Forecasting Direction of Trend of a Group of Analogous Time Series Using F-Transform and Fuzzy Natural Logic

Vilém Novák, Irina Perfilieva

University of Ostrava
Institute for Research and Applications of Fuzzy Modeling
NSC IT4Innovations, 30. dubna 22, 701 03 Ostrava 1, Czech Republic
E-mail: {*Vilem.Novak,Irina.Perfilieva*}*@osu.cz*

Received 28 April 2015

Accepted 26 October 2015

Abstract

We present an idea to group time series according to similarity of their local trends and to predict future direction of the trend of all of them on the basis of forecast of only one representative. First, we assign to each time series an adjoint one, which consists of a sequence of the F^1-transform components. Then, they are grouped together according to their similarity, a principal time series is selected in each group and its future course is forecasted. Finally, directions of trends of the other time series from the same group are computed using special methods of fuzzy natural logic.

Keywords: F-transform; time series; Fuzzy Natural Logic; evaluative linguistic expressions

1. Introduction

In economic applications we usually deal with many interrelated time series. This means that a dynamic behavior of one time series influences the dynamic behavior of the other one. More specifically, future values of one time series depend not only on its past values but also on past values of the other time series. For example, household expenditures may depend also on income, past repairs, interest rates, etc. Therefore, a plausible forecast must take values of the other time series into account.

A given time series can be taken as a result of measurements (in time) of some endogenous variable and so, to obtain its forecast we must compute forecast of values of the corresponding exogenous variables on which the given variable is dependent. The situation can be quite complex because the interrelations among time series may lead to a hierarchical structure and values of the given time series can be obtained using a structured aggregation of the other ones.

There is a vast literature on multiple time series analysis and forecasting. According to [24], these models have been shown (i) to be extremely flexible in capturing the dynamic inter-relationships between a set of variables, (ii) to be able to treat several variables endogenously, (iii) not to require firm prior knowledge on the nature of the different relationships, (iv) to be able to capture both short- and long-run inter-relationships, and (v) to outperform multivariate TSA models in parameter efficiency, goodness-of-fit measures as well as in forecasting performance.

There are many methods applied, namely statistical ones, e.g., multiple regression analysis, multivariate AR(I)MA models and other statistical methods, and also non-statistical methods, e.g., neural networks or genetic algorithms (cf. [11] and the citations therein).

All these methods assume that the time series are interrelated so that dependencies among them play important role and so, they should be discovered. Sometimes, however, groups of products (or services) are analogous in ways that make them follow similar time series patterns (cf. [4]). For example, it can happen that similar products may fall within the same sphere of influence, such as the same or similar consumer tastes, competition levels, local economic cycles, weather, regional trends, etc. The result is that the corresponding time series strongly correlate positively (or negatively) over time.

This suggests the idea that we can form a group of similarly behaving time series, forecast only one selected (principal) one and, on the basis of that, estimate the future behavior of the other time series from the given group. The problem is how similar behavior of a group of time series can be discovered and evaluated. In this paper, we suggest to use the technique of F-transform that was already proved to be very effective in the analysis of time series. In [20,19] it was mathematically proved and also demonstrated on the data that the trend-cycle that is a low-frequency trend component of time series can be estimated with high fidelity. Moreover, the F-transform provides not only values of the trend-cycle but also its analytic formula.

Further strong property of the F-transform (more precisely, the F^1-transform) is its ability to estimate slope of the time series in a specified area. This property is applied in this paper is combination with methods based on *Fuzzy Natural Logic* (FNL). This approach enables us to estimate future development of the trend-cycle and also to mine various kinds of information from a given time series. For example, in [17], we suggested a method for automatic generation of natural language comments to local trend of time series.

The paper is structured as follows. In the next section, we will briefly overview principal ideas of the used techniques, namely the fuzzy (F-)transform and fuzzy natural logic. Section 3 is the main part where we explain our method of forecasting of a group of time series as well as their linguistic characterization. The paper is finished by a conclusion where we outline future investigation.

2. Preliminaries

In this section we very briefly overview some of the main concepts of the techniques of Fuzzy Transform and Fuzzy Natural Logic [22,16]. These techniques were described in several papers and so, many details are omitted here.

2.1. Fuzzy transform

The main idea of this technique (often called simply as F-transform) is transform of a bounded real continuous function $f : [a,b] \longrightarrow [c,d]$ into a finite vector $\mathbf{F}[f] = (F_0[f], \dots, F_n[f])$ where $F_k[f]$ are components that can be either numbers or functions of a special form. This phase is called *direct F-transform*.

The transform from the space of vectors back to the space of real functions is called *inverse F-transform*. Its result is a (continuous) function $\widehat{f} : [a,b] \longrightarrow [c,d]$ that approximates f. Successful applications of this method are based on proper setting of parameters that lead to a function $\widehat{f}$ with desirable properties.

Depending on the form of the components we distinguish F-transform of various degrees. If $F_k[f]$ are numbers then we speak about zero degree F-transform (F^0-transform). If the components are linear functions then we speak about first degree F-transform (F^1-transform), etc. In the sequel, by F-transform we understand both F^0 as well as F^1-transform whenever the degree is clear from the context.

2.1.1. Fuzzy partition

The first step is to form a *fuzzy partition* of the domain $[a,b]$. We select nodes $a = c_0, \dots, c_n = b$, $n \geqslant 3$, in the interval $[a,b]$ and then define $n+1$ fuzzy sets $\mathscr{A} = \{A_0, \dots, A_n\}$ over $[a,b]$ (also called basic functions) having the following properties for all $x \in [a,b]$ and $k = 0, \dots, n$: $A_k(c_k) = 1$ (*normality*); $A_k(x) = 0$ for $x \notin (c_{k-1}, c_{k+1})$ (we formally put $c_{-1} = a$ and $c_{n+1} = b$); A_k is continuous, strictly increases on $[c_{k-1}, c_k]$ and strictly decreases on $[c_k, c_{k+1}]$, and $\sum_{k=0}^{n} A_k(x) = 1$ (orthogonality).

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

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

(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).

2.1.2. F^0-transform

Given a fuzzy partition $\mathscr{A}$, a *direct F-transform* of $f : [a,b] \longrightarrow [c,d]$ is a vector $\mathbf{F}[f] = (F_0[f], \ldots, F_n[f])$, where each k-th *component* $F_k[f]$ is equal to

$$F_k[f] = \frac{\int_a^b f(x) A_k(x)\, dx}{\int_a^b A_k(x)\, dx}, \qquad k = 0, \ldots, n. \quad (1)$$

Each $F_k[f]$ 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 function $\widehat{f} : [a,b] \longrightarrow \mathbb{R}$ such that*)

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

Fuzzy transform has many convenient properties. Let us only mention that the sequence $\{\widehat{f_n}\}$ of inverse F-transforms uniformly converges to f for $n \to \infty$. All the details and full proofs can be found in [22] and elsewhere.

2.1.3. F^1-transform

If we replace the components $F_k[f]$ in (1) by polynomials of some non-zero degree $m \geqslant 1$, we arrive at the F^m transform. This generalization has been in detail described in [23]. Besides better approximation properties, it enables us to estimate also derivatives of the given function f as average values over wider area. In this paper, we need only F^1-transform of a special form whose definition follows.

Definition 1. Let $f : [a,b] \longrightarrow \mathbb{R}$ be a continuous function and an h-uniform partition of $[a,b]$ be given by the triangular-shaped basic functions $\mathscr{A} = \{A_1, \ldots, A_{n-1}\}$ with the generating function $A_0 = 1 - |x|$. The vector of linear functions

$$\mathbf{F}^1[f] = (\beta_1^0 + \beta_1^1(x - c_1), \ldots, \beta_{n-1}^0 + \beta_{n-1}^1(x - c_{n-1})) \quad (2)$$

is called the direct F^1-transform of f with respect to the fuzzy partition $\mathscr{A}$, where

$$\beta_k^0 = \frac{\int_{c_{k-1}}^{c_{k+1}} f(x) A_k(x) dx}{h}, \quad (3)$$

$$\beta_k^1 = \frac{12 \int_{c_{k-1}}^{c_{k+1}} f(x)(x - c_k) A_k(x) dx}{h^3}, \quad (4)$$

for every $k = 1, \ldots, n-1$.

Remark 1. A pedantic notation would require to write components of (2) as $\beta_k^0[f] + \beta_k^1[f](x - c_k)$. However, if f is clear from the context, we will often omit the symbol $[f]$.

The following theorem, whose proof can be found in [10] plays an important role in our application to time series trend evaluation.

Theorem 1. *Let $\mathscr{A} = \{A_1, \ldots, A_{n-1}\}$ be an h-uniform partition of $[a,b]$, let functions f and $A_k \in \mathscr{A}$, $k = 1, \ldots, n-1$, be four times continuously differentiable on $[a,b]$. Finally, let $\mathbf{F}^1[f]$ be the F^1-transform (2) of f. Then, for every $k = 1, \ldots, n-1$, the following estimation holds true:*

$$|\beta_k^1 - f'(c_k)| \leqslant M h^2 \quad (5)$$

for some $M > 0$.

According to Theorem 1, the coefficients β_k^1, $k = 1, \ldots, n$ provide estimation of weighted averages of the first derivative of f over intervals $[c_{k-1}, c_{k+1}]$. We will use this result in the evaluation of the trend of time series.

*)By abuse of language, we call by direct F-transform both the procedure as well as its result $\widehat{f}$.

Similarly as in case of F^0-transform, the *inverse F^1-transform* of f is defined by

$$\widehat{f}^1(x) = \sum_{k=1}^{n-1} (F_k^1[f](x) \cdot A_k(x)). \qquad (6)$$

2.2. Selected methods of fuzzy natural logic

2.2.1. Fuzzy Natural Logic

Fuzzy natural logic (FNL) is a group of mathematical theories that extend mathematical fuzzy logic in narrow sense (cf. [14]). Its goal is to develop mathematical model of special human reasoning schemes that employ natural language but are independent on a concrete one. Hence, FNL includes also a model of the semantics of some classes of words and natural language expressions.

Constituents of FNL are the following:

(a) Theory of evaluative linguistic expressions.

(b) Theory of fuzzy/linguistic IF-THEN rules and logical inference from them.

(c) Theory of generalized and intermediate quantifiers and generalized Aristotle syllogisms.

All the theories are formal mathematical theories based on the formal system of higher-order fuzzy logic. We suppose that their list is not final and will be extended in the future. In this paper, we apply results of the theories (a) and (b).

2.2.2. Theory of evaluative linguistic expressions

Evaluative linguistic expressions are natural language expressions such as *small, very big, rather medium, extremely strong, roughly important*, etc. Their theory was in detail described in [13].

The set of evaluative linguistic expressions is denoted by *EvExpr*. We will write $Ev \in EvExpr$ for an arbitrary element from *EvExpr*. Its most important subclass is formed by *simple evaluative expressions*, i.e., natural language expressions of the form

$$\langle \text{linguistic hedge} \rangle \langle \text{TE-adjective} \rangle \qquad (7)$$

where ⟨TE-adjective⟩[†)] is a special adjective having a nominal form (e.g., *small*) and an *antonym* (.e.g, *big*).[‡)] In most cases, there is also a special expression in the position "a middle member". We say that these adjectives form a *fundamental evaluative trichotomy*. Typical representatives of ⟨TE-adjective⟩ are *gradable adjectives* (big, cold, deep, fast, friendly, happy, high, hot, important, long, popular, rich, strong, tall, warm, weak, young), *evaluative adjectives* (good, bad, clever, stupid, ugly, etc.), but also adjectives such as *left, middle, medium*, etc. Examples of an evaluative trichotomy are *low, medium, high*; *clever, average, stupid*; *good, normal, bad*, etc.

The ⟨linguistic hedge⟩ represents a class of adverbial modifications that includes a subclass of *intensifying adverbs* such as "very, roughly, approximately, significantly", etc. Two basic kinds of linguistic hedges can be distinguished in (7): *narrowing* hedges, for example, "extremely, significantly, very" and *widening* ones, for example "more or less, roughly, quite roughly, very roughly". Note that narrowing hedges make the meaning of the whole expression more precise while widening ones have the opposite effect. Thus, "very small" is more precise than "small", which, on the other hand, is more precise (more specific) than "roughly small". If on surface level the ⟨linguistic hedge⟩ is missing, then we take it as a presence of *empty linguistic hedge*. Consequently, all the simple evaluative expressions have the same form (7)[§)]which means that also expressions such as "small, long, deep", etc. are examples of (7).

In this paper, we will suppose that *EvExpr* always contains all the meaningful combinations of *small (Sm), medium (Me), big (Bi), zero (Ze)* with the hedges *extremely (Ex), significantly (Si), very (Ve), rather (Ra), more or less (ML), roughly (Ro), very roughly (VR), quite roughly (QR)* where in the brackets is a simple short that is used to simplify specification of evaluative expressions in applications. There is a special ordering of "sharpness" of evaluative expressions obtained on the basis of

†) The short "TE" means trichotomic evaluative.

‡) Note that *big* is antonym of *small* but also vice-versa, *small* is antonym of *big*.

§) As a special case of evaluative expressions, we also take *fuzzy numbers*. However, we do not deal with them in this paper.

the following two orderings: $Sm \lll Me \lll Bi$ and $Ex \lll Si \lll Ve \lll$ empty hedge $\lll ML \lll Ro \lll QR \lll VR$. This ordering is used in the PbLD method and also in the function of local perception (11) described below.

We will distinguish abstract evaluative expressions, i.e., expressions such as *small, weak, very strong*, etc., that alone do not talk about any specific objects and *evaluative linguistic predications* such as "temperature is high, expenses are extremely low, the building is quite ugly", etc. Since the concrete noun in applications is usually unimportant and only some specific numerical characteristics is considered the evaluative predication is simplified into a form

$$X \text{ is } \langle \text{simple evaluative expression} \rangle \tag{8}$$

where X is some variable taking values from $\mathbb{R}$.

Mathematical model of the semantics of evaluative expressions is developed on the basis of logical analysis of their meaning. The essential concept is that of (linguistic) *context*. In case of evaluative expressions, it is determined by a triple of real numbers $\langle v_L, v_S, v_R \rangle$ where $v_L < v_S < v_R$ $(\in \mathbb{R})$. These numbers represent the *smallest, typically medium*, and the *largest* thinkable values, respectively. The context is thus a set

$$w = \{x \mid v_L \leqslant x \leqslant v_R\} \tag{9}$$

together with three distinguished points $DP(w) = \langle v_L, v_S, v_R \rangle$. By W we denote the set of all contexts (9)

Each evaluative expression $Ev \in EvExpr$ is assigned the meaning which is a function

$$\text{Int}(Ev) : W \longrightarrow \mathscr{F}(\mathbb{R}).$$

We will call this function *intension* of the evaluative expression Ev. It assigns to each context $w \in W$ a fuzzy set $\text{Ext}_w(Ev) \mathrel{\underset{\sim}{\subset}} w$ called *extension* of the expression Ev in the context $w \in W$.

2.2.3. Linguistic descriptions and logical inference

The *linguistic description* is defined as a finite set of fuzzy/linguistic IF-THEN rules with common X and Y:

$$\begin{array}{c} \mathsf{IF}\ X \text{ is } \mathscr{A}_1\ \mathsf{THEN}\ Y \text{ is } \mathscr{B}_1 \\ \dots\dots\dots\dots\dots\dots\dots \\ \mathsf{IF}\ X \text{ is } \mathscr{A}_m\ \mathsf{THEN}\ Y \text{ is } \mathscr{B}_m \end{array} \tag{10}$$

Each rule in (10) is a conditional linguistic clause and the whole description can be taken as a *special piece of text* providing us with knowledge about some local situation, for example a decision-making problem, local characterization of behavior of some system, etc. In the reality, we usually need a complex knowledge that can be provided by a *knowledge base* consisting of a system of intercoonected linguistic descriptions.

We may also recognize the topic-focus articulation phenomenon in linguistic description (see more about this phenomenon, e.g., in [5]). Namely, the *topic* (the known information) is formed by the antecedents "X is $\mathscr{A}_j$" of the rules in (10) and the *focus* (the new information) is formed by the consequents "Y is $\mathscr{B}_j$", $m = 1, \dots, m$.

Inference from linguistic description is realized using the concept of *local perception*. This can be understood as a linguistic characterization of certain kind of "measurement" done by people in a concrete situation. Mathematically, we define a special function of *local perception* $LPerc : \mathbb{R} \times W \longrightarrow EvExpr$ assigning to each value $u \in \mathbb{R}$ for a given context $w \in W$ an intension

$$LPerc^{\text{LD}}(u, w) = \begin{cases} \text{Int}(X \text{ is } \mathscr{A}) & \text{if } u \in w, \\ \text{undefined} & \text{otherwise.} \end{cases} \tag{11}$$

of the *sharpest* evaluative predication where we consider the ordering $\lll$ of "sharpness" introduced above. For example, the meaning of *very small* is sharper than that of *small*, etc.. Moreover, the value $u \in w$ must be the *most typical* element for the extension $\text{Ext}_w(X \text{ is } \mathscr{A})$.

Both perception and linguistic description provide us with enough information and so, we can derive a conclusion on the basis of them. The procedure is called *perception-based logical deduction* (PbLD) and was in detail described in [12,18]. It is important to emphasize that PbLD works locally. This means that though vague, the rules are distinguished one from another one because they bring different local information about phenomena in concern. The function (11) makes it also possible to *learn* linguistic description from data. More details can be found in [15,21].

3. Time series analysis

By a time series we understand a stochastic process (see, e.g., [1,6])

$$X : \mathbb{T} \times \Omega \longrightarrow \mathbb{R} \tag{12}$$

where $\mathbb{T} = \{0, \dots, p\} \subset \mathbb{N}$ is a finite set of natural numbers interpreted as time moments and $\langle \Omega, \mathscr{A}, P \rangle$ is a probabilistic space. It follows from (12) that $X(t, \omega)$ for every $t \in \mathbb{T}$ and all $\omega \in \Omega$ is a random variable. If we fix $\omega \in \Omega$ then we obtain one realization of (12) and in this case, we will write $X(t)$ only.

3.1. Decomposition of time series

Important characteristics of the time series are its long-term *trend* component, medium-term *cyclic* component and short-term *seasonal* component. Hence, $X(t, \omega)$ can be decomposed as follows:

$$X(t, \omega) = T(t) + C(t) + S(t) + R(t, \omega), \quad t \in \mathbb{T}, \omega \in \Omega, \tag{13}$$

where $T(t)$ is a *trend*, $C(t)$ and $S(t)$ are cyclic and seasonal components, respectively. The these are assumed to be ordinary real functions. The $R(t, \omega)$ is a random noise, i.e., each $R(t)$ for any $t \in \mathbb{T}$ is a random variable with zero mean value and a finite variance.

In the literature (cf. [3,6,9,7] and elsewhere), trend and cycle are usually considered together as one *trend-cycle* component $TC(t)$ because the data may not be sufficiently long to be able to distinguish trend from the cycle.

The seasonal component $S(t)$ in (13) is assumed to be a sum of periodic functions

$$S(t) = \sum_{j=1}^{r} P_j \sin(\lambda_j t + \varphi_j), \qquad t \in \mathbb{T}, \tag{14}$$

for some finite r where λ_j are frequencies, φ_j phase shifts and P_j are amplitudes.[†)]

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

3.2. Estimation of the trend-cycle and evaluation of the local trend

3.2.1. Estimation of the trend-cycle

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

Theorem 2. *Let $X(t)$ be realization of the stochastic process (13) and $S(t)$ be its seasonal component (14). Put $\overline{\lambda} = \min\{\lambda_1, \dots, \lambda_r\}$. Let us now construct an h-uniform fuzzy partition $\mathscr{A}$ over nodes $c_0, \dots, c_n$ with $h = d\overline{T}$ where $\overline{T} = \frac{2\pi}{\overline{\lambda}}$ for some $d \in \mathbb{R}$. If we compute a direct F-transform $F[X]$ then there is a certain small number D converging to 0 for $d \to \infty$ such that the following holds for the corresponding inverse F-transform $\widehat{X}$ of X:*

$$|\widehat{X}(t) - TC(t)| \leqslant D, \qquad t \in [c_1, c_{n-1}]. \tag{15}$$

The precise formula for D and the proof of this theorem can be found in [19]. It follows from this theorem that if we set the fuzzy partition $\mathscr{A}$ in correspondence with the largest periodicity of a periodic function occurring in the seasonal component S then the latter is almost "wiped down" and the noise is also significantly reduced. What remains is the trend-cycle that is thus well estimated by the inverse F-transform $\widehat{X}$ of X.

3.2.2. Estimation of the trend

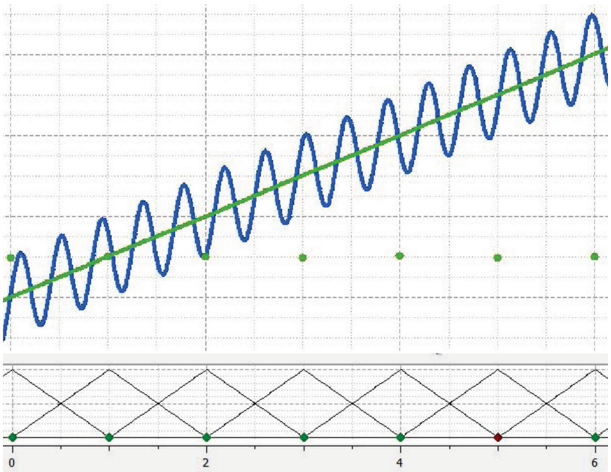
Trend $T(t)$ is a long term tendency of the time series that can be increasing, decreasing, or stagnating. There are many methods for its estimation, for example, regression analysis, exponential smoothing, low-pass filters and some other ones. None of them, however, is sufficiently reliable. In this paper, we will apply the F-transform because Theorem 2 can be applied also to extraction of the trend when specifying $\overline{T}$ equal to the length of the medium-term cyclic component. In practice, this is usually taken as length of the longest period detected in the time series.

In applications we are often interested only in the direction (slope, tendency) of the time series. This is mathematically characterized by a tangent.

Definition 2. Let $\overline{\mathbb{T}} \subset \mathbb{T}$ be a time interval and $X|\overline{\mathbb{T}}$ be a time series X restricted to $\overline{\mathbb{T}}$. Then *direction of the trend* of the time series $X|\overline{\mathbb{T}}$ is a real number $\mathrm{DTr}(X|\overline{\mathbb{T}}) \in \mathbb{R}$ being a slope of X over the area $\overline{\mathbb{T}}$.

The time series, however, can be largely volatile and so, how can we determine direction of its trend when looking at its graph? This task can be difficult and embarrassing even for people. One possibility was suggested in [8] where the authors approximate direction of the general trend by means of a properly inclined line. The algorithm, however, is heuristic without mathematical justification and the algorithm is quite complicated.

A more convenient solution is suggested by the F^1-transform because by Theorem 1, it provides estimation of a weighted average value of the first derivative of a function f over a given area. Note that if the function f is linear, i.e., $f(x) = px + q$, then it is easy to show that $\beta^1[f] = p$. If the function is more complicated, then to estimate its trend we have to decompose it into a linear component and a non-linear one. As an example, take the function $f(x) = x + \sin(15x)$ whose graph is below (blue line):



One can see that its trend should be given by the linear component $f(x) = x$. Let us now consider a triangular fuzzy partition with basic functions of width 2 that is depicted below the graph of the function. Then the straight line in the graph is the inverse F-transform $\widehat{f}(x) \approx x$. As mentioned, the tangent (slope; direction of the trend) of f should be close to 1 within any interval. Indeed, the coefficients (4) are $\beta^1[f|[0,2]] = 1.017$, $\beta^1[f|[1,3]] = 0.997$, $\beta^1[f|[2,4]] = 0.988$, $\beta^1[f|[3,5]] = 1.021$, $\beta^1[f|[4,6]] = 0.98$.[*)]

We conclude this section by the following definition.

Definition 3. Let A be a basic function defined over $\overline{\mathbb{T}}$. Then estimation of the direction of the trend of $X|\overline{\mathbb{T}}$ is

$$\mathrm{DTr}(X|\overline{\mathbb{T}}) = \beta^1[X|\overline{\mathbb{T}}]. \qquad (16)$$

3.2.3. Evaluation of trend using FNL

The course of time series, however, can be very complicated and also subject to random noise. Therefore, it may hardly have a sense to work with precise values of $\mathrm{DTr}(X|\overline{\mathbb{T}})$. In practice, it is usually sufficient to characterize direction of the trend by expressions of natural language only. In this paper, we will show how they can be automatically generated using methods of fuzzy natural logic. The core idea is to use the function of local perception (11).

The first step is to specify the context. This is clear because, for example, increase of temperature in winter by 3 C° during, say 2 hours, can be taken as "sharp increase" while the same in summer is a "very small increase". Therefore, we must start with specification of what does it mean "extreme increase (decrease)". More precisely, we must determine three distinguished values $w_{tg} = \langle v_L, v_S, v_R \rangle$ where v_R is the largest, v_S typically medium and v_L is the smallest acceptable (or meaningful) value of the slope. In practice, we determine the context by means of differences of values of the time series related to a given (basic) time interval (for example, 12 months, 31 days, 1 hour, etc.).

Definition 4. Let $X|\overline{\mathbb{T}}$ be a time series restricted to $\overline{\mathbb{T}} \subset \mathbb{T}$ and $\mathrm{DTr}(X|\overline{\mathbb{T}})$ be direction of its trend due

*) These coefficients can be precisely equal to 1 only in case that $f = x$.

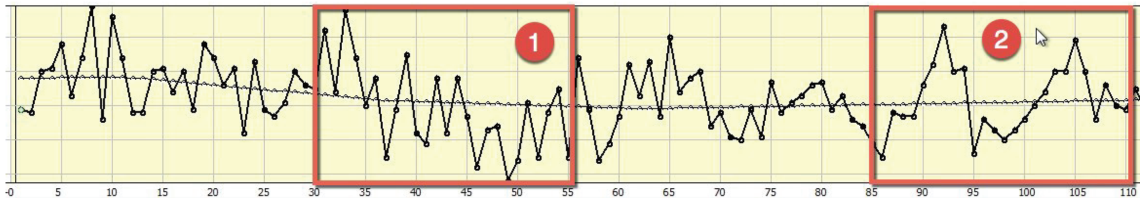


Figure 1: Example of the analysis of a real time series using F-transform: trend of the time series and its evaluation with respect to the context $w_{\mathrm{DTr}} = \langle 0, 860, 2150\rangle$ that was determined from the largest difference of values over 12 months. The trend was estimated using Theorem 2 by specifying the periodicity $\overline{T} = 51$ that is the longest periodicity detected by periodogram. Two rectangles mark areas in which local trend was linguistically evaluated using coefficients (4) as follows: *1 — clear decrease* and *2 — negligible increase*. Trend of the whole time series is *stagnating*.

to Definition 2. Let w_{DTr} be the corresponding context. Then intension of the evaluative predication "*Trend of* $X|\overline{\mathbb{T}}$ is $\langle$gradient$\rangle$"[†)] that evaluates direction of the trend of X in the time interval $\overline{\mathbb{T}}$ is

$$\mathrm{Int}(\textit{Trend of } X|\overline{\mathbb{T}} \text{ is } \langle\text{gradient}\rangle) = \mathit{LPerc}(\mathrm{DTr}[X|\overline{\mathbb{T}}], w_{tg}) \quad (17)$$

where

$$\langle\text{gradient}\rangle := \mathsf{stagnating}|\langle\text{hedge}\rangle\langle\text{sign}\rangle, \quad (18)$$

$$\langle\text{sign}\rangle := \mathsf{increasing}|\mathsf{decreasing}, \quad (19)$$

$$\langle\text{hedge}\rangle := \mathsf{negligibly}|\mathsf{slightly}|\mathsf{somewhat}| \mathsf{clearly}|\mathsf{roughly}|\mathsf{sharply}|\mathsf{significantly}. \quad (20)$$

Note that, because (17) is intension, extension of the corresponding evaluative predication in the context w_{DTr} is $\mathrm{Ext}_{w_{tg}}(\textit{Trend of } X|\overline{\mathbb{T}} \text{ is } \langle\text{gradient}\rangle) \underset{\sim}{\subseteq} \mathbb{R}$ that is a fuzzy set of real numbers being values of the tangent (4). Few more details can be found in [17].

Remark 2. Because we also distinguish increase and decrease of time series, we must distinguish positive and negative context w_{DTr} for its trend. The distinguished points are then ordered as follows: $-v_R < -v_S < -v_L \leqslant v_L < v_S < v_R$. We usually put $-v_L = v_L = 0$.

3.3. *Forecasting direction of trend of a group of related time series*

3.3.1. *The main principle*

Let us consider a group of time series $G = \{X_1, \ldots, X_m\}$ defined on the same time domain $\mathbb{T}$. Furthermore, let $\mathbb{T}' \supset \mathbb{T}$ be a new time domain. Our task is to extrapolate values of $X_j \in G$, $j = 1, \ldots, m$ to $X_j|(\mathbb{T}' \setminus \mathbb{T})$ on the basis of the known values of $X_j|\mathbb{T}$.[*)] The method for finding the former is called *forecasting*. There are many forecasting methods mostly formulated using probability theory (cf. [3,6,9]). In this paper, we consider the method described in [20] that are based on combination of the F-transform and fuzzy natural logic. Moreover, we will not discuss forecasting of the values of $X_j|(\mathbb{T}' \setminus \mathbb{T})$ but only forecasting of the trend $\mathrm{DTr}(X_j|(\mathbb{T}' \setminus \mathbb{T}))$. Recall that the trend is the basic characteristics of time series and so its forecast, if reliable, may be quite often sufficient.

Let us specify a fuzzy partition $\mathscr{A} = \{A_0, \ldots, A_n\}$ on $\mathbb{T}$. To each time series $X_j \in G$ we compute its F^1-transform $\mathbf{F}^1[X_j|\mathbb{T}]$ and form an adjoint time series

$$\mathbf{B}^1[X_j] = \langle \beta_1^1[X_j], \ldots, \beta_{n-1}^1[X_j]\rangle \quad (21)$$

where $\beta_i^1[X_j]$, $i = 1, \ldots, n-1$ are the coefficients (4).

Note that each time series $\mathbf{B}^1[X_j]$ in (21) characterizes course of trend of the corresponding time

[†)] The pedantic expression (17) should say something about "direction of trend", i.e., for example "direction of trend is increasing". The real language, however, is quite often not very precise and so, the expression "trend is increasing" is common.

[*)] This notation is with respect to the definition (12) superfluous. We use it, however, to keep the notation consistent.

series X_j. Our goal is to find time series with related course. This can be done using the well known *Spearman rank correlation coefficient* r_{jk}. The reason for using it is its ability to capture monotonous relations (increase-decrease) between variables, i.e., this coefficient not limited to measuring linear dependence only. It is also less sensitive to outliers.

First, we compute the Spearman correlation matrix $R = (r_{jk})_{\substack{j=1,\dots,m\\k=1,\dots,m}}$ where each $r_{j,k}$ is the Spearman correlation between the adjoint time series $\mathbf{B}^1[X_j]$ and $\mathbf{B}^1[X_k]$ (21) and estimate its significance (w.r.t. a chosen significance level α). To each significant r_{jk} we generate a linguistic description from the corresponding adjoint time series $\mathbf{B}^1[X_j], \mathbf{B}^1[X_k]$. This is done using the learning procedure developed in FNL (cf. [15,20]). Namely, each couple of values $(\beta_1^1[X_j], \beta_1^1[X_k]_i)$ generates a fuzzy/linguistic rule

$$\mathscr{R}_{jk} := \text{IF } \beta_1^1[X_j] \text{ is } \mathscr{C} \text{ THEN } \beta_1^1[X_k] \text{ is } \mathscr{D} \quad (22)$$

for all $i = 1,\dots,n-1$ where "β_i^{1,X_j} is $\mathscr{C}$", "β_i^{1,X_k} is $\mathscr{D}$" are evaluative linguistic predications generated using the procedure of *local perception* $LPerc(\beta_i^{1,X_j}, w_{tg,j})$, $LPerc(\beta_i^{1,X_k}, w_{tg,k})$ with respect to the respective linguistic contexts $w_{\mathrm{DTr},j}, w_{\mathrm{DTr},k}$.

3.3.2. *Forecasting procedure*

The overall procedure leading to forecast of trend of all time series from a given group G is the following:

1. Define a fuzzy partition $\mathscr{A} = \{A_0,\dots,A_n\}$ on the time domain $\mathbb{T}$ and for each time series $X_j \in G$ compute the F^1-transform $\mathbf{F}^1[X_j]$ and form the adjoint time series $\mathbf{B}^1[X_j]$ in (21) to all $X_j \in G$.

   The width of the basic functions A_i depends on volatility of the time series. If they are more volatile then the width should be shorter, if all time series are only slightly volatile then we can set A_i wider (more precise rules are not so far available).

2. Compute the Spearman correlation matrix R and divide G into subgroups $H_1,\dots,H_p$ of mutually dependent time series for all significant correlation coefficients.[†)]

3. In each subgroup H_l choose the principal time series X_{j_0} and generate a linguistic description consisting of the rules $\mathscr{R}_{j_0k}$ (cf. (22)) for all couples of adjoint time series $\mathbf{B}^1[X_{j_0}], \mathbf{B}^1[X_k]$ from the same subgroup H_l (i.e., for $X_{j_0}, X_k \in H_l$).

   Let us emphasize that the principal time series cannot be determined using formal means but only by expert. The reason is that they correspond to measurements of exogenous variables that depend on character of the modeled process.

4. Extend the time domain $\mathbb{T}$ to $\mathbb{T}'$ and forecast the principle time series X_{j_0} to the future time domain $\mathbb{T}' \setminus \mathbb{T}$. The forecast can be realized using arbitrary method; we suggest our method described in [20].

5. Determine a fuzzy set (basic function) A over $\mathbb{T}' \setminus \mathbb{T}$ and generate linguistic evaluation $\mathrm{Int}(\text{Trend of } X_{j_0}|(\mathbb{T}' \setminus \mathbb{T}) \text{ is } \langle\text{gradient}\rangle) = LPerc(\mathrm{DTr}(X_{j_0}|(\mathbb{T}' \setminus \mathbb{T})), w_{tg})$ of the direction of its future trend in the sense of Definition 4.

6. For each subgroup H_l, $l = 1,\dots,p$, do the following: Using the linguistic descriptions (22) generated for each couple of time series $X_{j_0}, X_k \in H_l$ compute estimation of the forecast $\mathrm{DTr}(X_k|(\mathbb{T}' \setminus \mathbb{T}))$ of the future trend of all $X_k \in H_l$ (over $\mathbb{T}' \setminus \mathbb{T}$) from the forecast $\mathrm{DTr}(X_{j_0}|(\mathbb{T}' \setminus \mathbb{T}))$ of the principle time series X_{j_0}. Finally, generate the linguistic evaluations

$$\mathrm{Int}(\text{Trend of } X_k|(\mathbb{T}' \setminus \mathbb{T}) \text{ is } \langle\text{gradient}\rangle),$$

   $X_k \in H_l$.

[†)] We must be careful because the time series X_j, X_k belong to the same subgroup H_l if $r_{jk} \geqslant r_{crit}$ where r_{crit} is a critical value. But we cannot assume transitivity when delineating these subgroups, i.e., $r_{jk} \geqslant r_{crit}$ and $r_{kl} \geqslant r_{crit}$ does not imply that $r_{jl} \geqslant r_{crit}$.

3.4. Experiment

The above described procedure was tested on a group of 10 time series $G = \{$NN101, NN102, NN103, NN105, NN109, NN111, N1757, ..., N1760$\}$ taken from the NN3 competition.[‡)] All the time series are defined on the domain $\mathbb{T}' = \{1,\dots,127\}$. First, we cut $\mathbb{T}'$ to $\mathbb{T} = \{1,\dots,114\}$ and defined on it a fuzzy partition $\mathscr{A} = \{A_0,\dots,A_{19}\}$. The reason is that the remaining part $\mathbb{T}' \setminus \mathbb{T}$ can be used as a *testing part* so that we can compare the computed data with the reality. Of course, the testing part is used only for comparison of results and never used in computation of the forecast.

The time series from the chosen group G are depicted in Figure 2. The procedure was the following:

1. For each time series $X \in G$, the direct F^1-transform $\mathbf{F}^1[X] = (F_0^1[X](t),\dots,F_{19}^1[X](t)$ was computed. Because the first and the last components are subject to big error (the reason is that the corresponding basic functions are incomplete), we omitted them from further elaboration.

2. To each time series $X \in G$ we formed the adjoint time series

$$\mathbf{B}^1[X] = \{\beta_1^1[X],\dots,\beta_{18}^1[X]\}.$$

3. To each couple of adjoint time series $\mathbf{B}^1[X_1], \mathbf{B}^1[X_2]$ we computed Spearman correlation coefficient $r_{X_1X_2}$.

4. Taking into account all significantly big correlation coefficients (with a significance level $\alpha = 0.05$), we formed the following two subgroups of mutually dependent time series:

$$\begin{aligned} H_1 &= \{\mathbf{NN102}, \text{NN101, NN103, NN111,}\\ &\qquad \text{N1757, N1758, N1760}\},\\ H_2 &= \{\mathbf{N109}, \text{N1759}\}. \end{aligned}$$

The principal time series in each subgroup is printed in boldface. Let us denote them by P_1, P_2, respectively. The time series NN105 does not depend on any other time series from G and so, it forms a one-element subgroup..

5. For each time series $\mathbf{B}^1[X]$, $X \in H_i$, $i = 1,2$ we specified the linguistic context $w_{tg,X}$:

Linguistic context for trend of time series		
Series	Negative $-w_{\mathrm{DTr}}$	Positive w_{DTr}
NN101	$\langle -55,-22,0\rangle$	$\langle 0,22,55\rangle$
NN102	$\langle -600,-240,0\rangle$	$\langle 0,240,600\rangle$
NN103	$\langle -5000,-2000,0\rangle$	$\langle 0,2000,5000\rangle$
NN111	$\langle -170,-68,0\rangle$	$\langle 0,68,170\rangle$
N1757	$\langle -344,-136,0\rangle$	$\langle 0,136,344\rangle$
N1758	$\langle -339,-135,0\rangle$	$\langle 0,135,339\rangle$
N1760	$\langle -180,-72,0\rangle$	$\langle 0,72,180\rangle$
NN109	$\langle -180,-72,0\rangle$	$\langle 0,72,180\rangle$
N1759	$\langle -219,-87,0\rangle$	$\langle 0,87,219\rangle$

6. Within each subgroup H_1, H_2, we generated a linguistic description charactering dependence of the trend of each time series $X \in H_i - \{P_i\}$ on the principal one $P_i \in H_i$, $i = 1,2$. The linguistic description has the form

$$\begin{array}{l} \mathsf{IF}\ \beta^1[P_i] \text{ is } \mathscr{C}_1 \ \mathsf{THEN}\ \beta^1[X] \text{ is } \mathscr{D}_1 \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ \mathsf{IF}\ \beta^1[P_i] \text{ is } \mathscr{C}_m \ \mathsf{THEN}\ \beta^1[X] \text{ is } \mathscr{D}_m \end{array} \tag{23}$$

and it is learned from the data, where $\mathscr{C}_k, \mathscr{D}_k$ are evaluative expressions.

For example, the linguistic description characterizing dependence of trend of NN103 on NN102 is the following:

IF $\beta^1[NN102]$ is ...	THEN $\beta^1[NN103]$ is ...
"– quite roughly big"	"– extremely big"
"more-or-less small"	"very roughly big"
"–typically medium"	"–roughly big"
"more-or-less medium"	"significantly big"

‡)http://www.neural-forecasting-competition.com/downloads/NN3/datasets/

IF $\beta^1[NN102]$ is ...	THEN $\beta^1[NN103]$ is ...
"quite roughly small"	"−roughly big
"−more-or-less small"	"rather medium"
"−rather medium"	"rather small"
"extremely small"	"−more-or-less small"
"roughly big"	"more-or-less small"
"−very roughly big"	"−rather medium"
"very roughly big"	"rather medium
"extremely big"	"more-or-less big
"−extremely big"	"−more-or-less big

Finally, we forecasted future courses of both principal time series NN102 and NN109 over the time period $\mathbb{T}' \setminus \mathbb{T} = \{115, \ldots, 127\}$ and linguistically evaluated their trends using the above described method (cf. [17]). Then, we computed forecast of trend of the other dependent time series from the forecast of the corresponding principal time series using the generated linguistic descriptions. Finally, we also generated the corresponding linguistic evaluation.

To be able to compare the forecasted and real trends, we computed the latter from the real data over the testing period $\mathbb{T}' \setminus \mathbb{T}$. Recall, that the testing parts of all the time series were detached from them in advance and were never used for the forecast. The computed trend was obtained from the corresponding learned linguistic descriptions (23) using the PbLD[§] method. The linguistic evaluations were obtained from values of β^1 using the procedure of local perception (8). The results are summarized in Tables 1 and 2.

As can be seen, all trends computed from the principal time series in the given subgroup using the generated linguistic description are in a good accordance with the forecasted trend, and also with the trend computed from the real data afterwards .

4. Conclusion

In this paper, we developed a new method for forecasting trends of a group of time series with similar course of their local trends. The method is based on application of the F-transform and selected methods of fuzzy natural logic.

The idea is based on estimation of the local trends using the F^1-transform components. The time series are then grouped together according to similarity of their local trends. The forecast of the future trend of all the time series is computed on the basis of a dependence of each time series from the given group on one principal time series whose forecast is known.

Further work will be focused on improvement of this method in the direction of making it more precise and also on the possibility to compare time series of various lengths. For this goal, we suppose also to use some other special methods, such as dynamic time warping [2].

Acknowledgments

The authors acknowledge that this paper was supported by the project LQ1602 IT4Innovations excellence in science.

References

1. J. Anděl, Statistical Analysis of Time Series, SNTL, Praha, 1976 (in Czech).
2. D. Berndt, J. Clifford, Finding patterns in time series: a dynamic programming approach, Advances in Knowledge Discovery and Data Mining (1996) 229–248.
3. A. Bovas, J. Ledolter, Statistical Methods for Forecasting, Wiley, New York, 2003.
4. G. Duncan, W. Gorr, J. Szczypula, Forecasting analogous time series, in: J. Armstrong (Ed.), Principles of Forecasting, Vol. 30 of International Series in Operations Research & Management Science, Springer US, 2001, pp. 195–213.
5. E. Hajičová, B. H. Partee, P. Sgall, Topic-Focus Articulation, Tripartite Structures, and Semantics Content, Kluwer, Dordrecht, 1998.
6. J. Hamilton, Time Series Analysis, Princeton, Princeton University Press, 1994.
7. R. Hyndman, G. Athanasopoulos, Forecasting: principles and practice, OTexts, http://otexts.org/fpp/, 2013.
8. J. Kacprzyk, A. Wilbik, S. Zadrożny, Linguistic summarization of time series using a fuzzy quantifier driven aggregation, Fuzzy Sets and Systems 159 (2008) 1485–1499.
9. B. Kedem, K. Fokianos, Regression Models for Time Series Analysis, J. Wiley, New York, 2002.
10. V. Kreinovich, I. Perfilieva, Fuzzy transforms of higher order approximate derivatives: A theorem,

[§] Perception-based Logic Deduction — see [18].

Table 1: Group G_1:

Series	Forecast of trend	Real trend	Computed trend
NN102	**rough decrease** ($\beta^1 = -428$)	**fairly large decrease** ($\beta^1 = -467$)	—
NN101	**clear decrease** ($\beta^1 = -25$)	**clear decrease** ($\beta^1 = -16$)	**rough decrease** ($\beta^1 = -36$)
NN103	**clear decrease** ($\beta^1 = -2460$	**quite large decrease** ($\beta^1 = -4430$)	**significant decrease** ($\beta^1 = -4967$)
NN111	**clear decrease** ($\beta^1 = -94$	**fairly large decrease** ($\beta^1 = -136$)	**significant decrease** ($\beta^1 = -168$)
N1757	**clear decrease** ($\beta^1 = -35$)	**clear decrease** ($\beta^1 = -106$)	**clear decrease** ($\beta^1 = -146$)
N1758	**clear decrease** ($\beta^1 = -83$)	**clear decrease** ($\beta^1 = -90$)	**clear decrease** ($\beta^1 = -135$)
N1760	**clear decrease** ($\beta^1 = -43$)	**clear decrease** ($\beta^1 = -61$)	**clear decrease** ($\beta^1 = -75$)

Table 2: Group G_2:

Series	Forecast of trend	Real trend	Computed trend
NN109	**clear decrease** ($\beta^1 = -30$)	**clear decrease** ($\beta^1 = -33$)	—
N1759	**slight decrease** ($\beta^1 = -10$)	**clear decrease** ($\beta^1 = -22$)	**clear decrease** ($\beta^1 = -23$)

Fuzzy Sets and Systems 180 (2011) 55–68.
11. H. Lütkepohl, New Introduction to Multiple Time Series Analysis, Springer, Berlin-Heidelberg, 2005.
12. V. Novák, Perception-based logical deduction, in: B. Reusch (Ed.), Computational Intelligence, Theory and Applications, Springer, Berlin, 2005, pp. 237–250.
13. V. Novák, A comprehensive theory of trichotomous evaluative linguistic expressions, Fuzzy Sets and Systems 159 (22) (2008) 2939–2969.
14. V. Novák, Reasoning about mathematical fuzzy logic and its future, Fuzzy Sets and Systems 192 (2012) 25–44.
15. V. Novák, On modelling with words, Int. J. of General Systems 42 (2013) 21–40.
16. V. Novák, Fuzzy Natural Logic: Towards Mathematical Logic of Human Reasoning, in: E. Trillas, R. Seising, J. Kacprzyk (Eds.), Fuzzy Logic: Towards the Future, Springer, 2015.
17. V. Novák, V. Pavliska, I. Perfilieva, M. Štěpnička, F-transform and fuzzy natural logic in time series analysis, in: Proc. Int. Conference EUSFLAT-LFA'2013, Milano, Italy, 2013, pp. 40–47.
18. V. Novák, I. Perfilieva, On the semantics of perception-based fuzzy logic deduction, International Journal of Intelligent Systems 19 (2004) 1007–1031.
19. V. Novák, I. Perfilieva, M. Holčapek, V. Kreinovich, Filtering out high frequencies in time series using F-transform, Information Sciences 274 (2014) 192–209.
20. V. Novák, M. Štěpnička, A. Dvořák, I. Perfilieva, V. Pavliska, L. Vavříčková, Analysis of seasonal time series using fuzzy approach, Int. Journal of General Systems 39 (2010) 305–328.
21. V. Novák, M. Štěpnička, J. Kupka, Linguistic descriptions: their structure and applications, in: H. Larsen, M. Martin-Bautista, M. Vila, T. Andreasen, H. Christiansen (Eds.), Flexible Query Answering Systems, Springer, Berlin, Heidelberg, 2013, pp. 209–220.
22. I. Perfilieva, Fuzzy transforms: theory and applications, Fuzzy Sets and Systems 157 (2006) 993–1023.
23. I. Perfilieva, M. Daňková, B. Bede, Towards a higher degree F-transform, Fuzzy Sets and Systems 180

(2011) 3–19.
24. N. Wagner, M. Z., S. Schellenberg, C. Chiriac, A. Mohais, Intelligent techniques for forecasting multiple time series in real-world systems, Int. J. of Intelligent Computing and Cybernetics 4 (3) (2011) 284–310.

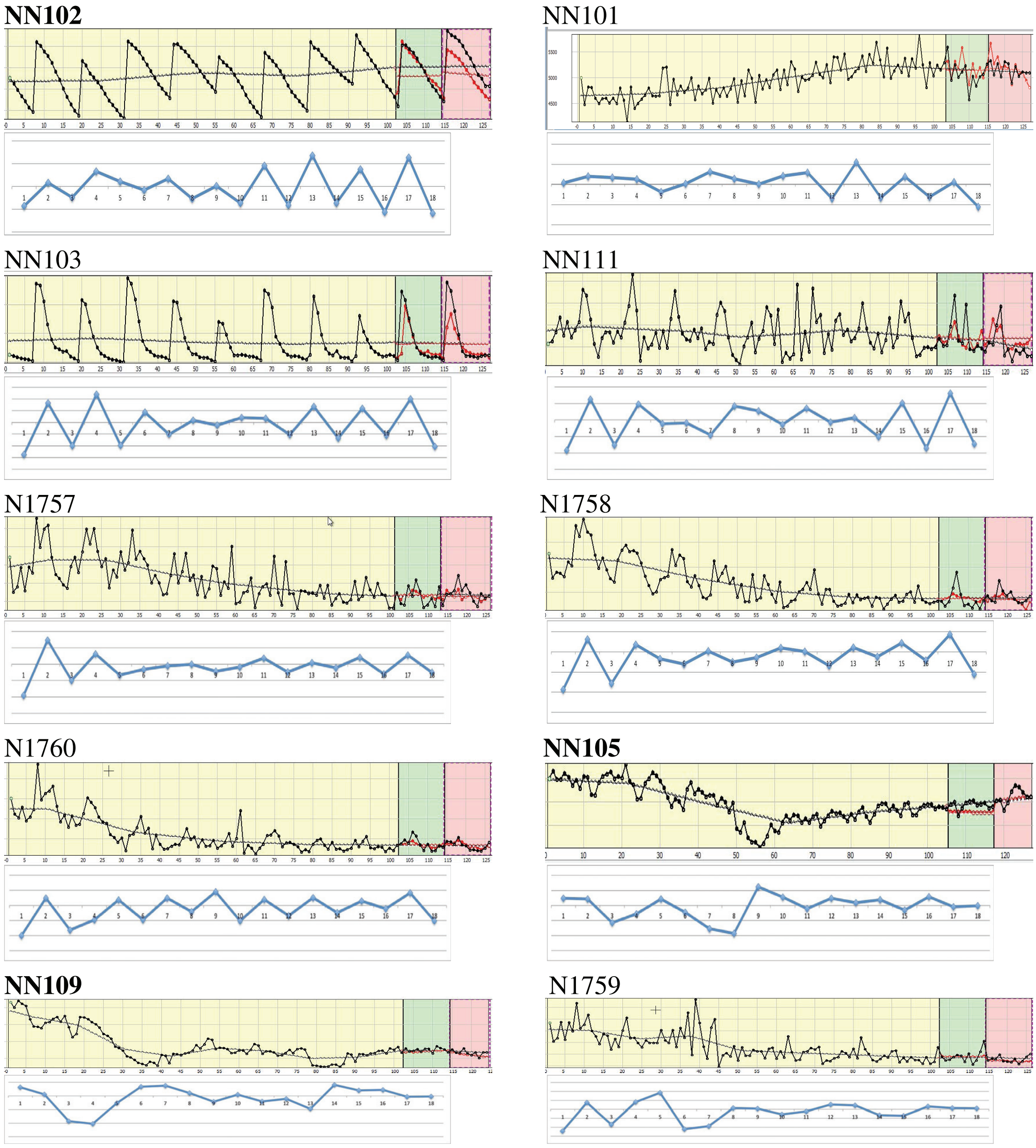


Figure 2: The time series chosen in the experiment. Each time series is depicted over the time domain $\mathbb{T}'$ together with the estimated trend-cycle. The time domain $\mathbb{T}'$ is divided into 3 parts: *learning, validation* and *testing* ones. Namely, $\mathbb{T}$ is divided into learning and validation part and $\mathbb{T}' \setminus \mathbb{T}$ is the testing part. The figure contains forecast of each time series over the testing part $\mathbb{T}' \setminus \mathbb{T}$ compared with the reality. Below each time series is depicted also the adjoint time series (21) over the time domain $\mathbb{T}$ (i.e., over learning and validation parts). The principal time series are marked in boldface letters.