Proceedings of the Federated Conference on
Computer Science and Information Systems pp. 321–327

ISBN 978-83-60810-22-4

A Hybrid Algorithm for Detecting Changes in Diagnostic Signals Received From Technical Devices

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

***Abstract*—In this paper, a hybrid two-level algorithm of the original changes in diagnostic signals received from multiple technical devices is presented. Research are aimed at identification of the changes, deviations or patters (events), through concurrent diagnostic signals processing, which occur in one selected signal (proposed algorithm is adjusted to omit concurrent and time-lagged changes). In the first stage, detection is based on non-stationarity detection with the short-term prediction comparison. In the second stage, dedicated distance-like measure is employed. Detection results obtained for sample random signals including simulated large deviations are presented.**

I. Introduction

TECHNICAL systems, for example Intelligent buildings ones, are often consist of technical networked devices supervised by a central unit or controller. Efficient system working depends on the reliable installed devices operation which may be successfully provided by the diagnostic signals concurrent monitoring. Such processing of available set of diagnostic signals, in particular in real time systems, may indicate the work status of devices or results in alarm notification resulting from defective device work, faults, power instability, loss of network security or generated signal reliability. On the other hand, selected signals may contain random, temporary changes, statistically insignificant, being response to the irrelevant devices interferences, networked and supervised by a central unit. Such changes can trigger false alarms or can directly influence the work of other connected devices (provide incorrect input signal) thus endanger unstable operation of the whole system.

Event detection from diagnostic signals is usually based on implementation of a detection algorithm capable of identify the well-defined, expected unusual behavior of the processed signal (short or long-term non-stationarities) [5]. Detection algorithms are often based on a statistical and frequency domain data characterization [2], [21], [8], adapted to sampling frequency and consistency of available dataset. In such cases, complex system monitoring is based on the individual signal processing which is not sufficient to identify coincidence of events in other analyzed time-series.

Considering diagnostic signals produced by multiple system devices, a reliable detection requires the use of dedicated algorithms based on concurrent or parallel processing of all available signals to capture non-random and relevant changes in selected (real-time monitored) one, strictly including time lags between occurring events as a possible effect of transmission delays, data queuing, assumed real-time regime disturbances.

The aim of this paper is to describe and present research progress on detection algorithms dedicated to capture original changes, deviations, or patterns, i.e. which are not an effect of the same external factors, thus occurring in the only one diagnostic signal – including time-delays between possible changes (avoiding changes presented in both signals – concurred or lagged). In the paper, a hybrid detection algorithm is presented, based on concurring processing of pairs of signals in a moving window of a fixed length where detection task splits into two levels: (1) a non-stationarity detection and (2) its confirmation with distance-like similarity method. Proposed approach is focused on reducing false alarms and early efficient detection of emergency situations in multiple diagnostic signal sets.

This idea is a development of earlier work aimed at the significant event detection from time series based on statistical signal analysis [5], distance-like methods [6], short-term prediction efficiency comparison [16],[18],[15], immune paradigm employed to event detection support [17], [20], [10], [22] and two-level algorithms dedicated to process signals in real time systems [19].

II. Event Detection From Signals Received From Multiple Devices

Event detection from time series is based on the processing of subsequent samples to identify an unusual process behavior, i. e. non-stationarities caused by external non-random factors. To achieve reliable results, the signal analyzes should be performed in a moving window of a fixed length [3], [5], [4] depending on intended impact of historical data. Event detection may be viewed as the unsupervised classification task where one class is described (one–class classification) and a formal method dedicated to distinguish between normal and anomaly class [10], [19].

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

Assuming that signals are generated by technical devices (for example intelligent building ones), during “normal” system behavior stationary is identified (lack of long-term and short-term abrupt changes of the mean value or variance); an appearance of non-stationarity may be recognized as the alarm signal or a specific event, i.e. pattern or unexpected change. Such sequences of changes may be visible as the deviations from the short/long term mean value exceeding a fixed threshold as the standard deviation multiplicity of arbitrary selected value. Depending on event detection task, changes of the same sign (positive or negative ones), different sign (or its absolute values) or patters – i.e. original configuration of deviations are tested (lack of changes between series of deviations of fixed length is often assumed).

To detect short and long-term changes in a selected diagnostic signal, one may use classical, robust procedures (for example, Page-Hinkley one [1]) which reasonable employing is often limited by the computation time or a need of long data sets processing which results in usually unacceptable time-delays [14]. Implementation of the detection algorithm strongly depends on the signal characterization (statistical, frequency), properties (dimensionality, completeness), attributes of possible events (amplitude, duration, periodicity, delay) or the assumed detection error [19].

Change detection from one signal is produced with implementation of the detection algorithm and usually it is sufficient for a short-tem changes identification, especially when one diagnostic signal from one device is obtained. For multiple signals, such approach may results in a number of false alarms or undetected events as a result of random signal changes (for example, short-term power differences, time delays during wire/wireless data transmission) or inability to identify a complex alarm situation (i.e. to detect a real device damage, multiple signals are often needed – temperature, input/output, revolutions per minute of central processing unit fan etc.). Moreover, significant changes may be announced by the short-term deviations in another signal. Thus, processing of a diagnostic series sets seems to be a promising way to capture such changes, deviations or patterns in a target signal.

Event detection from a selected diagnostic signal from the large available signals set (including heterogeneous ones) can based on the concurrent, parallel or distributed time-series processing, to identify both single changes and dependencies between events detected from individual signals (with suitable algorithms). Such complex processing may be a way to faster detection through the short–term "announcing" events recognition which can precede long-term changes of time series the statistical properties [14].

There is a number of methods dedicated to time series quantitative analyzes [19], like the statistical analysis of the frequency of events [12], trends, deviants and outliers, the patterns and characteristics similarity comparison [14]. Algorithms based on neural networks [7], genetic algorithms [13] and other data mining techniques including similarity measures and distance ones [23], [9] also may be employed [8], [12]. Although such results can provide the reliable detection; their efficiency depends on dimensionality of the set of signals, the estimation of statistical parameters (mean value, standard deviations, trend parameters), dynamics [21] and for selected ones – the learning period [17], [19].

For specific, not well-defined changes or patterns, processing of the set of signals with such standard algorithms are often not sufficient because of existing different signal properties, attributes of events, time-delays between changes in different signals and dependencies between them. Moreover, in many cases the aim is to detect only changes in one selected signal, excluding the synchronous changes in other signals and time-lagged ones. Such situations are encountered – for example – when reported statuses from the monitored devices don’t match.

III. A Hybrid Event Detection Algorithm

In this paper, a hybrid event detection algorithm is proposed. It is based on the processing of pair of signals in two levels (see listing 1): first – the preliminary non-stationarity detection in individual signals, and then – the second one – the confirmation with distance-like similarity method, computed for both signals. Such approach relies on an the assumption that all signals of available dataset (or some of them) can be paired. Thus, summarized detection results (at a time instant) will indicate the complex dependencies within data set. Moreover, such computation is suitable for the distributing and employing one of many computational intelligence paradigms (like multi-agent systems).

In the first level (for sample n) the non-stationarity is identified with the short-term prediction errors comparison obtained from the one-step-ahead zero-order-prediction (ZOP)/zero-order-hold (ZOH) model [2], [15] and the adaptive Holt predictor [11] in a moving window of constant length, suitable for non-stationary data analysis (in particular, trended time series). This preliminary detection procedure is suitable for time series which consist of the non-random components [19].

When short-term non-stationarity is recognized for a fixed number of samples (arbitrary adjusted threshold value), the second level detection is triggered (see listing 1, lines 9-15). To confirm change initial signal, the distance-like method is proposed.

The distance-like detection method – denoted as measure Z (or dZ) and first mentioned in [6] – is dedicated to two signals similarity monitoring. It is based on the synchronous processing of two signals (denoted as x and y) of fixed length in a moving window. For both signals x and y, the mean values of sub-sequences of absolute deviation values are calculated (the positive (x_{pm}, y_{pm}) and the negative (x_{nm}, y_{nm}) signs). The deviation is recognized when the absolute value exceeds a fixed threshold ρ_{zd} at the time-instant. For deviations smaller then threshold, zero value is assumed. As a result, two pairs of the positive values (x_{pm}, y_{pm}) and negative (x_{nm}, y_{nm}) are obtained.

```
1 n = 1
2 REPEAT
3 // level 1
4 perform ZOH prediction for both
    signals
```

```
perform Holt prediction for both
  signals
produce and compare short-term
  prediction errors
 IF(short-term non-stationarity
   recognized)
 {
   // level 2
   calculate dZ
   IF(dZ is greater than fixed
     threshold)
   {
     // change detected
     alert/raport the user/system
   }
 }
UNTIL n = N
```

Listing 1. Pseudo-code of the proposed event-detection hybrid algorithm for two signals of the length *N*. A fixed threshold ρ_Z value is assumed.

The temporary similarity d*Z* is calculated as follows:

$$dZ = \sqrt{(x_{pm} - y_{pm})^2 + (x_{nm} - y_{nm})^2} \qquad (1)$$

Processed data received from technical devices may contain the phased (time-lagged) events. To avoid such time delays between events/changes, during computing a final value of the measure d*Z*, a tolerance (denoted as L_{tol}) is assumed. Such tolerance may be viewed as a permissible time delay between occurring changes in both signals. For each subsequent sample *n*, to compute dZ_n a L_{tol} number of the measures dZ_t is calculated in a moving window ($t = n - L_{tol} + 1 \ldots n$) giving a set of the d*Z* values. Finally, dZ_n is chosen as the smallest dZ_t value between two sub-series:

$$dZ_n = min_{t=n-L_{tol}+1 \ldots n} dZt \qquad (2)$$

“Distance measure” term is used (instead of “distance” or “metric”) [19] because all formal definition of metric conditions are not satisfied (i.a. symmetry condition).

IV. Calculation Results

To verify the effectiveness of the proposed hybrid detection procedure, the algorithm was implemented and tested on the random-simulated data. Signals were generated as time series of the length 144 containing pseudo-random values drawn from the standard normal distribution *N*(0;1). Such relatively short length of the signal allows for a reliable analysis of the algorithm efficiency using data visualization.

To obtain data sets similar to real diagnostic signals (i.a. avoid negative values), a constant $c = 30$ was added to each sample. In the next step, diagnostic signals were modified by adding the simulated changes for random generated time instants (randomly selected samples were increased). The length of simulated change was fixed to 5 in an experimentally way.

In this paper, presented detection results are chosen from among many possible ones, and they are depended on arbitrary parameter adjusting. In particular, in the first stage (the short-term prediction comparison computed for differences between adjacent elements of processed series), the moving window width was fixed to 22. The second detection stage is triggered after recognition of series of – at least – three consecutive samples for which the non-stationary was found. To process the data in the same order of magnitude, signals were unified with their dispersions (in the moving window).

In the presented case, it was assumed that the preliminary detection performing in the first stage causes no need to use of the threshold value of the method d*Z*. Therefore, event detection is based on choosing the final d*Z* value from the set of length L_{tol} which was fixed to 5.

In the paper, four sample diagnostic signals are examined (see fig. 1-4) representing different event sets illustrated in the appropriate parts of the figures 1-16.

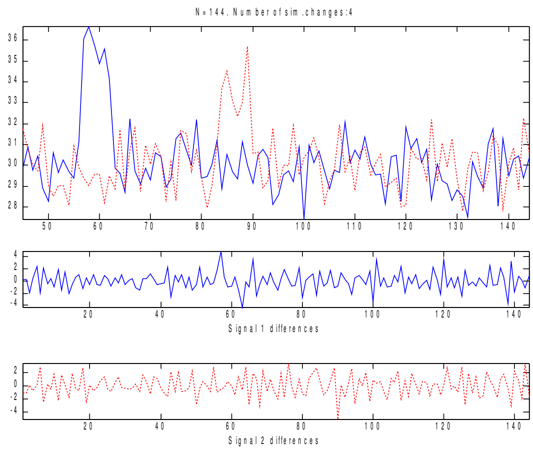

Fig. 1. Random-generated signals – part 1 (depicted with the solid and dotted lines), containing four simulated changes – the input (original) data (upper large sub-figures) and the differences between adjacent elements of the processed signals (2nd and 3rd sub-figure rows).

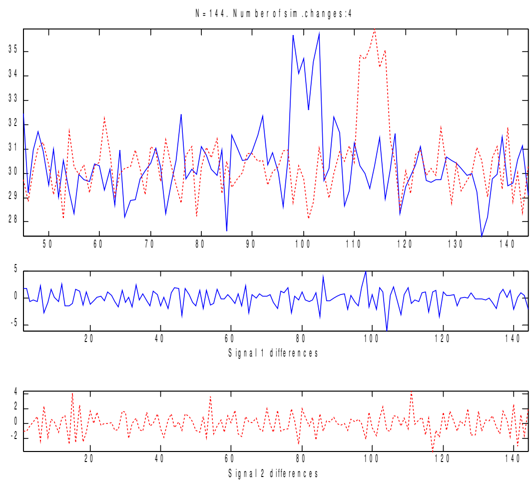

Fig. 2. Random-generated signals – part 2. Description of the symbols and lines – see Fig. 1.

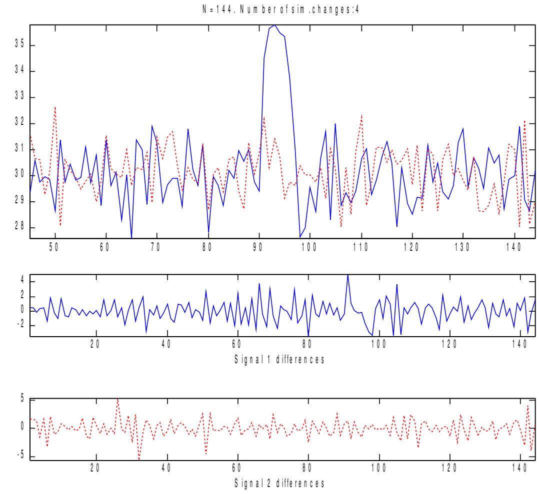

Fig. 3. Random-generated signals – part 3. Description of the symbols and lines – see Fig. 1.

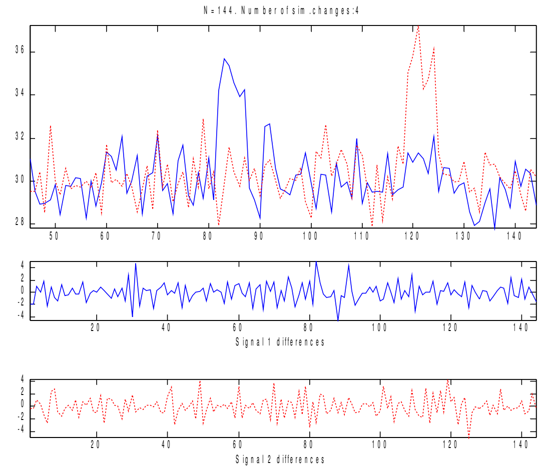

Fig. 4. Random-generated signals – part 4. Description of the symbols and lines – see Fig. 1.

It may be seen in fig. 5-8 that employing the detection performed in the first step results in the identification of the large deviations.

Fig. 9-12 illustrate the dZ value changes in the moving window. The basic assumption of the method is illustrated – especially – in fig. 10 (100-120th sample) where the dZ value is strongly depend on the samples included in the moving window. If changes in both signals are covered, the dZ value is near to zero; otherwise, abrupt dZ changes are visible.

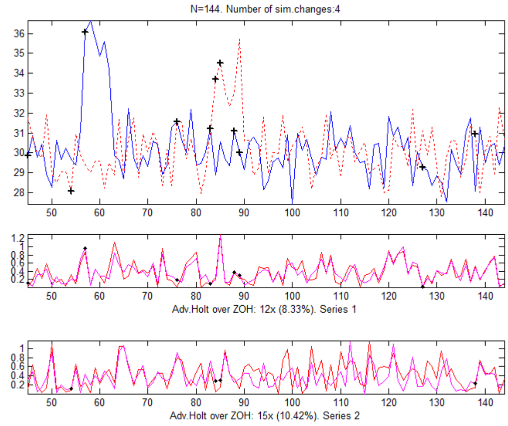

Fig. 5. Detection results obtained with the short-term prediction comparison (part 1) depicted in the row 2 and 3. Single non-stationarity detection (advantage Holt over ZOH) is denoted as black 'dots'.

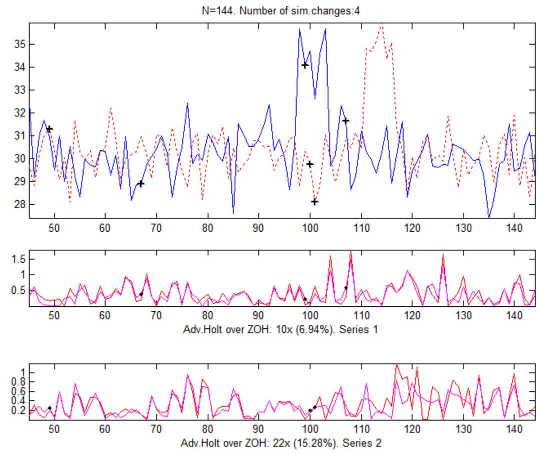

Fig. 6. Detection results obtained with the short-term prediction comparison (part 2). Description of the symbols and lines – see Fig. 5.

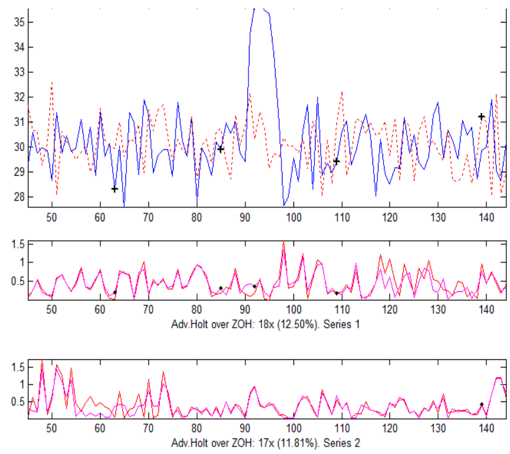

Fig. 7. Detection results obtained with the short-term prediction comparison (part 3). Description of the symbols and lines – see Fig. 5.

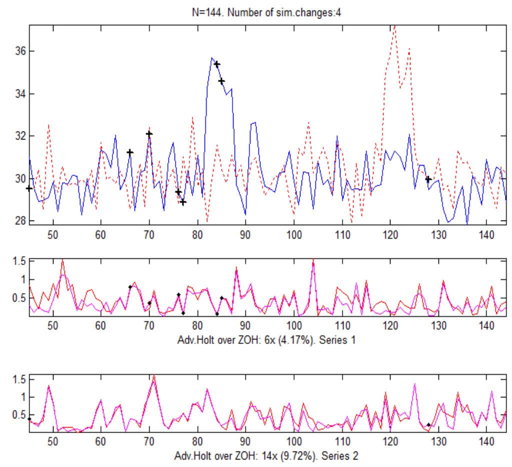

Fig. 8. Detection results obtained with the short-term prediction comparison (part 4). Description of the symbols and lines – see Fig. 5.

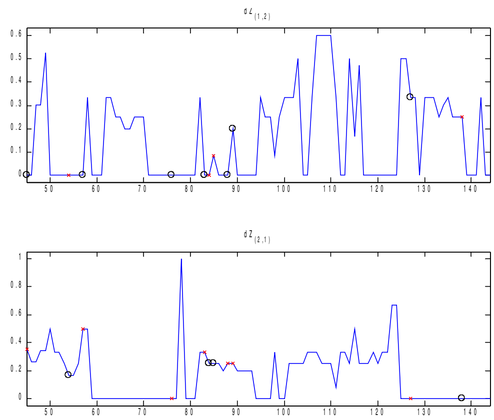

Fig. 9. Change detection performed with the proposed event-based similarity method d*Z* (part 1). Changed detected with short prediction comparison depicted as ‘circles’ and ‘asterisk’.

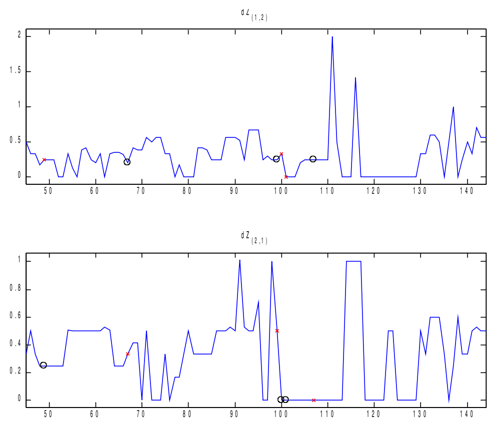

Fig. 10. Change detection performed with the proposed event-based similarity method d*Z* (part 2). Additional description – see Fig. 9

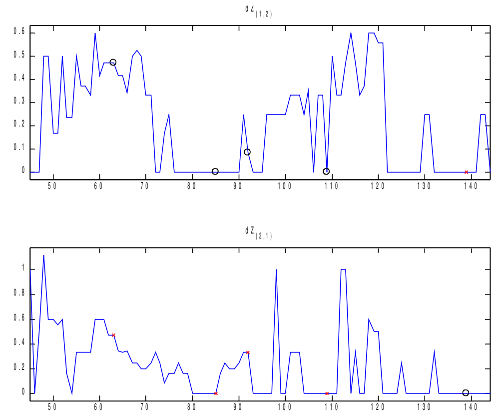

Fig. 11. Change detection performed with the proposed event-based similarity method d*Z* (part 3). Additional description – see Fig. 9

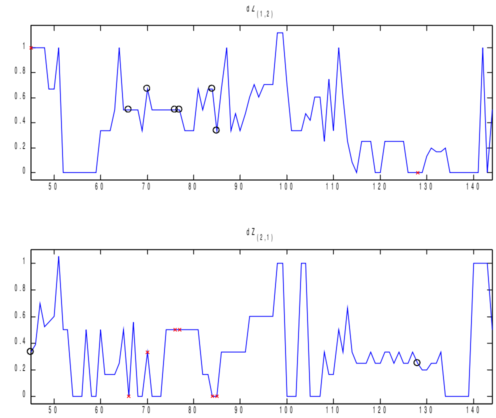

Fig. 12. Change detection performed with the proposed event-based similarity method d*Z* (part 4). Additional description – see Fig. 9

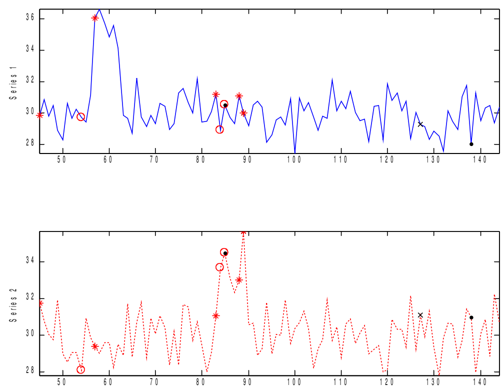

Fig. 13. Change detection from processed signals (part 1): 1 (depicted in the upper subfigure with solid line) and 2 (dotted line, the lower subfigure). Description of used symbols: detected non-stationarities from signal no. 1 confirmed with d_{Z1} (H_1/d_{Z1}) denoted as ‘cross’, confirmed with d_{Z2} (H_1/d_{Z2}) depicted as ‘dot’; detected non-stationarities from signal no. 2 confirmed with d_{Z1} (H_2/d_{Z1}) denoted as ‘asterisk’, confirmed with d_{Z2} (H_2/d_{Z2}) – as ‘circle’.

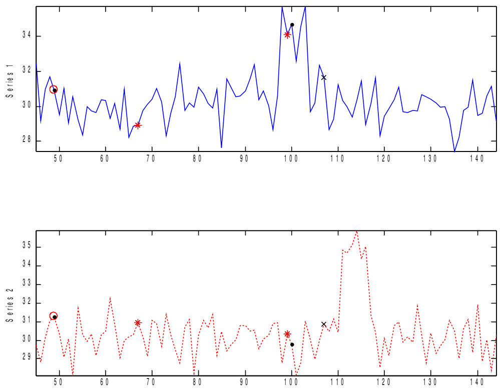

Fig. 14. Change detection from processed signals (part 2). Extended description – see Fig. 13.

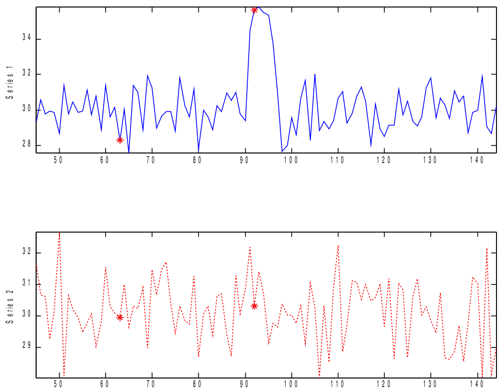

Fig. 15. Change detection from processed signals (part 3). Extended description – see Fig. 13.

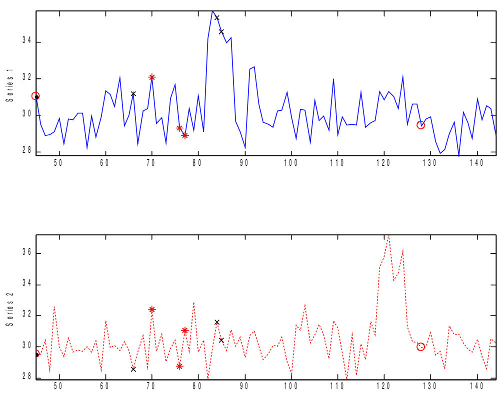

Fig. 16. Change detection from processed signals (part 4). Extended description – see Fig. 13.

As mentioned in the earlier part of this paper, research are aimed at detection of the original changes in the one processed signal. Figures 13-16 show such event selection. Proposed hybrid algorithm is capable of detecting large, single changes (see fig. 13 – for about 60th and 80th sample; in fig. 14 – about 100th sample, fig. 15 – about 95th sample). Notice, that few detection results may be viewed as false alarms (for example – fig. 16, about 100th sample), nevertheless, in such cases the interpretation will be connected with the analysis of small deviations that may be invisible in assumed figure resolutions.

V. Conclusions

Event detection from signals received from computer systems can be focused on – depending on detection task – the independent signal processing (the identification of short and long-term changes in signals separately) or the available diagnostic signals set processing – which is more complex and allows to identify the changes in the analyzed (target) signal environment, including implicit events. The proposed detection idea is related to the second detection approach and it is a part of wide (realized and planned) research on event detection from time series through concurrent and parallel environment of analyzed signal monitoring. Summarized detection results will indicate the significant and insignificant changes and the dependencies within processed data set.

In this paper it was shown that proposed two-level algorithm is suitable to detect the changes in diagnostic signals which occur in the one signal only. The procedure efficiency was tested on the random generated signals containing simulated changes, however, for real signals, peer analysis of available signals set may be valuable.

The intention was to show the most relevant properties of the hybrid algorithm rather than data acquisition process mapping and further processing according to the real computing conditions and limitations.

Presented detection results are obtained using the algorithm whose parameters were adjusted in an experimentally way. Therefore, the parameters adapting is a promising way to effectiveness improvement. Further research will be focused on algorithm adaptation (corresponding to processed signals properties) and the stage 2 modifications towards the elimination of the false alarms (in particular, changes of the moving window width will effect in the detection resolution).

The proposed idea of multiple signal processing can be developed (especially taking into account the detection task complexity when processing large datasets) with the distributed processing of signals from networked devices as i.e. multi-agent systems or artificial immune systems (such paradigm appears to be helpful for detection viewed as an unsupervised classification, especially for signals contain many random and non-random components).

References

[1] Benveniste A., Basseville M.: Detection of Abrupt Changes in Signals and Dynamical Systems. Lecture Notes in Control and Information Sciences, Vol. 77, Springer-Verlag, 1984

[2] Box G. E. P., Jenkins G.M.: Time Series Analysis: Forecasting & Control, Prentice Hall, 1994

[3] Brockwell P. J., Davis R. A.: Time Series: Theory and Methods. Springer Series in Statistics, 1991

[4] Cole R., Shasha D., Zhao X.: Fast Window Correlations over Uncooperative Time Series. ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, 2005

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

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

[7] Guh R., Zorriassatine F., Tannock J. D. T., O'Brien C.: On-line Control Chart Pattern Detection and Discrimination - a Neural Network Approach. Artificial Intelligence in Engineering, Vol. 13, Issue 4, Elsevier 1999

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

[9] Goldin D., Mardales R., Nagy G.: In Search of Meaning for Time Series Subsequence Clustering: Matching Algorithms Based on a New Distance Measure. ACM International Conference on Information and Knowledge Management, 2006

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

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

[12] Keogh E., Lonardi S., `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, Edmonton, Alberta, Canada 2002, pp. 550-556

[13] Kingdon J.: Intelligent Systems and Financial Forecasting, Springer, 1997

[14] Mahoney M., 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

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

[16] Pełech T, 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

[17] Pełech T., 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

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

[19] Pełech-Pilichowski T., Duda J. T.: A two–level algorithm of time series change detection based on a unique changes similarity method. IMCSIT Proceeding, Wisła, PL, 18-20 October, 2010

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

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

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

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