

Research Article

Change Point Detection in Time Series Using Higher-Order Statistics: A Heuristic Approach

Constantinos S. Hilas,¹ Ioannis T. Rekanos,² and Paris Ast. Mastorocostas¹

¹ Department of Informatics and Communications, Technological Educational Institute of Serres, Terma Magnisias, 62124 Serres, Greece

² Physics Division, School of Engineering, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

Correspondence should be addressed to Constantinos S. Hilas; chilas@teiser.gr

Received 10 November 2012; Accepted 20 May 2013

Academic Editor: Dane Quinn

Copyright © 2013 Constantinos S. Hilas et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Changes in the level of a time series are usually attributed to an intervention that affects its temporal evolution. The resulting time series are referred to as interrupted time series and may be used to identify the events that caused the intervention and to quantify their impact. In the present paper, a heuristic method for level change detection in time series is presented. The method uses higher-order statistics, namely, the skewness and the kurtosis, and can identify both the existence of a change in the level of the time series and the time instance when it has happened. The technique is straightforwardly applicable to the detection of outliers in time series and promises to have several applications. The method is tested with both simulated and real-world data and is compared to other popular change detection techniques.

1. Introduction

Univariate time series analysis, in which only a single observational unit (or subject) is being studied, is very common in scientific experiments as well as in commercial applications. When an event that affects the evolution of a specific phenomenon occurs, the resulting time series is referred to as an interrupted one. This event may act as an intervention in or an interruption of the normal evolution of the response time series, which, in the absence of the intervention, is usually assumed to be a pure ARIMA process. The intervention breaks the time series into two segments, namely, the preintervention and the postintervention ones. The affected time series usually present a change in level and/or slope, which can be either permanent or temporary. The analysis of interrupted time series, or time series quasi-experiments [1], has many applications in various areas of interest, such as in testing and measuring the impacts of new traffic laws, the impacts of gun control laws, the impacts of air pollution, and the impacts of psychological treatments. Moreover, intervention models can be used to model and

forecast the response series and/or to analyze the impact of the intervention.

The analysis of interrupted time series is performed by statistically comparing the two segments of the series before and after the intervention. Most of the interrupted time series analysis methods require a priori knowledge of the time instance of the event. Then, the pre- and postintervention parts of the series are compared in order to identify changes in the underlying model of the time series. One usually uses the preintervention data (also called training, estimation or Phase I data) to predict the evolution of the time series after the intervention. The derived prediction is then compared to the actual postintervention part [2, 3].

Furthermore, specific characteristics of the expected intervention, such as its time dependence (abrupt or gradual in onset) as well as its time duration (permanent or temporary), are crucial parameters for the performance of the analysis. McDowall et al. [4] have proposed three intervention components, each associated with a distinct pattern of impact. The three intervention components permit the analyst to test rival impact hypotheses.

The most problematic part of the above process is the model identification step [5]. This is due to three reasons, namely, the need of adequate number of data points, the need for an adequate degree of mathematical sophistication, and the fact that even if the above are being satisfied there is still a chance of incorrect model identification. In order to eliminate the model identification step, several suggestions have been made [6–8]. All of them assume that most time series can be represented satisfactorily by a single autoregressive ARIMA model of some order. The application of other traditional statistical methods such as two sample t -test comparisons of the pre- and postintervention parts of the time series is invalid because they ignore the dependence between successive observations of the time series [9]. Under the assumption of normality, generalized likelihood ratio tests may also be applied to detect change points when the process parameters, that is, the mean value and the variance, differ before and after the change point. This approach is more sensitive to the normality assumption than the t -test. A discussion on such generalized likelihood tests, their ability to cope with successive change points, and the impact of initial parameter estimation (i.e., estimation or training or Phase I data) can be found in [10].

Intervention analysis has also been used as a method for financial fraud detection [11], as well as an outlier detection method for improving ARIMA forecasting [12]. It is noted that all the above methods need a priori knowledge of the time of intervention. If this is unknown, then more sophisticated methods may be applied, which iteratively model probable interventions to all the successive points in the time series and try to identify the position that causes the observed behavior [11].

There are also methods that can identify an unknown change point based on the normality assumption. The assumption of these methods is that the series follow a normal distribution $N(\mu_1, \sigma_1^2)$ up to a time instance τ ; the change-point. Then, after the change point, the series follow another normal distribution differing in mean $\mu_1 \neq \mu_2$, in variability ($\sigma_1^2 \neq \sigma_2^2$), or in both mean and variability. The Shewhart \bar{X} charts and the cumulative sum (Cusum) charts fall in this category [10].

The detection of change points in time series has, also, drawn much attention in data mining. It is recognized as event change detection [13] and is closely related to activity monitoring [14]. Outlier detection is also of great interest, since it is closely related to fraud detection and rare event discovery. Yamanishi and Takeuchi [15] propose a unifying framework for dealing with both outlier and change point detection. In this framework, a probabilistic model of the data source is incrementally learned using an online discounting learning algorithm, and then the score of any given data is calculated to measure its deviation from the observed model. We should also mention that Matsumoto and Yosui use a sequential Monte-Carlo scheme to detect changes [16].

In the present paper, an automated method based on higher-order statistics for the identification of an intervention in a time series is presented. A preliminary version of this research has been published in [17], and parts of the original

publication are included here under permission by IEEE. The method includes the identification of both the existence of the intervention and the time point when it has happened. The proposed framework can also be applied for the detection of outliers in time series. The method is based on the behavior of higher-order statistics, namely, the skewness and the kurtosis of the data, computed in a sliding window over the data sequence. This approach has been recently used in the separation of weak biomedical signals from noise [18], in the identification of the arrival of important seismic events [19], and in the compression of signals for telemedicine applications [20]. The method is demonstrated by means of both simulated and real-world data. The initial motivation for this work came from a problem of fraud detection in a telecommunications network. In fact, telecommunications data are usually represented as vectors of features. Any or all of the features may serve as useful indicators of a user's deviation from a hypothesized normal behavior (profile). Nevertheless, we wanted to test if there would be possibility to detect such deviations from simple univariate representations of a user's behavior, for example, the number of phone calls per day. The examples in Section 4 show that such an approach is feasible. However, as long as telecommunications fraud detection is concerned, interrupted time series analysis can only serve as an extra tool within a larger fraud detection framework. This is because only special cases of fraud may be detected from each indicator.

The paper proceeds as follows. In Section 2, an introduction to higher-order statistics (HOS) is given. In Section 3, the description of the proposed method is presented. Experimental results are presented in Section 4, which include both simulated and real-world examples. A comparison between the proposed method and previous ones is carried out in Section 5. Finally, the conclusions are discussed in Section 6.

2. Higher-Order Statistics (HOS)

In this study, two higher-order statistics (HOS), namely, the skewness and the kurtosis, are used to detect changes in the evolution of a stochastic process. The values of these statistics may indicate a deviation from Gaussianity and may also reflect the existence of a nonstationary process. These attributes are used, first for the detection of a change in the level of the series and then for the estimation of the exact time instance of the intervention.

Skewness is a measure of the asymmetry of the data around the sample mean. If skewness is negative, the data are spread out more to the left of the mean than to the right. If skewness is positive, the data are spread out more to the right. The skewness is defined by

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{E(X - \mu)^3}{[E(X - \mu)^2]^{3/2}}, \quad (1)$$

where μ is the mean of the random variable X , μ_i is the i th central moment, and $E(\cdot)$ represents the expected value of the quantity in the parenthesis. We note that the skewness of a random variable with normal or, in general, with symmetric probability distribution function equals zero.

Kurtosis is a measure of how outlier-prone a distribution is. The kurtosis of the normal distribution is equal to 3. Distributions that are more outlier-prone than the normal distribution have kurtosis greater than 3; distributions that are less outlier-prone have kurtosis less than 3. Kurtosis is given by

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} = \frac{E(X - \mu)^4}{[E(X - \mu)^2]^2}. \quad (2)$$

Usually the kurtosis is normalized (*kurtosis excess*) by subtracting 3 in order to have a value of 0 for normal distributions. If the kurtosis is highly positive (or $\gamma_2 > 3$), the activity distribution is highly peaked and the data is likely to contain an outlier.

In the presence of sample data, skewness and kurtosis are calculated by means of their unbiased estimators g_1 and g_2 , respectively. For an N -sample sequence $\mathbf{x} = \{x(n) := n = 1, 2, \dots, N\}$, the estimators g_1 and g_2 are given by [21]

$$g_1 = \frac{\kappa_3}{\kappa_2^{3/2}} = \frac{N \cdot \sum_{n=1}^N (x(n) - \widehat{m})^3}{(N-1) \cdot (N-2) \cdot \widehat{\sigma}^3}, \quad (3)$$

$$g_2 = \frac{\kappa_4}{\kappa_2^2} = \frac{N \cdot (N+1) \cdot \sum_{n=1}^N (x(n) - \widehat{m})^4}{(N-1) \cdot (N-2) \cdot (N-3) \cdot \widehat{\sigma}^4} - \frac{3 \cdot (N-1)^2}{(N-2) \cdot (N-3)}, \quad (4)$$

respectively, where \widehat{m} and $\widehat{\sigma}$ are the estimates of the mean and the standard deviation of $x(n)$, respectively, and κ_i the i th κ -statistic. For a normal distribution, the expected values of g_1 and g_2 are 0 and $-6/(N-1)$, respectively, while the variances of the aforementioned statistics are given by [22]

$$\begin{aligned} \text{var}(g_1) &= \frac{6 \times N \times (N-1)}{(N-2) \times (N+1) \times (N+3)}, \\ \text{var}(g_2) &= \frac{(24 \times N \times (N-1)^2)}{(N-3) \times (N-2) \times (N+3) \times (N+5)}. \end{aligned} \quad (5)$$

3. The Proposed Heuristic Method

As already mentioned, the aim is to detect any significant changes in the level of a time series. The approach presented here is actually a heuristic method that yielded after the observations that follow. So, a brief example will be used to demonstrate it.

Let us examine the case of a time series $x(t)$ with an intervention occurring at time point $t = 40$ (Figure 1(a)). The intervention has the effect of raising the level of the series. In particular, the whole sequence is a white gaussian noise signal, $N(0, 1)$, superimposed on a step function $u(t - 40)$. The skewness and the kurtosis of $x(t)$ are calculated by means of a sliding window of size W , and the behavior of the two statistics in the vicinity of the intervention is examined. The proper size of the window needs to be identified, and it will be discussed later. The next three

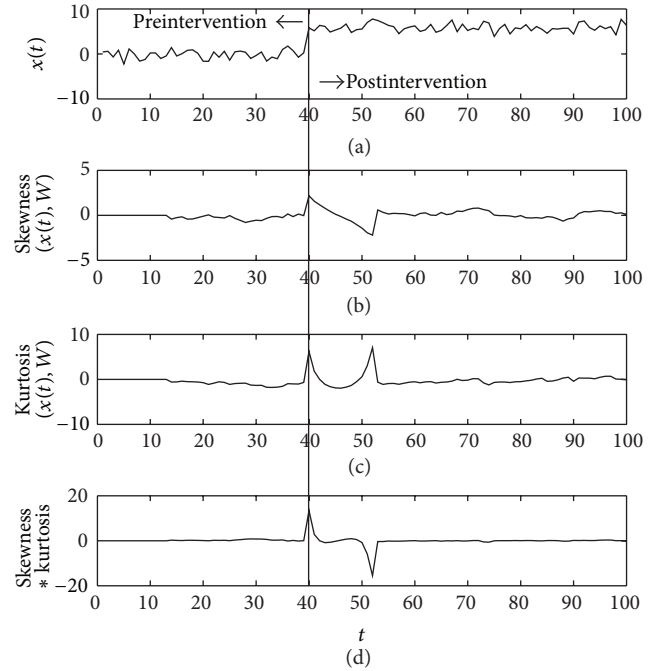


FIGURE 1: A time series $x(t)$ with an intervention occurring at time point $t = 40$, its skewness (b), kurtosis (c), and their product (d) are estimated for a window size $W = 14$.

subplots (Figures 1(b), 1(c), and 1(d)) depict the skewness, the kurtosis, and the product of the two statistics, respectively, which are derived by means of a moving window of size $W = 14$. The presence of prevalent peaks at the time point of the intervention is clear. In fact, this point behaves as an outlier among the data inside the window. There is also another prevalent extreme point $W - 2$ time lags after the first. This is due to the sliding window which has now only one point from the low (preintervention) level of the series and all others form the postintervention set. The low point behaves as an outlier for the new window position.

In the skewness plot (Figure 1(b)), one can see a maximum at the time point of the intervention as well as a minimum at $W - 2$ time points after that. In the kurtosis plot (Figure 1(c)), there are two successive maxima, the first at the time point of the intervention and the second $W - 2$ time points later. Last, their product plot (Figure 1(d)) has obviously a maximum at the first point and a minimum $W - 2$ points later. In fact the skewness-kurtosis product embodies the behavior of both statistics. In all cases the second peak, either positive or negative, confirms the significance of the first one and differentiates it from random fluctuations.

Under the considerations of the previous paragraphs and if we can assume that the time series under study is generated from a normal process, $N(\mu, \sigma)$, then the proposed method involves the following steps.

- (1) The size, W , of the sliding window is decided.
- (2) The W -sample sliding window is moved over the time series $x(t)$ with one sample difference between successive windows. For each new position, j , the two

statistics, g_{1j} and g_{2j} , in the window $w^j = \{x(k) : k = j-W+1, j-W+2, \dots, j\}$, ($j = W, W+1, \dots, N-W+1$) are estimated.

- (3) The values of the two statistics are thresholded by means of the Chebychev-Bienaymé inequality [23]:

$$P\left(\mu - \frac{\sigma}{\sqrt{1-q}} < X < \mu + \frac{\sigma}{\sqrt{1-q}}\right) > q, \quad (6)$$

which means that the random variable X lies within a confidence interval with bounds $\mu \pm \sigma/\sqrt{1-q}$ and probability at least q . From (6), the following confidence intervals for γ_1 and γ_2 , respectively, are derived:

$$\left\{ E(g_1) - \frac{\sqrt{\text{Var}(g_1)}}{\sqrt{a}}, E(g_1) + \frac{\sqrt{\text{Var}(g_1)}}{\sqrt{a}} \right\}, \quad (7)$$

$$\left\{ E(g_2) - \frac{\sqrt{\text{Var}(g_2)}}{\sqrt{a}}, E(g_2) + \frac{\sqrt{\text{Var}(g_2)}}{\sqrt{a}} \right\},$$

where $a = 1 - q$ is the level of significance. Its value is set equal to 0.05, throughout our experimentations, which means a confidence level of 95%. The thresholding procedure actually means that all the values of skewness and kurtosis that fall within the aforementioned intervals are set equal to their expected values.

- (4) The values of the two statistics form two new time series, $g_1(t)$ and $g_2(t)$, ($t = W, W+1, \dots, N$). In these series, all the local minima and maxima are identified.
- (5) The product of each estimated value of the $g_1(t)$ with the corresponding value of $g_2(t)$ is also computed. (As the expected value of skewness is 0 and we have already thresholded it, in step 3, the product of the two statistics is suppressed to 0 whenever skewness is 0.)
- (6) All pairs of successive minima and maxima at a distance of $W - 2$ are identified, and the position as well as the type (minimum or maximum) of the first one is recorded.
- (7) All positions recorded in the previous step are the identified points of intervention.

As stated earlier, the proposed method can be applied to signals for which the underlying process (i.e., the one that generates the time series) can be assumed Gaussian. If this is not the case, the method is altered by omitting step 3 and the bracketed part of step 5. A voting procedure is then applied which uses the minima and maxima identified in the three statistics (skewness, kurtosis, and their product). During the voting procedure, the time instances of each minimum (or maximum) are compared with those from the other series and the one with the more hits is voted as the most probable position of the intervention. If none of the identified points is equal to any of the others, then the absence of any intervention is decided. Both approaches gave similar results.

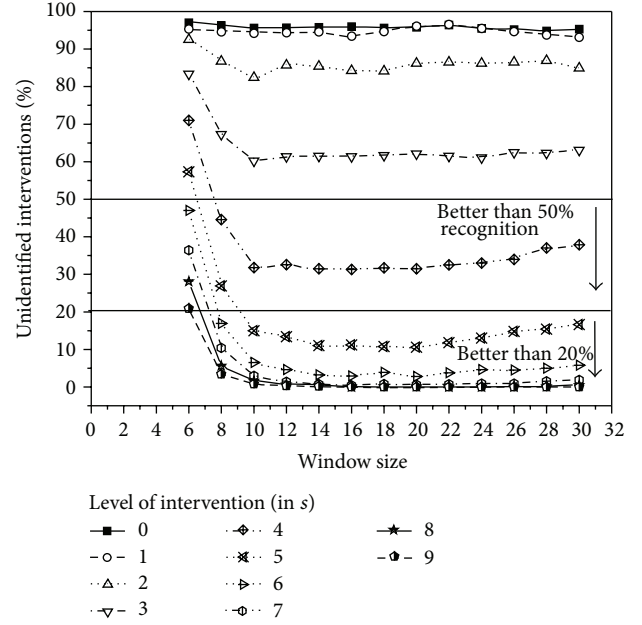


FIGURE 2: Unidentified interventions in regard to the level of the intervention and the size of the window used.

The approach to be followed depends on the specific application. In all the simulations that follow, the decision is considered correct if a detected change point is located within one time point before or after the true change point.

4. Experimental Results

4.1. Dependence on the Size of the Intervention. First, we examine the dependence of the identification ability of the method on the size of the intervention and the size of the sliding window used. It is obvious that the bigger the size of the window, the better the normal approximation of the sample would be. However, large windows will affect the “online” detection ability of the method. Moreover, real-world applications usually suffer from small sample sizes. So, first 1000 white noise $N(0, 1)$ sequences, of length 100 time points, were generated and for each one of them an intervention was created at a specific time point. The algorithm was applied to each sequence using window sizes that incremented, by 2, from $W = 6$ to $W = 30$. The level of the intervention was incremented by s , where s is the standard deviation of the original signal. The percentage of unidentified interventions for sizes of intervention ranging from 0 to $9s$ is shown in Figure 2. One can see that the identification ability stabilizes after $W = 10$. On the contrary, the ARIMA approach demands at least 50 points before and 50 after the intervention. Additionally, the method has an identification percentage better than 65% for intervention sizes equal to $4s$, which improves dramatically for higher interventions.

4.2. Dependence on the Window Size. In order to further explore the trade-off for various window sizes, we considered

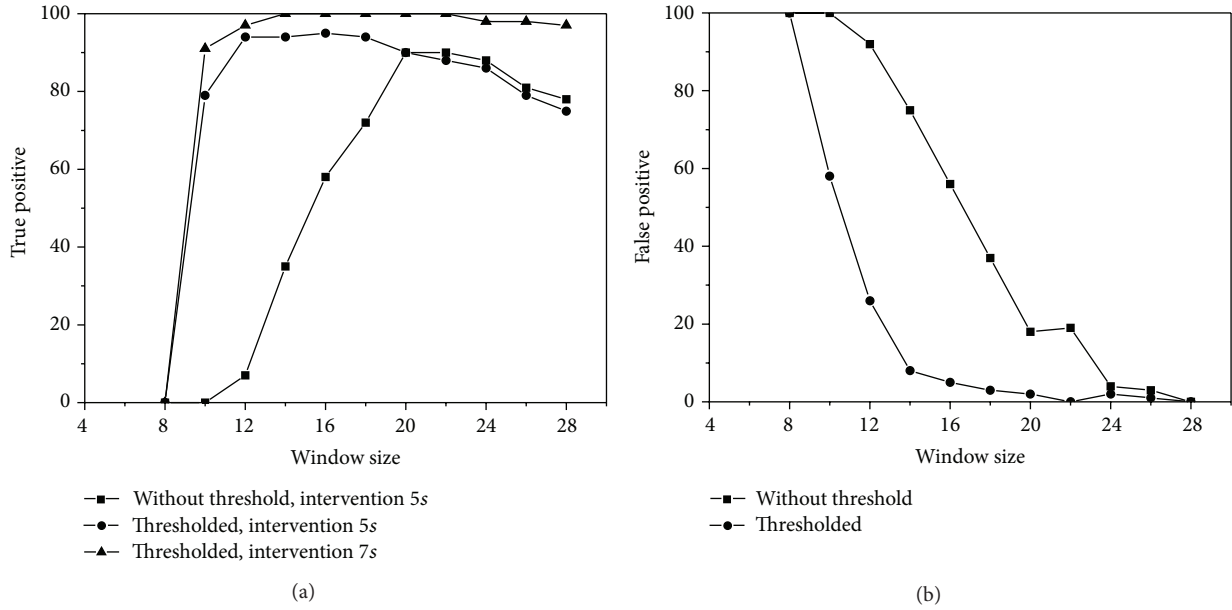


FIGURE 3: (a) True positive rate versus the window size and (b) false positive rate versus window size, with and without the use of thresholds.

two sets (A and B) of 1000 time series each. Set A consisted of 1000 time series with interventions, whereas set B consisted of 1000 time series without interventions. Each series consists of 100 time instances. We use two plots to show how the true positive (TP) and the false positive (FP) rates depend on the window size and on the thresholding procedure that is described in steps 3 and 5 of the proposed algorithm. In Figure 3(a), the true positive rate versus the window size is given. The line depicted with squares was derived by applying to set A an intervention of size 5s and by omitting step 3 of the method. That is without skewness and kurtosis thresholding. The line which is depicted with circles was produced from the same set of time series using thresholding with $a = 0.05$. For the third line (triangles), the intervention in the series had a level of 7s, where s is the standard deviation of the signal before the intervention. All lines were plotted for window sizes ranging from 8 to 28 with increments of 2. In Figure 3(b), the false positive rate versus the window size is depicted. Since this plot was produced from set B (i.e., without real interventions), only two lines are drawn, one derived by applying thresholding and one without. A dramatic improvement in the TP percentage with the implementation of step 3 (thresholding) is observed. For window size $W = 12$, we get a good percentage of TP hits while the FP rate reduces significantly. By incrementing the window size, the outcome improves. In fact, the trade-off is not only between the sensitivity and the specificity of the method but also in the size of the window used. Larger window sizes lead to better performance but have two countereffects. First, if one uses large windows then he loses the “real-time” detection of the intervention. Second, many real-world time series are small in size, so large window sizes may even cover the whole data set. Adding to this, small intervention sizes will not produce statistically significant

changes inside large windows, which is also evident by the decline in the TP rate for larger window sizes.

4.3. Performance with Stationary Time Series. Additional simulations were done in order to test the performance of the method with autoregressive data. Eleven sets of 1000 ARIMA(1,0,0) time series each were simulated. Each set of the time series was generated for different values of the autoregressive coefficient φ_1 ranging from $-1^* = -1 + e$ to $1^* = 1 - e$, where $e > 0$ is chosen arbitrarily small, with steps of 0.2. An intervention of size 5s was added at a specific time point and the identification ability of the method was tested. The percentage of unidentified interventions for each data set, for increments of size 2 to the window size, is depicted in Figure 4. One can see that the performance degrades near the stability boundaries $|\varphi_1| = 1$.

To further investigate the performance of the proposed method, we consider time series that follow two alternative AR(2) models. The preintervention part follows the AR model

$$y(k) = 0.4 \cdot y(k-1) - 0.5 \cdot y(k-2) + a(k), \quad (8)$$

where $a(k)$ is Gaussian random variable with mean 0 and variance 1. An intervention of size 4s occurs at time point $k = 101$ and after that the process follows a different AR model, given by

$$y(k) = 0.2 \cdot y(k-1) - 0.3 \cdot y(k-2) + a(k). \quad (9)$$

The resulting data sequence along with the skewness-kurtosis product for window size 16 is presented in Figure 5. One may observe that the proposed method correctly identifies the intervention.

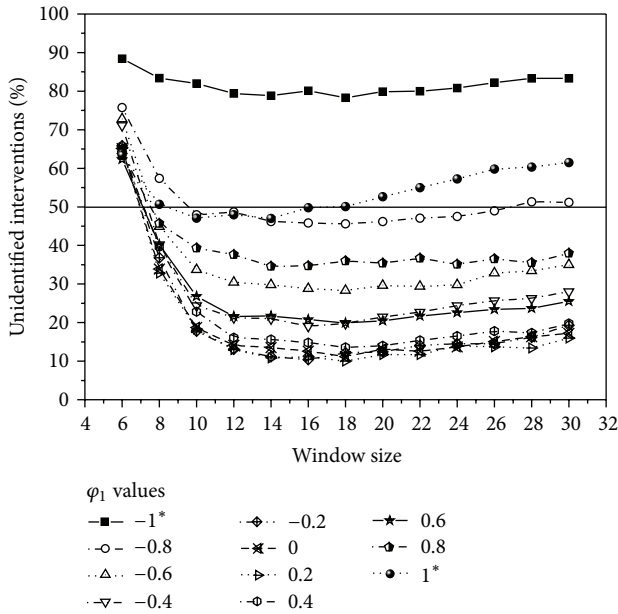


FIGURE 4: Unidentified interventions of size 5 for different ARIMA(1,0,0) processes.

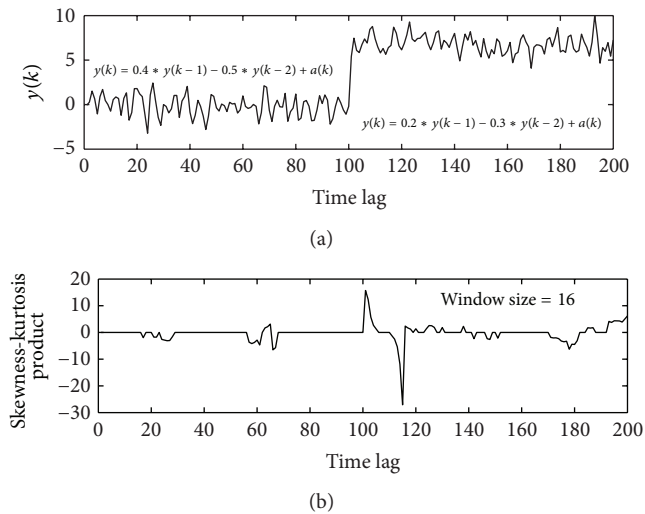


FIGURE 5: Intervention identification for a process that changes after the intervention.

4.4. Multiple Change Point Detection. The method was also evaluated by a numerical simulation, which involved a dataset with many change points. The data set consists of 1000 records. The changes in level occur at time points $t = k * 100, k = 1, 2, \dots, 9$ and are equal in size to 4s. Figure 6 shows the simulated time series as well as the skewness-kurtosis product, derived by use of a sliding window of size 18. The arrows in the plot point out the local maxima, which are followed by a minimum at a distance of -2 . These maxima indicate the change points. All maxima are followed by minima at distances dictated by the size of the window used. Random fluctuations do not have a minima follower. It should be noted that the window size sets a limit to the

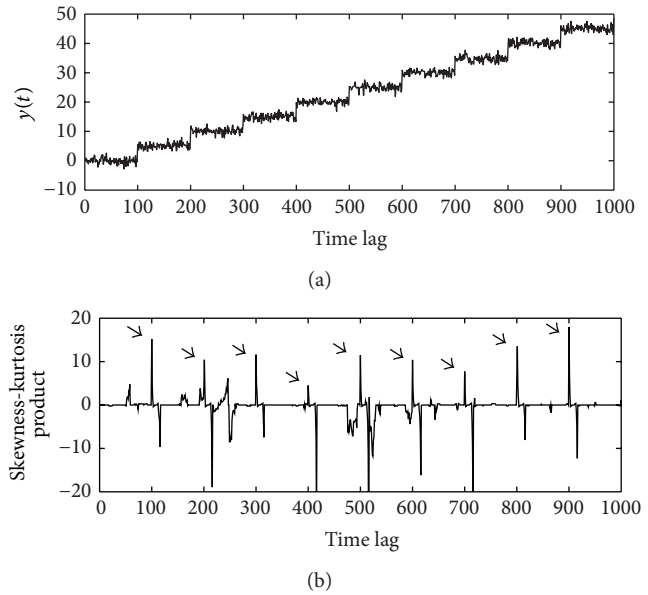


FIGURE 6: Change level detection for a simulated dataset with multiple change points.

proximity of successive change points that can be identified. Thus, change points that are closer to each other than the window size may go unidentified.

4.5. Outlier Detection. The following example emphasizes the ability of the proposed method to detect outliers. It is drawn from a case of superimposed fraud in the telecommunications network of an organization and demonstrates how the different modeling approaches can affect a data mining problem.

The daily behavior of the user under study is plotted in Figure 7. All data correspond to eight years (from Jan 1998 to Dec 2005) of user behavior. Figure 7(a) shows the number of all chargeable outgoing daily calls of a particular user (calls). The corresponding call duration per day (dur) is depicted in Figure 7(b), while the charging units per day (units) are given in Figure 7(c). Days with zero number of calls are not taken into account. This is because one way to identify cases of fraudulent behavior is to spot any statistical differences of the user's behavior before and after the fraud has been perpetrated. In this sense, days with zero number of calls do not hold neither legitimate nor fraudulent usage, and of course not any combination of them, so these days do not contribute any additional knowledge to the problem. Moreover, if these days were taken into account, they would have mitigated the effect of the fraudster's behavior.

Visual observation of each plot reveals different features; that is, in Figure 7(a) there is an interestingly high number of outgoing calls at day 144 while in Figure 7(c) there is an apparent outbreak in the charging units after day 175. In Figure 7(d), the inverse of the mean cost per sec per day is plotted. Here, two points (p1 and p2) stem out as outliers. It is interesting to notice that p1, day 42, stands in a position with no particular interest in the three previous plots. Thorough examination of the Call Detail Record (CDR)

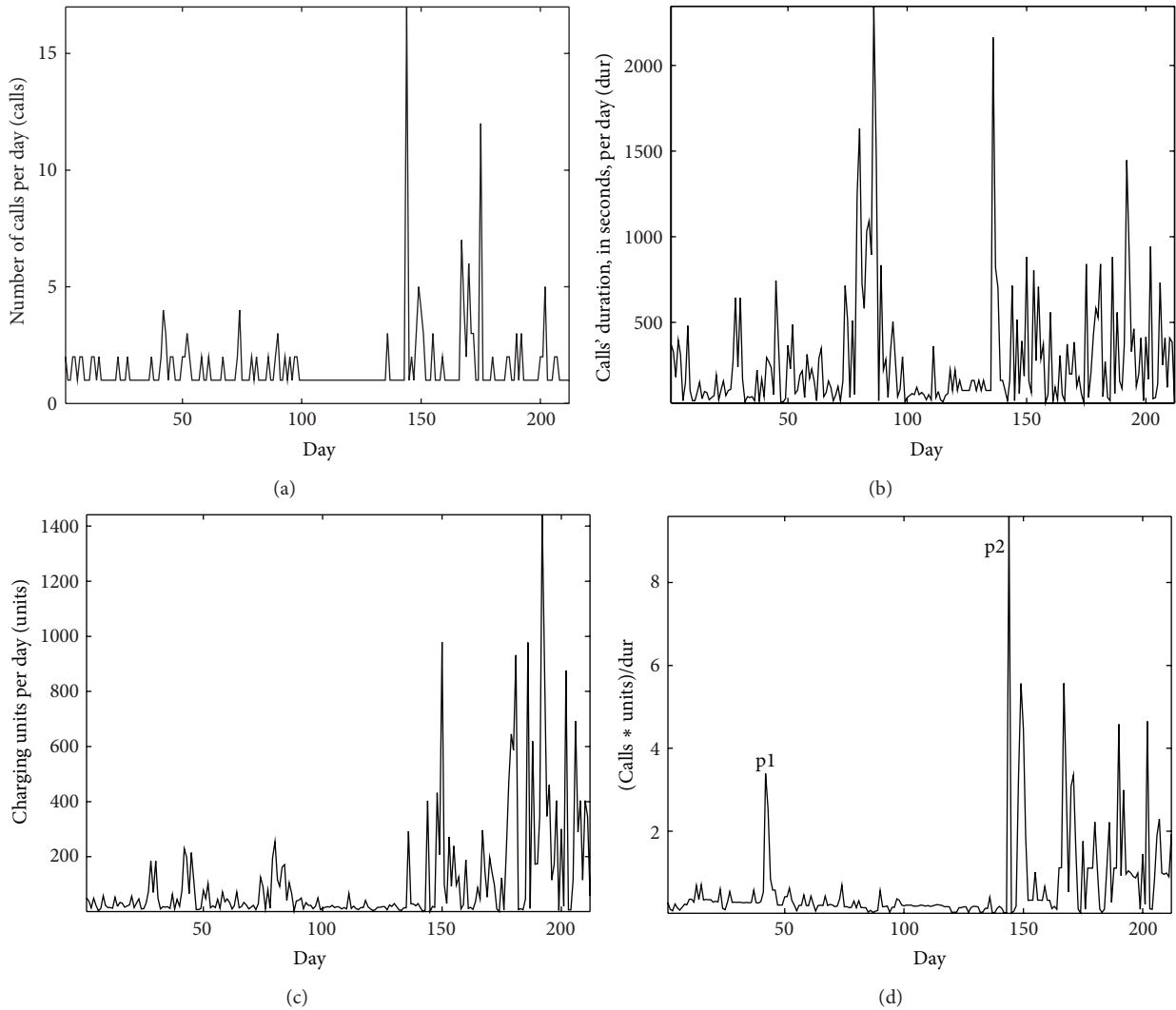


FIGURE 7: Several views of the daily behavior of a telecom user.

output of the organization’s Public Branch Exchange (PBX) for the particular user revealed that at p1 the user placed four calls to mobile destinations (since then he had never made more than 2 calls per day). This is an example of a case that would have caused a false alarm in a fraud detection system [14]. Nevertheless, it is a deviation from the user’s normal behavior. At p2 some calls are made to satellite services that are not associated with the user’s own telephone set. In this case, another person, that is, the fraudster, got access to the legitimate user’s PIN and used it for his own expensive calls. All behavior following p2 is dominated by the fraudster’s activity which is superimposed over the legitimate user’s one. It is, also, obvious that the parametric structure of this sequence changes after p2. The process of Figure 7(d) is differentiated as presented in Figure 8(a), in order to make the points of extreme behavior more prevalent. Differentiation, also, helps to make the process stationary about a constant mean level. The following are two plots of the skewness-kurtosis product for window sizes $W = 10$ (Figure 8(b)) and $W = 18$ (Figure 8(c)). Both give exact identification of all

the time points of interest. The larger the window used, the smaller the random fluctuations of the statistics are.

5. Comparison with Other Approaches

In order to further illustrate the method, the following comparisons were considered.

5.1. *The Cincinnati Data.* First, our method was applied on the Cincinnati data which were also used by McDowall et al. to introduce intervention analysis [4]. The daily number of calls to Cincinnati Directory Assistance was recorded for each month from January 1962 to December 1976. In March 1974, the 147th month of the series, Cincinnati Bell initiated a 0.20 USD charge for each call to Directory Assistance. There was no charge for these calls prior to this time. This intervention is exactly identified by our method at time position 147 (Figure 9).

5.2. *Comparison with the Continuous Wavelet Transform.* The performance of our method is also compared to the

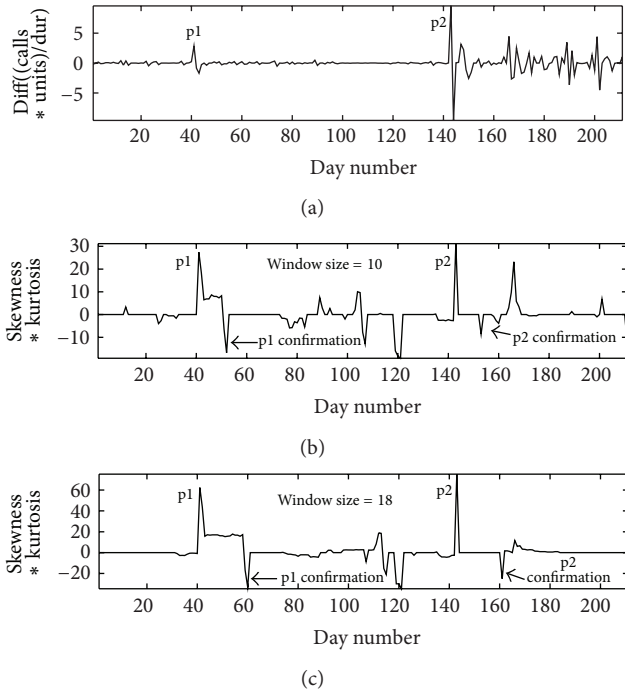


FIGURE 8: Fraud detection through outlier detection.

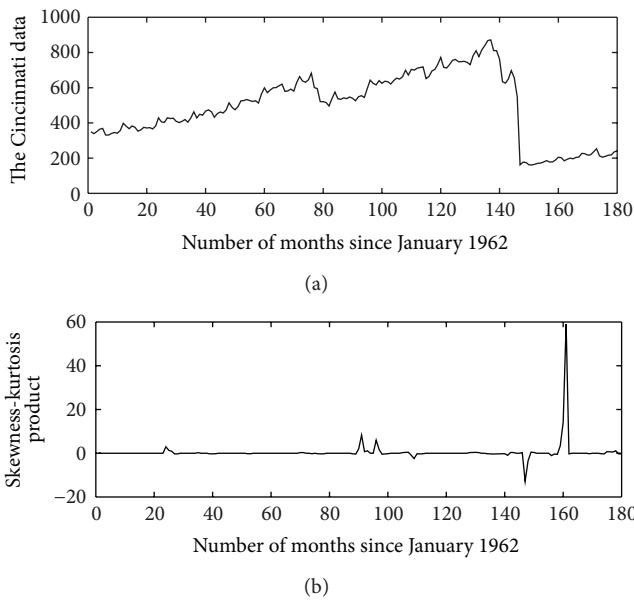


FIGURE 9: The Cincinnati data.

Continuous Wavelet Transform (CWT). The CWT has been used in numerous signal processing problems ranging from discontinuity detection and change point identification [24] to signal compression [20]. Given a *wavelet function* $\psi(t)$, the CWT is the transformation of a signal $f(t)$ to the wavelet coefficients $c(\alpha, b)$ according to

$$c(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(t) \psi^* \left(\frac{t-b}{a} \right) dt. \quad (10)$$

The variables a and b , scale and position, are the new dimensions after the wavelet transform and the $*$ represents

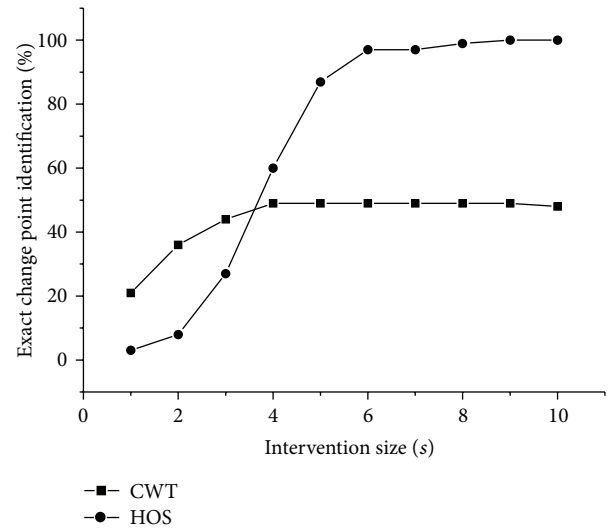


FIGURE 10: Exact change point detection. Comparison with the CWT. The CWT performs better for smaller intervention sizes while the proposed heuristic method is better for larger interventions.

operation of the complex conjugate [25]. The wavelet coefficients give a measure of the signal's similarity (or correlation) at some position b with a scaled version of the basis wavelet. Given the fact that several families of wavelets exist, each one with varying qualities, the success of a method based on wavelets relies on the selection of the appropriate wavelet.

For the purpose of our comparisons, we employ the *Haar* wavelet, which is discontinuous and resembles a step function. Decomposing a time series with a change in the level by means of the *Haar* wavelet is expected to yield wavelet coefficients with a minimum (or a maximum, depending on whether the change in the level is positive or negative, resp.) at the change point. Figure 10 gives the percentage of identification over the same set of 1000 time series for various intervention sizes. The window size that was used in the implementation of our proposed method (HOS) is $W = 18$. Wavelets have better performance for small interventions but for larger ones our method gives a more precise output.

It should be noted that the overall performance of the CWT was better but only with the use of the *Haar* wavelet. Other wavelets were also used, for example, the *Daubechies D4*, the *Daubechies D10*, and the *Mexican Hat*, but they could not detect the change in the level.

As regards outlier detection, the *Haar* wavelet is not appropriate. In this case, one should use another wavelet. On the contrary, the proposed heuristic method needs not to be altered, as it can cope with both changes in the level and outlier detection, see Section 4.5.

5.3. Comparison with the SDAR Algorithm. We also compare our method with the SDAR algorithm that was introduced by Yamanishi and Takeuchi [15]. In [15], a probabilistic model of the data source has been proposed that is incrementally learned using an online discounting learning algorithm, and then the score of any given data is calculated to measure its deviation from the observed model. This framework may deal

with both outlier and change point detection. In order to evaluate their method, Yamanishi and Takeuchi used a data sequence whose data between the change points was drawn according to the following AR(2) model:

$$x(k) = 0.6 \cdot x(k-1) - 0.5 \cdot x(k-2) + a(k), \quad (11)$$

where $a(k)$ is Gaussian random variable with mean 0 and variance 1. The dataset consists of 10000 records. The change points occur at time $j \times 1000$, ($j = 1, 2, \dots, 9$). The difference between the value of the $(j-1)$ th change point and that of the j th change point equals j . We use the same data sequence to compare the SDAR algorithm with our heuristic HOS approach.

The simulated dataset is depicted in Figure 11(a). Figure 11(b) shows the scores calculated by SDAR. In Figure 11(c), the skewness-kurtosis product is plotted. The arrows on the upper half of the plot (Figure 11(c)) point out the detected change points while the arrows on the lower half part of the plot point out the minima that follow and confirm the change point detection. One can see that both methods detect the same change points.

A drawback of the SDAR algorithm, against our approach, is that it cannot perform on time series with few data points. This is because the algorithm tries to learn the model parameters incrementally and needs several data in order to converge to the parameter values. Moreover, its implementation implies that the data generating process is AR(2), as its name also implies, which may not always be the case. The heuristic HOS approach makes no such assumption. However, it is interesting to note how the SDAR scores reflect the degree of changes in the level of the time series which cannot be achieved with the HOS.

6. Conclusions and Discussion

In the present paper, a method for the detection of changes in the level of a time series is presented. The method is a heuristic method based on an observation on how certain higher-order statistics, namely, the skewness and the kurtosis of the process, behave in the presence of a change in the level of the series. It can also be applied to detect outliers in the evolution of a time series. The method is tested with both simulated and real-world data. The latter are examples taken from the telecommunications industry. In particular, one is an interrupted time series which resulted after a change in service pricing policy, while the other is an example of superposition fraud. The proposed method maintains high specificity with the proper selection of the window size and can be applied to time series with only few data points. Adding to these, the method works with one scan of the data so it can be used for online detection of outliers or of changes in the level.

There was some consideration regarding the size of the window which in fact defines the sample size used and how this size affects the efficiency of the tests. The reader is referred to the work of Shapiro et al. [26] for an extensive study of various tests for normality. There, results are included which exhibit that a combination of skewness and kurtosis usually provides a sensitive judgment of nonnormality which

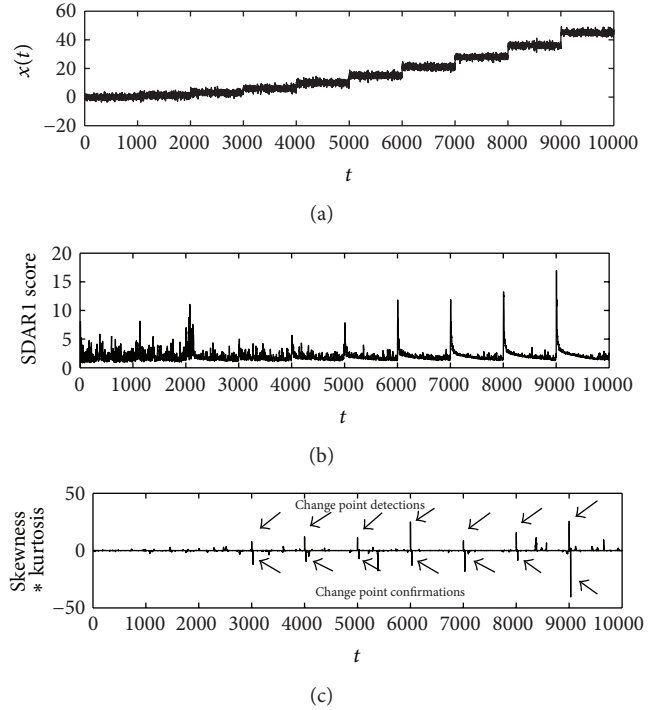


FIGURE 11: Detection of multiple change points in a time series. Comparison of the proposed heuristic HOS approach with the SDAR algorithm.

can be achieved with samples sizes less than 20. Compared with other intervention detection methods, our approach shows promise as it works on time series without any prior knowledge of the underlying process, the time and type of the intervention, or the size of the time series. It may also cope with time series with few data points.

Further research may include the study of the behavior of the method with different ARIMA processes and experimentation with more real-world problems. More work should also be added to check the behavior of the method in the case of interventions due to impulsive or exponential noise, while an appropriate extension to the approach may enhance its sensitivity for small interventions. It would also be interesting to extend the method in order to cope with multidimensional data.

Acknowledgments

The authors would like to thank the staff of the Telecommunications Center of the Aristotle University of Thessaloniki, Greece, for their contribution to real-world data. The authors wish to acknowledge the financial support provided by the Research Committee of the Technological Educational Institute of Serres under Grant SAT/IC/260111-2/6.

References

- [1] D. T. Campbell, "From description to experimentation: interpreting trends as quasi-experiments," in *Problems of Measuring Change*, C. W. Harris, Ed., University of Wisconsin Press, Madison, Wis, USA, 1963.

- [2] G. E. P. Box and G. C. Tiao, "A change in level of a non-stationary time series," *Biometrika*, vol. 52, pp. 181–192, 1965.
- [3] G. V. Glass, V. L. Wilson, and J. M. Gottman, *Design and Analysis of the Time Series Experiments*, Colorado Associated University Press, Boulder, Colo, USA, 1975.
- [4] D. McDowall, R. McCleary, E. E. Meidinger, and R. A. Hay, *Interrupted Time Series Analysis*, Sage, Newbury Park, Calif, USA, 1980.
- [5] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, Calif, USA, 2nd edition, 1976.
- [6] D. K. Simonton, "Cross-sectional time-series experiments: some suggested statistical analyses," *Psychological Bulletin*, vol. 84, no. 3, pp. 489–502, 1977.
- [7] W. F. Velicer and R. P. MacDonald, "Time series analysis without model identification," *Multivariate Behavioral Research*, vol. 19, pp. 33–47, 1984.
- [8] J. W. Harrop and W. F. Velicer, "A comparison of alternative approaches to the analysis of interrupted time series," *Multivariate Behavioral Research*, vol. 20, pp. 27–44, 1985.
- [9] W. L. Padia, *Effect of autocorrelation on probability statements about the mean [M.S. thesis]*, Laboratory of Educational Research, University of Colorado, 1973.
- [10] D. M. Hawkins, P. Qiu, and C. W. Kang, "The changepoint model for statistical process control," *Journal of Quality Technology*, vol. 35, no. 4, pp. 355–366, 2003.
- [11] D. P. Reilly, P. S. Foote, D. H. Lindsay, and A. Campbell, "Fraud detection using autoboxstm automatic intervention detection," in *Proceedings of the International Symposium on Forecasting*, Dublin, Ireland, 2002.
- [12] L. Bianchi, J. Jarrett, and R. Choudary Hanumara, "Improving forecasting for telemarketing centers by ARIMA modeling with intervention," *International Journal of Forecasting*, vol. 14, no. 4, pp. 497–504, 1998.
- [13] V. Guralnik and J. Srivastava, "Event detection from time series data," in *Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '99)*, pp. 33–42, 1999.
- [14] T. Fawcett and F. Provost, "Activity monitoring: noticing interesting changes in behavior," in *Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '99)*, pp. 53–62, 1999.
- [15] K. Yamanishi and J. Takeuchi, "A unifying framework for detecting outliers and change points from non-stationary time series data," in *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '02)*, pp. 676–681, July 2002.
- [16] T. Matsumoto and K. Yosui, "Adaptation and change detection with a sequential Monte Carlo scheme," *IEEE Transactions on Systems, Man, and Cybernetics B*, vol. 37, no. 3, pp. 592–606, 2007.
- [17] C. S. Hilas, I. T. Rekanos, S. K. Goudos, P. A. Mastorocostas, and J. N. Sahalos, "Level change detection in time series using higher order statistics," in *Proceedings of the 16th International Conference on Digital Signal Processing (DSP '09)*, pp. 1–6, July 2009.
- [18] I. T. Rekanos and L. J. Hadjileontiadis, "An iterative kurtosis-based technique for the detection of nonstationary bioacoustic signals," *Signal Processing*, vol. 86, no. 12, pp. 3787–3795, 2006.
- [19] C. D. Saragiotis, L. J. Hadjileontiadis, I. T. Rekanos, and S. M. Panas, "Automatic P phase picking using maximum kurtosis and κ -statistics criteria," *IEEE Geoscience and Remote Sensing Letters*, vol. 1, no. 3, pp. 147–151, 2004.
- [20] R. S. H. Istepanian, L. J. Hadjileontiadis, and S. M. Panas, "ECG data compression using wavelets and higher order statistics methods," *IEEE Transactions on Information Technology in Biomedicine*, vol. 5, no. 2, pp. 108–115, 2001.
- [21] J. H. Zar, *Biostatistical Analysis*, Prentice Hall, New Jersey, NJ, USA, 2nd edition, 1984.
- [22] J. F. Kenney, *Mathematics of Statistics*, Van Nostrand, Toronto, Canada, 2nd edition, 1948.
- [23] A. Papoulis, *Probability and Statistics*, Prentice-Hall, 1990.
- [24] M. Lavielle, "Detection of multiple changes in a sequence of dependent variables," *Stochastic Processes and their Applications*, vol. 83, no. 1, pp. 79–102, 1999.
- [25] G. Kaiser, *A Friendly Guide to Wavelets*, Birkhäuser, New York, NY, USA, 1994.
- [26] S. S. Shapiro, M. B. Wilk, and H. J. Chen, "A comparative study of various tests for normality," *Journal of the American Statistical Association*, vol. 63, pp. 1343–1372, 1968.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

