Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/100109

Fuzzy Cognitive Map Reconstruction: Dynamics Versus History

Wladyslaw Homenda[1,2,*], *Agnieszka Jastrzebska*[1] *and Witold Pedrycz*[3,4,5]

[1] Faculty of Mathematics and Information Science, Warsaw University of Technology, ul. Koszykowa 75, 00-662 Warsaw, Poland
[2] Faculty of Economics and Informatics in Vilnius, University of Bialystok, Kalvariju G. 135, LT-08221 Vilnius, Lithuania
[3] Department of Electrical & Computer Engineering, University of Alberta, Edmonton T6R 2G7 AB Canada
[4] Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University Jeddah, 21589, Saudi Arabia
[5] Systems Research Institute, Polish Academy of Sciences, ul. Newelska 6, 01-447 Warsaw, Poland

Received: 12 Jun. 2015, Revised: 10 Aug. 2015, Accepted: 11 Aug. 2015
Published online: 1 Jan. 2016

Abstract: This study is concerned with a fundamental issue of time series representation for modeling and prediction with Fuzzy Cognitive Maps. We introduce two distinct time series representation schemes for Fuzzy Cognitive Map design. The first method is based on the temporal relationships, namely time series amplitude, amplitude change, and change of amplitude change (dynamics perspective). The second scheme is based on three consecutive historical observations: present value, past value and before past value (history perspective, 2nd order relationships). Introduced procedures are experimentally verified and compared on several synthetic and real-world time series of various characteristics. The history-oriented time series representation turned out to be more advantageous. Quality of FCM-based time series models and one-step-ahead predictions were measured in terms of Mean Squared Error. We have shown that models designed with history-oriented time series representation generally require less FCM nodes to be of comparable quality to models built on dynamics-oriented time series representation. As a result, with the history-oriented time series representation scheme we are able to construct simpler and better models.

Keywords: Fuzzy Cognitive Maps, concepts, time series modeling and prediction

1 Introduction

Fuzzy Cognitive Maps are an alternative modeling framework for complex phenomena. Since their introduction by B. Kosko in [2], Fuzzy Cognitive Maps have been in the scope of interest of both theoretical and applications-oriented researchers. Almost three decades of intensive studies have resulted in efficient methodologies for Fuzzy Cognitive Maps learning and applications.

Time series modeling and prediction with Fuzzy Cognitive Maps is a relatively new stream of studies. It emerged around 2008, when Stach et al. published their research in [9]. The original methodology for time series modeling with Fuzzy Cognitive Maps laid the groundwork for present studies in this area.

In brief, Stach et al. proposed a modeling method based on time series amplitude and its change elevated by fuzzification to concepts.

The material presented in this article refers to the original approach of Stach et al. We present a research on time series representation for modeling with Fuzzy Cognitive Maps. Also, we introduce two new methods for Fuzzy Cognitive Map design. We compare and test proposed procedures in a series of experiments on both synthetic and real-world time series.

The objectives of this article are to discuss and compare two different time series representation schemes for time series modeling and prediction with Fuzzy Cognitive Maps. The first method exploits a representation space formed by on time series amplitude, amplitude change and change of amplitude change. Hence, we use here the name ”dynamics” perspective. The second is based on three successive values of time series: time series present values (amplitude), past values and before past values. Hence, we have named it ”history” perspective.

* Corresponding author e-mail: homenda@mini.pw.edu.pl

The paper is structured as follows. In Section 2 we introduce time series representations for modeling with Fuzzy Cognitive Maps. Section 3 covers discussion on Fuzzy Cognitive Maps design and presents experiments' schemes that we conducted to compare and illustrate proposed approach. In Section 4 we discuss the results of the experiments. Section 5 covers conclusion and identifies future research directions.

2 Concepts

2.1 Cognitive Maps

Fuzzy Cognitive Map (FCM) is an abstract soft computing model that can be described by a weighted directed graph. It comprises of nodes and weights connecting the nodes, c.f. [2]. Nodes represent concepts. Relations between concepts in a cognitive map are expressed through weighted edges between the nodes. An example of a 3-node FCM is in Figure 3. In practical applications, the nodes correspond to various concepts present to describe a certain phenomenon, for example: unemployment, skilled human resources, fuel prices, air pollution, high concentration of hydrogen ions, and so on.

Fuzzy Cognitive Map is represented by a matrix of weights $W = \left[w_{ij}|w_{ij} \in [-1,1], i, j = 1,2,,n\right]$, where n is the number of nodes. Each weight w_{ij} corresponds to the edge connecting the node $node_j$ to the node $node_i$. FCM exploration is based on activation $X = [x_1, x_2, \ldots, x_n]^T$, which is presented to an FCM, i.e. every value x_i of an activation X is presented to the $node_i$. The response of the FCM $Y = [y_1, y_2, \ldots, y_n]^T$ is computed according to the formula:

$$Y = f\big(W \cdot X\big) \tag{1}$$

where $W \cdot X$ is matrix product and f is a sigmoid function, i.e.,

$$f(x) = \frac{1}{1+e^{-\tau \cdot x}} \tag{2}$$

endowed with given parameter τ. The sigmoid function is applied separately to every element of matrix product. More specifically we have:

$$y_i = f\big(\sum_{j=1}^{n} w_{ij} \cdot x_j\big) \tag{3}$$

The unknown and searched element is the aforementioned weight matrix. It describes linkages between concepts in the map. The essence of the design of the FCM is to construct the weight matrix that modells phenomena. The entries of the weight matrix could be either given by experts or learned from data. In our research we focus on the latter approach.

Let us assume that for FCM training we have:

– N activations, namely: $X_1, X_2 \ldots, X_N$, where $X_i = [x_{1i}, x_{2i}, \ldots, x_{ni}]^T$ for $i = 1,2,\ldots,N$,
– N corresponding goals (targets), namely: $G_1, G_2, \ldots, G_N$, where $G_i = [g_{1i}, g_{2i}, \ldots, g_{ni}]^T$ for $i = 1,2,\ldots,N$,
– N map responses corresponding to activations, namely: $Y_1, Y_2 \ldots, Y_N$, where $Y_i = [y_{1i}, y_{2i}, \ldots, y_{ni}]^T$ for $i = 1,2,\ldots,N$,

The aim is to design (reconstruct) a Fuzzy Cognitive Map. We use verb "reconstruct" to highlight that the assumption of modelling with FCMs is that given phenomena could be perfectly represented with a map of linked nodes-concepts and the training goal is to reconstruct the strength of these linkages.

FCM reconstruction for given activations and goals is based on adjustment of (usually randomized at the beginning) weight matrix W in such a way that FCM responses are *as close as possible* to goals. The term *as close as possible* boils down to the minimization of a mean square error:

$$MSE = \frac{1}{n \cdot N} \sum_{i=1}^{N} \sum_{j=1}^{n} \big(y_{ji} - g_{ji}\big)^2 \tag{4}$$

Naturally, we can consider other optimization criteria, but this topic is out of the scope of the article. As for technical aspects of FCM learning, in this study authors took the benefit of Particle Swarm Optimization strategy, which is one the viable optimization alternatives to be engaged here. The value of parameter τ of the sigmoid function was set to 5, based on literature review, [3,9], and experimental studies carried out here.

The design of an FCM comprises of two fundamental development steps, namely structure design and further parametric optimization. In what follows, we focus on representation issues of time series and then we elaborate on key design facets of the overall process.

2.2 Time series and their representation

Representation of a time series, in terms of an FCM-based model architecture, could be decomposed into two essential design elements, namely a way to capture the dynamics of the system and extract of concepts - nodes in the map.

Concepts can be seen as aggregates of information, whose specificity determines the accuracy of phenomena description. In this light, FCMs align with the idea of granular computing, c.f. [5], which is focused on knowledge granules - abstract units of information. Nodes are knowledge granules, in FCMs conventionally realized with fuzzy sets.

The proposed approach is illustrated here using several synthetic time series. We deliberately start with synthetic time series, because of their better illustrative properties of such example. The considered time series are built involving two sets of numbers of cardinality 3 and 5: $\{2,5,8\}$ and $\{1,3,5,7,9\}$ and one time series is built on three numbers set $\{2,6,8\}$. A procedure for synthesizing a time series is:

1. select base sequence, which is aimed to cover different numbers of concepts in the dimensions of values, values changes and changes of values changes. This aim is gained by different order of base sequence elements, compare Table 4,
2. replicate base sequence so that the total length of a time series is 3000,
3. add random distortion drawn from the normal (Gaussian) probability distribution with mean equal to 0 and standard deviation equal to 0.7.

With this procedure we can obtain time series of varying complexity, depending on the complexity of the base sequence assumed in the first step. In this way, we capture quantitative performance of a given FCM for different configurations of in put time series. Such time series exerts both regular, periodic, variations and random variations, thanks to the Gaussian noise added at the end.

Figure 1 visualizes the idea behind the proposed approach. Both plots concern one of the easiest time series we addressed in this study. It was constructed based on a sequence 2,6,8. As a result in this particular time series we can clearly distinguish 3 one-dimensional concepts that could ideally represent its amplitude. Visual inspection of Figure 1 confirms that even without any knowledge about the specifics of the 3-step procedure listed above the dataset is clearly based on 3 dense clusters of data.

To enhance human-centric of an FCM model we usually assign linguistic labels. In the case above we can use Small to represent values centered around 2, Moderately High to represent cluster related to 6 and High to represent values close to 8. For convenience we abbreviated the linguistic variables to S, MH and H, respectively.

A dot-plot of amplitude in time (3000 time points) is shown in upper part of Figure 1, while the bottom plot shows first 200 points at the amplitude axis with centers of amplitude groups.

It is apparent that in real-world data it is much harder to detect such regularities. Hence, among biggest challenges in the domain of FCM-based modelling is to propose, so to say, bulletproof method for concepts extraction that will work even when data is highly irregular. Please note, that concepts extraction, though at the first sight it resembles clustering, is a domain-specific task. Concepts in an FCM generalize knowledge. A basic parameter of an FCM related to the number of its concepts is its specificity. The larger the map the higher its specificity. We expect specific maps to be numerically more accurate than less specific maps. At the same time very specific maps are much harder to interpret than less specific maps.

Let us continue with the main course: time series representation.

2.2.1 Dynamics based representation

Time series is a sequence of numbers, say *amplitudes*. Let us assume that we have a time series of length $N+2$

$$a_{-1}, a_0, a_1, a_2, a_3, \ldots, a_N$$

In order to capture dynamics of the system, we propose switching to coordinates *amplitudes/amplitude changes/changes of amplitude change* for consecutive time points forming the following series of triples:

$$(a_1, da_1, dda_1), (a_2, da_2, dda_2), \ldots, (a_N, da_N, dda_N)$$

where for any triple (a_i, da_i, dda_i) change of amplitude is $da_i = a_i - a_{i-1}$ and change of amplitude change is $dda_i = da_i - da_{i-1} = a_i - 2 \cdot a_{i-1} + a_{i-2}$ for $i = 1,2,\ldots,N$.[1] Note that the original time series are padded with the preceding elements a_{-1}, a_0 to allow computing amplitude change and change of amplitude change for all time points $i = 1,2,\ldots,N$.

An example of such series representation is in Table 1. It concerns already mentioned 2,6,8-based synthetic time series.

In Table 1 we display the values affecting the data by some noise with random distortions. The first row refers to consecutive time points, i.e. enumerated numbers. The second row contains numerical amplitude values. Notice that amplitudes are padded with the first values 6 and 8 (corresponding to time -1 and 0) in order to allow computation of amplitude change and change of amplitude change in time 1. The third and fourth rows correspond to amplitude changes and changes of amplitude change. Therefore, we have the following triples in the dynamics based representation: (2,-6,-8), (6,4,10), (8,2,-2), (2,-6,-8), (6,4,10), (8,2,-2), ...

2.2.2 History based representation

Dynamics-based representation of time series obviously is conceptually equivalent to the series of triples of *present/past/before past* amplitudes:

$$(a_1, a_0, a_{-1}), (a_2, a_1, a_0), \ldots, (a_N, a_{N-1}, a_{N-2})$$

This representation of the series based on the sequence 2, 6 and 8 is in Table 2. By analogy, we have the following triples in the history based representation:
(2,8,6), (6,2,8), (8,6,2), (2,8,6), (6,2,8), (8,6,2), ...

Theoretically, both representations are equivalent as long as they represent time span of the same length, i.e.:

– dynamics-based representation with (amplitudes, amplitude changes) pairs is equivalent to (present, past) pairs in history-based representation,

[1] Note, this representation corresponds to *function/first derivative/second derivative* in differential calculus.

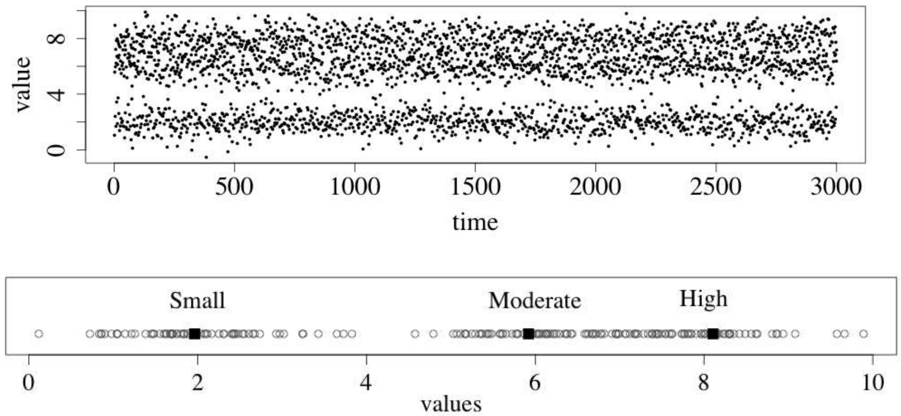

Fig. 1: The synthetic time series based on period 268 in time/amplitude coordinates (upper part) and first 200 elements on the amplitude axis (bottom part)

Table 1: An example of time series dynamics based representation built on the sequence 268 (not distorted values)

time	-1	0	1	2	3	4	5	6	7	8	...	2998	2999	3000
amplitudes	6	8	2	6	8	2	6	8	2	6	...	2	6	8
amplitude changes	∼	2	-6	4	2	-6	4	2	-6	4	...	-6	4	2
changes of amplitude change	∼	∼	-8	10	-2	-8	10	-2	-8	10	...	-8	10	-2

Table 2: An example of time series history based representation built on the sequence 268 (not distorted values)

time	-1	0	1	2	3	4	5	6	7	8	9	...	2998	2999	3000
present amplitudes	6	8	2	6	8	2	6	8	2	6	8	...	2	6	8
past amplitudes	∼	∼	8	2	6	8	2	6	8	2	6	...	8	2	6
before past amplitudes	∼	∼	6	8	2	6	8	2	6	8	2	...	6	8	2

– dynamics-based representation with (amplitudes, amplitude changes, changes of amplitude change) triples is equivalent to (present, past, before past)triples in history-based representation,
– and so on.

In this paper we elaborate on time series representation capturing 3 data points.

Note that the longer the time span we capture, the more complex the model gets. If the objective of modelling is to build a model to be presented and interpreted by a human being, time span of 3 data points is among highest reasonable values. Representation based on triples corresponds to a 3-dimensional space. It means, that with such representation we are able to plot the time series in a 3-dimensional coordinates system of either (amplitudes, amplitude changes, changes of amplitude change) coordinates system or in (present, past, before past) coordinates system. 3-dimensional space is the highest system a human being can immediately visualize.

We elaborate on time series and concepts representation in such 3-dimensional spaces in the following subsection.

Our past research (not a topic of this paper though) showed that the higher dimension we use for time series representation, the lower the numerical error of prediction. At the same time, increasing dimensionality entails important drawbacks: computational and interpretational. Hence, we have decided to base this research in 3-dimensional space as a reasonable compromise between numerical accuracy and interpretability.

2.3 Transforming time series into concepts space

Let us now elaborate on granular/conceptual representation of the time series given in Figure 1. Let us recall that for the considered example, first we have a scalar synthetic time series, which is expressed as a sequence of numbers - amplitudes. For each amplitude in a given time point we compute the corresponding amplitude change and change of amplitude change. Formally, amplitude and the corresponding dynamics specify a point in the amplitude/amplitude change/change of amplitude change system coordinates. For one-dimensional knowledge granules on amplitudes, amplitude changes and on changes of amplitude change we separately perform Fuzzy C-Means clustering and we obtain three, at this point separate, sets of granules. Next, using Cartesian product we determine three-dimensional concepts. Now, three-dimensional knowledge granules describe each pair amplitude/amplitude change/change of amplitude change. Finally, we attach linguistic variables to enhance interpretability of extracted concepts.

As the clustering has been completed for the individual variables (dimensions), it is very likely that some of the Cartesian product prototypes might not have any supportive experimental evidence and such combinations (viz. the nodes of the map) could be easily eliminated. Let us plot extracted concepts and the underlying time series in Figure 2. What strikes immediately is that concepts are of significantly differing quality.

Left plot is for dynamics-based representation. Hence, coordinates system is (amplitudes, amplitude changes, changes of amplitude change). Right plot concerns history-based representation. See labels axes: (present, past, before past).

Visual inspection confirms that synthesized time series is highly regular - observe three dense clouds of points in each plot. The aim is to find concepts to represent the data. Naturally, in such case the task is much easier than if we would have analyzed real data. Nevertheless, we continue with this example for its superior illustrative abilities.

2.3.1 Dynamics based granulation

In Figure 2 the left plot shows granulation into 3 concepts/granules each of dynamics dimension:

– amplitudes: concepts S+, MH+, H+ with corresponding granule centers 2, 6 and 8,
– amplitude changes: concepts MH–, S+ and MS+ with corresponding granule centers -6, 2 and 4,
– changes of amplitude change: concepts H–, S– and VH+ with corresponding granule centers -8, -2 and 10.

Note, abbreviations should be understood as follows: S+ - small positive, MH– - moderately high negative, VH+ - very high positive etc.

Next, using Cartesian product we determine 27 three-dimensional concepts:

$$\begin{aligned}&\{\text{S+,MH+,H+}\} \times \{\text{MH–,S+,MS+}\} \times \{\text{H–,S–,H+}\}\\&= \{(\text{S+,MH–,H–}),(\text{S+,MH–,S–}),(\text{S+,MH–,H+}),\ldots,\\&\quad (\text{H+,S–,H+}),(\text{H+,MS+,H+})\}\end{aligned}$$

Now, three-dimensional knowledge granules are used to linguistically describe each triple amplitude/amplitude change/change of amplitude change. Derived three-dimensional granules/concepts perfectly match clouds of the time series. Every granule in every dimension corresponds to one cloud of points. Intuitively, this granulation matches given time series. More exactly, centers of 3 granules fall into clouds of time series points while 24 other granules are not tied to this time series. Three concepts matching the time series are (S+,MH–,H–), (MH+,MS+,VH+), (H+,S+,S–) in dynamic coordinates.

The right plot of this Figure shows matching granulation for history-based representation: centers of 3 granules fall into clouds of respective clouds of 3D point, while 24 granules are not tied to this time series. Namely, the following concepts match this time series: (S+,MH+,H+), (MH+,S+,H+), (H+,MH+,S+) in dynamics coordinates.

2.3.2 History based granulation

The right plot of Figure 2 presents granular representation based on three-steps history, i.e. based on present, past and before past granular representation of time series. By analogy to dynamics based representation, we get 27 three dimensional concepts:

$$\begin{aligned}&\{\text{S+,MH+,H+}\} \times \{\text{S+,MH+,H+}\} \times \{\text{S+,MH+,H+}\}\\&= \{(\text{S+,S+,S+}),(\text{S+,S+,MH+}),(\text{S+,S+,H+}),\ldots,\\&\quad (\text{H+,H+,MH+}),(\text{H+,H+,H+})\}\end{aligned}$$

Three concepts matching the time series are (S+,H+,MH+), (MH+,S+,H+), (H+,MH+,S+) in history coordinates.

2.4 Granular representation

Time series is a sequences of numbers, which we have to elevate to terms of concepts. We form a collection of descriptors of amplitude, change of amplitude and change of amplitude change that are viewed formally as some information granules, say fuzzy sets. More specifically, we form a family (vocabulary) of fuzzy sets $A_1, A_2, \ldots, A_{c1}$ expressed over the space of amplitude of the time series, family of fuzzy sets $dA_1, dA_2, \ldots, dA_{c2}$ over the space of amplitude change and another family of fuzzy sets $ddA_1, ddA_2, \ldots, ddA_{c3}$ over the space of change of amplitude change. Then a time series $a_1, a_2, \ldots, a_N$ (along with its amplitude changes $da_1, da_2, \ldots, da_N$ and changes of amplitude changes $dda_1, dda_2, \ldots, dda_N$) is seen as a series of degrees of activation of the elements of the three vocabularies. More specifically, a_k, da_k and dda_k gives rise to the membership values of the corresponding fuzzy sets of the

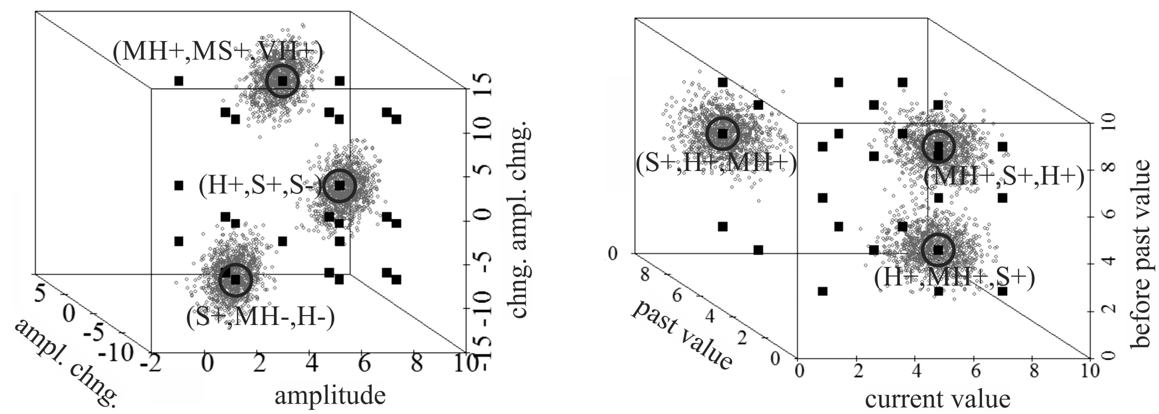

Fig. 2: The synthetic time series based on period 268 in two different coordinates systems: dynamics and history

vocabularies, c.f. Table 1 for the time series built on the sequence 2, 6 and 8.

Alike, we form a collection of descriptors of amplitude $A_1, A_2, \ldots, A_{c1}$ and then apply them to present and former amplitude values with given retrospective length. Namely, for the backward length 3, a_k, a_{k-1} and a_{k-2} gives rise to the membership values of the corresponding fuzzy sets of the vocabulary. Unlike in the case of dynamics, we have only one vocabulary, which is equally applied for present and former amplitude values, compare Table 2.

2.4.1 Building unknown granules

As we have noted before, for real-life problem we neither know cluster centers' coordinates, nor have any knowledge about a number of clusters. In this subsection we discuss, how to form granules of knowledge for any given time series.

Let us recall, that the outcome of the proposed concepts' design methods is a set of three-dimensional concepts. In the case of dynamics-based time series representation each concept is rooted in the amplitude/amplitude change/change of amplitude change space. In the case of history-oriented perspective each concept is in the present value/past value/before past value coordinates system. This is a natural consequence of the proposed concept design strategies and time series representation models.

The procedure of concepts' design is aided first with Fuzzy C-Means, second with ternary Cartesian product. Cartesian product elevates one-dimensional concepts to the three-dimensional spaces of amplitude/amplitude change/change of amplitude change or value/past value/before past value. The corollary of the application of the Cartesian product is that we obtain $c_1 \times c_2 \times c_3$ concepts for the case of dynamics-oriented time series representation and $c \times c \times c$ concepts for the case of history-oriented perspective. c_1, c_2, and c_3 is the number of 1-dimensional concepts on amplitude, amplitude change, and change of amplitude change respectively, while c is the number of 1-dimensional concepts on time series values. We have already marked that not all of the designed concepts have an empirical support in the form of underlying data points from the time series. Let us come back to Figure 2. Concepts with empirical support are the ones that truly generalize the underlying data. These are concepts marked in circles.

Here comes a vital problem: how to propose enough concepts to represent the data with required precision? The task of building unknown granules is in fact a task of finding a good balance between model's complexity and generality. It will be shown later that the number of concepts directly corresponds to the number of nodes in the FCM. With an increase of the number of nodes in the map the number of arcs grows quadratically and the time required to train such map grows exponentially. Apart from practical premises, large models are very inconvenient in use. Large FCMs are very difficult to interpret and to apply.

In the case of real-world data, selection of proper number and location of concepts is very difficult. Real data is typically irregular, with many outlying data points.

To overcome this obstacle the proposed design scheme at first extracts relatively large set of concepts from which we choose the best few. The criterion for best concepts selection is membership index defined as follows:

$$M(v_j) = \sum_{i=1}^{N} x_{ji} \tag{5}$$

Where $M(v_j)$ is membership index for v_j-th concept, x_{ji} is the level of membership to this j-th concept calculated with the use of standard Fuzzy C-Means objective function for an i-th data point:

$$\mathbf{x}_{ij} = \frac{1}{\sum_{k=1}^{n} \left(\frac{||\mathbf{a}_i - \mathbf{v}_j||}{||\,\mathbf{a}_i - \mathbf{v}_k||} \right)^{2/(m-1)}} \tag{6}$$

where n is the number of concepts, m is the fuzzification coefficient $(m > 1)$ and $||\cdot||$ is the Euclidean distance, $\mathbf{a}_i$ is a triple $\mathbf{a}_i = [a_i, da_i, dda_i]$ in the case of dynamics-oriented representation or a triple: $\mathbf{a}_i = [a_{i-2}, a_{i-1}, a_i]$ in the case of history-oriented representation. $\mathbf{v}_j$ describes j-th concept's coordinates in a 3-dimensional space of amplitude/amplitude change/change of amplitude change or present/past/before past values of the time series.

Concepts with high membership index are better. We select n best concepts based on their membership index. n could be determined by an inspection of a plot of concepts ordered by their membership index on OX axis and their respective membership index on OY axis. A knee-point in such plot roughly separates bad and good concepts.

With the membership criterion, a very intuitive measure of quality for the deign procedure that starts with Fuzzy C-Means, we are able to select a subset of best nodes and design a Fuzzy Cognitive Map based only on selected concepts.

2.4.2 Fuzzy Cognitive Map design

The final step is Fuzzy Cognitive Map design. Fuzzy Cognitive Maps operate on an abstract level of concepts - nodes of the map, which are connected with weighted arcs. Hence, we have to extract a set of concepts that becomes a set of FCM's nodes.

With the use of Formula 6 we calculate activation levels for each selected concept, we transpose it, and in this way we obtain $n \times N$ activations matrix $\mathbf{X}$. In other words, activations are membership values, levels of belongingness of time series observations to extracted concepts.

Goals are equal to activations shifted by one element forward. So that activations corresponding to an i+1-th time point are set together with goals for an i-th time point. Such representation is designed especially for time series processing with FCMs.

At this point we have all the elements necessary to train weight matrix for this time series.

3 Experiment

In this section we discuss evaluation and comparison procedures of Fuzzy Cognitive Map design methods based on the two time series representation schemes.

We have constructed 12 synthetic time series with different characteristics to test quality and stability of proposed approaches. As a quality measure we used Mean Squared Error, compare formula 4. We assess both modeling and prediction quality. Predictions for the time series are for one-step-ahead. The ratio of train/test data is 7:3. We use train data to learn the FCM, while test data is only for prediction. As a learning procedure we have chosen Particle Swarm Optimization. We have used an implementation of PSO from package "pso" in R language with default parameters compliant with the PSO 2007 implementation, detailed parameter list is under [12].

First series of experiments was on the synthetic time series. Second, on real-world time series. Real-world time series were downloaded from publicly available repositories under [13,14]. Time series can be identified by their names. Selected data sets are of different character: lack/different parameters of seasonality, trends, etc. We did not conduct any preprocessing procedures.

Note, that the experiments are conducted on the level of granules - aggregates of knowledge. Proposed modeling technique operates on a high level of abstraction, where instead of scalar values we operate on concepts and linguistic terms. Such knowledge representation and processing offers a compelling human-centered interface. The focus is on granules of knowledge that correspond to phenomena and relations within these phenomena gathered in trained weights matrix of FCMs. If we want to compare such model with classic time series modeling techniques it is necessary to perform a defuzzification procedure. Because such comparison is not in the scope of interest of this paper, we do not perform defuzzification.

For each time series, both synthetic and real-world, we have extracted and trained FCMs based on $n = 27, 22, 17, 12, 10, 8, 6, 4$ concepts. The $n = 3^3 = 27$-nodes FCMs are full-architecture maps, where no concepts were removed. Smaller maps are built on n best concepts selected with the membership criterion. We have trained FCMs both for dynamics-oriented time series representation and for history-oriented representation. With such selection of FCMs we are able to thoroughly review and compare the characteristics of our methods.

In the remainder of this section we introduce the synthetic and real-world time series. The following Section 4 presents experiments results.

3.1 Synthetic Series

Proposed approach is illustrated with synthetic time series. We deliberately used synthetic time series, because of better illustrative properties of such examples. The considered time series are built on two sets of numbers of cardinality 3 and 5: $\{2, 5, 8\}$ and $\{1, 3, 5, 7, 9\}$. One time series is built on a set of three numbers: $\{2, 6, 8\}$. Each time series is built on repeated sequence of numbers to reach length 3000. Numbers are distorted by values of the normal (Gaussian) probability distribution with mean equal to 0 and standard deviation equal to 0.7. The procedure for synthetic time series construction was already mentioned in Section 2.2.

For instance, the series built on the sequence $\{2, 5, 8, 5, 8, 2\}$ with, of course, every value (amplitude) altered by mentioned above random distortion is in the

Table 3: Dynamics based representation of a series built on the sequence 258582

time	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	...
amplitude:	8	2	2	5	8	5	8	2	2	5	8	5	8	2	...
amplitude change:	∼	-6	0	3	3	-3	3	-6	0	3	3	-3	3	-6	...
chng. of ampl. chng.:	∼	∼	6	3	0	-6	6	-9	6	3	0	-6	6	-9	...

time	...	2995	2996	2997	2998	2999	3000
amplitude:	...	2	5	8	5	8	2
amplitude change:	...	0	3	3	-3	3	-6
chng. of ampl. chng.:	...	6	3	0	-6	6	-9

Table 4: Summary of synthetic time series constructions

time series	period	number of concepts		
		amplitude	ampl. chng	chng. of chng.
258	3	3	2	3
268	3	3	3	3
258852	6	3	3	3
258582	6	3	4	5
225825558822885	15	3	5	7
268682826286286	15	3	6	9
15937	5	5	2	3
15739	5	5	5	5
1573993751	10	5	7	5
1573971593	10	5	6	9
153791377195395	15	5	9	9
157393975117359	15	5	8	11

first row of Table 3. We present corresponding amplitude changes and changes of amplitude change in the second and in the third row of this Table. The ∼ denotes missing amplitude change and change of amplitude change for the first values of amplitude. It is worth drawing attention that there are 3 different amplitude concepts/granules, 4 different concepts/granules of amplitude change and 5 different concepts/granules of change of amplitude change for this series. We summarize all processed synthetic time series in Table 4.

Figure 3 illustrates selected 2 synthetic time series elevated to three-dimensional spaces of amplitude, amplitude change and change of amplitude change. One can observe distinct regularities in presented plots.

Analyzed time series represent a spectrum of different structures, i.e. a variety of the following parameters: length of the repeated period and numbers of concepts/granules in dimensions of amplitudes, amplitude changes and changes of amplitude change. Table 4 outlines these parameters.

3.2 Real-world time series

To complement the experiments' section of our research on the two approaches to time series modeling with Fuzzy Cognitive Maps we have selected 6 different real-world time series, named Equiptemp, Kobe, Sunspots, Wave1, Wave2, and Well.

Figure 3 illustrates selected synthetic and real-world time series in three-dimensional spaces of amplitude, amplitude change and change of amplitude change (left column) and present, past, and before past values. Observe that real data has substantially different characteristics than synthetic. Points do not form regular alike, dense clusters.

4 Results

The results of conducted experiments are gathered in the form of figures. Plots follow a uniform convention. The OY axis refers to the MSE. OX axis informs about FCM

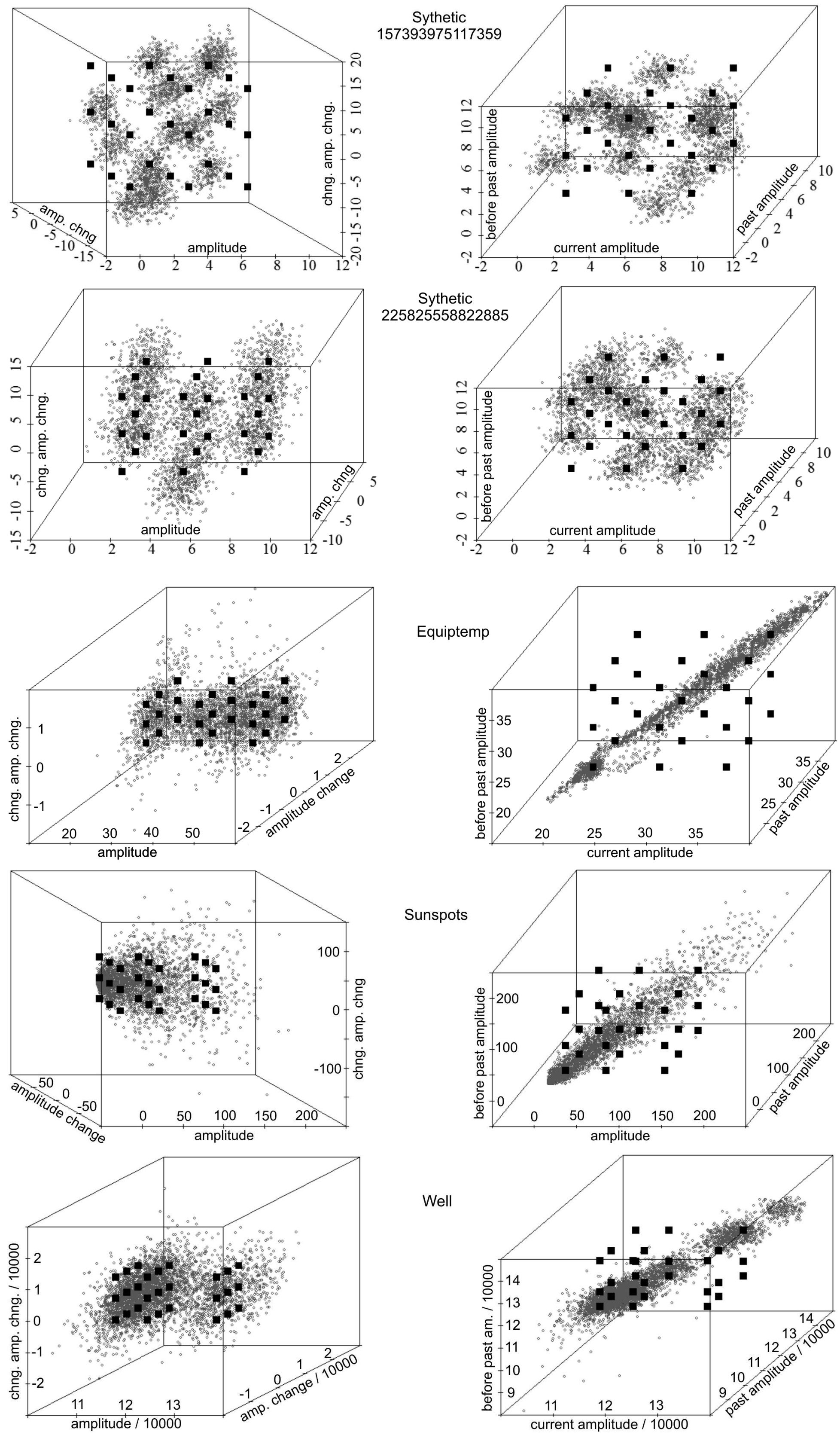

Fig. 3: Synthetic (upper two rows) and real-world (bottom three rows) time series elevated to spaces of amplitude, amplitude change and change of amplitude change (left column) and present, past, and before past amplitudes (right column).

architecture. Starting from the left side of each plot (the intersection of OX and OY) we plot MSE for Fuzzy Cognitive Maps of size $n = 4, 6, 8, 10, 12, 17, 22, 27$. In each group we plot MSE for time series model and one-step-ahead forecasts. Input data was partitioned into train and test. Train data was used to learn the FCM. Prediction data part is called test and it was not involved in the learning procedure. Two first bars, in darker color, refer to MSE for the model and prediction based on history-oriented time series representation. Third and fourth bar in each group (gray colors) refer to models built on dynamics-oriented time series representation.

Figure 4 illustrates MSE for synthetic time series models and predictions for FCMs designed according to history and dynamics oriented schemes.

Plots convincingly illustrate the advantage of history-oriented time series representation over dynamics-oriented. In almost each case for FCMs of the same size errors for models built on history are lower than for models constructed on dynamics-based representation.

The superiority of history-based representation scheme is especially clear for small maps. This is very important notice, because smaller maps are more convenient from practical points of view. Smaller maps are less burdensome to train and much easier to interpret and to apply in practice. With the growth of the number of nodes in the map, the advantage of history-oriented time series representation decreases, but still history-based overperforms the dynamics-based models. For full maps, with $n = 27$ nodes the quality of model and prediction are almost comparable.

At the same time a good FCM design is a balance between modeling quality (expressed as possibly low errors) and simplicity of the map. In this light, the question which map should be chosen to model given time series depends on the modeling purpose. In practice, small maps, for example with $n = 6$ nodes and 36 arcs, are favorable, because of the ease of interpretation and application of such model. The cost of simplicity in in numerical precision. Observe in Figure 4 that smaller models have higher errors.

There is a common characteristics for FCM designed on dynamics and on history-oriented time series representation. The larger the map, the smaller the MSE. The rate at which the error decreases with the increase of the number of nodes in nonlinear. There is an inflection point, which may be used as an indicator, stressing that adding more concepts will not result in a substantial improvement in modeling quality. This is even more clear for the history-oriented models. The inflection point in this case is close to the length of time series period. In other words, FCM-based models stabilize faster if we use history-oriented time series representation.

Figure 5 illustrates experiments results for 6 real-world time series.

The proposed modeling and prediction technique performs very well on real-world data. The experiments show that in several cases results on real data are even better than for synthetic ones, see OY scale in Figure 5. Errors are low both for models and for one-step-ahead predictions.

There is a clearly visible relation between complexity of time series and modeling quality. Simpler time series are much easier to learn and to predict. Refer to low bars for synthetic time series with period of 3 points in Figure 4. A corollary of this observation is that when data are irregular, like in case of real-world datasets, if formal methods give us inconclusive decision about the number of nodes, then typically we shall choose the highest reasonable (interpretable) number.

Conclusions from a comparison of the two methods, viz. oriented on dynamics and on history are analogical to the conclusions produced for the synthetic data. History-oriented models are better than dynamics-oriented. Even for very small maps we were able to model time series. The MSE decreased with the growth of the number of nodes in the map. The proposed membership index help select the most relevant concepts .

5 Conclusion

The article discusses and compares two time series representation schemes that can be applied in Fuzzy Cognitive Maps-based models. First method is oriented on dynamics. It elevates given scalar time series into a 3-dimensional space of amplitude, amplitude change and change of amplitude change. Second method uses unprocessed time series values and elevates scalar time series representation into a 3-dimensional space of present, past, and before past observations. Both perspectives are a start point for Fuzzy Cognitive Map design for time series modeling and prediction.

The experiments show that there exist some commonalities between these two methods. First of all, with the number of nodes in the map, we can more precisely model the time series. There is only a little difference in modeling quality for dynamics and history-oriented time series representations.

At the same time, modeling on the concept level is not oriented on accuracy, but on the interpretability of the models. Hence, focusing on the interpretability of the models, preference is given to a model that is relatively small. Therefore, obtaining such model, we need to extract relevant concepts and select only several best ones. In the article we have used membership index to extract best models. We have shown that for small maps, history-oriented time series representation outperforms the dynamics-based ones. We have illustrated that history-oriented method allows to build more accurate small models. There is a clear inflection point in the characteristic of the history-oriented method that indicates when adding more concepts stops substantial improvement of model's quality. Such an inflection point is not that visible in dynamics-based maps.

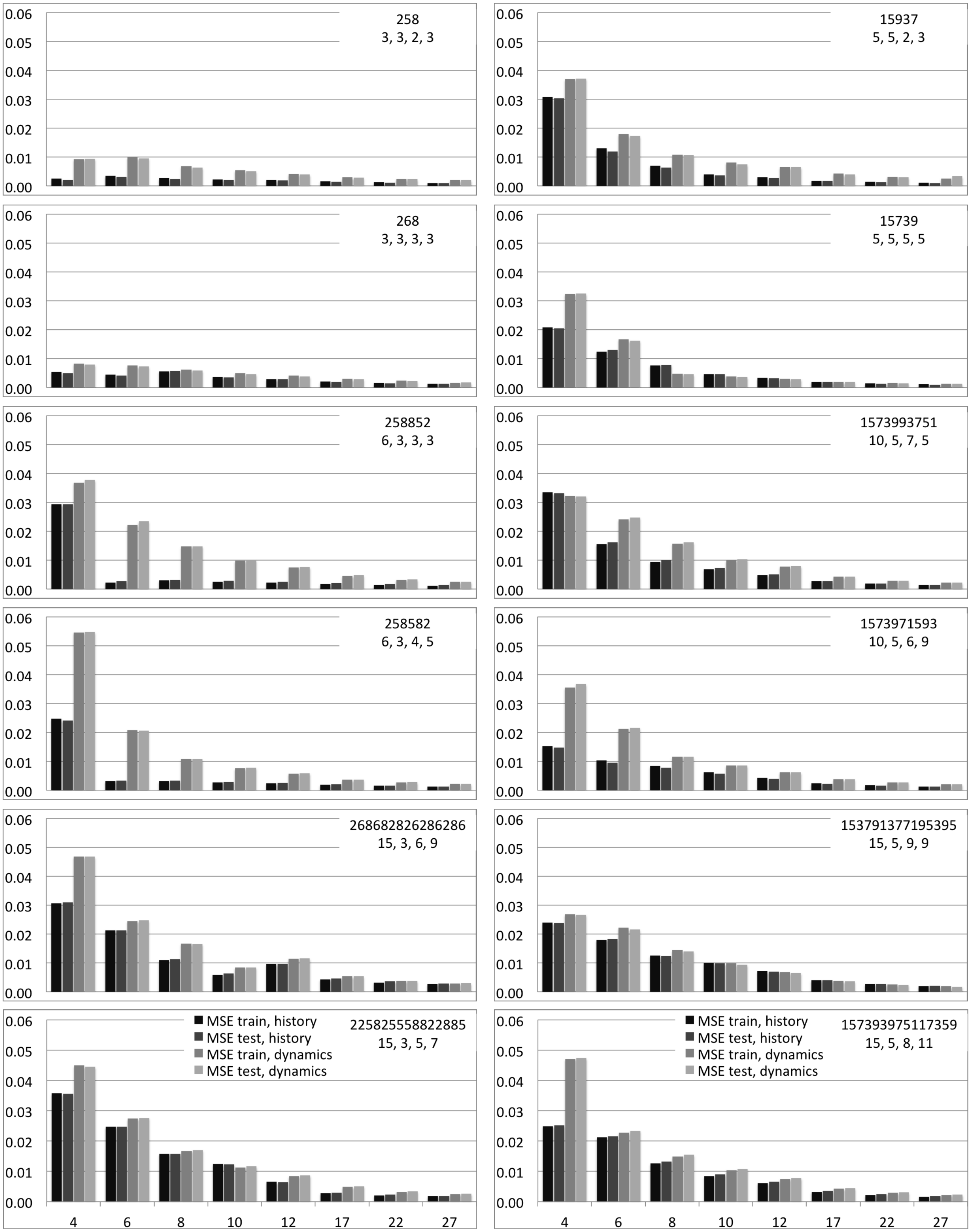

Fig. 4: Synthetic time series: performance of history and dynamics based representation

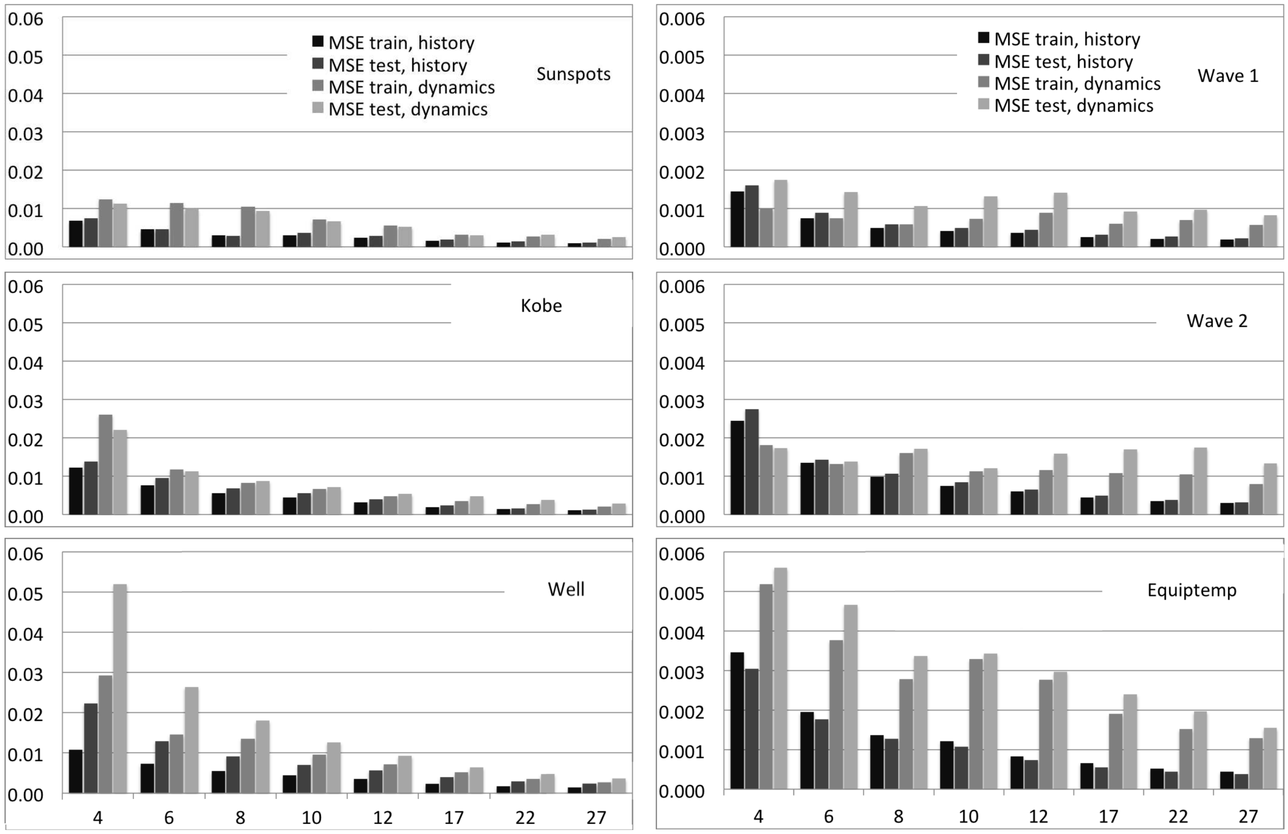

Fig. 5: Real time series: performance of history and dynamics based representation

In the future it could be worth investigating other selection criteria to choose relevant concepts. We also plan to extend experiments to higher dimensional spaces (higher order history-oriented temporal relationships), for example: present value, past value, before past value, before-before past value, etc.

Acknowledgement

The research is supported by the National Science Center, grant No 2011/01/B/ST6/06478, decision no DEC-2011/01/B/ST6/06478.

References

[1] Axelrod R., Structure of Decision: The Cognitive Map of Political Elites, Princeton University Press, 1976.

[2] Kosko, B., Fuzzy cognitive maps. In Int. J. Man Machine Studies 7, 1986.

[3] Mohr S. T., The use and interpretation of fuzzy cognitive maps, Master's Project, Rensselaer Polytechnic Institute, 1997.

[4] Pedrycz W., Granular Computing: An introduction, Proc. of the Joint 9th IFSA World Congress and 20th NAFIPS, Vancouver, Canada, 2001.

[5] Pedrycz W., The design of cognitive maps: A study in synergy of granular computing and evolutionary optimization, In: Expert Systems with Applications 37, 2010, pp. 7288-7294.

[6] Pedrycz W., Bargiela A., Granular Mappings, IEEE Transactions on Systems, Man, and Cybernetics - part A vol. 35 (2), 2001.

[7] Pedrycz W., Bargiela, A., Granular clustering: A granular signature of data, IEEE Trans. Syst. Man And Cybernetics - part B vol. 32 (2), 2002, pp. 212-224.

[8] Pedrycz W., The design of cognitive maps: A study in synergy of granular computing and evolutionary optimization, In: Expert Systems with Applications 37, 2010, pp. 7288-7294.

[9] Stach W., Kurgan L., Pedrycz W., Numerical and Linguistic Prediction of Time Series, in: IEEE Transactions on Fuzzy Systems 16(1), 2008, pp. 61-72.

[10] Stach, W., Kurgan, L., Pedrycz, W., and Reformat, M., Genetic learning of fuzzy cognitive maps, Fuzzy Sets and Systems 153, 2005, pp. 371-401.

[11] Tolman E.C., Cognitive maps in rats and men, Psychological Review 55(4), 1948, pp. 189-208.

[12] http://cran.r-project.org/web/packages/pso/pso.pdf

[13] http://lib.stat.cmu.edu

[14] http://robjhyndman.com/tsdldata.

Wladyslaw Homenda is an Associate Professor in the Faculty of Mathematics and Information Science at the Warsaw University of Technology, Warsaw, Poland. He is also with the Faculty of Economics and Informatics in Vilnius, Lithuania, at the University of Bialystok, Poland. He collaborates with several universities and research institution in Poland and abroad. His main research interests are in theoretical foundations of computer science, knowledge representation and processing and intelligent computing technologies, specifically in the areas of man-machine communication and human-centric computing, fuzzy modeling and Granular Computing, knowledge discovery and data mining. His teaching activities are mainly focused on theoretical foundations of computer science. He has rich experience in practical applications of research results in different areas of human activities including software support for visually impaired people and intelligent processing of music information.

Agnieszka Jastrzebska received her M.Sc. Eng. degree in Computer Engineering from the Rzeszow University of Technology, Poland. She also holds MA degree in Economics from the University of Rzeszow, Poland. She is a Ph.D. student in Informatics and Computational Modelling at the Systems Research Institute of Polish Academy of Sciences. Agnieszka is research and teaching assistant at the Faculty of Mathematics and Information Science at the Warsaw University of Technology. She has a strong background in data mining. Most of her research centers around Fuzzy Cognitive Maps and pattern recognition. In the past 3 years she was awarded prestigious scholarships from The Institute of Computer Science of Polish Academy of Sciences, The Systems Research Institute of Polish Academy of Sciences, and The Center for Advanced Studies of Warsaw University of Technology.

Witold Pedrycz is Professor and Canada Research Chair (CRC - Computational Intelligence) in the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada. He is also with the Systems Research Institute of the Polish Academy of Sciences, Warsaw, Poland. His main research directions involve Computational Intelligence, fuzzy modeling and Granular Computing, knowledge discovery and data mining, fuzzy control, pattern recognition, knowledge-based neural networks, relational computing, and Software Engineering. He has published numerous papers in this area. Prof. W. Pedrycz is intensively involved in editorial activities. He is an Editor-in-Chief of Information Sciences and served as Editor-in-Chief of IEEE Trans. on Systems, Man, and Cybernetics - Systems. He currently serves as an Associate Editor of IEEE Trans. on Fuzzy Systems and is a member of a number of editorial boards of other international journals.