

Fuzzy vs. Probabilistic Techniques in Time Series Analysis

Vilém Novák^(✉)

Institute for Research and Applications of Fuzzy Modeling, NSC IT4Innovations,
University of Ostrava, 30. dubna 22, 701 03 Ostrava 1, Czech Republic
Vilem.Novak@osu.cz

Abstract. In this paper, we discuss the difference between probabilistic and fuzzy techniques used in time series analysis. First, we focus on the fundamental difference between vagueness and uncertainty phenomena. Then we briefly describe probabilistic view of time series. In the main part, we demonstrate how special fuzzy techniques, namely the fuzzy natural logic and fuzzy transform can be applied in the analysis of time series and what is their outcome in comparison with the probabilistic approach. We argue that fuzzy techniques enable to obtain knowledge that is either more difficult or impossible to obtain using probabilistic techniques.

Keywords: Vagueness · Fuzzy transform · F-transform
Fuzzy natural logic · Components of time series

1 Introduction

In this section, we will demonstrate that probabilistic and fuzzy techniques are based on modeling of different phenomena, namely vagueness and uncertainty. Both phenomena are usually present and require different mathematical principles. Hence, in the reality both kinds of techniques are complementary rather than competitive.

2 Uncertainty and Vagueness

Two phenomena whose importance in science raised especially in 20th century are *uncertainty* and *vagueness* (cf. [2, 28]). Both of them characterize situations in which the amount, character and extent of knowledge we have at disposal is essential. It is important to stress that both uncertainty as well as vagueness form two complementary facets of a more general phenomenon called *indeterminacy*¹. In the reality, we often meet indeterminacy *with both its facets* present, i.e., *vague* phenomena can be at the same time also *uncertain*.

¹ This phenomenon is sometimes called “uncertainty in wider sense”.

2.1 Potentiality and Uncertainty

When observing the surrounding world, we encounter events of two kinds: those that already occurred and *potential* ones that can, but need not, occur. For example, consider a company producing tires. We know that today it produced, say 300 of them. But the number of tires produced the next day is not known. We may expect production of, e.g., 350 tires but the concrete number is uncertain because, for example, technical or personal problems on the production line may appear. From it follows that the *uncertainty phenomenon* emerges when there is a *lack of knowledge* about *occurrence* of some *event* (e.g., the production of tires). In general, we may state that uncertainty is encountered when a certain kind of experiment (process, test, etc.) is to proceed, the result of which is not known to us. It may refer to variety of potential outcomes, ways of solution, choices, etc.

Specific form of uncertainty is *randomness* which is uncertainty raising in connection with time. There is no randomness (uncertainty) after the experiment was realized (the event has occurred) and the result is known to us. Note that it is connected with the question whether a given event may be regarded within some time period, or not. This becomes apparent on the typical example with tossing a player's cube. The phenomenon to occur is *the number of dots on the cube* and it occurs after the experiment (i.e. tossing the cube one times) has been realized. Thus, we refer here to the future, to events that are *potential*; not yet existing.

Let us remark, however, that the variety of potential events may raise even a more abstract uncertainty that is less dependent on time. We may, for example, analyze uncertainty in potentiality (that is, lack of knowledge) without necessary reference to time, or with reference to the past (such as a posterior Bayesian probability).

The mathematical model (i.e. quantified characterization) of the uncertainty phenomenon is provided especially by the *probability theory*. In everyday terminology, probability can be thought of as a numerical measure of the likelihood that a particular event will occur. There are also other mathematical theories addressing the mentioned abstract uncertainty, for example possibility theory, belief measures and others.

2.2 Vagueness and Actuality

The *vagueness phenomenon* raises when we try to *group* together all objects that have a certain property φ . The result is a grouping of objects

$$X = \{o \mid o \text{ is an object having the property } \varphi\}. \quad (1)$$

We see the grouping X as one object consisting of objects o that all are at our disposal at once because we have already grouped them together. We say that X is *actualized*.

In general, however, X *cannot be taken as a set* since the property φ may be of such a character that when checking that a given object \hat{o} has the property φ ,

we hardly obtain a definite answer. For example, consider the property $\varphi =$ ‘to be expensive’ and let the total amount of money we have at disposal for all our expenses be 50,000 \$. Let o_1 be a car for 20,000 \$, o_2 a car for 48,000 \$, and o_3 a car for 35,000 \$. Then o_1 is not expensive at all, i.e., $\varphi(o_1)$ is false and $\varphi(o_2)$ is true. But what about $\varphi(o_3)$? This car is not really expensive but also not too cheap. Hence, we cannot say that the grouping X in (1) is a set because a set is formed only of objects that we *unambiguously* know that they have the property φ . Hence, we say that φ is *vague*. There can exist *borderline* elements o for which it is unclear whether they have the property φ (and thus, whether they belong to X), or not. On the other hand, it is always possible to characterize, at least some *typical objects* (prototypes), i.e. objects having typically the property in concern. For example, everybody can show a “blue sweater” or “huge building”, “expensive car” but it is impossible to show “all expensive cars”.

Vagueness is opposite to exactness and we argue that it cannot be avoided in the human way of regarding the world. Any attempt to explain an extensive detailed description necessarily leads to using vague concepts since precise description can contain such abundant number of details that we will be lost when learning all of them. To understand it, we must group them together — and this can hardly be done precisely. This idea was formulated by Zadeh in [30] as the *incompatibility principle*. The problem consists in the way how people regard the phenomena around them. This would be impossible without presence of vagueness.

The (so far) best mathematical concept that can be used to model vague groupings is that of a *fuzzy set*. Formally, a fuzzy set A is a function

$$A : U \longrightarrow L$$

where U is some universal set containing all the elements (objects) that may be considered to fall into the considered vague grouping and L is a set of *membership degrees* which is a special lattice. The function A is also called the membership function. Note that the fuzzy set is *identified* with its membership function. Sometimes we use the symbol $A \underset{\sim}{\subseteq} U$ to emphasize that A is a fuzzy set in the universe U . The value $A(x) \in L$ for any $x \in U$ is called the *membership degree* of the element x in A .

2.3 Actuality vs. Potentiality

In the discussion above we touched two phenomena: actuality and potentiality. A classical set is always understood as being *actual*², i.e. we take all its elements as already existing and at our disposal in one moment. Therefore, our reasoning about any set stems from the assumption that it is at our disposal as a whole. Of course, when a set is infinite then only God is able to see it as a whole while we can see only a part of it. It should be emphasized that the set theory (and so, the modern mathematics) can deal with *actualized sets* only!

² Cf. the analysis by Vopěnka in [29].

On the other hand, most events around us are only *potential*, i.e. they may, but need not, occur or happen. Thus, to create a grouping of objects, we may have only a method how a new element can be created but all of them will never exist together. For example, if a machine has on its input one piece of metal, then it can produce various products of it but only one will actually be finished. It is even impossible to imagine all products produced by the machine from one piece of metal. Note that the same we observe at the company producing tires considered above. In one day it produces from the given amount of material only one number of tires.

As already mentioned, there are two kinds of events: those that already happened and those that have not yet happened. We know the first ones because they are at our disposal and we know that they have a given property φ . However, we do not know the second ones and we even do not know whether some new events having the property φ will occur or not. We encounter uncertainty; we speculate about the whole X (1), but only part of it indeed exists. But as noted, mathematical description of X is possible only if it is actualized. The only solution thus is to imagine all (or, at least some) still not existing elements of X as existing. The “added” part may be, or may be not, possible to happen but we search for methods providing us with the estimation of the information about their possible occurrence.

For example, we can imagine all dots on a dice that can be tossed, i.e., we imagine the tossed numbers $X = \{1, \dots, 6\}$ as already existing (though they cannot be tossed all of them together). For example, let the numbers $\{1, 3, 5\}$ be already tossed. Then they already exist (this is the actualized part of X) and now we may try to guess whether another number will indeed be tossed (i.e., whether the given element of X will indeed occur). The measure of information about such possibility is modeled using the *probability theory*. As a mathematical theory, however, it works with the whole X , i.e., the problem that X is not yet created is disregarded.

Note that the vagueness phenomenon is not related to occurrence of whatever event. It concerns the question *how* is the given grouping X formed, i.e., what is the character of the property φ in (1) determining it. If for any object, either $\varphi(o)$ holds or not then φ is sharp. If it allows borderline cases then it is vague. Vagueness applies to an *actualized* non-sharply delineated grouping. Once an actualized (i.e. already existing) grouping of objects X is at our disposal, we may speak about *truth* of the fact that an object o has the property φ ; that is, we *know the truth of* $o \in X$.

In probability theory we introduce the concept of a probabilistic space $\langle \Omega, \mathcal{A}, P \rangle$, where Ω is a set of elementary random events, \mathcal{A} is a σ -algebra of subsets of Ω and $P : \mathcal{A} \rightarrow [0, 1]$ is a probabilistic measure. With respect to the discussion above, Ω is a sharp grouping of objects that is *actualized*. Moreover, we deal with the actualized set (σ -algebra) \mathcal{A} of subsets of Ω . Any element $Y \in \mathcal{A}$ is a mathematical model of an event that may, or may not occur. From the mathematical point of view, in fact, Y already exists but we pretend that

\mathcal{A} it is only potential and take P as the measure of information about possible occurrence of Y .

3 Probabilistic View on Time Series

The mathematical model of time series is based on the assumption that a probabilistic space $\langle \Omega, \mathcal{A}, P \rangle$ is given. A time series is then a stochastic process (see [1, 7])

$$X : \mathbb{T} \times \Omega \longrightarrow \mathbb{R} \tag{2}$$

where \mathbb{T} is a set of time moments. In general, it can be $\mathbb{T} = [a, b] \subset \mathbb{R}$ but in economy and elsewhere we usually take $\mathbb{T} = \{1, \dots, p\} \subset \mathbb{N}$ being a finite set of natural numbers. These are usually construed as hours, days, weeks, months, or years. Instead of the general form (2) we usually write time series as a system of random variables

$$\{X(t) \mid t \in \mathbb{T}\} \tag{3}$$

where each $X(t)$ is a random variable $X(t) : \Omega \longrightarrow \mathbb{R}$, $t \in \mathbb{T}$, i.e., it is a measurable function w.r.t. Borel sets on \mathbb{R} and \mathcal{A} . This enables us to define a function

$$F_t(x) = P\{\omega \in \Omega \mid X(t)(\omega) < x\}.$$

called the *distribution function*, which characterizes the probability distribution of values of the random variable $X(t)$. More generally, we may consider a multidimensional distribution function

$$F_{t_1, \dots, t_n}(x_1, \dots, x_n) = P\{\omega \in \Omega \mid X(t_1)(\omega) < x_1, \dots, X(t_n)(\omega) < x_n\}, \tag{4}$$

where $t_1, \dots, t_n \in \mathbb{T}$. When speaking about time series, we will usually write it simply as X without marking the time variable t .

This model assumes existence of a distribution function of each $X(t)$, $t \in \mathbb{T}$, or a joint distribution function (4) of a finite set of them. Let us realize that by this model, the time series is considered to be a sequence of values being measurements of outcomes of some real process that proceeds in time. We do not know which outcome really occurs but we assume to have information about probability of its occurrence. Such information, however, is very rough and does not enable us to penetrate into the substance of the considered process.

Something more we can learn from the following characteristics.

(a) *Mean value* of the time series:

$$\mathbf{E}(X(t)) = \int_{\mathbb{R}} x dF_t(x). \tag{5}$$

(b) *Covariance function* of the time series:

$$R(s, t) = \mathbf{E}((X(s) - \mathbf{E}(X(s)))(X(t) - \mathbf{E}(X(t)))). \tag{6}$$

Additional used characteristics is *variance*

$$\mathbf{D}(X(t)) = \int_{\mathbb{R}} [x - \mathbf{E}(X(t))]^2 dF_t(x).$$

The behavior of these characteristics gives rise to specific kinds of time series. We say that the time series is *strictly stationary* if

$$F_{t_1+h, \dots, t_n+h}(x_1, \dots, x_n) = F_{t_1, \dots, t_n}(x_1, \dots, x_n) \tag{7}$$

holds for all $t_1, \dots, t_n \in \mathbb{T}$ and $h \in \mathbb{R}$ such that $t_1+h, \dots, t_n+h \in \mathbb{T}$. This means that the joint probability distribution does not depend on time. Such time series behaves in a dully uniform way.

We say that the time series is *weak-sense stationary* if the following holds for all $t, s \in \mathbb{T}$:

- (i) $\mathbf{E}(X(t)) = \mu,$
- (ii) $R(s, t) = R(t - s).$

This means that the mean value remains the same independently on time and the covariance function is determined by the distance between time moments but not on the position in time.

It is important to emphasize that if we fix $\omega \in \Omega$ then the time series (1) becomes an ordinary function $X : \mathbb{T} \rightarrow \mathbb{R}$. We call it *realization* of the time series. Note that in practice, we always have *one* realization at disposal only. This fact, however, makes the assumption (3) not fully sound. In extreme case it means that we derive conclusions about time series in a given time moment on one measurement only. But this contradicts the basic assumptions of the probability theory, especially the mass scale, i.e., that its predictions are the more reliable the more measurements of a given random variable are at disposal. We are thus implicitly forced to assume that the real process does not (significantly) change during the time, i.e., whenever we measure its outcome, we measure the same random variable more or less independently on time.

Probabilistic methods, however, led to amazingly well working methods for analysis and prediction of time series. The best known is the *autoregressive moving-average model* ARMA(p, q) (also referred to as Box-Jenkins model) whose general formula is the following:

$$X(t) = \alpha_1 X(t - 1) + \dots + \alpha_p X(t - p) + Z(t) + \beta_1 Z(t - 1) + \dots + \beta_q Z(t - q) \tag{8}$$

where $\{Z(t) \mid t \in \mathbb{T}\}$ is a simple strictly stationary time series with zero mean value and bounded variance. The α_i are autoregressive coefficients and β_j are moving-average coefficients. This model, however, assumes that the time series is stationary, which is rarely the case. In practice, trends and periodicity exists in many datasets, so there is a need to remove these effects before applying such models. This is the fertile ground for application of fuzzy techniques to the analysis of time series.

Let us mention one more important concept, namely the *periodogram*. This is a function of frequencies

$$I(\lambda) = \frac{1}{2\pi N} \left| \sum_{t=1}^N X(t)e^{-it\lambda} \right|^2, \quad -\pi \leq \lambda \leq \pi. \tag{9}$$

This function makes it possible to identify distinguished frequencies contained in the time series X . Using the well known formula $T = \frac{2\pi}{\lambda}$ we can compute characteristic periodicities in X .

4 Fuzzy Techniques for Time Series Analysis

In this section we will describe basic techniques that are based on the concept of a fuzzy set and that turned out to be very useful in the analysis and prediction of time series. We will very briefly describe the main concepts. More details can be found in the book [24] and the other cited literature.

4.1 Fuzzy Transform

The fuzzy (F-)transform is a universal technique introduced by Perfilieva in [26, 27] that has many kinds of applications. Its fundamental idea is to map a bounded continuous function $f : [a, b] \rightarrow \mathbb{R}$ to a finite vector of numbers and then to transform it back. The former is called a *direct F-transform* and the latter an *inverse one*. The result of the inverse F-transform is a function \hat{f} that *approximates* the original function f . The advantage of this approach consists in the possibility to set the parameters of the F-transform in such a way that the approximating function \hat{f} has desired properties.

The power of the F-transform stems from its approximation abilities, from its ability to filter out high frequencies and from the ability to reduce noise [14, 15, 25]. Another outcome is the ability to estimate values of first and second derivatives in an area given approximately (cf. [11]).

4.1.1 Fuzzy Partition

The first step of the F-transform procedure is to form a *fuzzy partition* of the domain $[a, b]$. It consists of a finite set of fuzzy sets

$$\mathcal{A} = \{A_0, \dots, A_n\}, \quad n \geq 2, \tag{10}$$

defined over nodes

$$a = c_0, \dots, c_n = b. \tag{11}$$

The properties of the fuzzy sets from \mathcal{A} are specified by five axioms, namely: *normality, locality, continuity, unimodality, and orthogonality* that is formally defined by

$$\sum_{i=0}^n A_i(x) = 1, \quad x \in [a, b] \tag{12}$$

(Equation (12) is sometimes called *Ruspini condition*).

A fuzzy partition \mathcal{A} is called *h-uniform* if the nodes c_0, \dots, c_n are *h-equidistant*, i.e., for all $k = 0, \dots, n - 1$, $c_{k+1} = c_k + h$, where $h = (b - a)/n$ and the fuzzy sets A_1, \dots, A_{n-1} are shifted copies of a *generating function* $A : [-1, 1] \rightarrow [0, 1]$ such that for all $k = 1, \dots, n - 1$

$$A_k(x) = A\left(\frac{x - x_k}{h}\right), \quad x \in [c_{k-1}, c_{k+1}]$$

(for $k = 0$ and $k = n$ we consider only half of the function A , i.e. restricted to the interval $[0, 1]$ and $[-1, 0]$, respectively). The membership functions A_0, \dots, A_n of fuzzy sets forming the fuzzy partition \mathcal{A} are usually called *basic functions*.

Let us emphasize that the concept of fuzzy partition is crucial for the F-transform. Moreover, it is a typical concept used in many fuzzy techniques. Its main advantage for applications consists in the possibility that the neighboring fuzzy sets can overlap, which is not the case of the classical partition of a set.

4.1.2 Zero Degree F-transform

Once the fuzzy partition $A_0, \dots, A_n \in \mathcal{A}$ is determined, we define a *direct F-transform* of a continuous function f as a vector $\mathbf{F}[f] = (F_0[f], \dots, F_n[f])$, where each k -th *component* $F_k[f]$ is equal to

$$F_k[f] = \frac{\int_a^b f(x)A_k(x) dx}{\int_a^b A_k(x) dx}, \quad k = 0, \dots, n. \tag{13}$$

Clearly, the $F_k[f]$ component is a *weighted average* of the functional values $f(x)$ where weights are the membership degrees $A_k(x)$. The *inverse F-transform* of f with respect to $\mathbf{F}[f]$ is a continuous function³ $\hat{f} : [a, b] \rightarrow \mathbb{R}$ such that

$$\hat{f}(x) = \sum_{k=0}^n F_k[f] \cdot A_k(x), \quad x \in [a, b].$$

Theorem 1. *The inverse F-transform \hat{f} has the following properties:*

- (a) *The sequence of inverse F-transforms $\{\hat{f}_n\}$ determined by a sequence of uniform fuzzy partitions based on uniformly distributed nodes with $h = (b - a)/n$ uniformly converges to f for $n \rightarrow \infty$.*
- (b) *The F-transform is linear, i.e., if $f(x) = \alpha u(x) + \beta v(x)$ then $\hat{f}(x) = \alpha \hat{u}(x) + \beta \hat{v}(x)$ for all $x \in [a, b]$.*

All the details and full proofs can be found in [26, 27].

³ By abuse of language, we call by direct as well as inverse F-transform both the procedure as well as its respective results $\mathbf{F}[f] = (F_0[f], \dots, F_n[f])$ and \hat{f} .

4.1.3 Higher Degree F-transform

The F-transform introduced above is F^0 -transform (i.e., zero-degree F-transform). Its components are real numbers. If we replace them by polynomials of arbitrary degree $m \geq 0$, we arrive at the higher degree F^m transform. This generalization has been in detail described in [27]. Let us remark that the F^1 transform enables to estimate also derivatives of the given function f as weighted average values over a vaguely specified area.

The direct F^1 -transform of f with respect to A_1, \dots, A_{n-1} is a vector $F^1[f] = (F^1_1[f], \dots, F^1_{n-1}[f])$ where the components $F^1_k[f]$, $k = 1, \dots, n - 1$ are linear functions

$$F^1_k[f](x) = \beta^0_k + \beta^1_k(x - c_k) \tag{14}$$

with the coefficients β^0_k, β^1_k given by

$$\beta^0_k = \frac{\int_{c_{k-1}}^{c_{k+1}} f(x)A_k(x)dx}{\int_{c_{k-1}}^{c_{k+1}} A_k(x)dx}, \tag{15}$$

$$\beta^1_k = \frac{\int_{x_{k-1}^{x_{k+1}}} f(x)(x - c_k)A_k(x)dx}{\int_{c_{k-1}}^{c_{k+1}} (x - c_k)^2 A_k(x)dx}. \tag{16}$$

Note that $\beta^0_k = F_k[f]$, i.e. the coefficients β^0_k are just the components of the F^0 transform given in (13). The F^1 transform has also the properties stated in Theorem 1 (see [27]).

We will also use the F^2 transform. Its components are the functions

$$F^2_k[f](x) = \beta^0_k + \beta^1_k(x - c_k) + \left(\beta^2_k(x - c_k)^2 - \frac{h^2}{6} \right)$$

(provided that the basic functions are triangles).

Theorem 2 ([11]). *If f is four-times continuously differentiable on $[a, b]$ then for each $k = 1, \dots, n - 1$,*

$$\beta^0_k = f(c_k) + O(h^2), \tag{17}$$

$$\beta^1_k = f'(c_k) + O(h^2). \tag{18}$$

$$\beta^2_k = \frac{f''(c_k)}{2} + O(h^2). \tag{19}$$

Thus, the F-transform components provide a weighted average of values of the function f in the area around the node c_k (17), and also a weighted average of slopes (27) of f and that of its second derivatives (19) in the same area.

Remark 1 (important). It should be noted that only the nodes c_1, \dots, c_{n-1} should be considered when dealing with the F-transform and the edge nodes c_0, c_n should be omitted. The reason is that the areas $[c_0, c_1]$ and $[c_{n-1}, c_n]$ are covered by halves of the basic functions A_0, A_n , respectively and so, the approximation of f in these areas is subject to too large error. Hence, we should consider the function \hat{f} on the interval $[c_1, c_{n-1}]$ only.

4.2 Fuzzy Natural Logic

This is a special formal logical theory whose goal is to model the reasoning of people for which it is specific to use natural language. So far, it is not a unified theory but a bunch of the following theories:

- (a) A formal theory of evaluative linguistic expressions explained in detail in [18] (see also [17, 24]).
- (b) A formal theory of fuzzy IF-THEN rules and approximate reasoning [16, 22–24].
- (c) A formal theory of intermediate and generalized fuzzy quantifiers [5, 13, 19] and elsewhere.

4.2.1 Evaluative Linguistic Expressions

The central role in all these theories is played by the theory of *evaluative linguistic expressions*. These are expressions with the general form

$$\langle \text{linguistic modifier} \rangle \langle \text{TE-adjective} \rangle \quad (20)$$

where $\langle \text{TE-adjective} \rangle^4$ is one of the adjectives “small, medium, big” (and possibly other specific adjectives, especially the so called gradable or evaluative ones), or “zero” as well as arbitrary symmetric fuzzy number. The $\langle \text{linguistic modifier} \rangle$ is a special expression that belongs to a wider linguistic phenomenon called *hedging* and that specifies more closely the topic of utterance. In our case, the linguistic modifier makes the meaning of the $\langle \text{TE-adjective} \rangle$ more specific. Quite often it is represented by an intensifying adverb such as “very, roughly, approximately, significantly”, etc. The linguistic modifiers can have narrowing (“extremely, significantly, very, typically”) and widening effect (“more or less, roughly, quite roughly, very roughly”) on the meaning of the $\langle \text{TE-adjective} \rangle$.

If $\langle \text{linguistic hedge} \rangle$ is not present (expressions such as “weak, large”, etc.) then we take it as presence of *empty linguistic hedge*. Thus, all the simple evaluative expressions have the same form (20). Since they characterize values on an ordered scale, we may consider also scales divided into two parts that are usually interpreted as *positive* and *negative*. Hence, the evaluative expressions may have also a sign, namely “positive” or “negative”.

Simple evaluative expressions of the form (20) can also be combined using logical connectives (usually “and” and “or”) to obtain *compound* ones. A limited usage of the particle “not” is also possible. Let us emphasize, however, that syntactic and semantic limitations of natural language *prevent the compound evaluative expressions to form a boolean algebra!*

We distinguish abstract evaluative expressions from more specific *evaluative predications*. The latter are expressions of natural language of the form ‘ X is \mathcal{A} ’ where \mathcal{A} is an evaluative expression and X is a variable which stands for objects, for example “degrees of temperature, height, length, speed”, etc. Examples are “temperature is high”, “speed is extremely low”, “quality is very high”, etc.

⁴ The “TE” is a short for “trichotomic evaluative”.

In general, the variable X represents certain features of objects such as “size, volume, force, strength,” etc. and so, its values are often real numbers (Fig. 1).

Important notion is that of *linguistic context*. In our theory it is an interval $w = [v_L, v_S] \cup [v_S, v_R]$ determined by a triple of (real) numbers $w = \langle v_L, v_S, v_R \rangle$ where v_L is the leftmost typically small value, v_S is typically medium value and v_R is the rightmost typically big value. For example, when speaking about temperature of water, we may set $v_L = 15^\circ\text{C}$, $v_S = 50^\circ\text{C}$ and $v_R = 100^\circ\text{C}$. In the sequel, we will consider a set of all linguistic contexts

$$W = \{w = \langle v_L, v_S, v_R \rangle \mid v_L, v_S, v_R \in \mathbb{R}, v_L < v_S < v_R\}. \tag{21}$$

The element x belongs to a context $w \in W$ if $x \in [v_L, v_R]$. Then we write $x \in w$.

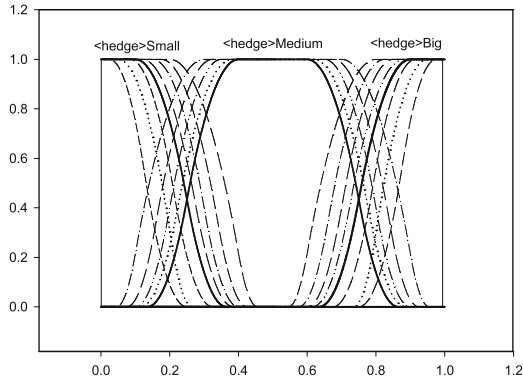


Fig. 1. Shapes of extensions of some evaluative expressions in the context $\langle 0, 0.5, 1 \rangle$. The hedges are {Extremely, Significantly, Very, empty hedge} for “small” and “big” and {More-or-Less, Roughly, Quite Roughly, Very Roughly} for “small”, “medium”, and “big”.

The meaning of an evaluative linguistic expression \mathcal{A} (as well as of a predication) is represented by its *intension*

$$\text{Int}(X \text{ is } \mathcal{A}) : W \longrightarrow \mathcal{F}(\mathbb{R}) \tag{22}$$

where $\mathcal{F}(\mathbb{R})$ is a set of all fuzzy sets on \mathbb{R} . For each context $w \in W$, the *extension* $\text{Ext}_w(X \text{ is } \mathcal{A})$ is a specific fuzzy set on \mathbb{R} . Example of extensions of several evaluative linguistic expressions is in Fig. 7. Let us emphasize that their shapes have been established on the basis of logical analysis of the meaning of the corresponding evaluative expressions (for the details, see [18]).

4.2.2 Linguistic Description

The evaluative linguistic predications are basic constituents of fuzzy/linguistic IF-THEN rules that are special conditional clauses of natural language. A set of

such rules is called a *linguistic description*, that is, a finite set of fuzzy/linguistic IF-THEN rules

$$\begin{aligned}
 \mathcal{R}_1 &= \text{IF } X \text{ is } \mathcal{A}_1 \text{ THEN } Y \text{ is } \mathcal{B}_1, \\
 &\dots\dots\dots \\
 \mathcal{R}_m &= \text{IF } X \text{ is } \mathcal{A}_m \text{ THEN } Y \text{ is } \mathcal{B}_m
 \end{aligned}
 \tag{23}$$

where “ X is \mathcal{A}_j ”, “ Y is \mathcal{B}_j ”, $j = 1, \dots, m$ are evaluative linguistic predications. The linguistic description can be understood as a specific kind of a (structured) text that can be used for description of various situations and processes.

4.2.3 Perception-Based Logical Deduction

Linguistic description taken as a special text requires a special inference method, namely the *Perception-based Logical Deduction* (PbLD). This inference method works with genuine evaluative linguistic expressions and it is based on formal properties of mathematical fuzzy logic (see [16, 17, 23]). The method is based on local properties of the linguistic description, so that we distinguish the rules as such but at the same time deal with them as vague expressions of natural language. The PbLD has nothing in common with the classical Mamdani’s inference (cf., e.g., [10]) (Fig. 2).

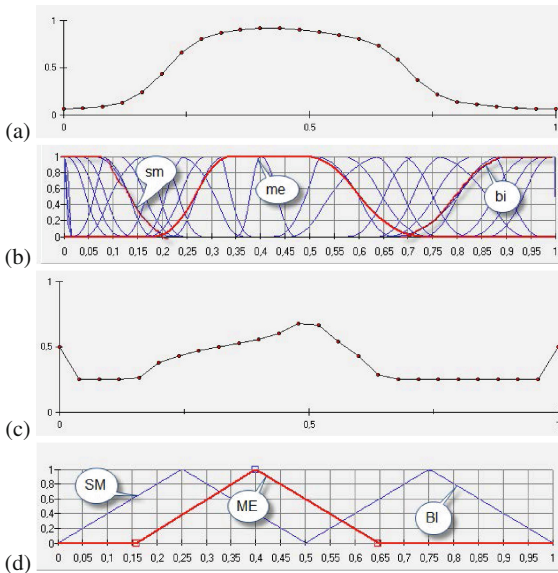


Fig. 2. (a) A function obtained from the simple linguistic description (24) using the PbLD method with smooth DEE defuzzification. (b) Extensions of the used evaluative expressions “small–medium–big” in the context $\langle 0, 0.4, 1 \rangle$. (c) A function obtained using Mamdani’s-COG method from linguistic description of the form (24) interpreted as fuzzy relation constructed using triangular membership functions depicted in (d).

The PbLD requires a defuzzification method called DEE (Defuzzification of Evaluative Expressions). Its variant realized using the F-transform is called *smooth DEE* (see [23]).

To demonstrate PbLD, let us consider the following linguistic description:

$$\begin{aligned} \mathcal{R}_1 &= \text{IF } X \text{ is } \textit{small} \text{ THEN } Y \text{ is } \textit{small}, \\ \mathcal{R}_2 &= \text{IF } X \text{ is } \textit{medium} \text{ THEN } Y \text{ is } \textit{big}, \\ \mathcal{R}_3 &= \text{IF } X \text{ is } \textit{big} \text{ THEN } Y \text{ is } \textit{small}. \end{aligned} \tag{24}$$

This description characterizes linguistically a function that has small functional values on the left and right side of the graphs and big ones in the middle. The result using PbLD method is depicted in part (a) of Fig. 3. In part (b) are extensions of the used evaluative expressions in the context $\langle 0, 0.4, 1 \rangle$.

To see that PbLD method cannot be replaced by the Mamdani’s method that is often used in various kinds of applications, we depicted in Fig. 3(c) and (d) the result obtained from (24) using it the basis of triangular fuzzy sets often (incorrectly) considered in literature as extensions of evaluative expressions. The reason why Mamdani’s method does not work in this case is the fact that it provides very good approximation of a function, but it is not logical inference suitable for manipulation with linguistic expressions.

4.2.4 Learning of Linguistic Description

In applications of the methods describe above, very important is the possibility to use a learning procedure developed in FNL (cf. [24]). If the data and a context w are given, we can learn linguistic description of the form (23) that linguistically characterizes the data. Using the PbLD inference method, we can obtain various kinds of specific information.

The learning procedure is realized by implementing a function of *local perception*

$$LPerc(x, w) = \mathcal{A} \tag{25}$$

where $w \in W$ is a given context and $x \in w$ is a given value. The evaluative linguistic expression \mathcal{A} characterizes the value x in the given context w . For example, the value $x = 0.15$ in a context $w = \langle 0, 4, 10 \rangle$ is evaluated by the evaluative expression “very small”.

Using this simple idea, we can transform data in the form

$$\begin{bmatrix} u_{11} & u_{12} & \dots & u_{1c} & v_1 \\ u_{21} & u_{22} & \dots & u_{2c} & v_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ u_{m1} & u_{m2} & \dots & u_{mc} & v_m \end{bmatrix}$$

into a linguistic description consisting of m fuzzy/linguistic IF-THEN rules of the form IF X_1 is \mathcal{A}_1 AND \dots AND X_c is \mathcal{A}_c THEN Y is \mathcal{B} .

The outcome of this procedure is twofold: first, it provides us with the succinct information understandable to people about the content of the data. Second, we can obtain answers to many “what if” questions and, on the basis of that, make proper decisions.

5 Analysis and Forecasting of Time Series Using Fuzzy Techniques

As discussed above, the fuzzy set theory (and fuzzy logic) is the mathematical model of vaguely determined actualized groupings of objects. No occurrence of any event is considered. In this section we will describe how fuzzy techniques can be applied when dealing with time series. This requires a slightly different view of time series. We will show that these techniques are able to compete with the probabilistic ones in forecasting not only future values of time series, but also to fit well the idea of their trend or trend cycle. But even more, the fuzzy models have the potential to bring new hints for analysis of time series that are not possible in the probabilistic approach. We have in mind especially applications of the model of the semantics of natural language using which we can obtain additional information about the behavior of time series which is, moreover, well understandable to people.

5.1 Decomposition of Time Series

In the probabilistic model, a time series is a sequence of random variables $\{X(t), t \in \mathbb{T}\}$ without considering their structure. A more apt model is the following: the time series is decomposed into several components

$$X(t) = Tr(t) + C(t) + S(t) + R(t), \quad t \in \mathbb{T}, \quad (26)$$

where Tr is the *trend*, C is a *cyclic* component, S is a *seasonal* component that is a mixture of periodic functions and R is a random *noise*, i.e., a sequence of independent random variables $R(t)$ such that for each $t \in \mathbb{T}$, the $R(t)$ has zero mean and finite variance.

The seasonal component S in (26) is assumed to be a sum of periodic functions

$$S(t) = \sum_{j=1}^r P_j \sin(\lambda_j t + \varphi_j), \quad t \in \mathbb{T}, \quad (27)$$

for some finite r where λ_j are frequencies, φ_j are phase shifts and P_j are amplitudes⁵.

In the practice, it is often difficult to distinguish trend and cycle. Therefore, these two components are often joined into one component called *trend-cycle*

$$TC(t) = Tr(t) + C(t), \quad t \in \mathbb{T}.$$

⁵ Because $\cos x = \sin(x + \pi/2)$, it is sufficient to consider only \sin .

Hence, we will replace the decomposition (26) by the simpler one

$$X(t) = TC(t) + S(t) + R(t), \quad t \in \mathbb{T}.$$

The difference between trend and trend-cycle was informally summarized by the following OECD definitions.

The *trend* is a component of a time series that represents variations of low frequency in a time series, the high and medium frequency fluctuations having been filtered out.

The *trend-cycle* is a component that represents variations of low and medium frequency in a time series, the high frequency fluctuations having been filtered out. This component can be viewed as those variations with a period longer than a chosen threshold (usually 1 year is considered as the minimum length of the business cycle). Form the mathematical point of view, we assume that both trend as well as trend-cycle are (continuous) functions with *small modulus of continuity*⁶.

Note that the decomposition model keeps the idea that the time series is a sequence of random variables but randomness is present only at the noise which is an unpredictable random component with specific properties. The rest are non-random components with clear interpretation. We argue, that fuzzy techniques provide more powerful means for extracting these components and, moreover, they make it possible to extract also additional information about time series. This information is usually vaguely specified, provided often in natural language and, therefore, it that cannot be obtained using the probabilistic methods.

The following theorem assures us that we can find a fuzzy partition enabling us to estimate either the trend *Tr* or the trend cycle *TC* with high fidelity.

Theorem 3. *Let $\{X(t) \mid t \in \mathbb{T}\}$ be a continuous realization of the stochastic process*

$$X(t) = Tr(t) + \sum_{j=1}^r P_j \sin(\lambda_j t + \varphi_j) + R(t), \quad t \in \mathbb{T}$$

where $\mathbb{T} = [0, b]$, $Tr : \mathbb{T} \rightarrow \mathbb{R}$ is a function with small modulus of continuity, $\lambda_1 \leq \dots \leq \lambda_r$ are frequencies and R is the noise from (26).

Let us construct an h -uniform fuzzy partition \mathcal{P} over nodes c_0, \dots, c_n with $h = d\bar{T}_1$, where $\bar{T}_1 = \frac{2\pi}{\lambda_1}$ and $d \geq 1$ is a real number. Let us compute the direct F -transform $F[X]$. Then there exists a number $D(d)$ such that $D(d) = 0$ for $d \rightarrow \infty$ and

$$|\hat{X}(t) - Tr(t)| \leq D(d), \quad t \in [c_1, c_{n-1}]$$

where \hat{X} is the corresponding inverse F -transform of X .

This theorem holds both for triangular as well as raised cosine fuzzy partition. The precise expressions for D in both cases and the proof of this theorem can be found in [14, 25]. It can also be proved that $D(d)$ is minimal if $d \in \mathbb{N}$.

⁶ Let $f : [a, b] \rightarrow \mathbb{R}$ be a continuous function. Then $\omega(h, f) = \max_{\substack{|x-y| < h \\ x, y \in [a, b]}} |f(x) - f(y)|$ is the modulus of continuity of f .

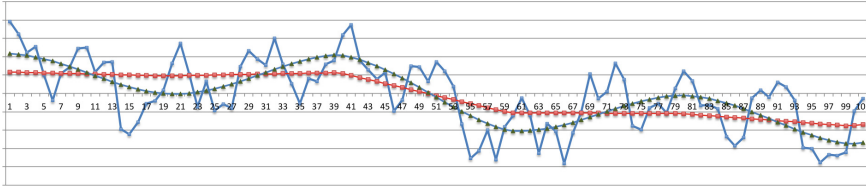


Fig. 3. Real trend and trend-cycle of the artificial time series.

Corollary 1. *Let $\{X(t) \mid t \in \mathbb{T}\}$ be a continuous realization of the stochastic process (26), Tr its trend and TC its trend-cycle. Then there exist numbers $D_1(d), D_2(d)$ such that $D_k(d) = 0$ for $d \rightarrow \infty, k = 1, 2$ and*

- (a) $|\hat{X}(t) - Tr(t)| \leq D_1(d),$
- (b) $|\hat{X}(t) - TC(t)| \leq D_2(d)$

for corresponding inverse F -transform \hat{X} of X and all $t \in [c_1, c_{n-1}]$.

It follows from this corollary that we can form the h -uniform fuzzy partition \mathcal{P} with h corresponding to the largest periodicity of a periodic constituent occurring either in the cyclic or the seasonal component of $S(t)$. Then all the subcomponents with shorter periodicities (i.e., higher frequencies) are almost “wiped down”, and also, the noise is significantly reduced. In other words, either the components C, S and R in (26) are almost completely removed and we obtain estimation of the trend

$$Tr(t) \approx \hat{X}(t), \tag{28}$$

or the components S and R are removed and we obtain estimation of the trend-cycle

$$TC(t) \approx \hat{X}(t), \tag{29}$$

$t \in [c_1, c_{n-1}]$.

To demonstrate the above outlined theory for estimation of trend Tr and trend-cycle TC of a time series, we constructed an artificial time series using the following formula:

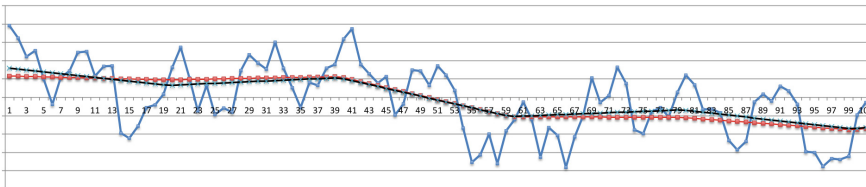


Fig. 4. Real trend (dotted line) of the artificial time series and its estimation (crossed line) using the F -transform with $h = 40$.

$$X(t) = \underbrace{Tr(t) + \sin(0.157t + 1.5)}_{TC} + \sin(0.283t + 0.34) + \sin(0.628t + 1.12) + \sin(1.57t + 0.79) + R(t) \quad (30)$$

where R is a random noise. The frequencies ω in this time series correspond to the following periodicities T , respectively: 40, 22.2, 10, 4. The trend $Tr(t)$ is determined explicitly by data and has no predefined shape (Fig. 4).

Using Periodogram, we found in the artificial time series (30) the following periodicities T : 36.9, 22.7, 16.6, 14.2, 9.9, 4. Note that Periodogram found two more not existing periodicities and also, that estimation of the periodicity $T = 40$ is not too precise (Fig. 5).

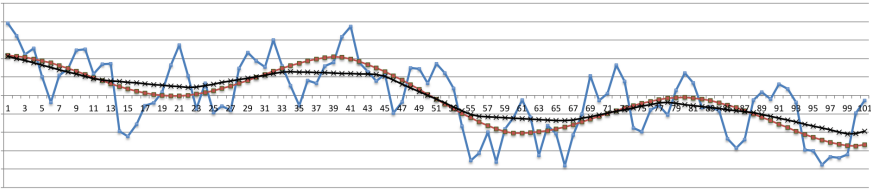


Fig. 5. Real trend-cycle (dotted line) of the artificial time series and its estimation (crossed line) using the F-transform with $h = 22$.

Finally, let us remark that the method is very robust towards missing values and outliers, i.e., there is no visible change of the trend or trend-cycle if omit some values of the time series.

5.2 Forecasting Future Course of Time Series

The linguistic description and PbLD inference method mentioned in Sect. 4.2.2 can be applied to forecasting of the trend Tr or trend-cycle TC . The method was described in detail in [24], and so, we will only briefly review its main ideas. Let $\bar{\mathbb{T}} \subset \mathbb{T}$. Then by $X|_{\bar{\mathbb{T}}}$ we denote the restriction of X to $\bar{\mathbb{T}}$. For the consistency of notation, we will write the time series (26) as $X|_{\mathbb{T}}$.

Let $\mathbb{T}' \supset \mathbb{T}$ be a new time domain. Our task is to extrapolate values of X to $X|_{(\mathbb{T}' \setminus \mathbb{T})}$ on the basis of the known values of $X|_{\mathbb{T}}$. The method for finding the former is called *forecasting*. As noted above, there are many forecasting methods mostly formulated using probability theory (cf. [3, 7, 9]). In this section, we present methods based on fuzzy techniques.

Recall that trend or trend-cycle are obtained using the F-transform on the basis of an h -uniform fuzzy partition \mathcal{P} . The result of the direct F-transform is a vector of F-transform components

$$\mathbf{F}[X] = (F_1[X], \dots, F_{n-1}[X]), \quad (31)$$

where each component $F_i[X]$ represents a weighted average of values of $X(t)$ in the area of width $2h$. The components (31) can be used as data for learning of a linguistic description. Then, using it and the PbLD method, we can forecast future F-transform components

$$F_n[X], \dots, F_{n+l}[X] \tag{32}$$

and from them, we can compute estimation of the future development either of trend or trend cycle using the inverse F-transform \hat{X} .

The learned linguistic description consists of fuzzy/linguistic rules of the form, for example,

$$\text{IF } \Delta^2 X_{i-1} \text{ is } \mathcal{A}_{\Delta^2 i-1} \text{ AND } \Delta X_{i-1} \text{ is } \mathcal{A}_{\Delta i-1} \text{ AND } X_i \text{ is } \mathcal{A}_i \text{ THEN } X_{i+1} \text{ is } \mathcal{B}_{i+1}. \tag{33}$$

where

$$\Delta F_i[X] = F_i[X] - F_{i-1}[X], \quad i = 1, \dots, n - 1 \tag{34}$$

$$\Delta^2 F_i[X] = \Delta F_i[X] - \Delta F_{i-1}[X], \quad i = 2, \dots, n - 1 \tag{35}$$

are the first and second differences, respectively. Let us remark that in practice, all kinds of combinations of the F-transform components and their first and second differences can occur both in the antecedent as well as in the consequent of (33). Example of such description is the following:

Rule	$F_i[X]$	$\Delta F_i[X]$	$\Rightarrow F_{i+1}[X]$
1	ex bi	ra me	qr bi
2	ro bi	-ml me	qr bi
3	ro bi	-ex sm	vr sm
4	ze	-ex bi	vr sm
5	si sm	si sm	ra me
⋮			

(the used shorts: ze-zero, sm-small, me-medium, bi-big, ex-extremely, ro-roughly, qr-quite roughly, vr-very roughly, ra-rather, si-significantly, ml-more or less).

Note that forecasting of the future values, the learned linguistic description provides us also with information in linguistic form (i.e., understandable to people) explaining how the forecast was obtained, i.e., what are the inner characteristics of the time series that led to the forecast. The differences (34) and (35) characterize dynamics of the time series as well as logical dependencies of the trend-cycle changes (hidden cycle influences).

5.3 Mining Knowledge on Time Series

5.3.1 Linguistic Evaluation of the Local Trend

If a certain time interval is given, it may be interesting to learn what trend (tendency) of the time series can be recognized in it. Surprisingly, recognition

of trend in by no means is a trivial task even when watching the graph. Recall that by Theorem 23, the F^1 -transform provides estimation of the average slope (tangent). Therefore, it is a convenient tool for estimation of the course of trend of the given time series. Such estimation can be, moreover, expressed in natural language. For example, we can say “fairly large decrease (huge increase) of trend”, “the trend is stagnating (negligibly increasing)”, etc. These expressions characterize trend (tendency) of the time series in an area specified by the user. It is quite important achievement of the fuzzy techniques that it provides algorithms using which it is possible to generate automatically this kinds of linguistic evaluations. The method is based on the theoretical results in fuzzy natural logic and was described in more detail in [20,21]. Its idea is outlined below.

First, we must specify, what does it mean “extreme increase (decrease)”. In practice, it can be determined as the largest acceptable difference of time series values with respect to a given (basic) time interval (for example 12 months, 31 days) that is, a minimal and maximal tangent. In practice, we set only the largest tangent v_R while the smallest one is usually $v_L = 0$. The typically medium value v_S is determined analogously as v_R . The result is the context $w_{tg} = \langle v_L, v_S, v_R \rangle$. Furthermore, we must specify the time interval $I \subset \mathbb{T}$ interesting for the inspection. The next step is to compute a basic function A with the support I (cf. Subsect. 4.1.1) and compute the coefficient β^1 using formula (16).

Finally, we generate a linguistic evaluation of the trend of the time series X in the area characterized by A with respect to the context w_{tg} . The required evaluative expression \mathcal{A} is obtained using the function of local perception

$$\mathcal{A} = LPerc(\beta^1, w_{tg}). \tag{36}$$

Demonstration of the principle of evaluation is in Fig. 6⁷.

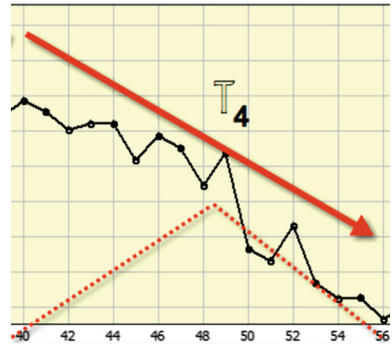


Fig. 6. The principle of linguistic of evaluation of direction of trend: *clear decrease*. The necessary parameter is the context w_{tg} specifying the lowest, typically medium and the largest value of the tangent. The triangle above the x -axis is the basic function of the F^1 -transform.

5.3.2 Mining More Kinds of Knowledge

Fuzzy techniques suggest more methods for mining knowledge from time series. One of them is finding *perceptionally important points* (PIP). According to [6], these are points where the time series essentially changes its course. Because

⁷ The results were obtained using the experimental software LFL Forecaster (see http://irafm.osu.cz/en/c110_lfl-forecaster/) which implements the described method. Its author is Viktor Pavliska.

of the complicated character following from the presence of various frequencies and noise, we cannot expect that this is just one isolated time point but better a certain area that cannot be precisely determined. Therefore, a very suitable method is based on the higher-degree F-transform because it makes it possible to estimate the first and second derivatives of a function with complicated course in a vaguely specified area. Namely, this can be done by looking for small values of the β^1 coefficient (27).

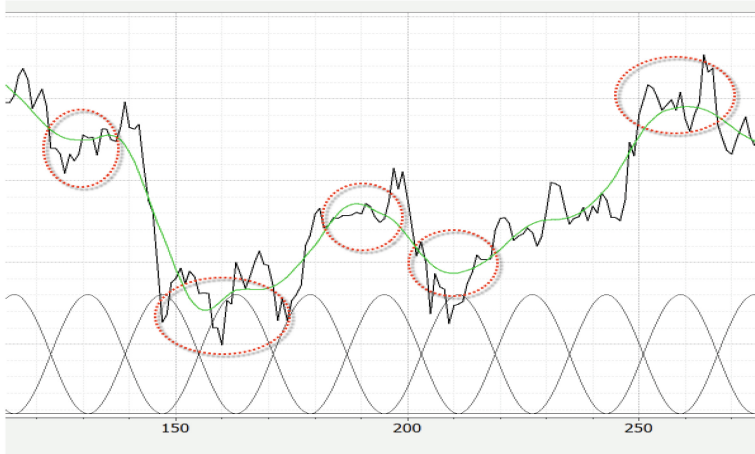


Fig. 7. Time series with marked perceptually important points and F^1 -approximation of its course. Along the x -axis is also depicted the fuzzy partition.

Demonstration of the result of searching PIP in a part of the Monthly Closing of Dow-Jones index is in Fig. 7⁸. The points are found in areas covered by the corresponding basic functions of the fuzzy partition. They correspond to values of β^1 close to zero⁹. To find the points we must shift the fuzzy partition to localize β^1 with minimal values. Note that it can be equal to zero only in case of ideal line parallel with x -axis.

Other interesting possibility is to find time intervals in which the trend of the time series X exhibits monotonous behavior which is also characterized linguistically. This means that we must decompose the time domain \mathbb{T} into a set of time intervals $\mathbb{T}_i \subseteq \mathbb{T}$, $i = 1, \dots, s$, with monotonous trend of X (increasing, decreasing, stagnating). Each interval \mathbb{T}_i is a union of one or more adjacent intervals $\bar{\mathbb{T}}_j$. As a final result, direction of the trend of $X|\mathbb{T}_i$ is linguistically evaluated similarly as is outlined above. The detailed algorithm can be found in [20].

An area that is becoming still more attractive is automatic summarization of knowledge about time series. This task is addressed by several authors (see,

⁸ The points were obtained using the experimental software FT-Studio whose author is Radek Valášek.

⁹ The actual values range between 0.07 and 1.4.

e.g., [4, 8]). The fuzzy natural logic suggests sophisticated formal theory of *intermediate quantifiers*. The summarized information may address either one time series or a set of time series. The theory includes a formalism on the basis of which we can develop a model of the meaning of linguistic statements containing quantified information, as is usual in natural language, but also human-like syllogistic reasoning that is based on the formal model of *generalized Aristotle's syllogisms*. For more details, see [12, 20].

6 Conclusion

In this paper, we discussed the difference between vagueness and uncertainty phenomena and their role in the fuzzy and probabilistic techniques applied to time series analysis. While probabilistic techniques assume that the time series is a stochastic process consisting of random variables, fuzzy techniques stem from the decomposition of the time series into deterministic components, assuming that only noise is random. We argue that, because both techniques have at disposal one realization of the time series only, statistically relevant processing is possible only under quite strong assumptions on the origins of the time series. While such assumptions are in the case of noise natural, for the whole time series they are too strong.

As fuzzy techniques are based on the model of the vagueness phenomenon, they are robust which means that they are little sensitive to changes of the data. Moreover, they enable to obtain information that cannot be obtained using probabilistic techniques. This concerns especially the area of mining knowledge from time series. On one hand, this knowledge can be obtained in an easier and straightforward way than classical methods (e.g., finding perceptually important points). On the other hand, the knowledge can be often obtained directly in expressions or even sentences of natural language, which is the form well understandable to people.

References

1. Anděl, J.: Statistical Analysis of Time Series. SNTL, Praha (1976). (in Czech)
2. Black, M.: Vagueness: an exercise in logical analysis. *Philos. Sci.* **4**, 427–455 (1937). reprinted in *Int. J. Gen. Syst.* **17**, 107–128 (1990)
3. Bovas, A., Ledolter, J.: *Statistical Methods for Forecasting*. Wiley, New York (2003)
4. Castillo-Ortega, R., Marín, N., Sánchez, D.: A fuzzy approach to the linguistic summarization of time series. *Multiple-Valued Logic Soft Comput.* **17**(2–3), 157–182 (2011)
5. Dvořák, A., Holčapek, M.: L-fuzzy quantifiers of the type $\langle 1 \rangle$ determined by measures. *Fuzzy Sets Syst.* **160**, 3425–3452 (2009)
6. Fu, T.-C.: A review on time series data mining. *Eng. Appl. Artif. Intell.* **24**, 164–181 (2011)
7. Hamilton, J.: *Time Series Analysis*. Princeton University Press, Princeton (1994)
8. Kacprzyk, J., Wilbik, A., Zadrożny, S.: Linguistic summarization of time series using a fuzzy quantifier driven aggregation. *Fuzzy Sets Syst.* **159**, 1485–1499 (2008)

9. Kedem, B., Fokianos, K.: *Regression Models for Time Series Analysis*. Wiley, New York (2002)
10. Klir, G., Bo, Y.: *Fuzzy Set Theory: Foundations and Applications*. Prentice Hall, Upper Saddle River (1995)
11. Kreinovich, V., Perfilieva, I.: Fuzzy transforms of higher order approximate derivatives: a theorem. *Fuzzy Sets Syst.* **180**, 55–68 (2011)
12. Murinová, P., Novák, V.: A formal theory of generalized intermediate syllogisms. *Fuzzy Sets Syst.* **186**, 47–80 (2012)
13. Murinová, P., Novák, V.: The structure of generalized intermediate syllogisms. *Fuzzy Sets Syst.* **247**, 18–37 (2014)
14. Nguyen, L., Novák, V.: Filtering out high frequencies in time series using F-transform with respect to raised cosine generalized uniform fuzzy partition. In: *Proceedings of International Conference FUZZ-IEEE 2015*. IEEE Computer Society, CPS, Istanbul (2015)
15. Nguyen, L., Novák, V.: Trend-cycle forecasting based on new fuzzy techniques. In: *Proceedings of the International Conference FUZZ-IEEE 2017*, Naples, Italy (2017)
16. Novák, V.: Perception-based logical deduction. In: Reusch, B. (ed.) *Computational Intelligence, Theory and Applications*, pp. 237–250. Springer, Berlin (2005)
17. Novák, V.: Mathematical fuzzy logic in modeling of natural language semantics. In: Wang, P., Ruan, D., Kerre, E. (eds.) *Fuzzy Logic - A Spectrum of Theoretical & Practical Issues*, pp. 145–182. Elsevier, Berlin (2007)
18. Novák, V.: A comprehensive theory of trichotomous evaluative linguistic expressions. *Fuzzy Sets Syst.* **159**(22), 2939–2969 (2008)
19. Novák, V.: A formal theory of intermediate quantifiers. *Fuzzy Sets Syst.* **159**(10), 1229–1246 (2008)
20. Novák, V.: Linguistic characterization of time series. *Fuzzy Sets Syst.* **285**, 52–72 (2016)
21. Novák, V.: Mining information from time series in the form of sentences of natural language. *Int. J. Approx. Reason.* **78**, 192–209 (2016)
22. Novák, V., Lehmke, S.: Logical structure of fuzzy IF-THEN rules. *Fuzzy Sets Syst.* **157**, 2003–2029 (2006)
23. Novák, V., Perfilieva, I.: On the semantics of perception-based fuzzy logic deduction. *Int. J. Intell. Syst.* **19**, 1007–1031 (2004)
24. Novák, V., Perfilieva, I., Dvořák, A.: *Insight into Fuzzy Modeling*. Wiley, Hoboken (2016)
25. Novák, V., Perfilieva, I., Holčapek, M., Kreinovich, V.: Filtering out high frequencies in time series using F-transform. *Inf. Sci.* **274**, 192–209 (2014)
26. Perfilieva, I.: Fuzzy transforms: theory and applications. *Fuzzy Sets Syst.* **157**, 993–1023 (2006)
27. Perfilieva, I., Daňková, M., Bede, B.: Towards a higher degree F-transform. *Fuzzy Sets Syst.* **180**, 3–19 (2011)
28. Russell, B.: Vagueness. *Aust. J. Phi.* **1**, 84–92 (1923)
29. Vopěnka, P.: *Mathematics in the Alternative Set Theory*. Teubner, Leipzig (1979)
30. Zadeh, L.A.: Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans. Syst. Man Cybern.* **SMC-3**, 28–44 (1973)