Proceedings of the International Multiconference on
Computer Science and Information Technology pp. 259–263
ISBN 978-83-60810-27-9
ISSN 1896-7094

# A two-level algorithm of time series change detection based on a unique changes similarity method

Tomasz Pełech-Pilichowski
AGH University of Science and Technology
Department of Applied Computer Science
Faculty of Management
ul. Gramatyka 10, 30–067 Krakow, Poland
Email: tomek@agh.edu.pl

Jan T. Duda
AGH University of Science and Technology
Department of Applied Computer Science
Faculty of Management
ul. Gramatyka 10, 30–067 Krakow, Poland
Email: jdu@agh.edu.pl

***Abstract*—In the paper, a novel two level algorithm of time series change detection is presented. In the first level, to identify non-stationary sequences in processed signals preliminary detection of events is performed with short-term prediction comparison. In the second stage, to confirm changes detected in first level a unique changes similarity method is employed. Detection of changes in non-stationary time series is discussed, implemented algorithms are described and results produced on exemplary four financial time series are showed.**

## I. Introduction

INFORMATION on changes in analysed time series is relevant to detect alarm situations, in particular, when signals are processed in real-time systems. Implementation of dedicated algorithms to event detection from non-stationary time series requires considering many factors, such as statistical and frequency-domain data characterisation, different type of short- and long-term changes, time lags between events or sampling frequency. Results given with such algorithms are helpful in advanced data analysis, prediction and finally—decision making process.

Our research are aimed at advanced algorithms of time series change detection, based on statistical approach, signal analysis [4] and employing immune paradigm to event detection support [15]. In particular, we proposed a method to gain information on early symptoms of significant changes by analysis of short-term prediction efficiency [14], [16], [13], [15] through comparison between zero-order-prediction (ZOP)/zero-order-hold (ZOH) model [1] and adaptive Holt predictor [9]. Proposed method is suitable for a particular type of changes in signals, for example occur in financial time series which consist of many non–random components. We also showed that application of artificial immune systems techniques [8], [21] is a way to improve event detection and thus prediction efficiency (especially prediction error variance). In the paper [15] we proposed original idea of implementation immune based approach to early detection of significant changes in time series (long term changes of mean value), where detection is decomposed into two stages: fast detection of nonselfs (employing fast statistical and prediction procedures) and then more accurate (and more time–consuming) recognition of the nonself type and the system recovery. In the papers [15], [18], [17] an improvement of this idea aimed at long term event detection were described.

The aim of this paper is to provide detection method of changes in diagnostic signals to early detection of emergency situations. It has been designed for monitoring a set of diagnostic signals (medical, technical, financial ones) to capture differences, i.e. unusual behavior of selected signal (process variables). Such solution may be implemented as detector of rapid changes in signal (for example alarm situations as dangerous changes of technical process temperature, liquid pressure) or exceeding the threshold valued of process limitations (technology losses minimisation).

We present and analyse two-level algorithm of change detection in time series, based on simultaneous processing of two time series of fixed length in a moving window, to capture unique changes—which occur only in one of two processed diagnostic signals, i.e. are not results of the same external factors. In the first stage, preliminary detection of events is performed with one-step-ahead prediction comparison given with methods dedicated to stationary or non-stationary data [19], [1], [13]. In the second stage, to confirm changes detected in first level, a unique changes similarity method is employed.

Proposed event detector is tested on financial time series—stochastic processes with unknown random inputs, which are hard-predictable and difficult to perform reliable statistical analysis, compared to technical ones. Therefore, efficient event detection in financial ones with proposed algorithm will indicate its applicability to technical time series (as a starting point for further implementation and adjustment in order to detect well-defined, specific changes).

In this paper, we describe the detection algorithm as a part of large detectors set capable of—through two signals

This work was supported by the European Regional Development Fund, Grant no. UDA–POIG.01.03.01–12–171/08/00.

concurrent processing—handling possible event patterns in processed time series. Such idea can be further developed as an autonomous event detector and implemented as one component (e.g. agent) of available set supervised by T-lymphocyte (immune-like event detector [15]).

## II. Event detection from time series

Event detection from time series is aimed at identifying short and long term periods of uncommon series behaviour, analysed in moving window, to detect a change in stationarity due to random external factors. Event detection task may be viewed as finding corresponding probability distribution [13], [14], [16] or as unsupervised classification task (one-class classification) where one may describe only one class and a method to distinguish between possible object (decision boundary between normal and anomaly class) with appropriate (training set tested) mapping function [19].

Considering long-term signal changes, detection of short-term "announcing" events may be a way to perform faster computation compared to classical, but robust statistical procedures of long-term changes detection. Such events may precede statistical properties changes in processed series and generate the need of adjust of assumed analysis window width, i.e. change of detection delay thus probabilities of false alarm and undetected events [12].

An implementation of given algorithmic method depends on input data properties (statistical, frequency, dimensionality, completeness), attributes of events (amplitude, duration, periodicity, applicability, coincidence, delay) or permissible detection error. Many algorithms require adjusting parameters and signals selection to specific conditions of dataset. In practice, usually the most important is appropriate selection of classification algorithm which directly affects the detection reliability and a possibility of implementation for heterogeneous data sets.

An application of a detection method suitable for dataset requires is to find specific attributes of events and the use of dependencies between processed signals and occurring events (often delayed). To achieve this aim, a number of methods can be exploited to time series quantitative analyses, such as analysis of the frequency of events [10], analysis of trends, patterns and characteristics similarity [12] or statistical methods for testing deviants and outliers, algorithms based on neural networks, genetic algorithms [11] and other data mining techniques applied to event detection from time series [7], [10], including methods based on similarity measures, in particular distance ones [22]. However, reliable results produced by such algorithms usually depend on length of series or learning [15] and highly accurate estimation of parameters such as mean value, information about trend, dynamics parameters and random component dispersion—which are an input for statistical tests [20]. In this paper we focus on methods based on similarity measures and signal analysis.

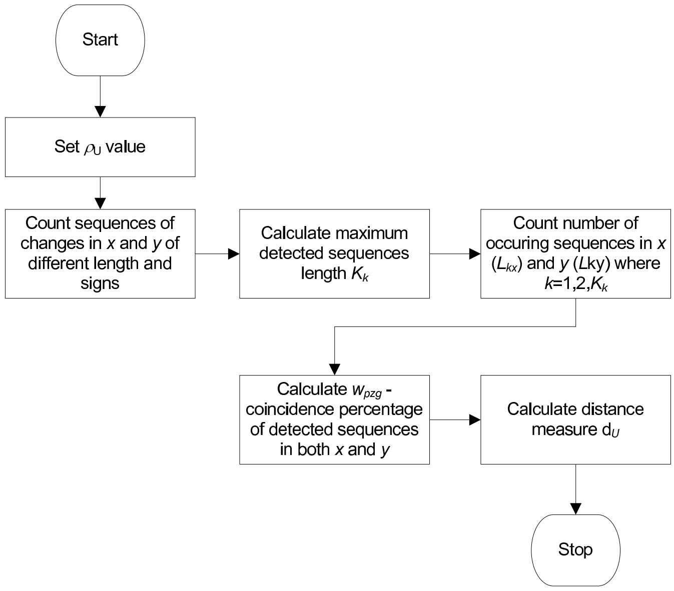


Fig. 1. A block diagram of unique changes similarity algorithm.

## III. A unique changes similarity method

Considering event detection from one time series, unusual behavior (non-stationarity) may be recognised as abrupt changes of mean value (visible as deviations exceeding fixed threshold), sequences of changes (of the same or different signs) or patterns understood as specific configuration of deviations. Considering a pair of time series, original (unique) event detection may be viewed as a selection of events occurring only in one processed dataset. To capture such events we propose distance-based similarity measures [5] applying in a moving window.

One of novel measures, designed and developed for a particular problem is an unique changes similarity method (distance-like one), first mentioned in our earlier papers [5], [17], denoted as method $U$ and measure d$U$. It is dedicated to identify unique changes which are considered as sequences of deviations of the same sign with different sequence length (1,2,...,$N$) exceeding an arbitrary fixed threshold—$\rho_U$ in two processed diagnostic signals of the length $N$ (constant moving window width $N$ is assumed).

The goal is to calculate a similarity coefficient—distance measure d$U$ (see equations 1–3) based on counting of sequences of fixed length (subsequences of different length or signs are treated as different ones). A block diagram of such detection for one sample (for assumed window width $N$) is shown in figure 1.

*'Distance measure'* term is used instead of *distance* or *metric* because all required conditions related to formal definition of metric are not satisfied, in particular symmetry condition.

To compute d$U$ value three partial calculations are performed (eq. 1–3). Coincidence percentage of detected sequences of deviations in analysed signals is calculated as:

$$w_{pzg} = w_p \sum_{k=1}^{K_k} k \cdot \min(L_{kx}, L_{ky}) \tag{1}$$

where $L_{ks}$ denotes a number of sequences of length $k$ in series $s$ ($x$ or $y$). To compute coefficient wp the following formula is employed:

$$w_p = 1/(\sum_{k=1}^{K_k} k \cdot \max(L_{kx}, L_{ky})) \tag{2}$$

Finally, distance measure $\mathrm{d}U$ is computed as:

$$\mathrm{d}U = 1 - w_{pzg} \tag{3}$$

Notice, in case of similar sequences in both series $\mathrm{d}U$ will have 0 value ($wpzg$ near 1). As threshold $\rho_U$ value, standard deviation in moving window or its multiplicity may be used, however, arbitrary adjusting may be valuable.

To avoid an impact of time delay between events, for instantaneous values of computed distance measures $\mathrm{d}U$, a tolerance (denoted as $L_{tol}$) is assumed as permissible delay between analysed signals. For each time instant $t$, a number ($L_{tol}$) of distance measures $\mathrm{d}U_n$ is calculated for time instants $n$ ($n = t - L_{tol} + 1, \ldots, t$). Finally, $\mathrm{d}U_n$ for which the lowest value was obtained is taken as the measure $\mathrm{d}U$ between two subseries for sample $t$.

## IV. Two-level change detection algorithm

Proposed algorithm is designed to different types of diagnostic signals. We have chosen financial time series as difficult ones to perform event detection or forecasting because of existing many non-random components. Notice, that financial time series are available from heterogeneous sources, however, causal-consecutive dependencies between time-lagged events are possible.

Proposed hybrid algorithm performs detection in two steps:

1) preliminary detection of short-tem non-stationarities in signals;
2) unique changes detection with method U (see eq. 1–3).

Considering detection **in the first level**, we have chosen an approach based on one-step-ahead prediction error comparison with methods efficient for stationary and non-stationary data [14], [16], performed for both signals separately.

Autocorrelation of daily increments of typical financial time series is statistically insignificant [3], thus during usual behavior no autocorrelation of daily returns may be expected [1]. The most efficient short-term prediction (the minimum variance error) is achievable by extrapolation of averaged returns [15], i.e. by the ZOP or ZOH model [1], [15]. To predict non-stationary data, three-parameters adaptive Holt method is employed (parameters $\alpha$, $\beta$, $\gamma$ adjusted for each sample) [13], which is more flexible than classical one [9]. The comparison allows to basic identification of non-stationarities, i.e. non-random components (advantage Holt method over ZOP/ZOH) in relatively short time and low computational resources consumption.

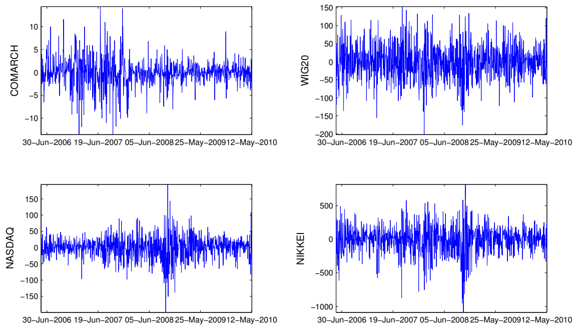


Fig. 2. Examined financial time series - daily returns (one-day increments).

**In the second level**, two signals are taken to calculate unique changes similarity measure ($\mathrm{d}U$) thus differences between found sequences of deviations. Finally, change in a signal is found after non-stationarity confirmation with $\mathrm{d}U$ value exceeding fixed threshold $\rho_U$. In this stage, only unique (original), non-stationary sequences are selected (accuracy depended on $\rho_U$).

Detection schema described above provides a possibility of minimizing false alarms as a result of identifying changes that are, in fact, random fluctuations. In addition, proposed two-level algorithm decrease computation time consuming which through switching similarity detection method U for relatively small amount of samples only.

## V. Calculation results

Four financial time series, individual (Comarch) and aggregated (WIG20, Nikkei, Nasdaq), has been used as exemplary dataset to analyse the proposed two-level detection algorithm. We have chosen local (Comarch, WIG20) and global time series (Nikkei, Nasdaq) from 2006–05–22 to 2010–05–14 (1040 samples).

Financial time series (see daily returns depicted in fig. 2) represent stochastic processes with unknown random inputs (in the simplest approach, such series can be considered as Wiener processes). Acceptable detection results obtain on such hard-predictable sequences of samples allow to assume that proposed solution applied to diagnostic signals generated by technical devices (which don't contain many variable components) will result in more accurate event detection.

For each processed time series, subsequent samples were taken to compute one-sample-ahead prediction with naive method (level 1; exemplary results of preliminary detection are shown in table I—for all analysed series non-stationarity identified for about 5%) while distance measure $\mathrm{d}U$ was calculated in moving window of constant length $N = 22$ samples (see fig. 1) between series 1 and 2 and in a reverse order (because of asymmetry of $\mathrm{d}U$ measure mentioned above). Processed signals were unified with standard deviation to achieve comparable orders of magnitude.

In the example discussed in this paper, to provide detection reliability, threshold sequence length was fixed to $\rho_{PC} = 3$.

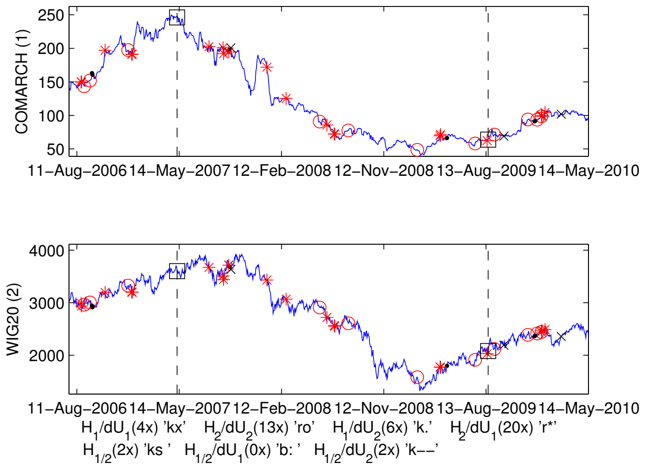


Fig. 3. Results of change detection for Comarch (1) and WIG20 (2)—depicted with solid lines—with proposed two-level algorithm. Recognized non-stationarity in series no.1 confirmed with d$U_1$ denoted as 'cross' ($H_1$/d$U_1$) and d$U_2$—as 'dot' ($H_1$/d$U_2$); non–stationarity in series no.2 confirmed with d$U_1$—'asterisk' ($H_2$/d$U_1$) and d$U_2$—'circle' ($H_2$/d$U_2$); changes $H_1$ and $H_2$ detected for the same samples—'square' ($H_{1/2}$), changes ($H_{1/2}$) confirmed with d$U_1$—'dotted line' and with d$U_2$—'dashed line'.

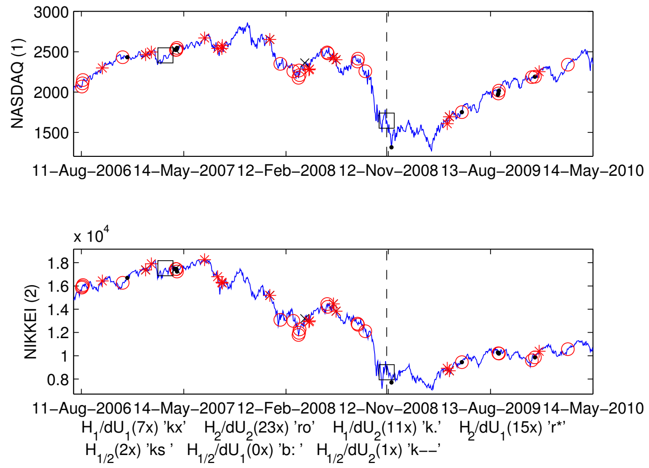


Fig. 4. Results of change detection for Nasdaq (1) and Nikkei (2)—depicted with solid lines—with proposed two-level algorithm. Description of symbols: see fig. 3.

The main parameter of the second-stage detection level ($\rho_U$) was set to 1 as an equivalent of signals standard deviation after performed unification. Values of these parameters were selected in an experimentally way (related to properties of four analysed time series) to achieve the smallest number of false alarms and undetected events.

Detection results produced with proposed two-level algorithm are depicted in figures 3–6 (with the sampled series in the background). The figures show confirmations of changes with short-term prediction comparison in the first level (denoted as $H_1$ and $H_2$) and distance measure d$U$ in the second level (eq. 1–3).

It may be seen in figures 3–6 that proposed method allows to detect local changes (denoted as 'cross', 'asterisk', 'circle' and 'dot'—see figure 3 caption), obtained with simple $H_1$ or $H_2$ confirmation with method U, and long-term changes, i.e. reversal in a trend (see vertical lines in fig. 3–6).

In the first case, a number of short-term changes were detected correctly, however, false alarms were found for some samples. Application of non-stationary detection confirmation (level 1) significantly reduces overall number of detected events (see table I and figures 3–6 caption).

In the second case, H1 and H2 changes are synchronised (non–stationarities detected in one signal only are rejected), which is illustrated as 'squares' in figures 3–6, and then—confirmed with d$U_1$ or d$U_2$ measure. This is relevant step to eliminate false alarms.

## VI. Conclusions

Proposed two-level algorithm of change detection in time series may be viewed as an inspiration for construction efficient detectors, adaptable to signal properties. It was found as

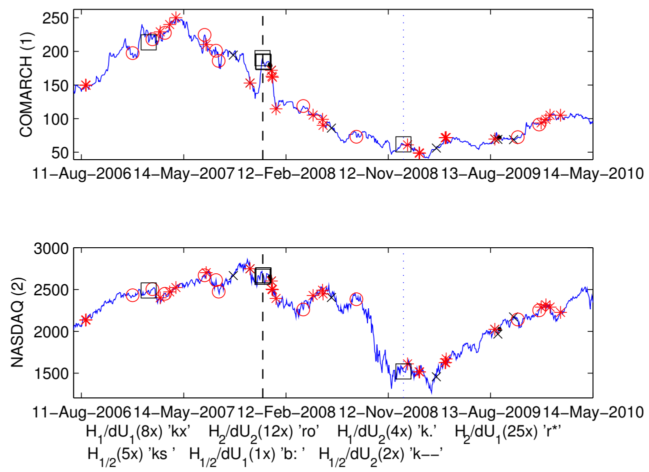


Fig. 5. Results of change detection for Comarch (1) and Nasdaq (2)—depicted with solid lines—with proposed two-level algorithm. Description of symbols: see fig. 3.

a suitable for short- and especially long-term changes of mean value through confirmation of non-stationary subsequences of processed signal. Acceptable detection results obtained on financial time series (stochastic processes with unknown random inputs, with many non-random components) allow to assume applicability end effectiveness for other types of time series, in particular, diagnostic signals generated by technical devices.

Calculation results presented in this paper show that to achieve reliable detection of symptoms of rapid changes in long-term trends in one processed time series, processing of series set of the same class (in the paper—financial ones) is valuable. Depending on the set (pair) of signals taken to calculate distance measure, detected changes may occur for different samples. This issue will be developed towards (1)

TABLE I
EXEMPLARY RESULTS OF DETECTION OF NON-STATIONARY SUBSEQUENCES IN PROCESSED TIME SERIES

| Time series (total number of samples: 1040) | Comarch | WIG20 | Nasdaq | Nikkei |
|---|---|---|---|---|
| Number of detected non-stationary subsequences with one-step-ahead prediction comparison | 60 | 50 | 42 | 59 |
| Percentage | 5.77% | 4.81% | 4.04% | 5.67% |

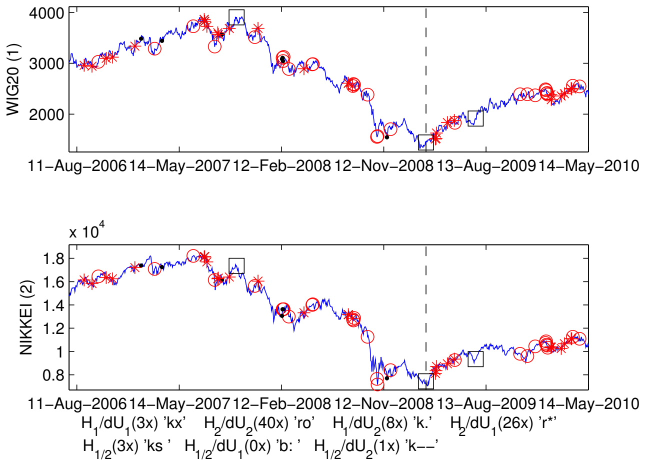


Fig. 6. Results of change detection for WIG20 (1) and Nikkei(2)—depicted with solid lines—with proposed two-level algorithm. Description of symbols: see fig. 3.

events recognition and (2) forecasting for one selected (based) signal through processing signals sets consist of many time series datasets.

Further research will be focused on implementing and testing of the proposed solution for technical diagnostic signals, especially processed in real-time systems. Studies will be required on implementation and testing of similarity methods dedicated to specific changes and patterns appearing in analysed signals, including time delays between events and elimination of false alarms. Further improvement appears to be obtained by modification of described distance measures towards variable length of subsequences taken to compute measure d$U$ (to this aim, fuzzy logic rules may be employed).

Possibility of adjusting parameters (threshold values, window widths, etc.) seems to be a promising way for further enhancement of the presented method which is suitable for the idea of immune-based adaptive detection, in particular as one of many detectors supervised by T-lymphocytes.

## REFERENCES

[1] Box G .E. P. and Jenkins G. M., *Analysis of Time Series* PWN, Warszawa, 1983

[2] Duda J. T., *Dobór parametrów algorytmu Page'a-Hinkleya przy ustalonych prawdopodobieństwach I i II Rodzaju* Unpublished paper, AGH–UST, 2005

[3] Duda J. T. and Augustynek A., *On possibilities of improvement of short-term predictions of stock indices with regression models* Company Management in Conditions of European Integration—Part 2. Economy, Informatics and Numerical Techniques, Ed. M.Czyz and Z.Cieciwa, AGH–UST University Press, 2004

[4] Duda J. T. and Pełech T., *Wykrywanie zdarzeń w szeregach finansowych z wykorzystaniem metod statystycznych* [In:] Inżynieria wiedzy i systemy ekspertowe, T.2, Edt. Grzech A., Oficyna Wydawnicza Politechniki Wrocławskiej, 2006

[5] Duda J. T. and Pełech-Pilichowski T., *Miary podobieństwa szeregów czasowych w detekcji zdarzeń* [In:] Systemy wykrywające, analizujące i tolerujące usterki, red. Kowalczuk Z., PWNT, 2009

[6] Farmer J. D., Packard N. and Perelson A., *The immune system, adaptation and machine learning* Physica D, vol.22, 1986, pp.187–204

[7] Guralnik, V. and Srivastava, J., *Event detection from time series data* [In]: Proceedings of the fifth ACM SIGKDD International Conference on Knowledge Discovery and Data mining, San Diego, California, USA, 1999, pp. 33–42

[8] Hofmeyr S. A. and Forrest S., *Immunity by Design: An Artificial Immune System* Proceedings of the Genetic and Evolutionary Computation Conference, San Francisco, 1999

[9] Holt C. C., *Forecasting seasonals and trends by exponentially weighted moving averages* Carnegie Institute of Technology, Pittsburgh, Pennsylvania, 1957

[10] Keogh E., Lonardi S. and 'Yuanchi' Chiu B., *Finding Surprising Patterns in a Time Series Database in Linear Time and Space* [In:] Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 550-556, Edmonton, Alberta, Canada, 2002

[11] Kingdon J., *Intelligent Systems and Financial Forecasting* Springer, 1997

[12] Mahoney M. and Chan P., *Trajectory Boundary Modeling of Time Series for Anomaly Detection* Workshop on Data Mining Methods for Anomaly Detection, Conference on Knowledge Discovery and Data Mining, 2005

[13] Pełech T., *Adaptive Holt's Forecasting Model Based on Immune Paradigm* Problemy osvoenia poleznyh iskopaemyh. Zapiski Gornogo Instituta, Sankt Petersburg State Mining Institute, 2006

[14] Pełech T. and Duda J. T., *Application of immune paradigm to monitoring of stock indices* Problems of Mechanical Engineering and Robotics, No.3, AGH-UST University Press, 2005

[15] Pełech T. and Duda J. T., *Event detection in financial time series by immune-based approach* Intelligent Information Processing and Web Mining. Advances in Soft Computing, Springer-Verlag, 2006

[16] Pełech T. and Duda J. T., *Immune Algorithm of Stock Rates Parallel Monitoring* Information Systems and Computational Methods in Management, Ed. Duda J.T., AGH-UST University Press, 2005

[17] Pełech-Pilichowski T, *Zastosowanie metod algorytmicznych w prognozowaniu szeregów finansowych* [In:] Zarządzanie organizacjami - finanse, produkcja, informacja, red. Howaniec H., Waszkielewicz W., Wyd. Akademii Techniczno-Humanistycznej, Bielsko-Biała, 2009

[18] Pełech-Pilichowski T. and Duda J. T., *Wykorzystanie podejścia immunologicznego do prognozowania szeregów czasowych* Automatyka: pølrocznik Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie, 2009

[19] Somayaji A. B., *Operating System Stability and Security through Process Homeostasis* Ph.D. thesis, University of New Mexico, 2002

[20] Wetherill G. B. and Brown D. W., *Statistical Process Control. Theory and Practice* Chapman and Hall, 1991

[21] Wierzchon S.T., *Sztuczne systemy immunologiczne. Teoria i zastosowania* Wyd. Exit, 2001

[22] Yang K. and Shahabi C., *A PCA-Based Similarity Measure for Multivariate Time Series*, ACM International Workshop on Multimedia Databases, 2004