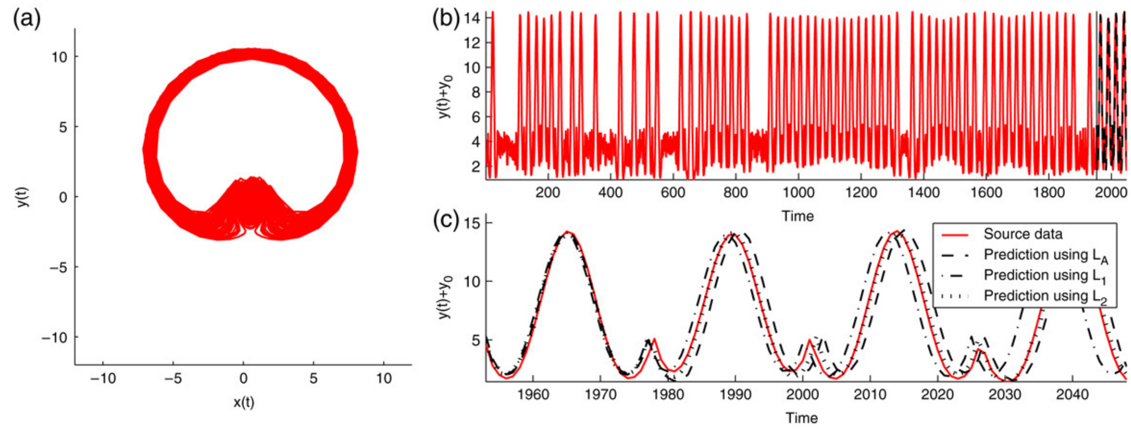

Fig. 6. (a) Chaotic time series based on the Ikeda map. (b) Vertical component of the time series, vertical line shows the prediction start. (c) Zoom of predicted values for three different metrics.

theorem, the embedding dimension m is equal to the dimension of the reconstructed space $m = 2D + 1$, D being the dimension of the attractor of the dynamical system; in this case $D = 2$. Since the underlying system is not translation and dilation invariant, in the optimal choice of the past history, two degrees of freedom are factorized out, which leaves us with an effective dimension of $5 + 2 = 7$. Therefore, we use the embedding dimension $m = 7$ to produce the Fig. 5(c) where solid line corresponds to the source time series, dashed line depicts the prediction using the adaptive metrics (9), dot–dashed and dotted lines represent cityblock (5) and Euclidean (6) metrics correspondingly.

(3) The Ikeda map may be given in terms of a mapping of the complex plane to itself. The coordinates of the phase space are related to the complex degree of freedom $z = x + iy$. The mapping itself is [2]

$$z_{n+1} = p + Bz_n e^{i\alpha - i\beta/(1+|z_n|^2)},$$
$$p = 1, B = 1, \alpha = 0.4, \beta = 6.$$

Fig. 6(a) is based on a time series of the Ikeda map, of length 2048. The prediction is done for the translated vertical component $y(t) + y_0$ for 5% of time series length (Fig. 6(b)). As in the previous example, we use the same embedding dimension $m = 7$ to calculate three predictions using three different metrics; these results are shown in Fig. 6(c).

From the one hand, the numerical evidence presented here shows that the proposed method could make the prediction on complicated chaotic time series at least as good as that by the conventional k-nearest neighbours algorithm. This could be a good justification to make some claims about presented method in the area of nonlinear/chaotic time series prediction.

From the other hand, these examples demonstrate that the adaptive method proposed in this contribution is only an improvement on the standard nearest neighbours method, which reduces problems of amplitude changing and trend determination. In the situation of detrended signals without amplitude changing, the prediction results, as expected, are similar.

4.3. *Real data*

In Figs. 7 and 8, the fragments of dirty bank data are presented. This data is related to the processing of bank-to-bank transactions and commercial transactions, which occur during weekdays. Calculations for both series are done for 10% of series length.

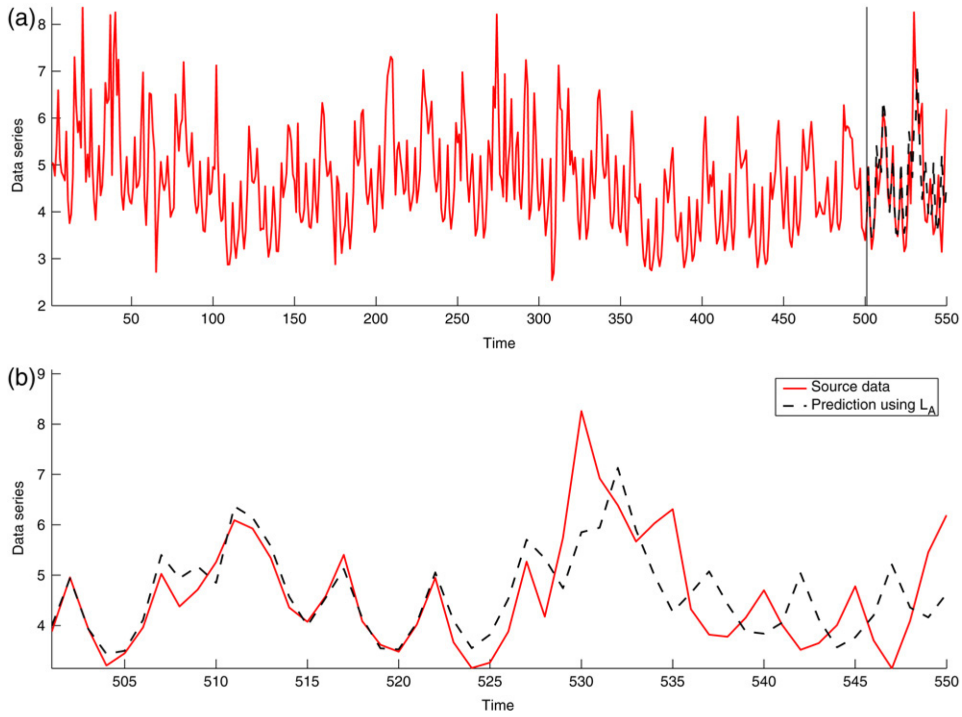

Fig. 7. (a) Real data—payments based on paper-documents, vertical line shows the prediction start. (b) Zoom of predicted values.

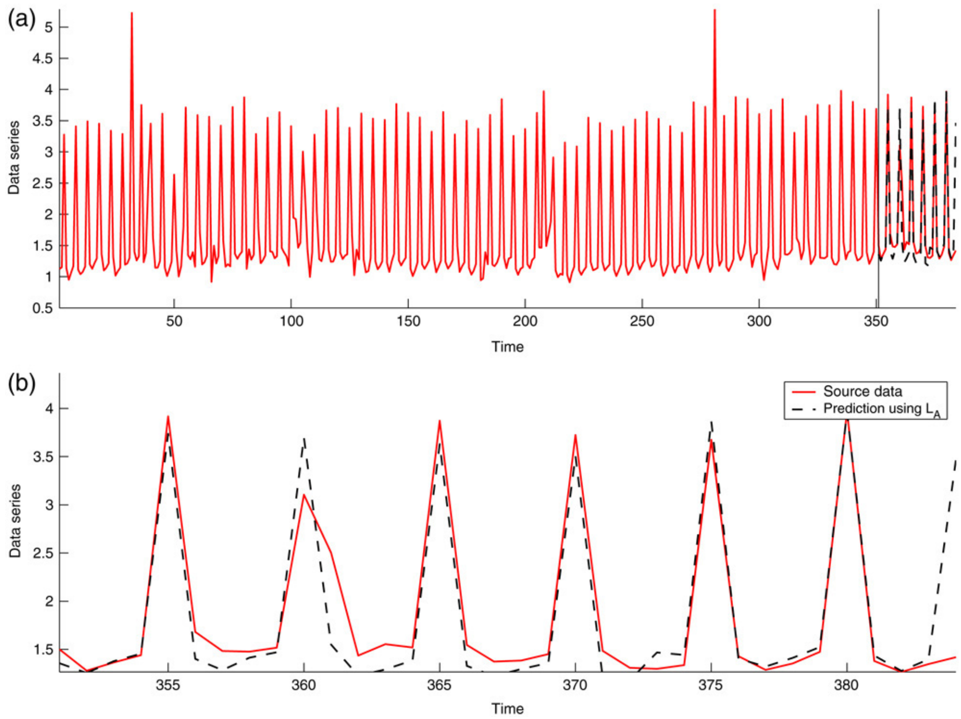

Fig. 8. (a) Real data—number of transactions, vertical line shows the prediction start. (b) Zoom of predicted values.

Data in Fig. 7 are the payments based on paper-documents (filled-in and send to the bank). Concerned data have appreciable seasonal component with sinusoidal trend. MAPE of prediction is 12,5% with following parameters of calculation: $m = 20$, $\tau = 1$, $k = 2$.

Daily numbers of automatic teller machine transactions is presented in Fig. 8. These data have perceptible periodical component and unnoticeable trend. Setting the parameters $m = 40$, $\tau = 1$, $k = 2$, MAPE of prediction appears 12,7%.

5. Conclusions

We proposed a method for nonlinear prediction of time series under the assumption that the underlying dynamics is translation and scale invariant. We show that our method is able to predict periodical time series with complicated structure. For synthetic and real signals forecast made by presented method gives us better fit to real data in comparison with standard method. However, like the standard nearest neighbours technique, proposed algorithm is particularly well suited to seasonal data. As distinct from the standard k-nearest neighbours, the algorithm presented in this paper offers an advantage of adapting to local variations of trend and amplitude. Therefore, the problem of trend determination and amplitude tendency observation is eliminated. Other kinds of invariance may be considered by some obvious modification of the presented method. We have shown, that this methods works even for chaotic time series. Although in the derivation of the method we have supposed some underlying invariance, the method even works for dynamical systems which do not have this invariance. In turn, we had to take a slightly larger embedding dimension to factorize out this invariance.

Acknowledgments

This project is supported by a grant from the Deutsche Forschungsgemeinschaft (DFG) within the framework of the priority program SPP 1114 "Mathematical methods for time series analysis and digital image processing". The third author acknowledges the scholarship of Helmholtz Institute for Supercomputational Physics. We also thank the anonymous reviewers for constructive criticism that greatly improved the presentation of the manuscript.

References

[1] R. Hegger, H. Kantz, T. Schreiber, Practical implementation of nonlinear time series methods: The TISEAN package, Chaos: An Interdisciplinary J. Nonlinear Sci. 9 (2) (1999) 413–435.

[2] D.B. Murray, Forecasting a chaotic time series using an improved metric for embedding space, Physica D Nonlinear Phenom. 68 (1993) 318–325.

[3] J.D. Farmer, J.J. Sidorowich, Predicting chaotic time series, Phys. Rev. Lett. 59 (8) (1987) 845–848.

[4] B. Brabec, R. Meister, A nearest-neighbor model for regional avalanche forecasting, Ann. Glaciol. 32 (2001) 130–134.

[5] V. Pérez-Muñuzuri, I.R. Gelpi, Application of nonlinear forecasting techniques for meteorological modeling, Ann. Geophys. 18 (2000) 1349–1359.

[6] F.J. Mulhern, R.J. Caprara, A nearest neighbor model for forecasting market response, Int. J. Forecast. 10 (2) (1994) 191–207.

[7] F. Fernandez-Rodrguez, S. Sosvilla-Rivero, J. Andrada-Felix, Exchange-rate forecasts with simultaneous nearest-neighbour methods: evidence from the EMS, Int. J. Forecast. 15 (1999) 383–392.

[8] J.M. Gimeno-Illa, J. Béjar-Alonso, M. Sánchez-Marré, Nearest-neighbours for time series, Appl. Int. 20 (1) (2004) 21–35.

[9] J. McNames, A nearest trajectory strategy for time series prediction, in: Proceedings of the International Workshop on Advanced Black-Box Techniques for Nonlinear Modeling, K.U. Leuven, Belgium, 1998, pp. 112–128.

[10] L. Cao, Practical method for determining the minimum embedding dimension of a scalar time series, Physica D Nonlinear Phenom. 110 (1997) 43–50.

[11] M. Small, C.K. Tse, Optimal embedding parameters: a modelling paradigm, Physica D Nonlinear Phenom. 194 (2004) 283–296.

[12] D. Kugiumtzis, Assessing different norms in nonlinear analysis of noisy time series, Physica D Nonlinear Phenom. 105 (1997) 62–78.

[13] N. Tanaka, H. Okamoto, M. Naito, An optimal metric for predicting chaotic time series, Jpn. J. Appl. Phys 34 (1) (1995) 388–394.

[14] P. García, J. Jiménez, A. Marcano, F. Moleiro, Local optimal metrics and nonlinear modeling of chaotic time series, Phys. Rev. Lett. 76 (9) (1996) 1449–1452.

[15] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, Optimization by simulated annealing, Science 220 (4598) (1983) 671–680.

[16] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipe in C: The Art of Scientific Computing, Cambridge University Press, 1992.

[17] N. Wirth, Algorithms + Data Structures = Programs, Prentice Hall PTR, Upper Saddle River, NJ, USA, 1978.

[18] D.J. Pedregal, P.C. Young, Modulated cycles, an approach to modelling periodic components from rapidly sampled data, Int. J. Forecast. 22 (1) (2006) 181–194.

[19] P.C. Young, D.J. Pedregal, W. Tych, Dynamic harmonic regression, J. Forecast. 16 (6) (1999) 369–394.

[20] L. Cao, Practical method for determining the minimum embedding dimension of a scalar time series, Physica D Nonlinear Phenom. 110 (1997) 43–50.

[21] B.P. Bezruchko, A.S. Karavaev, V.I. Ponomarenko, M.D. Prokhorov, Reconstruction of time-delay systems from chaotic time series, Phys. Rev. E 64 (5) (2001) 056216.