2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

Modeling Time Series with Fuzzy Cognitive Maps

Wladyslaw Homenda*†, Agnieszka Jastrzebska*, Witold Pedrycz‡§
*Faculty of Mathematics and Information Science, Warsaw University Technology, ul. Koszykowa 75, 00-662 Warsaw, Poland
†Faculty of Mathematics and Computer Science, University of Bialystok, ul. Sosnowa 64, 15-887 Bialystok, Poland
‡Systems Research Institute, Polish Academy of Sciences, ul. Newelska 6, 01-447 Warsaw, Poland
§Department of Electrical & Computer Engineering, University of Alberta, Edmonton T6R 2G7 AB Canada
e-mail:{homenda, A.Jastrzebska}@mini.pw.edu.pl, wpedrycz@ualberta.ca

***Abstract*—Fuzzy Cognitive Maps are recognized knowledge modeling tool. FCMs are visualized with directed graphs. Nodes represent information, edges represent relations within information. The core element of each Fuzzy Cognitive Map is weights matrix, which contains evaluations of connections between map's nodes. Typically, weights matrix is constructed by experts. Fuzzy Cognitive Map can be also reconstructed in an unmanned mode. In this article authors present their own, new approach to time series modeling with Fuzzy Cognitive Maps. Developed methodology joins Fuzzy Cognitive Map reconstruction procedure with moving window approach to time series prediction. Authors train Fuzzy Cognitive Maps to model and forecast time series. The size of the map corresponds to the moving window size and it informs about the length of historical data, which produces time series model. Developed procedure is illustrated with a series of experiments on three real-life time series. Obtained results are compared with other approaches to time series modeling. The most important contribution of this paper is description of the methodology for time series modeling with Fuzzy Cognitive Maps and moving windows.**

I. Introduction

Cognitive Maps are soft computing models, that allow to model phenomena and relations between phenomena. Cognitive maps (term fathered by E. Tolman, [11]) are present in sciences since 1940s. The beginnings of the field are associated with researches on hidden learning process observed among vertebrate animals. First application of Cognitive Maps to modeling was in 1976, when R. Axelrod, political scientist, represented a set of social scientific phenomena, [1].

Cognitive Map is a set of knowledge elements describing a phenomenon in its environment. Examples of such phenomena are (from, say, economics): unemployment, fuel prices, taxes, government expenditures, accumulation of capital, technological change, birth rate, etc. Cognitive Map is represented with a directed graph, where a set of nodes $\{nd_1, nd_2, \ldots, nd_n\}$ corresponds to phenomena of interest. Edges $\{w_{11}, w_{12}, \ldots, w_{nn}\}$ correspond to relations between the phenomena. Cognitive Maps may be perceived as spatially oriented knowledge representation models, where unique data structure and its access methods enhance system's learning possibilities.

In this paper authors apply Fuzzy Cognitive Maps to model time series. Our objective is to present a new methodology for time series modeling and forecast, which joins moving window technique with Fuzzy Cognitive Map learning. Approach introduced in this paper is not in the current stream of research on Fuzzy Cognitive Maps. We believe, that the proposed procedure is an original and valuable contribution to this area.

The paper is structured as follows. In Section II the authors present a brief literature overview on selected topics related to Fuzzy Cognitive Maps. Section III introduces and provides background of the proposed methodology for time series modeling with Fuzzy Cognitive Maps. Theoretical discussion is supported by a series of experiments presented in Section IV. Section V covers conclusions and future research plans.

II. Brief Literature Review

Time series modeling is well recognized area of knowledge mining. The main goal of time series modeling is to make forecasts that may be applied for decision making, such as making orders (if we expect that prices will raise), prevention (if we expect extremely high rainfalls), etc. Time series modeling and forecasting is a topic wide enough to provide study material for a year-long academic courses (as it is usually at universities) and this may not be enough to exhaust this subject. Therefore, we would like to mention only selected issues.

Time series is a sequence of observations, usually gathered at regular intervals. Examples of popular time series are economic indicators: prices, rates, meteorological phenomena: rainfall, water level, temperature, population-related phenomena: migration, birth rate, and so on.

Time series analysis usually considers several types of time series. Important differences in characteristics of time series (ts) are: linearity/nonlinearity, univariate series/multivariate series, stationarity/nonstationarity.

Simple methods decompose a time series into ([10]):

- trend - long-term shifts in the mean,
- seasonal effects - cyclical fluctuations related to the calendar,
- cycles - other cyclical fluctuations (such as a business cycles),
- residuals - other random or systematic fluctuations.

Typically, four aforementioned elements are combined either additively or multiplicatively into a single model. Alternatively, several nonparametric approaches to time series analysis have been introduced, i.e. spectral analysis or Kernel methods, [7].

Among popular classical approaches to time series modeling and forecasting are:

- linear stationary models called ARMA (for AutoRegressive Moving Average models),
- linear nonstationary models called ARIMA (for AutoRegressive Integrated Moving Average),
- nonlinear models called ARCH (for AutoRegressive Conditional Heteroskedasticity) and its generalized counterpart: GARCH,
- exponential smoothing models (with and without accounting a trend), exponential smoothing in its simplest form can be classified as ARIMA model with parameters par= (0,1,1),
- approaches to time series modeling rooted in frequency analysis, in particular Fourier analysis of stochastic processes,
- alternative methods based on other data mining techniques, i.e. regression or neural networks, rule-based systems, distances, etc.

We would like to highlight that due to space limitations we did not intend to prepare full literature review and we did not cover all models as we wish we could have done. Interesting readings in the area of models listed above is not only, but also in: [2], [3], [4] and [6].

In this article authors explore Fuzzy Cognitive Maps and their applicability to describe time series. This approach is not in the main stream of research on time series modeling. It is an interesting, soft computing model, a great alternative to existing methodologies.

The application of Fuzzy Cognitive Maps to time series modeling has been previously discussed not only but also in [5], [9] and [8]. Methodologies presented in these articles are fundamentally different from ours. Time series modeling discussed in the referenced papers is tightly related to fuzzy techniques in data mining. In the cited articles nodes represent certain constructs, decided beforehand by the researcher. For example, in [9] nodes represent 9 combinations of fuzzy assessments of current value of given data point point and the change (delta). Discussed are following 9 combinations: High-High, High-Medium, High-Small, Medium-High, ... , Small-Small. Value passed to the node representing a particular combination is determined with a membership function.

In general, existing methodologies to time series modeling are based on the following steps:

- fuzzification of input data,
- FCM learning,
- forecasting,
- output defuzzification.

We would like to highlight again the originality of our approach, which is different from methodologies discussed in the aforementioned articles. We do not perform the fuzzification/defuzzification steps, time series modeling and forecasting is based on moving window technique, input and output data does not undergo any special procedures. Fuzzification/defuzzification procedures described in the cited papers may be perceived as somehow subjective. They rely on evaluation of membership functions, which number and shape are determined by the modeling party.

III. Methodology

A. Fuzzy Cognitive Maps

1) Modeling: A Fuzzy Cognitive Map models a real phenomena of interest and relations between them, which are manifested by observed *causes* and *results*. Observed causes are presented to FCM, which should respond with results. Causes presented to a FCM are named *activations* and results are named *targets* or *goals*. Since perfect modeling is rather impossible, FCM *responses* are expected to be as close as possible to the *results*.

A Fuzzy Cognitive Map is characterized by its weights matrix W. It is the crux of the FCM - weights describe connections between the nodes in the map. Connections in a Fuzzy Cognitive Map are expressed as real numbers from the $[-1, 1]$ interval. Input and output information (activations and responses) in a Fuzzy Cognitive Map is scaled to the $[0, 1]$ interval.

Let us determine notation for Fuzzy Cognitive Maps based on up-to-date research and use abbreviations from here on:

- FCM - Fuzzy Cognitive Map,
- n, N - numbers of phenomena and observations,
- W - square matrix of weights with n rows and columns,
- $W_{i\cdot}$, $W_{\cdot j}$ - i-th row and j-th columns of W,
- w_{ij} - item of W in i-th row and j-th column,
- X and G - matrices of (N) observations with n rows and N columns, called activations (X) and targets or goals (G), respectively,
- Y - matrix of FCM responses, n rows and N columns,
- $X_{i\cdot}$, $X_{\cdot j}$, $G_{i\cdot}$, $G_{\cdot j}$, $Y_{i\cdot}$, $Y_{\cdot j}$, x_{ij}, g_{ij}, y_{ij} - rows, columns and items in corresponding matrices X, G and Y; for example, selection of the first row from X: $X_{1\cdot} = [x_{11}, x_{12}, \dots, x_{1N}]$ and selection of the first column from X: $X_{\cdot 1} = [x_{11}, x_{21}, \dots, x_{n1}]^T$,
- the pair $X_{\cdot j}$ and $G_{\cdot j}$ is called (j-th) observation,
- $X_{\cdot j}$, $G_{\cdot j}$ $Y_{\cdot j}$ are named (j-th) activation, target and response, respectively.

FCM exploration is based on activations X, which are processed with the weights W according to the formula:

$$Y = f(W * X) \quad (1)$$

where $*$ is an operation performed on matrices W and X. Matrix product is an example of such operation and it is utilized in this study. f is a sigmoid mapping applied individually to elements of $W * X$:

$$f(z) = \frac{1}{1 + e^{-\tau z}} \quad (2)$$

with positive value of the parameter τ. In this study the value of τ is arbitrarily set to 2.5, based on experiments.

Single (j-th) activation $A_{\cdot j}$ and response $Y_{\cdot j}$ are bind as follows:

$$Y_{\cdot j} = f(W * X_{\cdot j}) \tag{3}$$

and the individual i-th element of the j-th response is defined by the following formula:

$$y_{ij} = f(W_{i\cdot} * X_{\cdot j}) = f\Big(\sum_{k=1}^{n} w_{ik} \cdot x_{kj}\Big) \tag{4}$$

2) Quality of modeling: The better is the map, the closer are map responses to targets. Distance between targets and map's responses can be naturally measured by metrics ρ such as Chebyshev, Manhattan, Euclidean etc.

We can compute distance between given j-th map's response $Y_{\cdot j}$ and corresponding target $G_{\cdot j}$ as $\rho(Y_{\cdot j}, G_{\cdot j})$. Distance between all map's responses Y and targets G is $\rho(Y, G)$. Distances can be interpreted as errors' measures of modeling real phenomena by FCMs. In this article as a basic performance measure we use Mean Squared Error, which is defined as follows:

$$MSE_j = \frac{1}{n} \cdot \sum_{i=1}^{n} (y_{ij} - g_{ij})^2$$
$$MSE = \frac{1}{N} \sum_{j=1}^{N} MSE_j = \frac{1}{n \cdot N} \cdot \sum_{j=1}^{N} \sum_{i=1}^{n} (y_{ij} - g_{ij})^2 \tag{5}$$

The MSE_j error measure is defined for single observation and its modeled value. The MSE error measure is defined for the all N of observations and their models. MSE measures the average of the squared Manhattan distances between map responses and targets. It is a popular performance measure, sensitive to extreme prediction errors.

3) Reconstructing: In order to reconstruct a good FCM, we have to propose weights, which allow to produce map responses closest to targets. This is the fundamental rule for FCM reconstruction procedure.

The FCM reconstruction procedure aims at optimization of weights matrix W so that the distance between targets and map responses are minimized. In order to proceed with FCM construction we need appropriate learning data: activations and targets. In this article we minimize the MSE between targets and map responses with the use of Particle Swarm Optimization procedure. FCM reconstruction procedure has been adapted to build FCMs that model time series.

B. Time series modeling with Fuzzy Cognitive Maps

Let us consider a series of numbers (or data points) c_i for $i = 1, 2, \cdots$, which we call *time series*:

$$c_1, c_2, c_3, c_4, \ldots$$

Using naming convention presented at the beginning of this section, we say that we have a series of observations, where a single observation is just one value c_i.

1) Modeling: We attempt to process time series with FCMs in order to predict future values based on a given history. Let us assume that a FCM has n nodes. We construct the set of N FCM observations using $N + n$ points of time series. Processing with FCMs imposes input time series to be normalized to the $[0, 1]$ interval. Recall that observations are manifested as activations and targets (goals), which are matrices X and G having n rows and N columns. Columns of activations matrix consist of consecutive n time series points. Videlicet:

$$\begin{aligned} X_{\cdot 1} &= [c_1, c_2, \ldots, c_n]^T \\ X_{\cdot 2} &= [c_2, c_3, \ldots, c_{n+1}]^T \\ &\cdots \\ X_{\cdot N} &= [c_N, c_{N+1}, \ldots, c_{N+n-1}]^T \end{aligned}$$

where $[\cdot]^T$ denotes transposition.

We assume that for given activations FCM responds with the sequence of time series points in times shifted by one unit. That is to say, FCM responds with values corresponding to the target $Y_i = [c_{i+1}, c_{i+2}, \ldots, c_{i+n}]^T$ while it receives the activation $X_i = [c_i, c_{i+1}, \ldots, c_{i+n-1}]^T$. Therefore the matrix G of goals is almost directly derived from the matrix of activations: dropped is the first column $X_{\cdot 1}$ and attached is the the extra column $G_{\cdot N} = [c_{N+1}, c_{N+2}, \ldots, c_{N+n}]^T$.

The outcome of such data preprocessing step are matrices of activations and targets. Note that the concept illustrated with Table I is coherent with moving window technique. In the moving window technique each consecutive information chunk passed to processing is shifted forward by a fixed interval. In this paper time series in activations and targets is shifted by 1 data point (observe columns in Table I).

Subsequently, we can apply obtained input data to construct a weights matrix W. In this way we trained the FCM to model given time series. The performance of such FCM is described by Formula 1.

There are several attempts to process time series:

- values of weights are not limited, of course besides the general assumption that they come from the bipolar unit interval $[-1, 1]$,
- weights matrix W is lower triangular, i.e. all elements above the main diagonal are set to zero. Required lower triangular matrix implements an assumption that

X	$X_{\cdot 1}$	$X_{\cdot 2}$	$X_{\cdot 3}$	$\ldots$	$X_{\cdot N-1}$	$X_{\cdot N}$
$X_{1\cdot}$	c_1	c_2	c_3	$\cdots$	c_{N-1}	c_N
$X_{2\cdot}$	c_2	c_3	c_4	$\cdots$	c_N	c_{N+1}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$X_{n-1\cdot}$	c_{n-1}	c_n	c_{n+1}	$\cdots$	c_{N+n-1}	c_{N+n-2}
$X_{n\cdot}$	c_n	c_{n+1}	c_{n+2}	$\cdots$	c_{N+n-2}	c_{N+n-1}

G	$G_{\cdot 1}$	$G_{\cdot 2}$	$G_{\cdot 3}$	$\ldots$	$G_{\cdot N-1}$	$G_{\cdot N}$
$G_{1\cdot}$	c_2	c_3	c_4	$\cdots$	c_N	c_{N+1}
$G_{2\cdot}$	c_3	c_4	c_5	$\cdots$	c_{N+1}	c_{N+2}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$G_{n-1\cdot}$	c_n	c_{n+1}	c_{n+2}	$\cdots$	c_{N+n}	c_{N+n-1}
$G_{n\cdot}$	c_{n+1}	c_{n+2}	c_{n+3}	$\cdots$	c_{N+n-1}	c_{N+n}

TABLE I. THE MATRICES OF OBSERVATIONS (ACTIVATIONS X AND TARGETS G) FOR TIME SERIES PROCESSING WITH FCM OF SIZE n.

future cannot affect any past event. Therefore, the next time series point is computed based on former points,

- processing with biases. In this case, the sum of weighted activations is added on with a value independent on activations.

Let us discuss the second and third cases in details.

Lower triangular weights' matrix: Firstly, for the case of simplicity, assume that FCM has 5 nodes and that the activation $X_{\cdot i} = [c_{i+1}, c_{i+2}, c_{i+3}, c_{i+4}, c_{i+5}]^T$ is presented to the FCM. According to Table I, we expect to get the following targets $G_{\cdot i} = [c_{i+2}, c_{i+3}, c_{i+4}, c_5, c_6]^T$, c.f. Figure 1:

$$
\begin{aligned}
c_{i+2} &= f(w_{11} \cdot c_{i+1}) \\
c_{i+3} &= f(w_{21} \cdot c_{i+1} + w_{22} \cdot c_{i+2}) \\
c_{i+4} &= f(w_{31} \cdot c_{i+1} + w_{32} \cdot c_{i+2} + w_{33} \cdot x_{i+3}) \\
c_{i+5} &= f(w_{41} \cdot c_{i+1} + \ldots + w_{43} \cdot c_{i+3} + w_{44} \cdot c_{i+4}) \\
c_{i+6} &= f(w_{51} \cdot c_{i+1} + \ldots + w_{54} \cdot c_{i+4} + w_{55} \cdot c_{i+5})
\end{aligned}
$$

Responses as above are desired, but rather not possible. Therefore we will be looking for a vector of responses $Y_{\cdot i} = [y_{i+2}, y_{i+3}, y_{i+4}, y_5, y_6]^T$, which are close to targets rather than equal to.

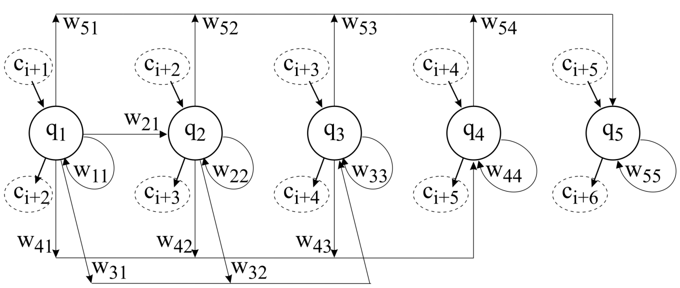

Fig. 1. FCM n=5 for time series modeling.

The general formula for FCM of n nodes and for j-th activation $X_j = [c_{j+1}, c_{j+2}, \ldots, c_{j+n}]$ can be formulated:

$$y_{kj} = f\big(\sum_{i=1}^{k} w_{ki} \cdot c_{j+i}\big) \text{ for } k = 1, 2, \ldots, n \qquad (6)$$

The y_{nj} is taken as forecast of c_{j+n+1} time series point. Other response' values $y_{n-1\,j}, y_{n-2\,j}, \ldots, y_{1\,j}$ may also be interpreted as forecast of time series points prior to c_{n+1}, i.e. $c_{j+n}, c_{j+n-1}, \ldots, c_{j+2}$. Since we have values for time not later than $j+n$, then these forecasts are of no interest.

FCM with bias: In this case, FCM response values are shifted by constant values called biases. The formula 3 takes the form:

$$Y_{\cdot j} = f\big(B + W \cdot X_{\cdot j}\big) \qquad (7)$$

where $B = [b_1, b_2, \ldots, b_n]^T$ is a vector of biases with $b_1, b_2, \ldots, b_n \in [-1, 1]$. In this case the Formula 4 goes to:

$$y_{ij} = f(b_i + W_{i\cdot} * X_{\cdot j}) = f\Big(b_i + \sum_{k=1}^{n} w_{ik} \cdot x_{kj}\Big) \qquad (8)$$

2) Forecasting: Now, we can apply Formulas 1, 3 and 4 to process time series. That is to say, the next time series point is computed (forecasted, predicted) based on the history of former n points:

$$y_{k+1} = \sum_{i=1}^{n} w_{ni} \cdot c_{k+i} \text{ for } k = 1, 2, \ldots, N - n - 1 \qquad (9)$$

In other words, n-th node of the FCM responses with a forecast of next time series point based on history of its former n points. Of course, these forecasts $y_{n+1}, \ldots, y_{n+N}$ may be compared with corresponding targets $c_{n+1}, \ldots, c_{n+N}$ in order to evaluate FCM performance.

IV. Time Series Modeling - Experiments

In this section we discuss a series of experiments, in which we apply FCMs to model time series. We study two important issues: the impact of bias and the map size on the quality of the model. Map reconstruction (optimization for the time series modeling) has been implemented in R. FCMs are optimized using Particle Swarm Optimization technique, implemented in package "pso".

The study has been performed on three different time series datasets, which describe real-life phenomena of varied character:

- annual rainfall in London from 1813-1912,
- number of births per month in New York city, from January 1946 to December 1959,
- Campito tree rings, which indicate tree growth, in years 1907-1960 .

Rainfall (Figure 2, upper part) time series is an additive model with a constant level without a trend. It is often cited in the literature as a benchmarking dataset for quality assessment.

The second data set (births per month in New York city) has seasonal and random fluctuations. *Births* time series (Figure 2, middle part) has very small amplitude.

The *tree rings* dataset has no trend and no seasonality. It is a time series with relatively the highest amplitude.

We have intentionally selected such variety of real-life datasets, to illustrate capabilities of our methodology. Chosen time series have substantially different characteristics. Moreover, they differ in size. Rain time series contains 100 data points, births in New York - 168 and the third dataset (tree growth) has only 66 observations. We have downloaded named datasets from: http://robjhyndman.com/tsdldata/.

In each dataset 90% of data points were used for model training, last 10% data points were left for testing purposes. All data has been normalized to $[0, 1]$ interval by dividing them per maximum value.

In the subsections below we present general properties of FCMs applied do time series modeling. We investigate which learning technique: with or without bias performs better. We also test the influence of the map size on the quality of modeling. Secondly, we investigate the quality of forecasting with FCMs. Finally, we compare our results with standard time series forecasting techniques.

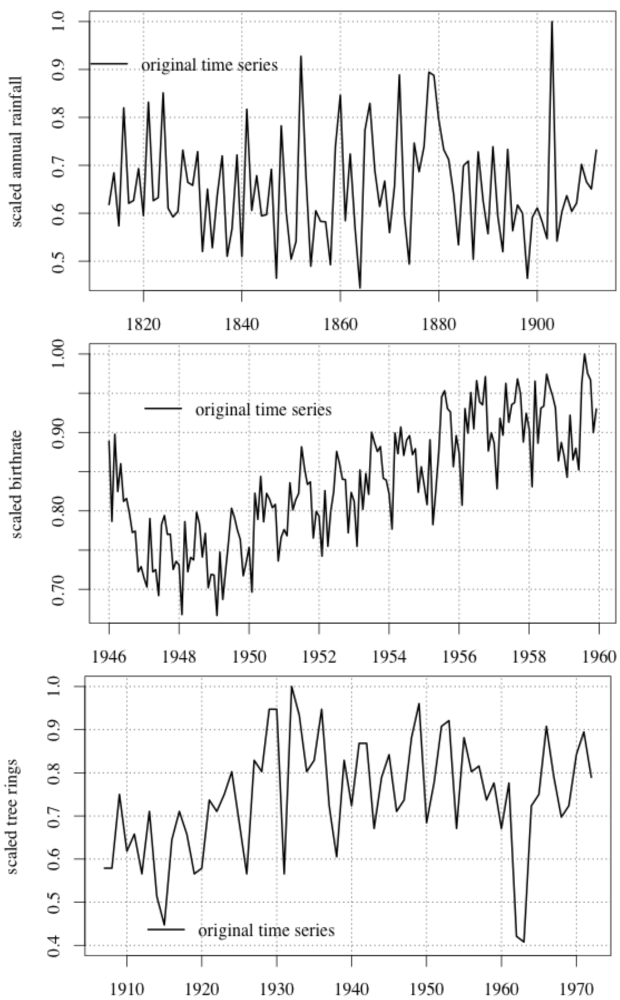

Fig. 2. Normalized *rain*, *births* and *tree rings* time series.

A. Time Series Modeling with Fuzzy Cognitive Maps - Map Size and Learning Technique

In this subsection we investigate map size and the impact of the training procedure on the accuracy of time series models. We apply FCMs of different sizes to describe chosen time series. Moreover, we study, if the technique based on learning with bias performs better than the technique without the bias. A statistical reference, that will allow to compare the quality of produced models is MSE.

We test the MSEs of time series models built with FCMs with $n = 3, 4, \ldots, 12$ nodes. We present MSE on train and test datasets. Train datasets were used to reconstruct FCMs. Size of the train dataset is $n \times N$. Test datasets are only for quality assessment. Activations from the test datasets were not involved in FCM training.

The proposed methodology for time series modeling and forecast uses technique of moving window. Moving window corresponds to the size of the FCM. On each node of the FCM we pass activations constructed from the input time series according to the scheme forced by the moving window. In this way, in each iteration we pass n history data points for periods, say, $k+1, k+2, \ldots, k+n$ and receive n map responses for periods $k+2, \ldots, k+n, k+n+1$. The larger the map, the longer history we use to train the model. Hence, we investigate the influence of map size on accuracy. Trained FCM allows to model and to forecast future values (unlimited number) for the time series.

Each node produces its own responses. We may treat nodes responses as n individual models of the time series. The most accurate response is for the $n-th$ node, because it is based on the longest history. We can also average nodes responses and treat the model as a whole. In this paper we treat the model in this way - as a whole. Such methodology, as one may suspect, achieves worse results. Withal, we have chosen to treat all map outputs as the model, to investigate the overall quality of reconstructed FCM. The objective of this paper is to discuss general suitability of FCMs to model time series. Model tuning, in order to achieve the highest performance, will be the topic of our further research.

Table II present MSEs for training and testing datasets for the three time series discussed in this paper. The modeling procedure produces n values for each data point in the time series. In other words, each data point is present n times in the moving window. As a result plotting all FCM outputs may be confusing. Therefore, we have plotted minimum, maximum and averaged of the modeled values for the time series. Due to space limitations, figures concern only FCMs with 5 nodes.

MSE errors for modeled *rain* time series are smaller for the training procedure with the bias. In general, the bigger the map, the smaller MSEs. Modeled values are based on longer history of observations, hence the improvement of accuracy.

Figures 3 present minimum, maximum and average of the modeled *rain* time series with FCM trained without and with bias respectively. Each map was of size $n = 5$, figures show training dataset.

For the *rain* time series there is no significant difference in models obtained with FCMs with and without the bias. Figures 3 confirm this conclusion. *Rain* time series is a typical dataset for methods evaluation and comparison. It has no trend and no seasonality; values oscillate around 0.65.

Modeled values have relatively smaller amplitude than the original series. FCMs outputs follow the trend. Local extrema of modeled values cover extrema of the time series. Nevertheless, one may expect modeled values to cover original time series in a better fashion. Unfortunately, for a lot of sudden peaks of high amplitude, like in *rain* time series, FCMs always produce outputs somehow moderated, averaged. Noteworthy is that the MSE for *rain* time series is rather low - see left part of Table 2. It will be shown at the end of this section, that other popular models also are not able to follow peaks in this time series.

MSEs for *births* time series train and test datasets modeled with different FCMs is presented in Table II. For maps trained with bias performed better in each case. We have also observed, that FCMs trained with bias have higher accuracy when they are larger.

Plots in Figures 5 illustrate models for *births* time series produced with FCMs of size $n = 5$ with and without bias. Figures show minimum, averaged and maximum of modeled values.

TABLE II. MSEs*100 OBTAINED FOR THREE TIME SERIES MODELS WITH DIFFERENT FCMs SIZES WITH AND WITHOUT BIAS.

	rain				births				tree rings			
map	no bias		with bias		no bias		with bias		no bias		with bias	
size	train	test	train	test	train	test	train	test	train	test	train	test
3	1.24	1.92	1.18	1.67	0.36	0.65	0.23	0.40	1.64	0.90	3.66	4.12
4	1.21	1.91	1.17	1.71	0.34	0.64	0.22	0.38	1.62	1.18	4.98	5.68
5	1.20	1.92	1.17	1.77	0.34	0.61	0.21	0.35	1.57	1.80	5.44	5.80
6	1.19	1.98	1.16	1.87	0.34	0.60	0.21	0.33	1.54	2.21	4.46	4.94
7	1.18	2.04	1.15	1.98	0.34	0.59	0.20	0.32	1.56	1.95	3.45	3.51
8	1.17	2.10	1.14	2.06	0.34	0.58	0.26	0.40	1.52	2.67	2.79	2.98
9	1.16	2.09	1.14	2.09	0.35	0.58	0.21	0.32	1.48	2.96	2.51	3.17
10	1.16	1.95	1.16	2.04	0.35	0.58	0.20	0.30	1.44	3.14	3.12	5.15
11	1.16	1.83	1.14	1.86	0.35	0.57	0.20	0.29	1.41	3.26	2.91	5.52
12	1.16	1.73	1.14	1.81	0.36	0.59	0.20	0.29	1.37	3.43	2.59	5.88

Births time series has a clearly visible seasonal fluctuations and a trend. Therefore, models were more accurate with the FCM trained with bias. Figures confirm this observation. Model trained with bias follows the trend and at the same time preserves hill-terrace-like characteristics of the seasonal fluctuations of the *births* time series. Model spans through the amplitude and follows local extrema. Model trained without bias preserves the characteristics of the time series, but does not follow the trend equally well.

Figure 4 illustrates FCM with 5 nodes, which has been reconstructed to describe *births* time series.

Table II contains MSEs for *tree rings* time series models built with FCMs of different sizes with and without the bias. For FCMs trained without bias, the size of the map influences accuracy. Larger maps model time series better, but smaller maps turned out to be better in forecasts. FCMs with bias achieved worse results.

Figures 6 illustrate models built with FCMs of size $n = 5$ with and without the bias. Though FCM without the bias achieved numerically better results, plots show that modeled time series does not reflect well the original. In contrast, model built with FCM trained with bias mimics *tree rings* time series better.

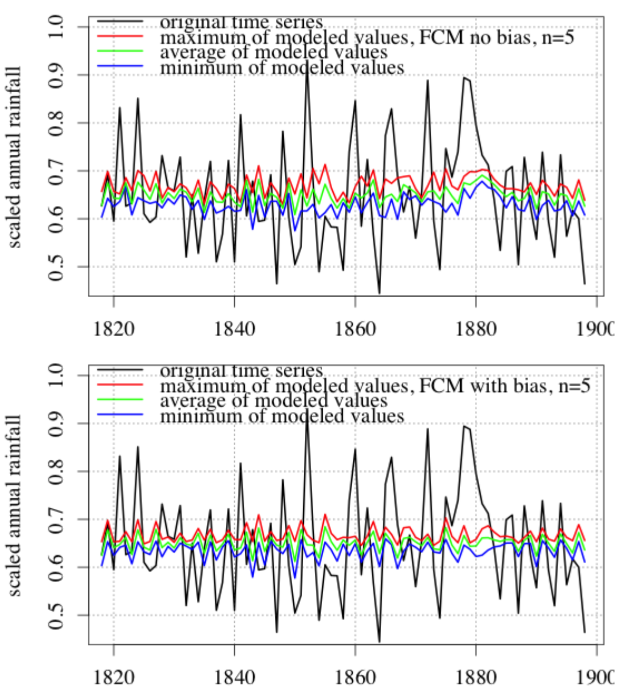

Fig. 3. Original and modeled *rain* time series, train and test dataset, FCM, n=5, no bias and with bias.

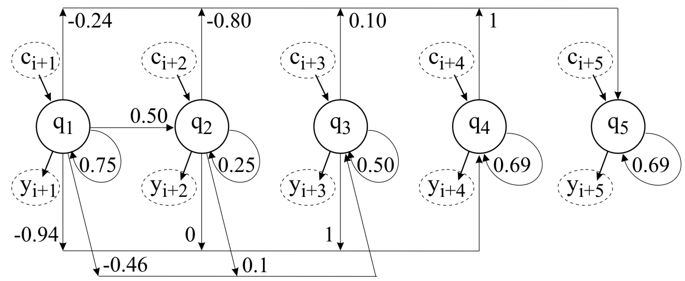

Fig. 4. FCM n=5 for *births* time series.

B. Forecasting with Fuzzy Cognitive Maps

FCMs can be applied to forecast future values of phenomena of interest. Prediction for one data point ahead is based on n available, observed, data. Forecast for the second data point is based on $n-1$ observed data points and 1 already predicted value. Prediction for the third data point is based on $n-2$ observed values and 2 forecasts and so on. Predictions for over $n+1$ data points ahead are based only on forecasts.

We have predicted future values for M data points ahead with FCMs built on N training observations. Train datasets contained 90% of time series. The quality of the forecasts are expressed with MSEs.

Table III contains MSEs for all three time series forecasts

TABLE III. MSEs*100 FOR TIME SERIES FORECASTS WITH DIFFERENT MAP SIZES WITH AND WITHOUT BIAS.

	rain		birth		tree rings	
map size	no bias	with bias	no bias	with bias	no bias	with bias
3	1.69	1.69	1.00	0.89	1.00	1.36
4	1.70	1.70	0.98	0.86	1.15	1.78
5	1.71	1.72	0.96	0.78	1.80	2.06
6	1.77	1.80	0.92	0.78	2.19	2.33
7	1.82	1.87	0.91	0.75	1.96	2.07
8	1.86	1.89	0.89	0.77	2.66	2.74
9	1.92	1.92	0.89	0.73	2.92	2.97
10	1.81	1.86	0.89	0.68	2.99	3.11
11	1.67	1.71	0.89	0.69	2.89	3.01
12	1.64	1.66	0.89	0.74	2.85	3.05

produced with different FCMs. There are no significant differences in these results for the *rain* time series in both models: without and with bias. For the *births* time series predictions made with FCMs trained with bias made are more accurate. We have observed, that the larger the map, the better are the forecasts. Finally, for the *tree rings* FCMs trained without the bias gave more accurate forecasts and map size in this case played less important role.

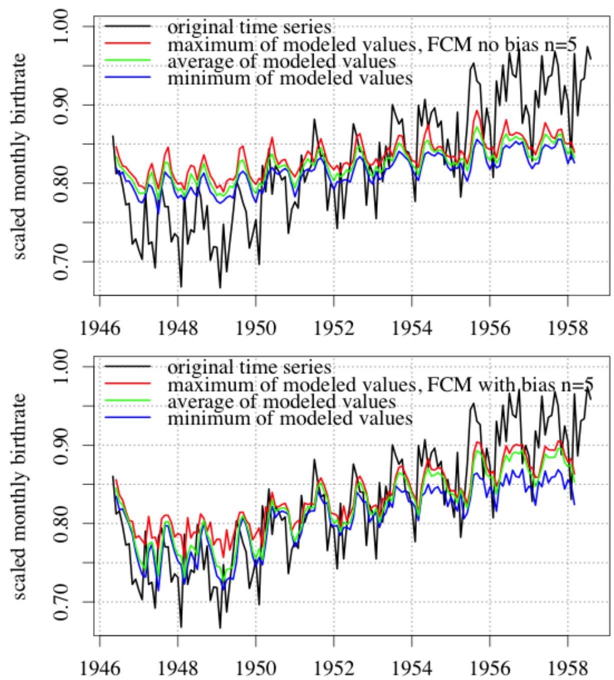

Fig. 5. Original and modeled *births* time series, train and test dataset, FCM, n=5, no bias and with bias.

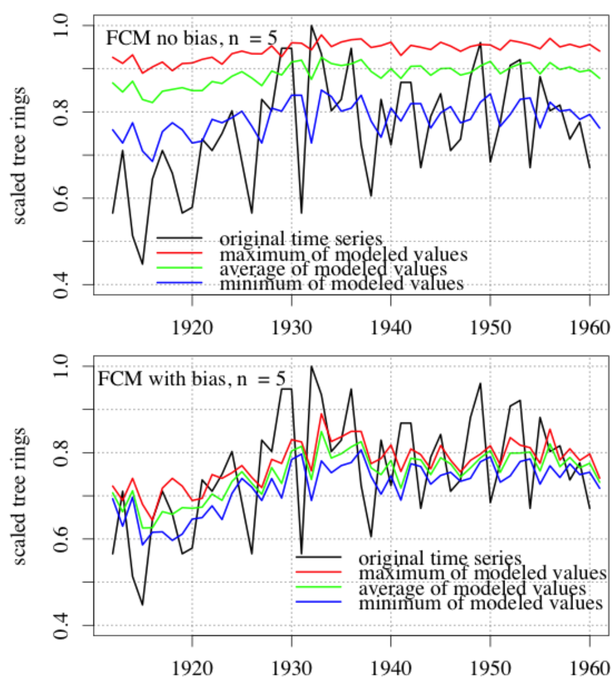

Fig. 6. Original and modeled *tree rings* time series, train and test dataset, FCM, n=5, no bias and with bias.

C. Time Series Modeling - Final Remarks

In this subsection we compare the quality of predictions made with FCMs with forecasts based on two popular approaches to time series modeling: Holt-Winters exponential smoothing and autoregressive integrated moving average (ARIMA) model, which is a generalization of an autoregressive moving average model.

Holt-Winters exponential smoothing method has been proposed by C.C. Holt in 1957 and then it has been updated by his student, Peter Winters in 1960. Over the years, this method has been successful adapted and explored. Holt-Winters method in its forecasts takes into account possibility of constant and non-constant trends. In this way the model in forecasts includes both the level and the slope.

ARIMA is time series modeling and forecasting tool, which is suitable also in some cases of non-stationarity of the time series. In such case, prior to forecasting, one has to do the differencing step to remove the non-stationarity. The ARIMA model is described by three parameters: p, d, and q. These are non-negative integers correspond to autoregressive, integrated, and moving averages parts of the model respectively.

In our study we have applied procedures for Holt-Winters and ARIMA (with p, d, q parameters selection) implemented in R. Parameter selection for ARIMA was also performed in R, with the use of "forecast" package.

Table IV contains MSEs for time series modeled and predicted using models build with ARIMA with parameters par=(1,0,0), Holt-Winters and FCM of size n=5 train with and without the bias.

Proposed time series modeling and forecasting methodology performs better, than both ARIMA and Holt-Winters. As we have shown before, for the *rain* time series, technique of training with and without the bias performs similarly good.

In the case of *births* time series our method obtains results comparable with the other two ones. Chosen time series contains seasonality and it has relatively very small amplitude. FCMs' models show better MSE than Holt-Winters does, but worse than ARIMA. Forecasts made with FCMs were worse. FCM with bias gave more accurate model and forecast.

For the *tree rings* time series quality of a model produced with FCM without the bias was comparable to results obtained with ARIMA. FCM trained with bias gave less accurate results.

Figure 7 compares original time series with models and forecasts built with the discussed four approaches: ARIMA, Holt-Winters and FCM with and without the bias. Vertical line divides model from the forecasts.

Proposed procedure allowed us to model *rain* time series with preservation of its characteristics. In comparison to

TABLE IV. COMPARISON OF MSES OBTAINED FOR THREE TIME SERIES WITH DIFFERENT MODELING AND FORECASTING METHODS.

MSE*100	rain		births		tree rings	
method	train	forecast	train	forecast	train	forecast
ARIMA par=(1,0,0)	1.18	2.15	0.13	0.59	1.30	3.51
Holt-Winters	1.27	1.77	0.37	0.18	1.31	4.60
FCM, n=5, no bias	1.20	1.71	0.34	0.96	4.03	3.71
FCM, n=5 with bias	1.17	1.72	0.21	0.78	5.44	2.33

ARIMA and Holt-Winters, our approach follows local extrema to the greatest extent. Models built with FCMs with and without the bias aim at preserving the original amplitude. Models built with ARIMA and Holt-Winters are slightly flatter. Figure 7 confirms that the differences between FCMs trained with and without the bias are not significant. Forecasts made with the four discussed approaches are similar. In all four cases, forecasts resemble straight line.

In the case of *births* time series all models allowed to follow time series characteristics. FCM without the bias is less satisfying - it does not preserve the trend. FCM with the bias follows the time series. Models and forecasts made with Holt-Winters and ARIMA also follow the trend.

The *tree rings* time series turned out to be the most challenging one. In this case, the best ARIMA model almost is convergent with Holt-Winters model, but far from a good fit. Time series models and forecasts with FCMs also differ from the original time series.

V. Conclusion

To sum up, in this paper we have presented a new approach to time series modeling and forecast, which combines FCMs and moving window approach. Theoretical discussion was supported by a series of experiments. We have tested our methodology on three real-life time series. We have also compared our approach with two popular time series modeling techniques.

Proposed methodology allows to describe time series well. We have distinguished two different FCM training procedures: with and without the bias. The choice of model with or without bias depends on the time series. For example, in the case of time series with trends FCM trained with bias is better. Time series with no trends and no seasonality are equally well modeled with both procedures.

Similarly as for most of data exploration approaches, our methodology performs very well, but it is not equally good for all kinds of data. We have shown that *births* time series has been modeled very satisfyingly. We were able to preserve time series characteristics and follow the trend (with FCM trained with bias). In contrast, *tree rings* time series turned out to be challenging. We have also shown, that *tree rings* were difficult t learn for standard models: ARIMA and Holt-Winters. The third time series - *rain* was a typical dataset, with no trend and no seasonality. We have shown, that in such case FCMs behave well and bias does not play very important role - built two models are similar.

The contribution in this paper is a new, not present in the literature, approach to time series modeling with FCMs and moving window. In our future research we plan to continue our research in this area. The results presented in this paper are encouraging. We plan to experiment with different learning techniques and apply other information representation schemes.

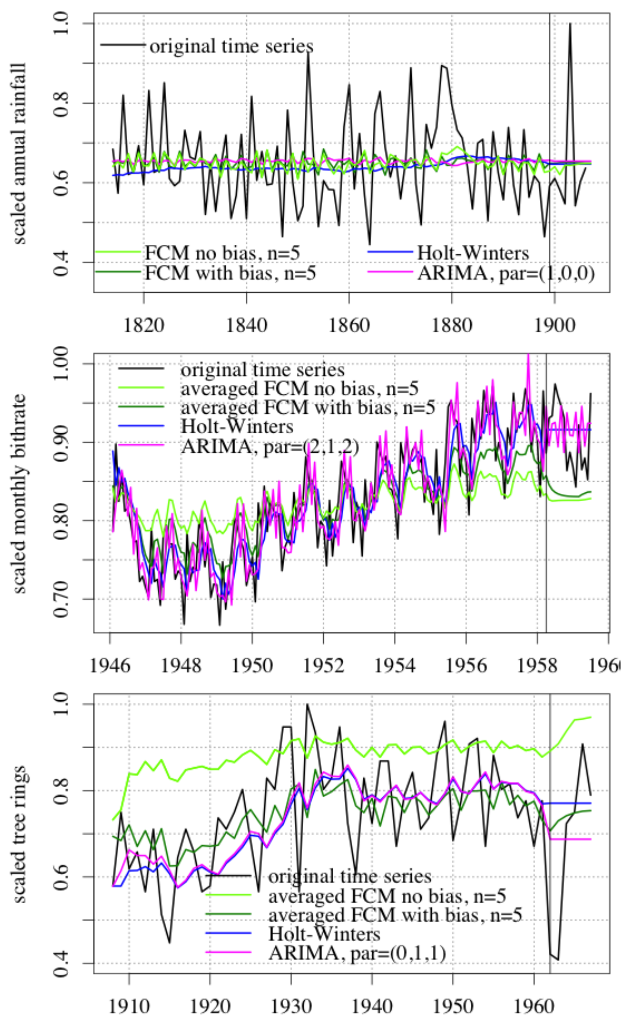

Fig. 7. Time series models and forecasts.

Acknowledgment

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 Maps of Political Elites*, Princeton University Press, Princeton, NJ, 1976.

[2] Box G., Jenkins G., *Time series analysis: forecasting and control*, rev. ed., Oakland, California: Holden-Day, 1976.

[3] Brockwell P.J., Davis R.A., *Time Series: Theory and Methods*, Springer Series in Statistics, 1986.

[4] Chatfield C., *Time-series Forecasting*, Chapman & Hall, 2001.

[5] Froelich W., Papageorgiou E. I., Samarinasc M., Skriapasc K., *Application of evolutionary fuzzy cognitive maps to the long-term prediction of prostate cancer*, in: Applied Soft Computing 12, 2012, pp. 3810-3817.

[6] Hamilton J.,*Time Series Analysis*, Princeton: Princeton Univ. Press, 1994.

[7] Hardle W., Lutkepohl H., Chen R., *A Review of Nonparametric Time Series Analysis*, in: International Statistical Review, Vol. 65, No. 1, 1997, pp. 49-72.

[8] Lu W., Yang J., Liu X., *The Linguistic Forecasting of Time Series based on Fuzzy Cognitive Maps*, in: proc. of IFSA/NAFIPS, 2013, pp. 649 - 654.

[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] Weber R., Course materials for Time Series course at Cambridge University, 2001.

[11] Tolman E. C., *Cognitive Maps in Rats and Men*, in: The Psychological Review, 55(4), 1948, pp. 189-208.