Neurocomputing 72 (2009) 2618–2626

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Non-uniform attractor embedding for time series forecasting by fuzzy inference systems

Minvydas Ragulskis *, Kristina Lukoseviciute

Research Group for Mathematical and Numerical Analysis of Dynamical Systems, Kaunas University of Technology, Studentu 50-222, Kaunas, LT-51638, Lithuania

ARTICLE INFO

Article history:
Received 16 April 2008
Received in revised form
1 August 2008
Accepted 15 October 2008
Communicated by M.T. Manry
Available online 11 November 2008

Keywords:
Time series forecasting
Fuzzy inference system
Attractor embedding

ABSTRACT

A new method for identification of an optimal set of time lags based on non-uniform attractor embedding from the observed non-linear time series is proposed in this paper. Simple deterministic method for the determination of non-uniform time lags comprises the pre-processing stage of the time series forecasting algorithm which is implemented in the form of a fuzzy inference system. Identification of embedding parameters of the underlying dynamical system includes not only optimization of time lags but also determination of optimal dimension of the reconstructed phase space. Experiments done with benchmark chaotic time series show that the proposed method can considerably improve the forecasting accuracy. The proposed method seems to be an efficient candidate for prediction of time series with multiple time scales and noise.

1. Introduction

Time series forecasting, especially long-term prediction, is a challenge in many fields of science and engineering. Many techniques exist for time series forecasting. In general, the object of these techniques is to build a model of the process and then use this model on the last values of the time series to extrapolate past behavior into future. Forecasting procedures include different techniques and models. Moving averages techniques, random walks and trend models, exponential smoothing, state space modeling, multi-variate methods, vector autoregressive models, cointegrated and causal models, methods based on neural, fuzzy networks or data mining and rule-based techniques are typical methods used in time series forecasting [1–4]. Although the search for a best time series forecasting method continues, it is agreeable that no single method will outperform all others in all situations.

The object of this paper is to demonstrate that a well established time series forecasting method based on fuzzy networks can be considerably improved if non-uniform sampling of the recent outcomes is used instead of standard uniform sampling (when time lags between the data points are all equal). We use a concept based on attractor embedding to construct a simple and deterministic rule for the selection of a non-uniform vector of time lags.

Attractor embedding is an important step in the process of computation of invariant quantities used to characterize the dynamics of a system represented by a scalar time series [5,6]. But computation of such quantities as Lyapunov spectrum is dependant on techniques used for the attractor reconstruction [7,8]. Time delay and embedding dimension are two factors determining the attractor reconstruction by a delay coordinate method. Embedding dimension determines the dimension of the phase space. One of the classical techniques used for the identification of the optimal embedding dimension is the false nearest neighbor (FNN) method [5]. The quantification of the time delay is usually performed using computational techniques based on the autocorrelation function [9], mutual information [10] or geometric considerations where the optimality is based on the principle of spreading out the attractor in the embedded space [11]. The object of delay selection methods is making components of reconstructed vectors independent as far as possible, yet not too far to keep the information about dynamic properties of the embedded time series.

Standard methods for the selection of the optimal time delay vector produce equal time lags between adjacent coordinates of the multi-dimensional phase space. Such uniform time delay vectors are widely exploited for time series forecasting using fuzzy and neural networks [12–15] (Fig. 1a).

Non-uniform attractor embedding when time lags between adjacent coordinates are not necessarily equal can be considered as an alternative to uniform embedding techniques. Schreiber [9] points out that non-uniform embeddings are better for the reconstruction of attractors in a multi-dimensional delay coordinate space when a time series involves several incommensurate

* Corresponding author.
E-mail address: minvydas.ragulskis@ktu.lt (M. Ragulskis).

0925-2312/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2008.10.010

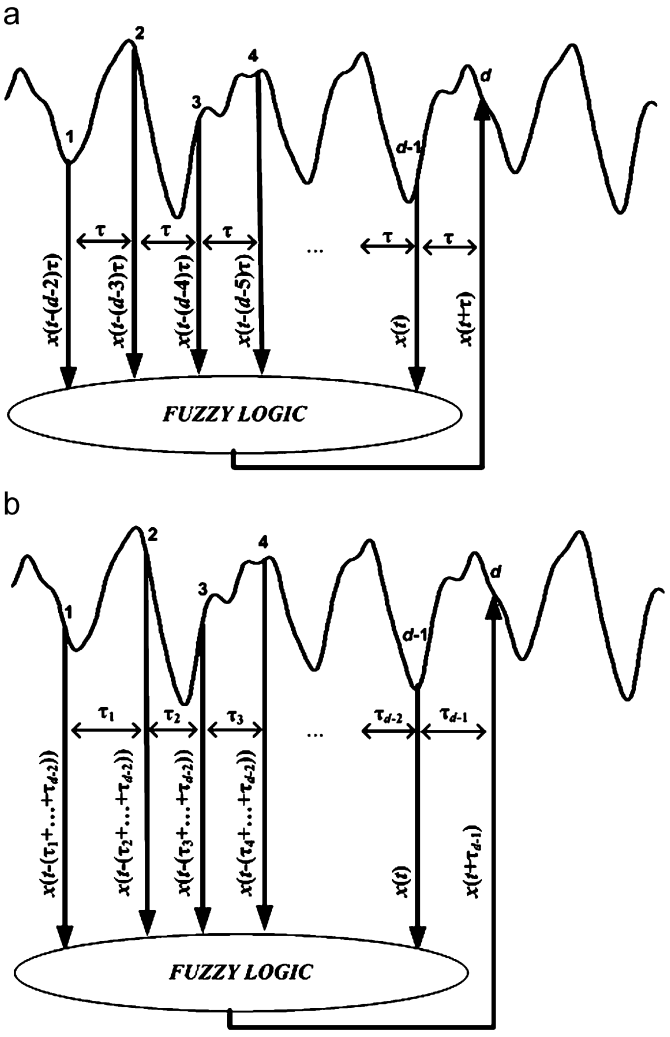

Fig. 1. Schematic diagram illustrating time series forecasting techniques based on: (a) uniform embedding—all time lags are equal; (b) non-uniform embedding—time lags can be different. The embedding dimension (also the number of channels) is d; the number of time lags is $d-1$. The fuzzification interface comprises $d-1$ input channels and 1 output channel. Function values are passed into the fuzzy inference system through the input channels; the predicted function value is delivered through the output channel.

frequencies. Recently, such non-uniform embeddings have been successfully exploited in time series forecasting [16]. Unfortunately, heuristic computational methods for determination of time lags for non-uniform embedding are usually preferred over autocorrelation or mutual information based methods. Particularly, a feed-forward neural network with structural learning is exploited in [16] for identification of a near-optimal set of time lags.

In this paper we propose a simple, deterministic method for the selection of optimal time lags for non-uniform embedding. This method emphasizes the multi-dimensionality of the problem, is able to cope with optimization problems in a multi-parameter space of arguments and is well applicable in time series forecasting routines.

This paper is organized as follows: Section 2 presents a method for the selection of the optimal set of time lags. Initially, a computational technique for the determination of uniform time lags is developed. Then it is generalized for non-uniform embedding. Simulation results for two time-series forecasting problems are presented in Section 3. Concluding remarks comprise the last section.

2. Selection of the optimal time lag vector for attractor embedding

2.1. Construction of the objective function

Suppose that a scalar time series is given as $x_1, x_2, \ldots, x_N$. Then time delay vectors produced by a non-uniform embedding can be reconstructed as follows ([6]):

$$y_p^{(d)} = (x_p, x_{p+\tau_1}, \ldots, x_{p+\tau_1+\cdots+\tau_{d-1}}), \quad p = 1,2,\ldots,N-\sum_{i=1}^{d-1}\tau_i, \tag{1}$$

where d is the embedding dimension; $y_p^{(d)}$ is the p-th reconstructed vector with the embedding dimension d; $\tau_i \in N$. If the time span between two adjacent points in the scalar time series is δ (the sampling rate is $1/\delta$), $\delta\sum_{i=1}^{d-1}\tau_i$ is the total embedding window.

2.1.1. Initial considerations—two-dimensional delay coordinate space

We will use geometrical considerations to construct the objective function. The goal is to spread the attractor in the phase space as much as possible. We need to construct an analytical criterion describing the magnitude of the attractor's spreading in the phase space. We start from a two-dimensional delay coordinate space and investigate a geometrical shape of the attractor produced by a harmonic time series when it is mapped into the phase plane by time delay embedding.

A discrete harmonic time series can be represented as

$$x_s = A\,\sin(\omega\delta(s-1)+\varphi), \quad s = 1,2,\ldots,N, \tag{2}$$

where A is the amplitude, ω is the cyclic frequency and φ is the phase of the harmonic function. A pair of function values x_i and $x_{i+\tau}$ $(i = 1,2,\ldots,N-\tau)$ is mapped into a point $(x_i,x_{i+\tau})$ in a phase plane X_10X_2, where τ is the time lag; X_1 and X_2 are axes of the delay coordinate plane. Elementary mathematical transformations help to show that a harmonic time series is mapped into an ellipse by two-dimensional time delay embedding; the equation of that ellipse reads

$$X_2 = X_1\,\cos(\omega\tau\delta) + \sqrt{A^2 - X_1^2}\,\sin(\omega\tau\delta). \tag{3}$$

It can be noted that main diagonals of the ellipse lie on diagonals $X_2 = X_1$ and $X_2 = -X_1$ due to the symmetry of the harmonic signal. Now, if the embedding window is equal to

$$\tau\delta = \frac{2\pi}{\omega}(n-1), \quad n = 1,2,\ldots, \tag{4}$$

the ellipse is compressed into a line segment on the diagonal $X_2 = X_1$. Similarly, when

$$\tau\delta = \frac{\pi}{\omega} + \frac{2\pi}{\omega}(n-1), \quad n = 1,2,\ldots, \tag{5}$$

the ellipse is compressed into a line segment on the diagonal $X_2 = -X_1$. Finally, if

$$\tau\delta = \frac{\pi}{2\omega} + \frac{\pi}{\omega}(n-1), \quad n = 1,2,\ldots, \tag{6}$$

the harmonic time series is embedded into a circle in the delay coordinate plane.

Now the geometrical shape of the attractor will be exploited for the construction of a parameter quantifying the quality of the embedding.

2.1.2. Definition of the quality of the embedding

We use geometrical considerations for the definition of parameter Q_1 which characterizes the spreading of the attractor in the embedded space:

$$Q_1 = \frac{E}{\pi A^2}, \tag{7}$$

where E is the area of the mapped ellipse in the reconstructed two-dimensional phase space. It is clear that $0 \leqslant Q_1 \leqslant 1$. $Q_1 = 1$ corresponds to the optimal scenario when the harmonic time series is mapped to a circle. $Q_1 = 0$ (the worst scenario) occurs when the ellipse is compressed into a line segment on one of diagonals. Radiuses of the mapped ellipses r_1 and r_2 can be expressed explicitly from Eq. (3):

$$r_1 = A\,\sin(\omega\tau\delta)/\sqrt{1-\cos(\omega\tau\delta)},$$
$$r_2 = A\,\sin(\omega\tau\delta)/\sqrt{1+\cos(\omega\tau\delta)}. \tag{8}$$

Thus the area of the ellipse is

$$E = \pi \cdot r_1 r_2 = \pi A^2 |\sin(\omega\tau\delta)|, \tag{9}$$

and the quality parameter Q_1 takes the form

$$Q_1 = |\sin(\omega\tau\delta)|. \tag{10}$$

2.1.3. Multi-dimensional delay coordinate space, uniform embedding

If the dimension of the delay coordinate space is d, there exist $d(d-1)/2$ different planar projections of the embedded attractor. If axes of the delay coordinate space are numerated as X_i, $i = 1,\ldots,d$, then a projection of the harmonic time series will be a planar ellipse in a X_i–X_j plane, here $1 \leqslant i \leqslant d$; $1 \leqslant j \leqslant d$; $i \neq j$. Uniform embedding yields

$$\tau = \tau_1 = \cdots = \tau_{d-1}. \tag{11}$$

We define the parameter of embedding quality for every possible combination of i and j:

$$Q_{ij} = |\sin(\omega(j-i)\tau\delta)|. \tag{12}$$

Q_{ij} depends on the difference between j and i; thus the following notations are introduced:

$$\begin{aligned} Q_1 &= Q_{ij}|_{|j-i|=1},\\ Q_2 &= Q_{ij}|_{|j-i|=2},\\ &\vdots\\ Q_{d-1} &= Q_{ij}|_{|j-i|=d-1}. \end{aligned} \tag{13}$$

We define the function of embedding quality as the average of parameters of embedding quality in all possible different planar projections:

$$\begin{aligned} Q(\tau,\omega) &= \frac{2}{d(d-1)}\sum_{k=1}^{d-1}(d-k)Q_k\\ &= \frac{2}{d(d-1)}\sum_{k=1}^{d-1}(d-k)|\sin(k\omega\tau\delta)|. \end{aligned} \tag{14}$$

It is clear that $0 \leqslant Q(\tau,\omega) \leqslant 1$. Equality $Q = 0$ means that a harmonic series is compressed to line segments in all possible projections. For a harmonic time series $Q = 0$ occurs at

$$\omega = \frac{\pi}{\tau\delta}m, \quad m = 1,2,\ldots . \tag{15}$$

It can be noted that those frequencies do not depend on d. Moreover, integral

$$\begin{aligned} &\int_{(\pi/\tau\delta)m}^{(\pi/\tau\delta)(m+1)} Q(\tau,\omega)\,\mathrm{d}\omega\\ &\quad= \frac{2}{d(d-1)}\sum_{k=1}^{d-1}\left((d-k)\int_{(\pi/\tau\delta)m}^{(\pi/\tau\delta)(m+1)}|\sin(k\omega\tau\delta)|\,\mathrm{d}\omega\right)\\ &\quad= \frac{2}{\tau\delta}, \quad m \in N \end{aligned} \tag{16}$$

is an invariant quantity in respect to d.

2.1.4. Multi-dimensional delay coordinate space, non-uniform embedding

The number of different planar projections in d-dimensional delay coordinate space is still $d(d-1)/2$, but generalizations used in Eq. (14) are no more valid due to the fact that time lags are different. The function of embedding quality now becomes

$$\begin{aligned} Q(\tau_1,\ldots,\tau_{d-1},\omega) = \frac{2}{\mathrm{d(d-1)}}\Bigg(&\sum_{i=1}^{d-1}|\sin(\omega\delta\tau_i)|\\ &+\sum_{i=1}^{d-2}|\sin(\omega\delta(\tau_i+\tau_{i+1}))|+\cdots+\left|\sin\left(\omega\delta\sum_{j=1}^{d-1}\tau_j\right)\right|\Bigg) \end{aligned} \tag{17}$$

2.1.5. Definition of the objective function

Every harmonic component of the original time series will be affected by the function of embedding quality when the original time series will be embedded into a d-dimensional delay coordinate space. Harmonic components with frequencies where Q is small will be suppressed (in average) in all plane projections of the delay coordinate space. Similarly, harmonic components with frequencies where Q is high will experience rich representation in the reconstructed phase space. Keeping those considerations in mind we define an objective function F which characterizes the magnitude of the attractor's spreading in the embedded space:

$$F(\tau_1,\ldots,\tau_{d-1}) = \frac{\pi}{2}\frac{\int_0^\infty A(\omega)Q(\tau_1,\ldots,\tau_{d-1},\omega)\,\mathrm{d}\omega}{\int_0^\infty A(\omega)\,\mathrm{d}\omega}, \tag{18}$$

where $A(\omega)$ is Fourier amplitude spectrum of the original time series. We calculate indefinite integrals over the whole frequency range (in practice one should compute definite integrals up to a preset upper frequency bound). The integral of $A(\omega)$ in the denominator is used in order to normalize the magnitude of the objective function (this integral can be computed once for a concrete time series before the optimization procedure is commenced). A constant $\pi/2$ is used in order to make the value of the objective function equal to one when the time series is white noise; detailed explanations are presented in the next section.

2.2. Properties of the objective function

(a) $F(0,\ldots,0) = 0$. That follows from the definition of the function of embedding quality.

(b) $F(\ldots,\tau_k,\ldots,\tau_l,\ldots) = F(\ldots,\tau_l,\ldots,\tau_k,\ldots)$. The objective function is insensitive to the permutation of its arguments. Function F is constructed as an average area of all possible plane projections in the reconstructed phase space. Clearly, the objective function does not depend from the numeration order of time lags in the time delay vector.

The objective function can be calculated explicitly for certain types of scalar time series.

(c) If $A(\omega)=L$ ("white noise") where L is a positive constant and embedding dimension $d=2$, then the function of embedding quality is $Q(\tau_1,\omega)=|\sin(\omega\tau_1\delta)|$.
The area of the first peak of a function $A(\omega)Q(\tau_1,\omega)$ can be calculated explicitly:

$$S_1=\int_0^{\pi/\tau\delta} L\sin(\omega\tau\delta)\,\mathrm{d}\omega=\frac{2L}{\tau\delta}. \tag{19}$$

Then,

$$F(\tau)=\frac{\pi}{2}\lim_{n\to\infty}\frac{\int_0^{n\cdot\pi/\tau\delta}L|\sin(\omega\tau\delta)|\,\mathrm{d}\omega}{\int_0^{n\cdot\pi/\tau\delta}L\,\mathrm{d}\omega}=\frac{\pi}{2}\lim_{n\to\infty}\frac{nS_1\tau\delta}{\pi nL}=1 \tag{20}$$

for all $\tau\geqslant 1$. Thus the objective function does not depend from the time lag if the scalar time series represents white noise data—the shape of the embedded attractor is the same at any time lag.

(d) Analogously, when $A(\omega)=L$; $d\geqslant 2$ and $\tau_1=\cdots=\tau_{d-1}=\tau$(uniform embedding):

$$\begin{aligned}F(\tau,\ldots,\tau)&=\frac{\pi}{2}\sum_{k=1}^{d-1}\left(\frac{2(d-k)}{d(d-1)}\frac{\int_0^\infty A(\omega)|\sin(k\omega\tau\delta)|\,\mathrm{d}\omega}{\int_0^\infty A(\omega)\,\mathrm{d}\omega}\right)\\&=\frac{\pi}{2}\sum_{k=1}^{d-1}\frac{2(d-k)}{d(d-1)}\frac{2}{\pi}=1.\end{aligned} \tag{21}$$

Here the indefinite integral in Eq. (21) is calculated using the same principle as in Eq. (20). Similar result holds also for non-uniform embedding:

$$\begin{aligned}F(\tau_1,\ldots,\tau_{d-1})&=\frac{2}{d(d-1)}\frac{\pi}{2}\frac{1}{\int_0^\infty L\,\mathrm{d}\omega}\\&\times\left(\sum_{i=1}^{d-1}\int_0^\infty L|\sin(\omega\delta\tau_i)|\,\mathrm{d}\omega+\cdots+\int_0^\infty L\left|\sin\left(\omega\delta\sum_{j=1}^{d-1}\tau_j\right)\right|\mathrm{d}\omega\right)\\&=\frac{2}{d(d-1)}((d-1)+\cdots+1)=1.\end{aligned} \tag{22}$$

Each indefinite integral in Eq. (22) is calculated separately applying the technique described by Eq. (20) (of course τ must be replaced there by an appropriate compound time lag).
It can be mentioned that a scalar time series with uniform distribution in interval $[a,b]$ will be mapped into a d-dimensional hyper-cube. The density of the embedded points in any location inside the hyper-cube is constant and does not depend on time lags.

(e) $A(\omega)$ is delta function; $d=2$. It is understood that

$$A(\omega)=\delta(\omega_0)=\begin{cases}\infty, & \text{when } \omega=\omega_0\\ 0, & \text{elsewhere}\end{cases}\quad\text{and}\quad\int_0^\infty\delta(\omega_0)\,\mathrm{d}\omega=1.$$

Then,

$$F(\tau)=\frac{\pi}{2}Q(\tau,\omega_0). \tag{23}$$

3. Experiments

The developed computational technique for the determination of the optimal set of time lags is applied in time series forecasting techniques based on non-uniform embedding and is demonstrated for two chaotic time series generated by Mackey–Glass and Rössler chaotic maps and one experimental time series describing daily brightness of a variable star on 600 successive midnights [17]. First we determine the optimal set of time lags for each time series and then use Matlab-based adaptive network-based fuzzy inference system (ANFIS) to predict these time series.

In general, fuzzy inference systems employing fuzzy if–then rules can model qualitative aspects of human knowledge and reasoning processes without employing precise quantitative analyses. Basically, a fuzzy inference system is composed of five functional blocks: a rule base containing a number of fuzzy if–then rules; a database which defines the membership functions of the fuzzy sets used in the fuzzy rules; a fuzzy reasoning unit which performs the inference operations on the rules; a fuzzification interface which transforms the crisp inputs into degrees of match with linguistic variables and a defuzzification interface which transforms the fuzzy results of the inference into a crisp output [14,15,18].

3.1. Application to Mackey–Glass time series

Chaotic Mackey–Glass delay differential equation is a paradigm problem that has been used and reported by a number of researchers for comparing the learning and generalization ability of different models [18–20]. The Mackey–Glass delay differential equation reads

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t}=\frac{ax(t-c)}{1+x^{10}(t-c)}-bx(t), \tag{24}$$

where a, b and c are constants; t is time. At $a=0.2$, $b=0.1$ and $c=17$, Mackey–Glass equation generates a chaotic time series.

3.1.1. Case 1

In order to make comparisons, first we use uniform embedding to predict the time series. Initially we use a benchmark example from Matlab Fuzzy Logic toolbox. We predict $x(t+6)$ from past values of this time series, that is, $x(t-18)$, $x(t-12)$, $x(t-6)$ and $x(t)$. In other words, the embedding dimension is $d=5$ and the uniform time lag is $\tau=6$ (Fig. 1a).

We construct 1000 data vectors of the above format from $t=119$ to 1118; the first 500 are used for ANFIS training while the others are used for checking the accuracy of prediction.

We do not plot the original and the predicted time series. Instead we calculate prediction errors and plot them in Fig. 2. Similar approach will be also used in the following sections. The basic structure of computations can be illustrated by the following steps. The first step is the identification of the set of time lags.

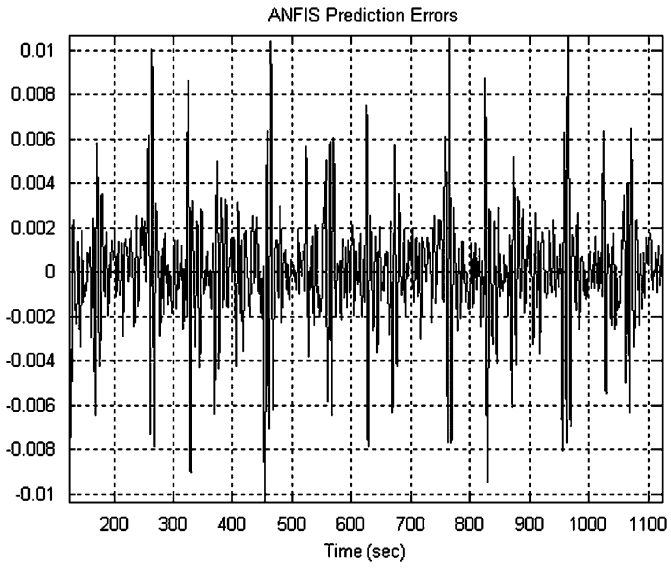

Fig. 2. ANFIS prediction errors (differences between the predicted and the original time series) for Mackey–Glass solution. The embedding dimension is 5; the set of time lags is {6,6,6,6}.

Next is the adaptation of ANFIS. Time lags between channels (input and output) must correspond to the set of time lags identified in the previous step (Fig. 1). Moreover, the number of ANFIS channels must be changed if the embedding dimension has changed in the process of the identification of the optimal set of time lags. As mentioned previously, ANFIS is a standard object in Matlab fuzzy logic toolbox and interested readers are referred to the documentation of the toolbox where all technical aspects are discussed in detail. Finally, the time series can be predicted when the adaptation of ANFIS to the optimal set of time lags is done.

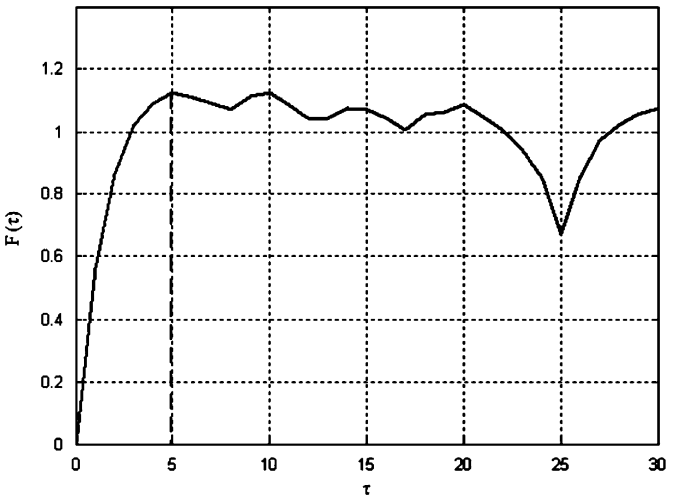

Fig. 3. Optimization of the objective function for Mackey–Glass time series. Uniform embedding; the embedding dimension is 5. Function values $F(\tau,\tau,\tau,\tau)$ are plotted for $0\leqslant\tau\leqslant30$. The maximum function value is reached at $\tau=5$; $F(5,5,5,5)=1.1249$.

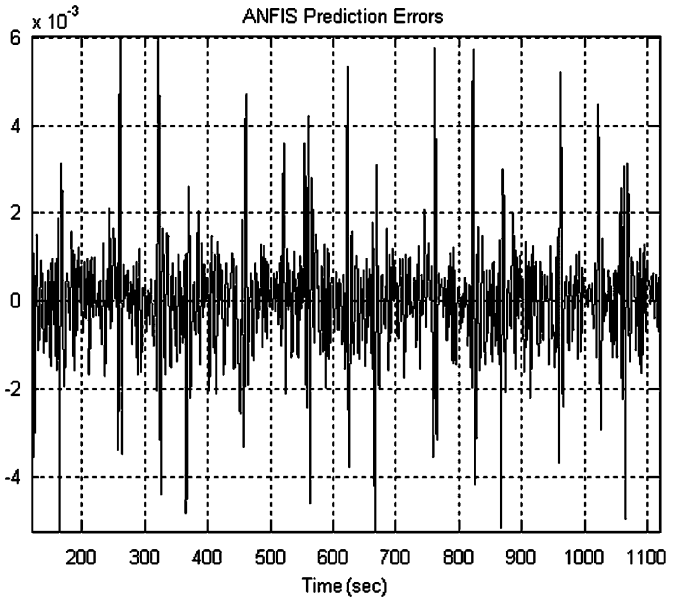

Fig. 4. ANFIS prediction errors for Mackey–Glass time series; the embedding dimension is 5; the set of time lags is (5,5,5,5).

3.1.2. Case 2

We continue with Mackey–Glass equation, use uniform embedding and keep $d=5$. But now we determine the optimal uniform time lag using the developed objective function in Eq. (18). It appears that the optimal time lag is $\tau=5$ (Fig. 3); $F(5,5,5,5)=1.1249$ (while $F(6,6,6,6)$ is only 1.1098). That is a definite indication that the attractor is more spread in the phase space at $\tau=5$ compared to $\tau=6$. One can observe several local maximums in Fig. 3, but we are interested in a global maximum in a bounded interval $0\leqslant\tau\leqslant30$. By the way, we do not calculate $F(0,0,0,0)$; it is equal to zero for any time series (property (a); Section 2.2).

We predict $x(t+5)$ from past values $x(t-15)$, $x(t-10)$, $x(t-5)$ and $x(t)$. We again construct 1000 data vectors from $t=116$ to 1115, use the first 500 ones for training, while the others for determining the accuracy of prediction. Prediction errors are presented in Fig. 4. It is clear that the quality of the predicted time series has improved—prediction errors have decreased more than 10 times compared to the benchmark example in Fig. 2 (Table 1).

3.1.3. Case 3

We still continue with Mackey–Glass equation and exploit uniform embedding. But now we use classical FNN algorithm [21] to determine the optimal embedding dimension first, which is determined to be $d=6$. Again we need to determine the optimal uniform time lag. Optimization of $F(\tau,\tau,\tau,\tau,\tau)$ produces $\tau=9$ (Fig. 5); $F(9,9,9,9,9)=1.0990$. It must be noted that one should not compare values of the objective function F for different embedding dimensions because functions of embedding quality are different then. Now this value is lower if compared to $F(5,5,5,5)$ (and even $F(6,6,6,6)$) from Case 2, but nevertheless, it is the best value so far for $d=6$. Again, there are several local maximums and one clearly expressed minimum at $\tau=25$, but we are interested in the global maximum in a bounded interval $0\leqslant\tau\leqslant30$ (what corresponds to maximum spreading of the attractor in the phase space).

The structure of the ANFIS fuzzy logic network must be modified before we can start predicting the Mackey–Glass time series; the number of inputs is greater by one if compared to the previous case (Fig. 1a). We will predict $x(t+9)$ from past values $x(t-36)$, $x(t-27)$, $x(t-18)$, $x(t-9)$ and $x(t)$. Training and checking procedures are repeated; ANFIS prediction errors are presented in Fig. 6. In average, prediction errors have decreased three times compared to Fig. 4 (Table 1).

3.1.4. Case 4

We continue with Mackey–Glass equation, but now the embedding is non-uniform. We have already determined that

Table 1
Improvement of predictions for Mackey–Glass chaotic time series.

Embedding method	Section no.	The embedding dimension d	The set of time lags $F(\tau_1\ldots\tau_{d-1})$	$F(\tau_1,\ldots,\tau_{d-1})$	Standard deviation of prediction errors σ
Uniform	3.1.1	5	(6,6,6,6)	1.1098	2.5125E−03
Uniform	3.1.2	5	(5,5,5,5)	1.1249	1.3112E−03
Uniform	3.1.3	6	(9,9,9,9,9)	1.0990	7.2862E−04
Non-uniform	3.1.4	6	(6,5,5,5,5)	1.1026	5.0686E−04

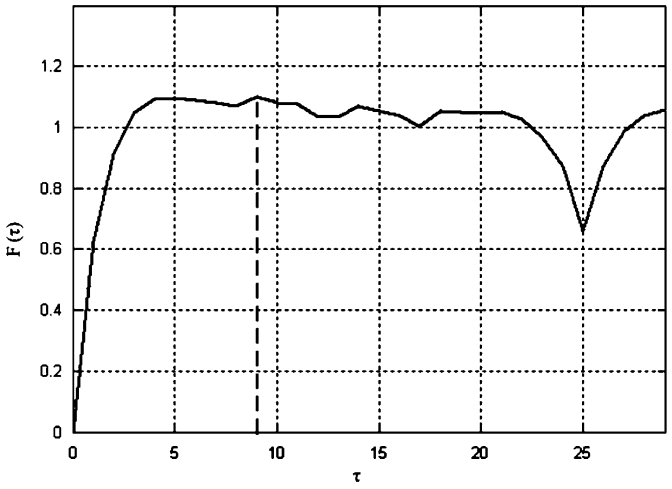

Fig. 5. Optimization of the objective function for Mackey–Glass time series. Uniform embedding; the embedding dimension is 6. Function values $F(\tau,\tau,\tau,\tau,\tau)$ are plotted for $0\leqslant\tau\leqslant30$. The maximum function value is reached at $\tau=9$; $F(9,9,9,9,9)=1.0990$.

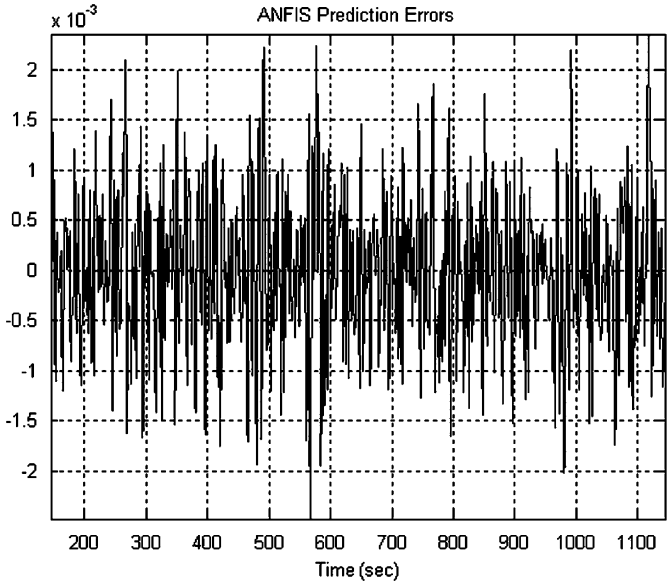

Fig. 6. ANFIS prediction errors for Mackey–Glass time series. The embedding dimension is 6; the set of time lags is {9,9,9,9,9}.

the optimal embedding dimension for this time series is $d=6$. Now we need to determine the optimal set of time lags $\{\tau_i\}$, $i=1,\ldots,5$. Full sorting yields such a set of optimal time lags: {6,5,5,5,5}; $F(6,5,5,5,5)=1.1026$. Unfortunately, we cannot plot a figure of the objective function because it is not feasible to visualize a d-dimensional space.

But one should pay attention to property (b) of the objective function (Section 2) before the training and checking procedures are commenced. The objective function is insensitive to permutation of its arguments, but data vectors constructed using non-uniform time lags are definitely sensitive to permutation (Eq. (1)). Thus the selection of the optimal numerated set of time lags is subject to additional optimization. All different permutations of the set $\{\tau_i\}$ must be tested before the best forecasting strategy is declared. In our case we have five different options: {6,5,5,5,5}; {5,6,5,5,5}; {5,5,6,5,5}; {5,5,5,6,5} and {5,5,5,5,6}. It appears, that the first variant is the best, when we predict $x(t+5)$ from past

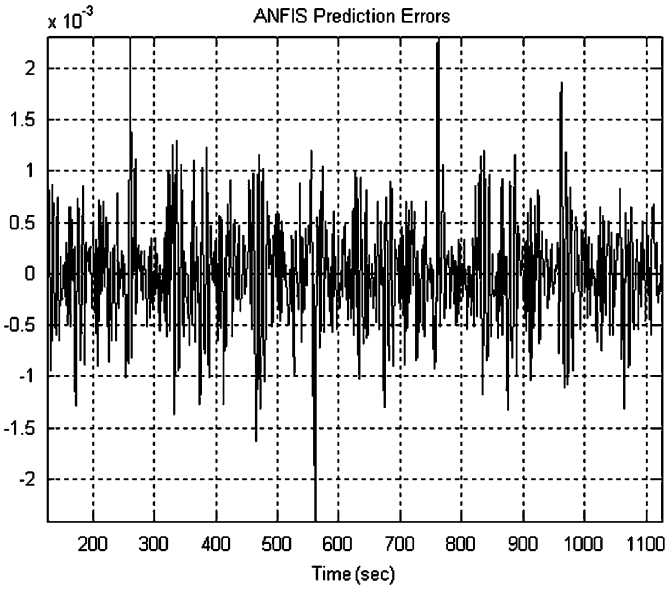

Fig. 7. ANFIS prediction errors for Mackey–Glass time series. The embedding dimension is 6; the set of time lags is {6,5,5,5,5}.

values $x(t-21)$, $x(t-15)$, $x(t-10)$, $x(t-5)$ and $x(t)$. ANFIS prediction errors are presented in Fig. 7. It can be noted that maximum difference between the original and the predicted time series in Figs. 7 and 6 are almost the same but standard deviations of errors are smaller: $\sigma=7.2862\times10^{-4}$ for errors in Fig. 6 and $\sigma=5.0686\times10^{-4}$ for errors in Fig. 7 (Table 1).

As mentioned earlier, full sorting of all possible permutations of the optimal set of time lags is a necessary step before the final forecasting strategy can be selected. This is an inherent feature of the proposed concept. The objective function (Eq. (18)) represents the spreading of the attractor in a multi-dimensional phase space, but it does not discriminate between concrete projections of the attractor. The objective function gives a strict analytical evaluation of the attractor's spreading; on the other hand associated computations are simple and effective. Computation of the optimal set of time lags can be even more effective if one keeps in mind that permutation of the same set of time lags has no effect to the value of the objective function. If all time lags are bounded in interval $1\leqslant\tau_i\leqslant R$; $i=1,\ldots,d-1$, then the total number of different sets of time lags is R^{d-1}. But if one considers the insensitivity to permutations, the number of different sets becomes only $(R+d-2)!/((d-1)!(R-1)!)$.

The spreading of the embedded attractor does not change when time lags are permuted. But delay vectors $y_p^{(d)}$ are different for different permutations. Thus all possible permutations need to be tested with ANSYS after the optimal set of time lags is identified. That guarantees the best possible result so far as the geometric concept is merged with ANSYS technology.

In general, the selection of the best set of time lags comprises two steps. The first one is optimization of the objective function. The second one is full sort of permutations of the optimal set of time lags. The second step is unavoidable, but the first step can be simplified if only approximate solutions are satisfactory (especially when d and R are very large). Then the recommendation would be to apply uniform embedding first (what would produce a lower bound for the value of the objective function). Then approximate discrete optimization techniques (for example genetic algorithms) can be used to produce a better set of time lags.

3.2. Application to Rössler time series

Rössler chaotic map is one of the paradigm problems thoroughly explored in non-linear dynamics [22]. Rössler system comprises three ordinary differential equations:

$$\begin{aligned}\frac{dx}{dt} &= -y - z,\\ \frac{dy}{dt} &= x + ay,\\ \frac{dz}{dt} &= b + z(x - c),\end{aligned} \tag{25}$$

where x, y and z are state variables; t is time; a, b and c are control parameters. The traditional bifurcation diagram for Rössler attractor is created by varying c with $a = b = 0.1$ [23]. This bifurcation diagram reveals that the attractor is periodic at low values of c, but quickly becomes chaotic as c increases. In the recent years Rössler attractor has experienced lots of attention from biomedical community [24,25]. Rössler map is well suited for characterization of different biological rhythms, especially neuron oscillations. We will test attractor non-uniform embedding based techniques to forecast chaotic Rössler attractor.

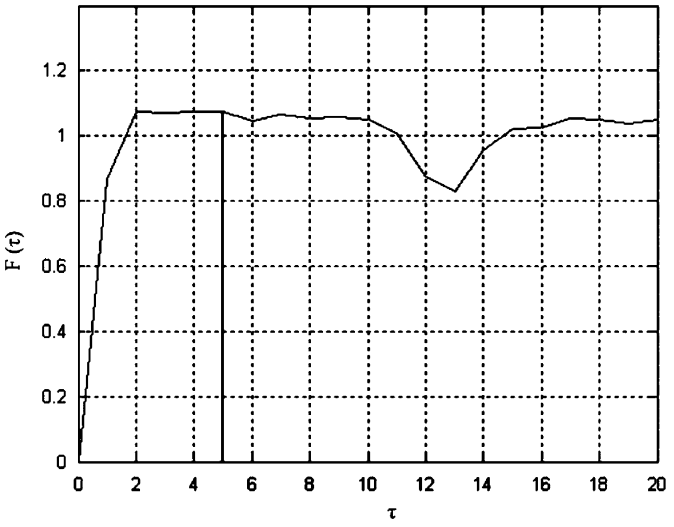

Fig. 8. Optimization of the objective function for Rössler time series. Uniform embedding; the embedding dimension is 6. Function values $F(\tau,\tau,\tau,\tau,\tau)$ are plotted for $0 \leqslant \tau \leqslant 20$. The maximum function value is reached at $\tau = 5$; $F(5,5,5,5,5) = 1.0760$.

3.2.1. Case 1

We select a post-transient Rössler attractor at $a = b = 0.1$ and $c = 20$. We use time marching techniques with constant time step 0.02 to construct a numerical solution of Eq. (25). Next, we select every twelfth data point from the numerical solution and construct a new data series which will be used for forecasting (what results into time step $\delta = 0.24$). Such an artificial shortening of the data series changes the time scale preserving the original accuracy of integration.

Initially we use uniform attractor embedding based forecasting techniques. We need to determine the embedding dimension first.

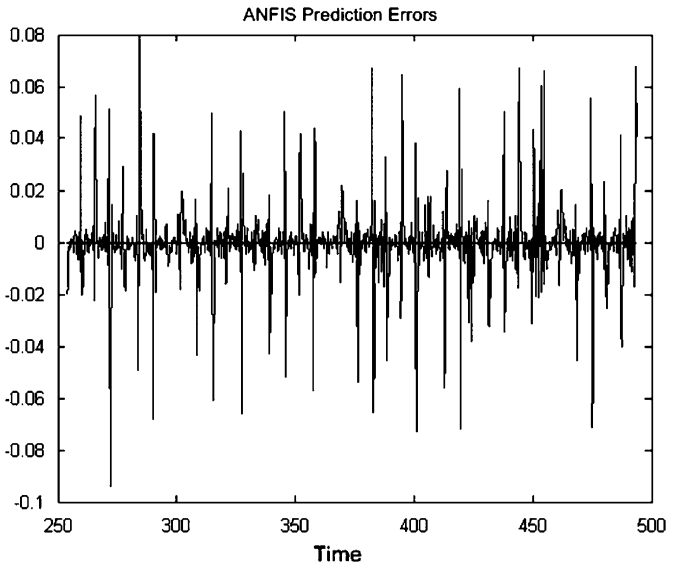

Fig. 9. ANFIS relative prediction errors for Rössler time series. The embedding dimension is 6; the set of time lags is {5,5,5,5,5}.

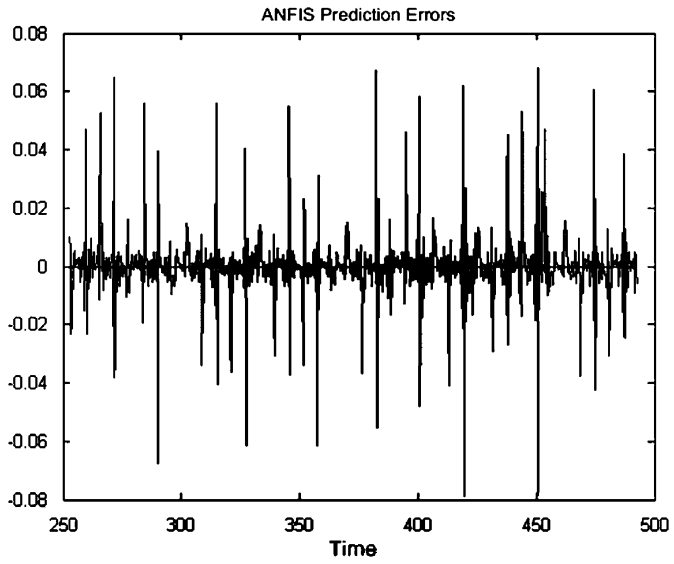

Fig. 10. ANFIS relative prediction errors for Rössler time series. The embedding dimension is 6; the set of time lags is {4,6,4,3,3}.

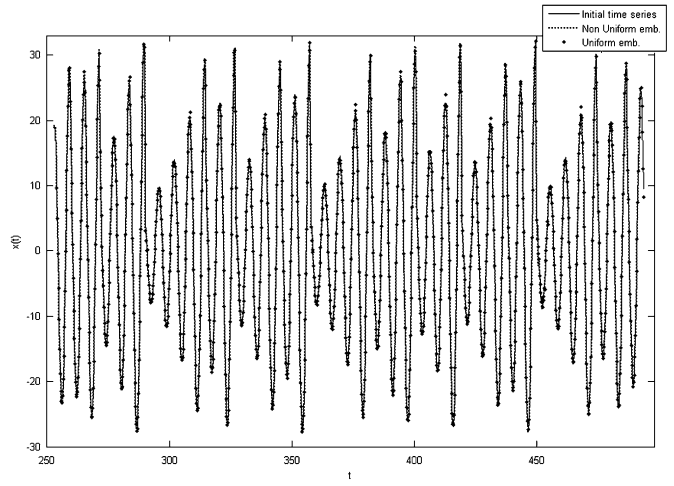

Fig. 11. A comparison between the original Rössler time series and time series predicted by forecasting techniques based on uniform and non-uniform embeddings. Predictions based on uniform embedding overestimate the time series at its local maximums. Non-uniform embedding produces predictions not separable from the original time series by a naked eye.

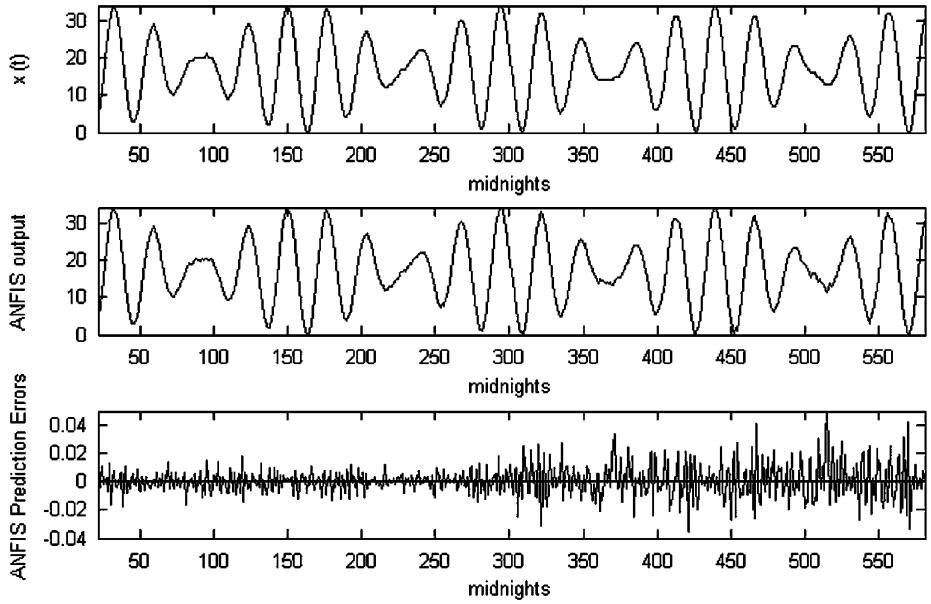

Fig. 12. ANFIS prediction of the daily brightness of a variable star: $x(t)$—a real world time series; ANFIS output—predictions using non-uniform embedding. The embedding dimension is 6; the set of time lags is {4,4,5,4,4}.

FNN algorithm suggests that $d=6$. Now, the optimal time lag is determined by optimizing the objective function (Fig. 8). Though there exist several local maximums (and some of them look almost equally high by a naked eye) but the global maximum is at $\tau=5$; $F(5,5,5,5,5)=1.0760$ (Fig. 8). We predict $x(t+5)$ from past values $x(t-20)$, $x(t-15)$, $x(t-10)$, $x(t-5)$ and $x(t)$. We construct 1000 data vectors from $t=252.80$ to 492.56, use the first 500 ones for training, while the others for determining the accuracy of the prediction. Relative prediction errors are presented in Fig. 9; the standard deviation of these errors is 0.4928.

3.2.2. Case 2

We keep the same embedding dimension, but choose non-uniform embedding. Full sorting yields such a set of optimal time lags: {6,4,4,3,3}; $F(6,4,4,3,3)=1.0823$. Such a set of optimal time lags produces 30 different permutations, all of which must be individually tested for the quality of forecasting. The best choice appears to be {4,6,4,3,3}. We predict $x(t+3)$ from past values $x(t-17)$, $x(t-13)$, $x(t-7)$, $x(t-3)$ and $x(t)$. We again construct 1000 data vectors from $t=252.08$ to 491.84, use the first 500 ones for training, while the others for determining the accuracy of prediction. Relative prediction errors are presented in Fig. 10; the standard deviation of the errors is 0.3901.

In order to visualize the difference between the forecasting techniques based on uniform and non-uniform embeddings we plot the numerical solution of the Rössler system (the original time series) together with predicted time series in Fig. 11. It can be seen that forecasting techniques based on uniform embedding overestimate values of the time series at points of its local extremes. One cannot see any differences between the original time series and the time series predicted by forecasting techniques based on non-uniform embedding by a naked eye at the scale used in Fig. 11.

3.3. Application to a real world time series

As mentioned earlier, we will use the developed technique to predict a real world time series describing the daily brightness of a variable star on 600 successive midnights [17]. We need to determine the embedding dimension of the time series first. FNN algorithm suggests that $d=6$. The optimal set of time lags for uniform embedding is determined to be {4,4,4,4,4}; $F=(4,4,4,4,4)=1.0719$; the standard deviation of relative prediction errors by ANFIS is 0.0125.

We continue with non-uniform embedding. Full sorting yields such a set of optimal time lags: {5,4,4,4,4}; $F=(5,4,4,4,4)=1.0747$. All different permutations of this set produces five different options: {5,4,4,4,4}; {4,5,4,4,4}; {4,4,5,4,4}, {4,4,4,5,4} and {4,4,4,4,5} which are tested individually. It appears, that the third variant is the best, when we predict $x(t+4)$ from the past values $x(t-17)$, $x(t-13)$, $x(t-9)$, $x(t-4)$ and $x(t)$. From $t=22$ to 581, we construct 560 data vectors $y_p^{(d)}$, use the first 280 ones for training, while the others for determining the accuracy of the prediction. The original time series, the predicted time series and relative prediction errors are presented in Fig. 12; standard deviation of relative errors is 0.0111. This is another good example illustrating the value of non-uniform embedding for time series prediction by fuzzy inference systems.

4. Conclusions

The object of this paper is to show that ANFIS-based time series forecasting techniques can be considerably improved if non-uniform attractor embedding is used to identify the optimal set of time lags. A simple deterministic algorithm is proposed for the selection of non-uniform time lags and comprises the pre-processing stage of the forecasting algorithm which is implemented in the form of a fuzzy inference system. Such an approach has many advantages over existing forecasting techniques in terms of the prediction accuracy. It is well applicable even for time series with multiple time scales when the state of the system changes in time and variable embedding is necessary to provide optimal system's description in the reconstructed phase space. Fuzzy networks do not require long time series for training (compared to neural networks); therefore, the presented method could be effectively used for forecasting systems with dynamically changing parameters, what is the object of future research.

References

[1] J.S. Armstrong, Principles of Forecasting: A Handbook for Researchers and Practitioners, Kluwer Academic Publishers, Norwell, 2001.
[2] E.M. Azoff, Neural Network Time Series: Forecasting of Financial Markets, Wiley, Chichester, 1994.
[3] R. Harris, Applied Time Series Modeling and Forecasting, Wiley, Chichester, 2004.
[4] S. Makridakis, C.C. Wheelwright, R.J. Hyndman, Forecasting Methods and Applications, Wiley, New York, 1998.
[5] T. Sauer, M. Casdagli, J.A. Yorke, Embedology, Journal Statistical Physics 65 (1991) 579–616.
[6] F. Takens, Detecting strange attractors in turbulence. Dynamical systems and turbulence, Lecture Notes in Mathematics 898 (1981) 366–381.
[7] H. Kantz, T. Schreiber, Nonlinear Time Series Analysis, Cambridge University Press, London, 1997.
[8] J.M. Nichols, J.D. Nichols, Attractor reconstructions for non-linear systems: a methodological note, Mathematical Biosciences 171 (2001) 21–32.
[9] T. Schreiber, Interdisciplinary applications of nonlinear time series methods, Physical Reports 308 (1999) 2–64.
[10] L. Cao, Practical method for determining the minimum embedding dimension of a scalar time series, Physica D: Nonlinear Phenomena (1997) 43–50.
[11] T. Buzug, G. Pfister, Optimal time delay and embedding dimension for delay time coordinates by the analysis of global statical and local dynamical behaviour of strange attractors, Physical Review A 45 (1992) 7073–7084.
[12] A.L. Cechin, D.R. Pechmann, L.P.L. Oliveira, Optimizing Markovian modeling of chaotic systems with recurrent neural networks, Chaos, Solitons and Fractals 37 (2008) 1317–1327.
[13] T. Hui-Kuang Yu, K.-H. Huarng, A bivariate fuzzy time series model to forecast the TAIEX, Expert Systems with Applications 34 (2008) 2945–2952.
[14] M. Khashei, S. Reza, M. Bijari, A new hybrid artificial neural networks and fuzzy regression model for time series forecasting, Fuzzy Sets and Systems 159 (2008) 769–786.
[15] M. Zounemat-Kermani, M. Teshnehlab, Using adaptive neuro-fuzzy inference system for hydrological time series prediction, Applied Soft Computing 8 (2008) 928–936.
[16] Y. Manabe, B. Chakrabortya, A novel approach for estimation of optimal embedding parameters of nonlinear time series by structural learning of neural network, Neurocomputing 70 (2007) 1360–1371.
[17] 〈http://www.robjhyndman.com/TSDL〉.
[18] H. Gu, H. Wang, Fuzzy prediction of chaotic time series based on singular value decomposition, Applied Mathematics and Computation 185 (2007) 1171–1185.
[19] C. Harpharm, C.W. Dawson, The effect of different basis functions on a radial basis function network for time series prediction: a comparative study, Neurocomputing 69 (2006) 2161–2170.
[20] L. Zunxiong, L. Jianhui, Chaotic time series multi-step direct prediction with partial least squares regression, Journal of Systems Engineering and Electronics 18 (2007) 611–615.
[21] M.B. Kennel, R. Brown, H.D.I. Abarbanel, Determining minimum embedding dimension using a geometrical construction, Physical Review A 45 (1992) 3403–3411.
[22] O.E. Rössler, An equation for continuous chaos, Physics Letters 57A (1976) 397–398.
[23] S.H. Strogatz, Nonlinear Dynamics and Chaos, Perseus Publishing, Massachusetts, 1994.
[24] T. Gedeon, M. Holzer, M. Pernarowski, Attractor reconstruction from interspike intervals is incomplete, Physica D: Nonlinear Phenomena 178 (2003) 149–172.
[25] J. Lian, J. Shuai, D.M. Durand, Control of phase synchronization of neuronal activity in the rat hippocampus, Neural Engineering 1 (2004) 46–54.

Minvydas Ragulskis received the Ph.D. in 1992 from Kaunas University of Technology, Lithuania. Since 2002 he is a Professor at the Department of Mathematical Research in Systems, Kaunas University of Technology. His research interests include nonlinear dynamical systems and numerical analysis.

Kristina Lukoseviciute received the M.Sc. degree in Mathematics in 2006 from the University of Technology, Lithuania. She is currently an Associate Lecturer and a Ph.D. degree student within the Department of Mathematical Research in Systems, Kaunas University of Technology. Her current areas of research interest are neuro-fuzzy systems and time series forecasting.