University of Ostrava

Institute for Research and Applications of Fuzzy Modeling

Linguistic Approach to Time Series Analysis and Forecasts

Martin Štěpnička, Antonín Dvořák, Viktor Pavliska and Lenka Vavříčková

Research report No. 146

2010

Submitted/to appear:

2010 IEEE World Congress on Computational Intelligence

Supported by:

Projects 1M0572 and MSM6198898701 of the MŠMT ČR.

University of Ostrava
Institute for Research and Applications of Fuzzy Modeling
30. dubna 22, 701 03 Ostrava 1, Czech Republic

tel.: +420-59-7091401 fax: +420-59-6120478
e-mail: martin.stepnicka@osu.cz

Linguistic Approach to Time Series Analysis and Forecasts

Martin Štěpnička and Antonín Dvořák and Viktor Pavliska and Lenka Vavříčková

***Abstract*— Linguistic approach of time series analysis is suggested. It adopts aspects of the decomposition and autoregression. The linguistic, i.e., interpretable and transparent, nature of the approach is emphasized. Precision of the suggested approach is demonstrated on real time series.**

I. Introduction

A. Time Series - State of the Art

Analysis and forecasting of time series have a wide practical use in economy, industry, meteorology, and other areas of application [6]. There is a vast variety of potential approaches to this task, among them two are, say, standard approaches. The first approach stems from the *Box-Jenkins methodology* [2] and it consists of autoregressive and moving average models. For instance, the ARMA(p,q) model is a typical representative of this methodology, assumes that every single value x_t of a given time series can be computed as follows:

$$x_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i x_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}, \qquad (1)$$

where $\varphi_1, \ldots, \varphi_p$ are parameters of the autoregressive model, $\theta_1, \ldots, \theta_q$ are parameters of the moving average model, c is a constant, ε_t is a white noise term and $\varepsilon_{t-1}, \ldots, \varepsilon_{t-q}$ are error terms.

The second approach is based on a *decomposition* of a given time series into *trend, cycle, season* and *noise* components. So, this approach assumes that a given times series is an additive or multiplicative composition of the above terms which have clear meanings. Hence, models decomposing a given time series into these components may be quite transparent.

Compared with the decomposition, the autoregressive and moving average models of the Box-Jenkins methodology are not as transparent and well interpretable since one cannot that easily see the influences of, e.g., trends or seasonal components. On the other hand, these models have been demonstrated to be very powerful and successful in forecasts.

ARMA model (1) as well as most of the other Box-Jenkins methods work under the stationarity assumption, i.e., assuming that the moments of x_t such as mean and variance do not change over time. This implicitly means that the time series should not contain any observable trend. To apply the standard Box-Jenkins approach to a trend containing time series and to use its powerful properties, one has to first *de-trend* a given time series or apply an *autoregressive integrated moving average* ARIMA(p,d,q) model where the integrated part newly added to ARMA process may model generally polynomial trends. The parameter d determines the trend polynomial order, e.g., $d = 1$ determines a constant trend (non zero average), $d = 2$ determines a linear trend, $d = 3$ determines a quadratic trend etc.

B. Fuzzy Approaches to Time Series Analysis

So far, a notable number of works aiming at fuzzy approach to time series modeling and prediction has been published. For instance, a study presenting Takagi-Sugeno rules [18] in the view of the Box-Jenkins methodology has been already published, see [1]. However, the Takagi-Sugeno rules use functional consequents without any linguistic meanings, their antecedents are usually determined by a cluster analysis and they do not employ any kind of logical implication. They can be considered as a special kind of regression model rather than a linguistic approach.

Analogously, various neuro-fuzzy approaches, which lie on the border between neural networks, Takagi-Sugeno models and evolving fuzzy systems, are very often successfully used [15], [8]. However, it happens quite often that Gaussian fuzzy sets are tuned to have the center, say, at node 5.6989 and the width parameter equal to 2.8893 (see [8]), which is obtained using some optimization technique. The interpretability of such fuzzy sets is undoubtedly far from the interpretability of systems using models of fragments of natural language.

Therefore, it may be stated that so far published approaches, although very effective and powerful, are closer to standard regression methods than to an interpretable linguistic approach. It may be also concluded that the above mentioned works are generally more motivated by the Box-Jenkins methodology than by the decomposition.

As mentioned above the decomposition assumes the privilege of the interpretable model where the interpretability is meant in a sense of "readability" for non-statisticians and non-mathematicians. Let us also recall, that the interpretable linguistic approach is very often mentioned as one of the basic advantages for fuzzy methods. Therefore, we find the decomposition idea to be very suitable to be investigated as distinct from so far preferred Box-Jenkins motivated fuzzy approaches.

We propose a new methodology for the analysis and forecasting of time series, which is based on a combination of two techniques: the *fuzzy transform* [12] and the *perception-based logical deduction* [9], [11]. Our approach employs both the decomposition as well as the autoregression idea. First, we decompose the time series into the so called *trend-cycle*, and the *seasonal component*. The fuzzy transform plays an essential role in this step.

All authors are with the Institute for Research and Applications of Fuzzy Modeling, University of Ostrava, 30.dubna 22, 701 03 Ostrava, Czech Republic (phone: +420 597091403; email: Martin.Stepnicka@osu.cz).

Second, we describe the trend-cycle by the so-called linguistic description comprised from fuzzy rules. Fuzzy rules in the linguistic description describe process of an autoregressive nature. The motivation stems from the well known ability of fuzzy systems to describe distinct logical or functional dependencies (in a robust way) using formal logics, fragments of fuzzy set theory, and sentences formulated in natural language. This ensures the transparency and interpretability of the autoregressive trend-cycle model, which is very important for further comprehension of the processes that led to a given time series.

Third, the linguistic description generated automatically from data is used together with a specific inference method - perception based logical deduction - to forecast future trend-cycle values.

Finally, the autoregressive model of the seasonal components is determined and used to forecast these components in the future. Both forecasted components are composed together to obtain the time series forecasts.

Let us stress the crucial importance of the clear interpretability of the whole approach (we allow only fuzzy/linguistic IF-THEN rules, i.e., rules taken as special conditional clauses of natural language). On the other hand, increasing the interpretability of a model should not dramatically decrease the precision of its forecasts. Therefore, we provide readers with a comparison study on several real time series.

II. Fuzzy (F)-Transform

Let us briefly recall one of the main tools employed in the suggested approach, the *fuzzy transform* (F-transform for short) [12], in particular.

The F-transform is a special technique that can be applied to a continuous function, defined on a fixed real interval $[a, b] \subset \mathbb{R}$. The essential idea is to transform a given function defined in one space into another, usually simpler space, and then to transform it back. The simpler space consists of a finite vector of numbers obtained on the basis of the well established *fuzzy partition* of the domain of the given function. The reverse transform then leads to a function approximately reconstructing the original one. Thus, the first step, sometimes called the *direct F-transform*, results in a vector of averaged functional values. The second step, called the *inverse transform*, converts this vector into another continuous function, which approximates the original one. In this section, we will briefly overview the main concepts. More details can be found in [12].

In the sequel, by a fuzzy set in the universe U we will understand a function $A : U \longrightarrow [0, 1]$. The $\mathcal{F}(U)$ denotes the set of all fuzzy sets on U.

The F-transform is defined with respect to a *fuzzy partition*, which consists of *basic functions*.

Definition 1: Let $c_1 < \cdots < c_n$ be fixed nodes within $[a, b]$, such that $c_1 = a, c_n = b$ and $n \geq 2$. We say that fuzzy sets $A_1, \ldots, A_n \in \mathcal{F}([a, b])$ are *basic functions* forming a fuzzy partition of $[a, b]$ if they fulfill the following conditions for $i = 1, \ldots, n$:

1) $A_i(c_i) = 1$;
2) $A_i(x) = 0$ for $x \notin (c_{i-1}, c_{i+1})$, where for uniformity of notation we put $c_0 = c_1 = a$ and $c_{n+1} = c_n = b$;
3) A_i is continuous;
4) A_i strictly increases on $[c_{i-1}, c_i]$ and strictly decreases on $[c_i, c_{i+1}]$;
5) for all $x \in [a, b]$,

$$\sum_{i=1}^{n} A_i(x) = 1. \tag{2}$$

Usually, the *uniform fuzzy partition* is considered, i.e., n equidistant nodes $c_i = c_{i-1} + h$, $i = 2, \ldots, n$ are fixed. Let us remark that the shapes of the basic functions are not predetermined and can be chosen on the basis of further requirements.

Definition 2: Let a fuzzy partition of $[a, b]$ be given by basic functions $A_1, \ldots, A_n, n \geq 2$, and let $f : [a, b] \longrightarrow \mathbb{R}$ be an arbitrary continuous function. The n-tuple of real numbers $[F_1, \ldots, F_n]$ given by

$$F_i = \frac{\int_a^b f(x)A_i(x)dx}{\int_a^b A_i(x)dx}, \qquad i = 1, \ldots, n, \tag{3}$$

is a *direct fuzzy transform (F-transform)* of f with respect to the given fuzzy partition. The numbers $F_1, \ldots, F_n$ are called the *components* of the F-transform of f.

In practice, the function f is usually not given analytically, but we are at least provided some data, obtained, for example, by some measurements. In this case, Definition 2 can be modified in such a way that the definite integrals in Formula (3) are replaced by finite summations.

Let n basic functions be given, forming a fuzzy partition of $[a, b]$, and let the function f be given at $T > n$ fixed points $x_1, \ldots, x_T \in [a, b]$. We say that the set of points $\{x_1, \ldots, x_T\}$ is *sufficiently dense with respect to the fuzzy partition* if for every $i \in \{1, \ldots n\}$ there exists $t \in \{1, \ldots, T\}$ such that

$$A_i(x_t) > 0.$$

Definition 3: Let a fuzzy partition of $[a, b]$ be given by basic functions $A_1, \ldots, A_n, n \geq 2$ and let $f : [a, b] \longrightarrow \mathbb{R}$ be a function that is known on a set $\{x_1, \ldots, x_T\}$ of points that is sufficiently dense with respect to the given fuzzy partition. The n-tuple of real numbers $[F_1, \ldots, F_n]$ given by

$$F_i = \frac{\sum_{t=1}^{T} f(x_t)A_i(x_t)}{\sum_{t=1}^{T} A_i(x_t)}, \qquad i = 1, \ldots, n, \tag{4}$$

is a *discrete direct F-transform* of f with respect to the given fuzzy partition. The $F_1, \ldots, F_n$ are the *components* of the (discrete) F-transform of f.

Since this paper deals with the application of the F-transform to the analysis and prediction of time series, which is a discrete problem, we will consider only the discrete fuzzy transform, and talk about the "F-transform" without explicitly specifying that it is the discrete one.

The F-transform of f with respect to $A_1, \ldots, A_n$ will be denoted by $F_n[f] = [F_1, \ldots, Fn]$. It has been proven

[12] that the components of the F-transform are weighted mean values of the original function, where the weights are determined by the basic functions.

The original function f can be approximately reconstructed from $F_n[f]$ using the following inversion formula.

Definition 4: Let $F_n[f]$ be the direct F-transform of f with respect to $A_1, \ldots, A_n \in \mathcal{F}([a, b])$. Then the function $f_{F,n}$ given on $[a, b]$ by

$$f_{F,n}(x) = \sum_{i=1}^{n} F_i A_i(x), \tag{5}$$

is called the *inverse F-transform* of f.

The inverse F-transform is a continuous function on $[a, b]$.

Let us recall the main two properties of the fuzzy transform. First, it should be stressed that for uniform fuzzy partitions the sequence of the inverse F-transform $\{f_{F,n}\}_n$ uniformly converges to the original function f for $n \to \infty$ [12]. Assuming certain additional properties, an analogous result is valid even for non-uniform fuzzy partitions [17].

Second, the F-transform components keep a certain optimality, particularly it minimizes the *piecewise integral least square criterion*.

Consequently, the direct F-transform may serve us as a discrete approximate representation of a function and may be successfully used to numerical integration of a function while the inverse F-transform is a suitable continuous approximation of a given function. For various properties of the F-transform and detailed proofs — see [12], [17], [14].

III. Evaluative Linguistic Expressions and Linguistic Description

Evaluative linguistic expressions are special expressions of natural language that are used whenever it is important to evaluate a decision situation, to specify the course of development of some process, to characterize manifestation of some property, and in many other specific situations. The expressions *very large, extremely expensive, roughly one thousand, more or less hot* are typical examples of evaluative (linguistic) expressions. Note that their importance and the potential to model their meaning mathematically have been pointed out by L. A. Zadeh (e.g., in [19], [20] and elsewhere). A formal theory of evaluative expressions is elaborated in detail in [10]. It includes a mathematical model of their semantics, which is also considered in this paper.

We will deal with simple forms of evaluative expressions with the following syntactic structure:

$$\langle \text{linguistic hedge} \rangle \langle \text{atomic evaluative expression} \rangle. \tag{6}$$

Atomic evaluative expressions comprise any of the *canonical* adjectives *small, medium, big**), abbreviated in the following as Sm, Me, Bi, respectively. It is important to stress that these words are in practice often replaced by other kinds of evaluative words, such as "thin", "thick", "old", "new", etc., depending on the context of speech.

*)In many situations, it is advantageous to extend the set of atomic evaluative expressions by the evaluative expression *zero* abbreviated as Ze.

Linguistic hedges are specific adverbs that make the meaning of the atomic expression more or less precise: we may classify hedges to those with *narrowing effects* and those with *widening effects*, see Tab I.

TABLE I
List of linguistic hedges with their abbreviations.

Narrowing effect		Widening effect	
Hedge	Abbreviation	Hedge	Abbreviation
very	Ve	*more or less*	ML
significantly	Si	*roughly*	Ro
extremely	Ex	*quite roughly*	QR
		very roughly	VR

Note that as a special case, the $\langle$linguistic hedge$\rangle$ can be empty. This enables us to identify atomic evaluative expressions with simple ones and develop a unified theory of their meaning. The evaluative expressions of the form (6) will generally be denoted by script letters $\mathcal{A}, \mathcal{B}$, etc. Note that for the sake of simplicity we have omitted numerals interpreted by fuzzy numbers from our considerations in this paper. However, this kind of linguistic evaluative expressions are not generally omitted from the theory, see [10].

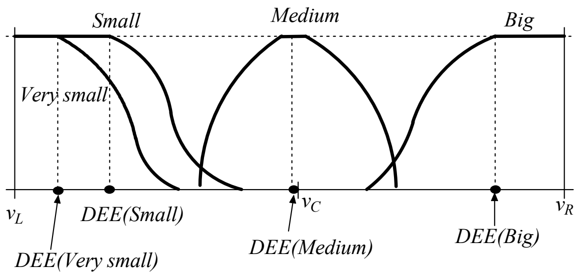

Fig. 1. Fuzzy sets that interpret intensions of some evaluative linguistic expressions.

Evaluative expressions are used to evaluate values of some variable X. The resulting expressions are called *evaluative (linguistic) predications*, and they have the form

$$X \text{ is } \mathcal{A}. \tag{7}$$

Examples of evaluative predications are "temperature is very high", "price is low", "pressure is rather strong", etc.

Our model of the meaning of evaluative expressions and predications makes distinction between *intensions* and *extensions* in various *contexts*. The mathematical representation of an intension is a function defined on a set of contexts, which assigns to each context a fuzzy set of elements. An intension leads to different truth values in various contexts, but is invariant with respect to them. An extension is a class of elements (i.e., a fuzzy set) determined by the intension when a particular context is specified. It depends on the particular context, and changes whenever the context is changed. For example, the expression "tall" is the name of an intension being a property of some feature of objects, i.e., their height. However, we can speak about the heights of various objects.

Then the meaning of "tall" can be, e.g., 30 cm when a beetle needs to climb a straw, 30 m when speaking about trees, or 200 m or more when speaking about skyscrapers, etc.

We see from the above example that in the case of evaluative expressions, the context characterizes a range of possible values. This range can be characterized by a triple of numbers $\langle v_L, v_M, v_R \rangle$, where $v_L, v_M, v_R \in \mathbb{R}$ and $v_L < v_M < v_R$[†)]. These numbers characterize the minimal, middle, and maximal values, respectively, of the evaluated characteristics (such as "height") in the specified context of use. Therefore, we will identify the notion of context with the triple $\langle v_L, v_M, v_R \rangle$. By $u \in w$ we mean $u \in [v_L, v_R]$. In the sequel, we will work with a set of contexts W

$$W \subset \{\langle v_L, v_M, v_R \rangle \mid v_L, v_M, v_R \in \mathbb{R}, v_L < v_M < v_R\}$$

that are given in advance.

The *intension* of an evaluative predication "X is $\mathcal{A}$" is a certain formula whose interpretation is a function

$$\mathrm{Int}(X \text{ is } \mathcal{A}) : W \longrightarrow \mathcal{F}(\mathbb{R}), \tag{8}$$

i.e., it is a function that assigns a fuzzy set to any context from the set W.

Given an intension (8) and a context $w \in W$, we can define the *extension* of "X is $\mathcal{A}$" in the context w as a fuzzy set

$$\mathrm{Int}(X \text{ is } \mathcal{A})(w) \underset{\sim}{\subseteq} [v_L, v_R]$$

where $\underset{\sim}{\subset}$ denotes the relation of fuzzy subsethood and v_L, v_R are the left and right bounds of the given context $w = \langle v_L, v_M, v_R \rangle$, respectively.

Evaluative predications occur in conditional clauses of natural language of the form

$$\mathcal{R} := \mathsf{IF}\ X \text{ is } \mathcal{A}\ \mathsf{THEN}\ Y \text{ is } \mathcal{B} \tag{9}$$

where $\mathcal{A}, \mathcal{B}$, are evaluative expressions. The linguistic predication "X is $\mathcal{A}$" is called the *antecedent* and "Y is $\mathcal{B}$" is called the *consequent* of the rule (9). Of course, the antecedent may consist of more evaluative predications, joined by the connective "AND". The clauses (9) will be called fuzzy/linguistic IF-THEN rules in the sequel.

The *intension of a fuzzy/linguistic IF-THEN rule* $\mathcal{R}$ in (9) is a function

$$\mathrm{Int}(\mathcal{R}) : W \times W \longrightarrow \mathcal{F}(\mathbb{R} \times \mathbb{R}). \tag{10}$$

This function assigns to each context $w \in W$ and each context $w' \in W$ a *fuzzy relation* in $w \times w'$. The latter is an *extension* of (10).

Fuzzy/linguistic IF-THEN rules are gathered in a *linguistic description*, which is a set $LD = \{\mathcal{R}_1, \dots, \mathcal{R}_m\}$ where

$$\begin{aligned} \mathcal{R}_1 &= \mathsf{IF}\ X \text{ is } \mathcal{A}_1\ \mathsf{THEN}\ Y \text{ is } \mathcal{B}_1, \\ &\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \mathcal{R}_m &= \mathsf{IF}\ X \text{ is } \mathcal{A}_m\ \mathsf{THEN}\ Y \text{ is } \mathcal{B}_m. \end{aligned} \tag{11}$$

[†)]Let us emphasize that the middle value v_M is not required to be in the exact center of the interval $[v_L, v_R]$.

Because each rule in (11) is taken as a specific *conditional sentence of natural language*, a linguistic description can be understood as a *specific kind of a (structured) text*. This text can be viewed as a *model* of specific behavior of the system in concern.

We also need to consider a linguistic phenomenon of topic-focus articulation (cf. [5], [16]), which in the case of linguistic descriptions requires us to distinguish the following two sets:

$$\begin{aligned} \mathit{Topic}_{LD} &= \{\mathrm{Int}(X \text{ is } \mathcal{A}_j) \mid j = 1, \dots, m\}, \\ \mathit{Focus}_{LD} &= \{\mathrm{Int}(Y \text{ is } \mathcal{B}_j) \mid j = 1, \dots, m\}. \end{aligned}$$

IV. Perception-based Logical Deduction

Let us describe-so called *perception-based logical deduction* (PbLD for short) which is a specific inference method. This method aims to attain intuitive behavior of fuzzy inference, i.e., it chooses these fuzzy/linguistic IF/THEN rules which would be chosen by human in a given context for a given observation.

A. Ordering of Linguistic Predications

First of all, a partial order of linguistic expressions has to be defined. Let us start with the ordering on the set of linguistic hedges. We may define the ordering $\leq_{\mathrm{H}}$ as follows:

$$\mathrm{Ex} \leq_{\mathrm{H}} \mathrm{Si} \leq_{\mathrm{H}} \mathrm{Ve} \leq_{\mathrm{H}} \langle \text{empty} \rangle \leq_{\mathrm{H}} \mathrm{ML} \leq_{\mathrm{H}} \mathrm{Ro} \leq_{\mathrm{H}} \mathrm{QR} \leq_{\mathrm{H}} \mathrm{VR}. \tag{12}$$

Let us stress that we may easily omit some of the hedges from the set of linguistic hedges or add new ones if it is required by further improvements or application requirements.

Based on $\leq_{\mathrm{H}}$ we may also define an ordering of linguistic expressions. In order to define the ordering, we have to define the following three subsets (categories) of pure linguistic expressions:

$$\mathrm{Ev}_{\mathrm{Sm}} = \{\langle \text{hedge} \rangle \mathrm{Sm}\}, \tag{13}$$
$$\mathrm{Ev}_{\mathrm{Me}} = \{\langle \text{hedge} \rangle \mathrm{Me}\}, \tag{14}$$
$$\mathrm{Ev}_{\mathrm{Bi}} = \{\langle \text{hedge} \rangle \mathrm{Bi}\}. \tag{15}$$

Then, we may define the ordering $\leq_{\mathrm{LE}}$ of evaluative linguistic expression. Let $\mathcal{A}_i, \mathcal{A}_j$ be two linguistic expressions such that $\mathcal{A}_i := \langle \text{hedge} \rangle_i \mathcal{A}'_i$ and $\mathcal{A}_j := \langle \text{hedge} \rangle_j \mathcal{A}'_j$. Then we write

$$\mathcal{A}_i \leq_{\mathrm{LE}} \mathcal{A}_j \tag{16}$$

if $\mathcal{A}'_i, \mathcal{A}'_j \in \mathrm{Ev}_{\mathrm{H}}$, $H \in \{\mathrm{Sm}, \mathrm{Me}, \mathrm{Bi}\}$ and $\langle \text{hedge} \rangle_i \leq_{\mathrm{H}} \langle \text{hedge} \rangle_j$. In other words, linguistic expressions of the same type are ordered according to their specificity (resp. generality) which is given by their hedges[*)].

It should be noted that usually the Topic_{LD} contains intensions of linguistic predications which are compound by a conjunction of more than one pure linguistic predications.

[*)]Note that as well as in the case of linguistic hedges, it will be possible to add new subsets of pure linguistic expressions in the future if necessary. It would mean that there would be additional subsets of pure linguistic expressions besides these of (13)-(15).

In other words, we usually meet the following multiple input situation

$$(X \text{ is } \mathcal{A}_i) = (X_1 \text{ is } \mathcal{A}_i^1) \textsf{ AND } \cdots \textsf{ AND } (X_K \text{ is } \mathcal{A}_i^K),$$
$$(X \text{ is } \mathcal{A}_j) = (X_1 \text{ is } \mathcal{A}_j^1) \textsf{ AND } \cdots \textsf{ AND } (X_K \text{ is } \mathcal{A}_j^K).$$

In this case, the orderings $\leq_{\mathrm{LE}}$ is preserved with respect to the components:

$$\mathcal{A}_i \leq_{\mathrm{LE}} \mathcal{A}_j \quad \text{if} \quad \mathcal{A}_i^k \leq_{\mathrm{LE}} \mathcal{A}_j^k \quad \text{for all } k = 1, \ldots, K. \tag{17}$$

Note that for all $k = 1, \ldots, K$, $\mathcal{A}_i^k$ and $\mathcal{A}_j^k$ should be in the same category of pure evaluative expressions, i.e., in one of (13)-(15), otherwise they cannot be ordered by $\leq_{\mathrm{LE}}$.

The extension of the compound linguistic predication $(\mathrm{Int}(X \text{ is } \mathcal{A}_i)(w_1, \ldots, w_K))(u_1, \ldots, u_K)$ is given by the Gödel conjunction of the intension of the pure predications forming the compound one, i.e., it equals to:

$$\bigwedge_{k=1}^{K} (\mathrm{Int}(X_k \text{ is } \mathcal{A}_i^k)(w_k))(u_k). \tag{18}$$

B. Perceptions, Deductions, Defuzzification

A perception is understood in our approach as a subset of evaluative expressions appearing in the antecedent parts of the fuzzy IF-THEN rules assigned to the given value in the given context. These rules are in some precisely defined sense optimal.

Therefore, the local perception is a mapping

$$\mathit{LPerc}^{LD} : w \times W^K \longrightarrow \mathcal{P}(\mathit{Topic}_{LD})$$

where $\mathcal{P}(\mathit{Topic}_{LD})$ denotes the potential set of Topic_{LD}.

The main principle is as follows. Given an observation $u_0 \in w \in W$, perception-selection algorithm chooses the most fired rule(s), i.e., antecedent(s) to which u_0 has the highest membership degree. If more than one such antecedent exists, it searches for the most specific one(s) for which the above introduced partial order $\leq_{\mathrm{LE}}$ has been defined.

For the final ordering $\leq_{(u_0,w)}$, we determine

$$\neg a_j = 1 - (\mathrm{Int}(X \text{ is } \mathcal{A}_j)(w))(u_0), \quad j = 1, \ldots, m$$

searching for the least one (having in mind the extension of the compound predication (18). In case of equal values, we determine the ordering $\leq_{\mathrm{LE}}$ of the expressions appearing in the components of the compound one as given by (17). It is important to keep in mind that $\leq_{(u_0,w)}$ is a *partial order*, and that especially in the case of more than one antecedent variable we often meet incomparable predications, resulting in a higher number of them selected as perceptions for a given observation.

In deduction, every element of the local perception LPerc^{LD} yields a fuzzy sets on w'. For instance, the following local perception:

$$\mathit{LPerc}^{LD}(u_0, w) \equiv \{\mathrm{Int}(X \text{ is } \mathcal{A}_{i_\ell}), \ell = 1, \ldots, L\},$$

means that all of the respective rules were fired in the same degree, say, $a_{i_\ell} \in [0, 1]$, i.e.,

$$(\mathrm{Int}(X \text{ is } \mathcal{A}_{i_\ell})(w))(u_0) = a_{i_\ell}, \quad \ell = 1, \ldots, L.$$

Then we get L fuzzy sets $C_{i_\ell}, \ell = 1, \ldots, L$ on w' such that each of them is given as follows

$$C_{i_\ell} = a_{i_\ell} \to (\mathrm{Int}(Y \text{ is } \mathcal{B}_{i_\ell})(w'))(v), \quad \text{for } v \in w' \tag{19}$$

where $\to$ is so-called *Łukasiewicz implication* defined as:

$$a \to b = \max(1 - a + b, 0) \quad a, b \in [0, 1].$$

Notice that C_{i_ℓ} in (19) is in a sense a projection of the observation u_0 through fuzzy relation

$$\mathrm{Int}(X \text{ is } \mathcal{A}_{i_\ell})(w))(u) \to (\mathrm{Int}(Y \text{ is } \mathcal{B}_{i_\ell})(w'))(v)$$

determined by the i_ℓ-th rule of the linguistic description *LD*.

Let us denote by $C \in \mathcal{F}(w')$ the intersection of all fuzzy sets deduced by r_{PbLD}, i.e.

$$C(v) = \bigwedge_{\ell=1,\ldots,L} C_{i_\ell}(v). \tag{20}$$

Then there is a defuzzification methods suggested to defuzzify fuzzy sets determined as conclusion deduced by (19). The *defuzzification of evaluative expressions* (DEE) method first classifies the intersection of inferred fuzzy sets into one of the following three classes:

$$S^- = \{C \in \mathcal{F}(w') \mid C \text{ is non-increasing}\},$$
$$S^+ = \{C \in \mathcal{F}(w') \mid C \text{ is non-decreasing}\},$$
$$\Pi = \mathcal{F}(w') \setminus (S^- \cup S^+),$$

and then it is given as follows

$$\mathrm{DEE}(C) = \begin{cases} \mathrm{LOM}(C) & \text{if} \quad C \in S^-, \\ \mathrm{MOM}(C) & \text{if} \quad C \in \Pi, \\ \mathrm{FOM}(C) & \text{if} \quad C \in S^+ \end{cases}$$

where $\mathrm{LOM}, \mathrm{MOM}, \mathrm{FOM}$ are the well-known defuzzifications *last of maxima*, *mean of maxima*, *first of maxima*, respectively.

Remark 5: It should be stressed that neither the suggested ordering approach for determination of the local perception nor the DEE defuzzification are the only applicable possibilities. Generally, we may consider whole class of, say, PbLD-like methods which differ by the perception procedure function. Analogously, distinct defuzzification methods can be taken into account.

The idea of assigning local perceptions is not restricted only to the topic. If we generalize it slightly, we can learn the linguistic description on the basis of the given data. More details about this method can be found in [3]. Let us remark that we have successfully implemented this method in the software system LFLC2000 (see [4]).

V. Time Series Analysis

A. Time Series Decomposition

Let

$$\{x_t \mid t = 1, \ldots, T\} \subset \mathbb{R}, \quad T \geq 3 \tag{21}$$

be a given time series. The task is to analyze it and to forecast its future development, i.e., to determine the values

$$\{x_t \mid t = T+1, \dots, T+\gamma\} \subset \mathbb{R}, \quad \gamma \geq 1. \tag{22}$$

The main idea of the decomposition model is to decompose each element x_t into the following components:

$$x_t = Tr_t + S_t + C_t + E_t \tag{23}$$

where Tr_t, S_t, C_t, E_t, $t = 1, \dots, T$ are the *trend, seasonal, cyclic* and *error* components of the time series, respectively.

Trend and seasonal components may be analyzed. The cyclic component C_t is a bit problematic. The name "cyclic" comes from the economic cycles, which are not regular and are dependent on many external factors to be analyzed from the past. The error component is a random noise that essentially cannot be forecast, and therefore is omitted from our further considerations. Therefore, the following simplified decomposed model is considered for further investigation:

$$x_t = Tr_t + S_t. \tag{24}$$

The traditional approach to the decomposition assumes the trend to be an a priori given function, e.g., linear, polynomial, exponential or a kind of a saturation function such as sigmoidal function. This approach simplifies the analysis, which consists of a regressive determination of parameters of the predetermined function, as well as the forecast, which is a simple prolongation, i.e., an evaluation of the determined trend function at time points $T+1, \dots, T+\gamma$.

Such an approach, however, is too restrictive and not always the most appropriate. The course of the trend can vary, especially in a case of a long time series, its forecasting is very difficult. Typical examples are equity indexes, where we cannot usually prolong the trend in a simple way because robust growth is often followed by dramatic fall, which can be followed by stagnation and then again by growth. This is precisely due to the influence of the cyclic component. Here, prolongation might in some cases be the worst thing to be applied in forecasting. For such cases, complicated adaptive trend changing models or models with changes in regime [6] are constructed.

Generally, we speak about the so called *trend-cycles*. For their estimation, we propose to use the F-transform method because it does not fix any shape of the curve and it has powerful approximation and noise reduction properties [14].

The time series $\{x_t \mid t = 1, \dots, T\}$ may be viewed as a function x defined on the interval $[0, T]$, which is not given analytically. Instead, measurements $x(t) = x_t$ at points $t = 1, \dots, T$ are provided.

Let us build a uniform fuzzy partition according to Definition 1 such that each of the basic functions $A_2, \dots, A_{n-1}$ "covers" the number of nodes equal to the number of nodes belonging to a season. For example, in the case of a time series on the monthly basis, each basic function covers 12 points, with the exception of A_1, A_n, which cover the first and the last 6 points, respectively. Consequently, the set of points x_t is sufficiently dense with respect to the fuzzy partition. From this point forward, we will consider a time series on the monthly basis since everything may be easily generalized for other cases.

Let

$$F_n[x] = [X_1, \dots, X_n] \tag{25}$$

be an F-transform of the function x w.r.t. the given fuzzy partition, and let $x_{F,n}$ be its inverse F-transform. The inverse F-transform will serve as a model of the trend-cycle. Recall that the shape of the trend-cycle function is not fixed a priori, which would enable us to capture the trend-cycle in a more realistic way.

Remark 6: Let us note, that the fact that the transparency plays an essential role throughout the whole methodology influenced also the above choice of constructing basic functions in order to cover one year of monthly time series. From this point of view, the F-transform components are easily interpretable as average year values. Therefore, technically possible fuzzy partition where one basic functions covers, e.g, 14 values makes no sense. However, further natural fuzzy partition, e.g., with basic functions covering 24 values may make sense and sometimes even improve results.

If we omit the error component E_t, then the seasonal component S_t from (23) can be obtained using the formula

$$S_t = x_t - x_{F,n}(t), \tag{26}$$

where $x_{F,n}(t) = Tr_t + C_t$. The trend-cycle may be further analyzed and described using autoregressive fuzzy rules; see Subsection VI-A.

It should be stressed here that the suggested approach is an alternative to the *model with changes in regime* [6] (also called the *regime switching model*), which is, unlike our approach, based on the theory of random processes and Markov chains. Our motivation is to obtain a transparent description in natural language.

VI. Time Series Forecast

In this section, we will describe the forecasting of time series on the basis of the analysis described above (see Section V).

A. Trend-Cycle Forecast

The classical approaches first model the trend only and then determine the seasonal components that are influenced by the cyclic irregular changes. Our approach treats the problem the other way around; the trend-cycle model $x_{F,n}$ primarily serves us to get pure seasonal components without the cyclic influences.

On the other hand, we cannot easily forecast such a trend-cycle model by the prolongation, i.e., by the evaluation of the predetermined fixed trend function at points $t = T+1, \dots, T+\gamma$. Due to the drawbacks of this traditional approach, this is not a disadvantage but an advantage, as will be explained below.

We follow the idea of [13], and for the trend-cycle forecast, we employ the perception-based logical deduction. As antecedent variables, we consider the F-transform components of the given time series X_i, $i = 1, \ldots, n-1$ as well as their first- and second-order differences:

$$\Delta X_i = X_i - X_{i-1}, \qquad i = 2, \ldots, n-1$$
$$\Delta^2 X_i = \Delta X_i - \Delta X_{i-1}, \qquad i = 3, \ldots, n-1$$

respectively.

The differences of the F-transform components expressing the time series trend-cycle of distinct orders are able to describe the dynamics of the time series better than the F-transform components themselves. Furthermore, due to the use of the differences, the time series does neither have to be de-trended as in case of the classical autoregressive approach using, e.g., ARMA model (1), nor integrated model has to be used as in case of ARIMA.

Fuzzy rules may describe logical dependencies of the trend-cycle changes (hidden cycle influences), which is highly desirable and suggested in comparison with the standard prolongation of the trend-cycle observed in the past. The advantage of the transparently interpretable form of the rules using fragments of natural language is unquestionable. This might be helpful in better understanding the functionalities and motive factors determining the changes in a process yielding the times series in question.

We already mentioned in Section I-B that fuzzy/linguistic IF-THEN rules can be understood as a description of an autoregressive process. Every rule describes the local dependence of X_{i+1} on previous values of X_i, X_{i-1} etc. expressed in the form of differences ΔX_{i-1}, $\Delta^2 X_{i-1}$ and the like. Perception-based logical deduction algorithm described in Section IV then chooses the best local rule or rules with respect to a given situation.

Let us mention that fuzzy rules are automatically generated by the *linguistic learning* algorithm [3] implemented in the software package LFLC 2000 [4] from the F-transform components of the time series and their differences.

Remark 7: Though the fuzzy/linguistic IF-THEN rules are also generated from the past, the suggested approach can learn, describe, and successfully predict the future of equity indexes mentioned as a motivation example at the beginning of Subsection V-A. Of course, this is possible, if similar progress has been observed and measured in the past. The prolongation of a trend function is generally not able to perform this task successfully.

B. More Steps Ahead Forecasts with Independent Models

Let us now consider the F-transform components (25) of the given time series. Fuzzy/linguistic rules and the perception-based logical deduction can be used for forecasting the next F-transform components

$$X_{n+1}, \ldots, X_{n+\zeta} \tag{27}$$

from which the trend-cycle of the time series can be determined as values of the inverse F-transform $x_{F,n+\zeta}(T+1), \ldots, x_{F,n+\zeta}(T+\gamma)$, where $\zeta < \gamma$.

It is difficult to forecast the course of the time series for a long time interval. First of all, note that the F-transform components X_1, X_n related to the first and the last basic functions are *singular* because they can be depreciated. This is caused by the fact that the corresponding basic function of X_1 as well as X_n is only a half of the regular one. Therefore, even the last F-transform component X_n, which may be otherwise calculated from the given data, is forecasted.

There are two principal ways to forecast the F-transform components:

(i) forecast the next component on the basis of the previous n components (or a subset) and their corresponding first and second differences;
(ii) forecast some of the following components (not necessarily the immediate next one) from some of the components (25) and their first and second differences.

In case (i), we consider the components $X_1, \ldots, X_{n-1}$ and their differences $\Delta X_2, \ldots, \Delta X_{n-1}$ and $\Delta^2 X_3, \ldots, \Delta^2 X_{n-1}$ and forecast the component X_n. Then, using the same linguistic description, we forecast X_{n+1} from $X_1, \ldots, X_n$, $\Delta X_2, \ldots, \Delta X_n$, $\Delta^2 X_3, \ldots, \Delta^2 X_n$, etc. Obviously, there is a danger of propagation of forecast errors, since we forecast from forecasted values. The longer the prediction term, the higher the damage.

Case (ii) overcomes this problem because we build a finite number of independent trend-cycle forecasting linguistic descriptions (models) using the technique described in Subsection VI-A. Such linguistic descriptions deal with consequent variables X_{n+j} or with consequent variables ΔX_{n+j}, for $j = 0, 1, 2, \ldots$. Each linguistic description is generated by the linguistic learning algorithm and each linguistic description may be used to forecast j-steps ahead.

On the basis of the forecasted F-transform components (27) we can compute the forecasted trend-cycle of the time series, where the latter consists of the values of the inverse F-transform:

$$x_{F,n}(T+1), \ldots, x_{F,n+\ell}(T+k).$$

C. Forecasting Seasonal Component

The seasonal components are forecasted as follows. Let

$$\mathbf{S}_\xi = [S_{p\cdot(\xi-1)}, S_{p\cdot(\xi-1)+1}, S_{p\cdot(\xi-1)+2}, \ldots, S_{p\cdot(\xi-1)+p-1}] \tag{28}$$

be the ξ-th vector of the seasonal components (26), where p denotes the seasonality period, i.e., in our case of a time series on a monthly basis, $p = 12$ and $\mathbf{S}_\xi$ is the vector of the January, February etc. to December measurements of the ξ-th year.

The assumption of stationarity of the seasonal component of the time series is considered, i.e., we assume that $\mathbf{S}_\xi$ is a linear combination of previous θ vectors $\mathbf{S}_{\xi-\theta}, \ldots, \mathbf{S}_{\xi-1}$. This means that we generate the following system of equations:

$$\mathbf{S}_\xi = \sum_{j=1}^{\theta} d_j \cdot \mathbf{S}_{\xi-j}, \quad \xi > \theta, \tag{29}$$

and search for its optimal solution with respect to the coefficients $d_1, \ldots, d_\theta$. The computed coefficients are then used to determine the $\mathbf{S}_\xi$. Let us mention that the stationarity is a standard assumption that can easily be checked, see [2].

The last step to get the overall time series forecast is composition of both forecasts, i.e., of the forecasted trend-cycle and the seasonal components. This is done inversely to the original decomposition which was either additive or multiplicative.

D. Optimization

There are some unknown parameters in the whole procedure that have to be determined individually for every single time series. Basically, these are the antecedent variables for the prediction of the F-transform components, the number of the antecedent variables, and the parameter θ from the previous subsection.

The time series is divided into a *learning set* and a *validation set* in such a way that the latter is given by the last values of the time series of the length equal to the forecast horizon. This means that the learning set is $\{x_1, \ldots, x_{T-\gamma-1}\}$ and the validation set is $\{x_{T-\gamma}, \ldots, x_T\}$ (provided that we need to forecast k values).

All the possible combinations of the antecedent variables up to the maximal number combined with seasonal components are determined. For these computations, only the learning set is used.

All the computed models are used to forecast $\{x'_{T-\gamma}, \ldots, x'_T\}$ and these forecasts are compared with the validation set $\{x_{T-\gamma}, \ldots, x_T\}$. As a suggestion to the user, all the models are ordered according to a pre-specified error criterion which may be, in general, arbitrary. The user may then employ any of the optimized and tuned models for forecasting the values $x_{T+1}, \ldots, x_{T+\gamma}$.

VII. Demonstration

A. Comparison Study

As mentioned in the introduction, the majority of fuzzy approaches to the time series analysis employ Takagi-Sugeno rules, distinct neuro-fuzzy systems or evolving fuzzy systems [1], [7], [8], [15], which are powerful and robust but less interpretable.

Therefore, we decided to provide a comparison study of our results with the results obtained by the ForecastPro® business software. The reason is that it is a standard professional package which employs nearly all the existing Box-Jenkins methods as well as exponential smoothing models, simple naïve methods etc. It also contains an expert settings which chooses the best method for every single time series. Consequently, we do not compare our single approach with one or two or more standard approaches. We compare our methodology with always the best standard method among the many ones usual user is provided with!

Moreover, we wish to emphasize that we found the interpretability of the whole approach crucial and so, we deal with fuzzy/linguistic IF-THEN rules with linguistically specified antecedent as well as consequents only. On the other hand, increase of the interpretability of a model should not dramatically decrease its precision in forecasting. Therefore, the comparison cannot be restricted to computational intelligence or even to fuzzy approaches. One would never choose a fuzzy approach for its advantages if it provided significantly worse results. This even strengthens the choice of the ForecastPro® package for the comparison.

TABLE II

SMAPE ERRORS

Time series	Linguistic approach	Forecast Pro®		
		Expert	Box-J	Winters
Passengers	6.941%	1.949%	1.949%	4.510%
Pigs	5.677%	5.969%	5.969%	7.332%
Cars	10.505%	12.722%	12.722%	9.666%
Average	7.708%	6.880%	6.880%	7.170%

The comparison has been be made on 3 monthly real time series suggested by the organizers of the *WCCI 2010* special session titled *Computational Intelligence in Forecasting*. These are the *Passengers* time series [2] containing the information about the number (in thousands) of passengers of international airlines (Jan'49-Dec'60); *Pigs* time series contains numbers of pigs slaughtered in Victoria (Jan'80-Aug'95) and *Cars* time series consisting of car sales in Quebec ('60-'68). All the time series are provided by the *Rob J Hyndman Time Series Data Library* [21].

The out samples used for testing the precision of the forecast were comprised of last 19 values of the Passengers time series and last 12 values of the latter two time series.

Forecasted values were compared with the out samples and the SMAPE error measure was computed. For the results, we refer to Table II. It can be seen, that in some cases, the suggested approach even outperformed the business software with classical methods. However, the average precision was higher in case of the standard methods provided by the commercial software. Generally, it may be stated that the results are fully comparable.

Besides the precision, there was another additional value brought by the suggested linguistic approach - the transparency and the interpretability. To demonstrate we provide readers with the linguistic description of winning prediction model for the Pigs time series.

For the sake of completeness, let us mention that for this winning model, fuzzy partition with basic functions covering 24 values. Figure 2 displays the time series including out samples and the forecasted values. Zoom-in forecasted values and out samples is displayed on Figure 3.

VIII. Conclusions

We have introduced a novel linguistic approach to analysis and forecast of time series. The approach combines the aspects two classical approaches - of the decomposition as well as of the autoregression. We have emphasized the linguistic nature of the suggested methodology motivated by the general requirement to have interpretable and transparent models. Besides the linguistic nature, we have demonstrated

TABLE III

FUZZY RULES GENERATED FOR THE DESCRIPTION AND PREDICTION OF PIGS TIME SERIES.

Nr.	Antecedents			Consequent
	X_i	X_{i-1}	$\Rightarrow$	X_{i+1}
1	Ro Bi	Me	$\Rightarrow$	QR Bi
2	Ro Bi	Ro Bi	$\Rightarrow$	QR Bi
3	ML Bi	Ro Bi	$\Rightarrow$	Bi
4	Ex Bi	ML Bi	$\Rightarrow$	Bi
5	Ex Bi	Ex Bi	$\Rightarrow$	QR Sm
6	Me	Ex Bi	$\Rightarrow$	Ze
7	Ze	Me	$\Rightarrow$	Si Sm
8	ML Sm	Ze	$\Rightarrow$	Ex Sm
9	Sm	ML Sm	$\Rightarrow$	ML Sm
10	QR Sm	Sm	$\Rightarrow$	VR Sm
11	ML Me	QR Sm	$\Rightarrow$	Ro Bi
12	ML Bi	ML Me	$\Rightarrow$	ML Me
13	QR Bi	ML Bi	$\Rightarrow$	QR Bi
14	ML Bi	QR Bi	$\Rightarrow$	Ro Bi

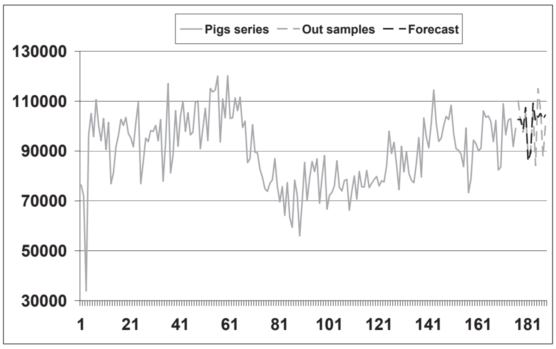

Fig. 2. Pigs time series including out samples and the forecasted values.

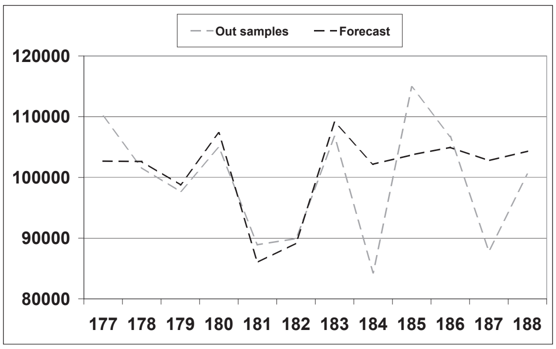

Fig. 3. Forecasted vales of the Pigs time series compared to the out sample values.

on the comparison study, that the approach does not lack precision and is comparable with the best standard methods.

ACKNOWLEDGMENT

We gratefully acknowledge partial support of the projects 1M0572 and MSM6198898701 of the MŠMT ČR.

REFERENCES

[1] Aznarte, J., Benítez, J. and Castro, J., "Smooth transition autoregressive models and fuzzy rule-based systems: Functional equivalence and consequences," *Fuzzy Sets and Systems*, vol. 158, pp. 2734-2745, 2007.

[2] Box, G. and Jenkins, G., *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day, 1976.

[3] Bělohlávek, R. and Novák, V., "Learning Rule Base of the Linguistic Expert Systems," *Soft Computing*, vol. 7, pp. 79–88, 2002.

[4] Dvořák, A., Habiballa, H., Novák, V. and Pavliska, V., "The software package LFLC 2000 - its specificity, recent and perspective applications," *Computers in Industry*, vol. 51, 269–280, 2003.

[5] Hajičová, E., Partee, B. and Sgall, P., *Topic-focus Articulation, Tripartite Structures, and Semantic Content*, Dordrecht: Kluwer, 1998.

[6] Hamilton, J.D., *Time Series Analysis*. New Jersey: Princeton University Press, 1994.

[7] Kasabov, N. and Song, Q., "DENFIS: Dynamic Evolving Neural-Fuzzy Inference System and Its Application for Time-Series Prediction," *IEEE Transactions on Fuzzy Systems*, vol. 10, pp. 144–154, 2002.

[8] Leng, G., McGinnity, T. and Prasad, G., "An approach for on-line extraction of fuzzy rules using a self-organising fuzzy neural network," *Fuzzy Sets and Systems*, vol. 150, pp. 211–243, 2005.

[9] Novák, V., "Perception-Based Logical Deduction," in *Computational Intelligence, Theory and Applications (Advances in Soft Computing)*, Edited by B. Reusch, Berlin: Springer, pp. 237–250, 2005.

[10] Novák, V., "A Comprehensive Theory of Trichotomous Evaluative Linguistic Expressions," *Fuzzy Sets and Systems*, vol. 159, pp. 2939-2969, 2008.

[11] Novák, V. and Perfilieva, I., "On the Semantics of Perception-Based Fuzzy Logic Deduction," *International Journal of Intelligent Systems*, vol. 19, 1007–1031, 2004.

[12] Perfilieva, I., "Fuzzy Transforms: theory and applications," *Fuzzy Sets and Systems*, vol. 157, pp. 993–1023, 2006.

[13] Perfilieva, I., Novák, V., Pavliska, V., Dvořák, A. and Štěpnička, M., "Analysis and Prediction of Time Series Using Fuzzy Transform," *Proc. IEEE World Congress on Computational Intelligence*, Hong Kong, 2008, pp. 3875–3879.

[14] Perfilieva, I. and Valášek, R., "Fuzzy Transforms in Removing Noise," in: *Computational Intelligence, Theory and Applications (Advances in Soft Computing)*, Edited by B. Reusch, Berlin: Springer, pp. 221–230. 2005.

[15] Rong, H.J., Sundararajan, N., Huang, G.B. and Saratchandran, P., "Sequential Adaptive Fuzzy Inference System (SAFIS) for nonlinear system identification and prediction," *Fuzzy Sets and Systems*, vol. 157, pp. 1260–127, 2006.

[16] Sgall, P., Hajičová, E., Panevová, J. and Mey, J., *The meaning of the sentence in its semantic and pragmatic aspects*, Boston: Kluwer, 1986.

[17] Štěpnička, M. and Polakovič, O., "A neural network approach to the fuzzy transform," *Fuzzy Sets and Systems*, vol. 160, pp. 1037–1047, 2009.

[18] Takagi, T. and Sugeno, M., "Fuzzy identification of systems and its applications to modeling and control," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 15, pp. 116–132, 1985.

[19] Zadeh, L.A., " Outline of a new approach to the analysis of complex systems and decision processes," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. SMC-3, pp. 28–44, 1973.

[20] Zadeh, L.A., "The concept of a linguistic variable and its application to approximate reasoning I, II, III," *Information Sciences*, vol. 8-9, pp. 199–257, pp. 301–357, pp. 43–80, 1975.

[21] Hyndman, R.J.(n.d.),"Time series data library," *http://www.robjhyndman.com/TSDL/*.