Just-in-Time Adaptive Classifiers—Part I: Detecting Nonstationary Changes

Cesare Alippi, *Fellow, IEEE*, and Manuel Roveri

***Abstract*—The stationarity requirement for the process generating the data is a common assumption in classifiers' design. When such hypothesis does not hold, e.g., in applications affected by aging effects, drifts, deviations, and faults, classifiers must react just in time, i.e., exactly when needed, to track the process evolution. The first step in designing effective just-in-time classifiers requires detection of the temporal instant associated with the process change, and the second one needs an update of the knowledge base used by the classification system to track the process evolution. This paper addresses the change detection aspect leaving the design of just-in-time adaptive classification systems to a companion paper. Two completely automatic tests for detecting nonstationarity phenomena are suggested, which neither require *a priori* information nor assumptions about the process generating the data. In particular, an effective computational intelligence-inspired test is provided to deal with multidimensional situations, a scenario where traditional change detection methods are generally not applicable or scarcely effective.**

***Index Terms*—Intelligent systems, learning systems, neural networks, pattern classification.**

NOMENCLATURE

$\Theta = \{\theta_1, \ldots, \theta_\lambda\}$	Parameter vector.
$p_\Theta(x)$	Probability density function (pdf) parameterized in Θ.
$R(t)$	Cumulative sum of the log-likelihood ratio at time t.
$m(t)$	Minimum of $R(\tau)$ in the $[0, t]$ interval.
$g(t)$	Difference between $R(t)$ and $m(t)$.
h	Threshold of the CUSUM test.
N	Number of samples of the test configuration sequence.
H_0	Null hypothesis.
H_1	Alternative hypothesis.
γ	Sensitivity of the test.
$\varphi(t)$	Feature vector.
$\varphi_{\mathrm{PCA}}(t)$	Feature vector φ transformed with PCA.
$\varphi_Y(t)$	Average of φ_{PCA} with respect to (w.r.t.) a time window.

Manuscript received November 13, 2006; revised February 27, 2007 and June 15, 2007; accepted November 9, 2007. First published March 12, 2008; last published July 7, 2008 (projected).

The authors are with the Dipartimento di Elettronica e Informazione, Politecnico di Milano, 20133 Milano, Italy (e-mail: cesare.alippi@polimi.it; alippi@elet.polimi.it; roveri@elet.polimi.it).

Digital Object Identifier 10.1109/TNN.2008.2000082

I. INTRODUCTION

ASSESSING the stationarity of a data generating process, i.e., verifying whether drifts, abrupt changes, smooth deviations, and aging effects have affected the data-generating mechanism or not, is a key issue in industrial, environmental, and medical applications [1]. Monitoring over time the validity of the stationary hypothesis allows the designer to do the following: 1) verifying if the stationarity hypothesis holds (many applications, e.g., classifier design, system identification, and fault detection are designed under the assumption that the process is stationary); 2) taking actions, e.g., by updating the classification systems to track the system evolution.[1]

The problem of assessing the stationarity hypothesis can be addressed with three main approaches: *data driven*, *analytical*, or *knowledge based* [1]. A data-driven approach directly inspects data coming from the process and assumes that the available data set is large enough to assess the validity of the stationarity hypothesis with large confidence [2], [3]. This solution guarantees a good drift detection ability without requiring any *a priori* information about the process under investigation. The analytical modality assumes that a mathematical description of the process generating the data is available: only few data are hence required to assess the hypothesis [4], [5]. The knowledge-based modality assumes instead that some *a priori* information about the process (but not the model) is available, e.g., derived from samples; in this class, we have causal analysis and expert systems [4], [6], [7].

We believe that an effective stationarity assessment test should be accurate and flexible enough to deal with a large class of applications without requiring any *a priori* information, as they are hardly available in the real world. As such, here, a data-driven approach, which assumes that *no critical a priori information is available*, is suggested.

In the neural network community, identification of the time instant associated with the loss in stationarity allows the designer to take actions, e.g., by updating the network weights to track the process evolution [9], [10], retraining the classifier [11], [12] exactly when needed, detecting or giving a diagnosis

[1]This second aspect is addressed in the companion paper [8].

about the presence of faults in the neural-network-based solution [13], [14], and assessing the effectiveness of adaptive preprocessing techniques designed for nonstationarity applications [15], [16].

A relatively large literature addresses classifiers in specific nonstationary applications, e.g., [17], [18] with detection of the process change implicitly assumed as known (here, a mechanism for detecting a change occurrence is provided hence completing their analysis). Conversely, if the change detection test asserts that the process remains stationary, supervised information coming from the field can be fully exploited and integrated into consistent neural network classifiers to asymptotically improve accuracy as delineated by the theory [19].

The most common data-driven techniques for change detection are parametric and nonparametric statistical tests [1], [2]. Simple parametric techniques for testing a signal change—or hypothesis—are the student t-test and the Fisher f-test [20], [21], which address changes in mean and variance, respectively. More articulated analyses, e.g., based on regression techniques [22], need to be considered to contemporarily deal with mean and variance (but are weakly effective in detecting small changes). Parametric tests generally require availability (or an estimate) of the probability density functions of the process generating the data before and after the change and/or information about the nature of the drift [20], [21]. Differently, nonparametric tests are somehow more general because they do not require strong *a priori* information [21], [23]. The Mann–Whitney U-test for independent samples [24] (which relies on the possibility to rank two independent samples of observations) and the Wilcoxon signed-rank test for related samples [25] (which can be considered as a nonparametric alternative to the t-test) are nonparametric tests originally designed for detecting a single point change (which also represents their main limit). The Mann–Kendall [26] (designed to analyze climatic changes) and the CUmulative SUM (CUSUM) [27], [28] (developed in the system control community to detect structural changes) nonparametric tests are particularly suitable in sequential analysis. In particular, CUSUM has been successfully used in several diversified applications such as fault detection, onset detection in seismic signal processing [29], [30], and changes in mechanical systems [31], [32]; it provides a relative simplicity, a graphical interpretation of results and the ability to detect unusual patterns whereas Mann–Kendall is characterized by a low computational complexity.

Another general, pdf-free method for detecting changes in dynamical systems has been suggested in [33] and [34]. Interestingly, the author suggests a heuristics to identify the required parameters hence aiming at a first automatic identification of the test parameters.

Change detection issues have also some affinity with design for testability, e.g., refer to [35]–[37]; however, a comprehensive analysis is outside the goal of this paper.

All traditional tests generally require a design-time configuration phase to fix the test parameters (e.g., Mann–Kendall requires a significance level for the test while CUSUM needs to fix some thresholds to detect changes); such parameters are identified by exploiting *a priori* information or through a trial-and-error approach.

The aim of this paper is to develop a novel, general, *a priori* information-free, automatic change detection test: general in the sense that it only requires temporal independence among data, and automatic because the designer can easily configure the needed parameters at design time. We paid attention to the computational complexity of the test to grant real-time execution also in embedded systems such as wireless sensor networks, generally characterized by energy constraints and low computational abilities. The results of the study are two CUSUM-inspired tests, both effective in detecting abrupt changes and smooth drifts, differentiating themselves on approach, validity, and computational complexity.

The structure of this paper is as follows. Section II presents the traditional CUSUM test and the suggested extended approach to CUSUM. Section III addresses the development of a computational intelligence-based test for signal change detection. Experimental results are finally given in Section IV.

II. Extended CUSUM Test

Let S be a stochastic process producing, over time, the sequence $X = \{x(t)\}$, $t = 1, \ldots, N$, of scalar, real, independent and identically distributed (i.i.d.) random samples subject to pdf $p_{\Theta}(x)$ parameterized in the parameter vector $\Theta = \{\theta_1, \theta_2, \ldots, \theta_\lambda\}$. Assume that S changes its statistical behavior at unknown time $\bar{t}$ without modifying its pdf model family; under such hypothesis, the change in S can be modeled as a parameter transition from Θ^0 to Θ^1.

To evaluate the discrepancy between the two pdfs at time t, CUSUM requires first computation of the cumulative sum of the log–likelihood ratio

$$R(t) = \sum_{\tau=1}^{t} \ln \frac{p_{\Theta^1}(x(\tau))}{p_{\Theta^0}(x(\tau))}$$

and, then, the minimum of the sequence $m(t) = \min_{1 \le \tau \le t} (R(\tau))$. Traditional CUSUM identifies a change in S at time $\bar{t}$ when the difference $g(\bar{t}) = R(\bar{t}) - m(\bar{t})$ overcomes a given threshold h.

CUSUM is rather effective in detecting changes but it requires availability of the pdf and parameters Θ^0, Θ^1, h at design time, information which is rarely available (in the best case, parameters can be estimated through a trial-and-error approach, e.g., by knowing that, following a drift or a fault, the process will move from Θ^0 into a new working point Θ^1).

By assuming that samples $x(t)$ are i.i.d., N is perfectly divisible by n, and n is sufficiently large to grant validity of the central limit theorem, a pdf-free CUSUM test can be derived. In particular, random variable

$$y(t) = \frac{1}{n} \sum_{\tau=(t-1)n+1}^{tn} x(\tau), \qquad t = 1 \ldots N/n \tag{1}$$

is ruled by a Gaussian pdf with mean and variance parameters $\Theta = \{\mu, \sigma^2\}$ which can be estimated on the transformed sequence $Y = \{y(i)\}$, $i = 1, \ldots, N/n$, and provide the null hypothesis $H_0 : \Theta^0 = \{\hat{\mu}_0, \hat{\sigma}_0^2\}$ requested by the traditional CUSUM. In different words, the suggested pdf-free mechanism of the extended CUSUM derives from the central limit theorem providing, for any process generating the data satisfying the i.i.d. hypothesis, a canonical Gaussian pdf family.

To complete the test, the designer has to provide the alternative hypothesis $H_1 : \Theta^1 = \{\hat{\mu}_1, \hat{\sigma}_1^2\}$ which, to address any type of nonstationarity changes, can be defined as not being in Θ^0, in contrast to being in Θ^1. The "not being in Θ^0" statement must be intended in statistical terms by developing a confidence neighborhood for Θ^0. Such a neighborhood can be generated by considering the $100(1-\alpha)\%$ confidence intervals $[\hat{\mu}_{\min}, \hat{\mu}_{\max}]$ and $[\hat{\sigma}^2_{\min}, \hat{\sigma}^2_{\max}]$ for $\hat{\mu}_0$ and $\hat{\sigma}_0^2$, where α is a confidence parameter.

Differently from traditional CUSUM, the method is not aiming at estimating the alternative hypothesis Θ^1 but, instead, at defining a zone of the hypothesis space that represents the "not being in Θ^0" situation.

By assuming that Θ^0 can be approximated with the ideal one (i.e., N/n is sufficiently large to allow the designer for considering the estimated parameters instead of the real ones), the alternative hypothesis $H_1 : \Theta^1 = \{\hat{\mu}_1, \hat{\sigma}_1^2\}$ can be obtained by identifying a neighborhood confidence for Θ^0 [20]. In particular, parameter $\hat{\mu}_1$ becomes

$$\hat{\mu}_1 = \hat{\mu}_0 \pm \gamma \sqrt{\frac{n}{N}} Z_{\alpha/2} \hat{\sigma}_0 \tag{2}$$

where Z and χ^2 are the normal and the chi-square distributions, respectively, and γ is the sensitivity of the test. Likewise, the confidence interval extremes for the variance are [20]

$$\begin{aligned}\hat{\sigma}^2_{1,\max} &= \hat{\sigma}^2_0 + \gamma\left(\hat{\sigma}^2_{\max} - \hat{\sigma}^2_0\right) = \hat{\sigma}^2_0 + \gamma\left(\frac{\frac{N}{n} - 1}{\chi^2_{1-\alpha/2;\frac{N}{n}-1}}\right)\hat{\sigma}^2_0\\ \hat{\sigma}^2_{1,\min} &= \hat{\sigma}^2_0 - \gamma\left(\hat{\sigma}^2_0 - \hat{\sigma}^2_{\min}\right) = \hat{\sigma}^2_0 - \gamma\left(1 - \frac{\frac{N}{n} - 1}{\chi^2_{\alpha/2;\frac{N}{n}-1}}\right)\hat{\sigma}^2_0.\end{aligned} \tag{3}$$

The presence of $\pm$ in (2) implies that we can both identify a change and its monotonical behavior (increasing $+$, decreasing $-$). To detect both changes, the designer has to configure and run two tests, one for detecting the increasing change and the other for detecting the decreasing change in μ. The same note holds for the variance (3), where we have two possible values for $\hat{\sigma}_1^2 = [\hat{\sigma}^2_{1,\min}, \hat{\sigma}^2_{1,\max}]$. Additional hypotheses for Θ^1 can be configured by considering combinations of $\hat{\mu}_{\min}$, $\hat{\mu}_{\max}$, $\hat{\sigma}^2_{\min}$, and $\hat{\sigma}^2_{\max}$.

By increasing γ, the test is less sensitive to changes affecting Θ^0; conversely, by decreasing γ, the test becomes very sensitive to small changes.

Once parameters Θ^0 and Θ^1 have been set, we compute $R(t)$ and $g(t)$ for $1 \leq t \leq N/n$; the threshold h to be used can be estimated as the maximum $g(t)$ in sequence Y

$$h = \max_{1 \leq t \leq N/n} g(t). \tag{4}$$

Once Θ^0, Θ^1, and h are estimated, the traditional CUSUM test becomes operational: the test detects a change in S at time $\bar{t} > N$ when fresh incoming data, transformed according to (1), provide a $g(\bar{t})$ above threshold h.

Results, here suggested for a monodimensional signal, can be suitably extended to cover multidimensional cases. Such an extension will be provided in Section III.

III. Computational Intelligence-Based CUSUM Test

The components of Θ can be intended as features used by the extended CUSUM to detect a nonstationarity condition. As such, in search for a more expressive, robust, and effective test embedded to the application, it is timely to think of extending basic features $\{\mu, \sigma^2\}$ with a further set extracted from the signal(s). Additional features can be extracted from input data by using suitable kernels (e.g., see [38]) and/or by considering features suggested in the specific change detection literature. In other words, designers can consider their favorite features to construct a new nonstationarity detection test.

In the following, the operative framework, notations and hypotheses are those leading to the extended CUSUM, i.e., inputs $x(t)$ are i.i.d., here with $x(t)$ extended to the D-dimensional case.

For each input, we compute mean μ, standard deviation σ, features derived from the pdf and the cumulative density function (cdf), as well as features inspired by Mann–Kendall (indices $4D+1$ to $8D$) and CUSUM (indices $8D+1$ to $12D$)

$$\begin{aligned}\varphi_j(t) &= \left|\hat{\mu}_{X^j} - \hat{\mu}_{X_O^j}\right|, \qquad j = 1, \ldots, D\\ \varphi_{D+j}(t) &= \left|\hat{\sigma}_{X^j} - \hat{\sigma}_{X_O^j}\right|, \qquad j = 1, \ldots, D\\ \varphi_{2D+j}(t) &= \int_{X^j} \left|\hat{\mathrm{pdf}}_{X^j} - \hat{\mathrm{pdf}}_{X_O^j}\right| dx, \qquad j = 1, \ldots, D\\ \varphi_{3D+j}(t) &= \int_{X^j} \left|\hat{\mathrm{cdf}}_{X^j} - \hat{\mathrm{cdf}}_{X_O^j}\right| dx, \qquad j = 1, \ldots, D\\ \varphi_j(t) &= \sum_{\tau=2}^{t} \mathrm{sgn}\left(\varphi_{j-4D}(\tau) - \varphi_{j-4D}(\tau-1)\right), \qquad 4D+1 \leq j \leq 8D\\ \varphi_j(t) &= \sum_{\tau=2}^{t} \ln\left(\frac{\varphi_{j-4D}(\tau)}{\varphi_{j-4D}(\tau-1)}\right), \qquad 8D+1 \leq j \leq 12D\end{aligned} \tag{5}$$

where $X = \{x(t) : 1 \leq t \leq N\}$ and $X_O = \{x(t) : t > N\}$ represent the parameter configuration sequence and the operational one, respectively; X^j and X_O^j denote the jth components of X and X_O; sgn is the sign function; and $\hat{\mathrm{pdf}}$ and $\hat{\mathrm{cdf}}$ are the estimated probability and cumulative density functions, respectively.

In order to consider information at a group level, we also envisage features obtained by correlating any two features of (5) through the linear correlation operator ρ

$$\varphi_{(11+i)D+(j-1)}(t) = \left|\hat{\rho}\left(\varphi_i(t), \varphi_j(t)\right)\right|, \quad \{j > i, 1 \leq i, j \leq 12D\}. \tag{6}$$

To sum up, there are $4D$ features from traditional change detection tests, $4D$ from Mann–Kendall, $4D$ from CUSUM, and $\binom{D}{2}$ features derived from the correlation operator for a total of $12D + \binom{D}{2}$ features. Because a supervised feature selection phase cannot be considered to reduce the complexity of the input space, a principal component analysis (PCA) technique is suggested which, applied to vector $\varphi(t)$, provides a reduced feature vector $\varphi_{\mathrm{PCA}}(t)$. The number of considered eigenvectors can be empirically identified by removing those eigenvalues whose sum is below a threshold, e.g., one thousandth of the sum of the others.

The $\varphi_{\mathrm{PCA}}(t)$ is then used to configure the parameters of the test and, because the pdf of φ_{PCA} is not available, we apply the same framework delineated in Section II for the extended CUSUM. In particular, by averaging each ith component of $\varphi_{\mathrm{PCA}}(t)$ as

$$\varphi_Y^i(t) = \frac{1}{n} \sum_{\tau=(t-1)n+1}^{tn} \left|\varphi_{\mathrm{PCA}}^i(\tau)\right| \tag{7}$$

and, by invoking the central limit theorem, transformed $\varphi_Y(t)$ vector follows a multivariate Gaussian distribution $N(M, C)$ of mean M and covariance matrix C.

Once estimated, mean $\hat{M}$ and covariance matrix $\hat{C}$ of $\varphi_Y(t)$ constitute the null reference hypothesis $H_0 : \Theta^0 = \{\hat{M}_0, \hat{C}_0\}$ needed by the change detection test. The alternative hypothesis $H_1 : \Theta^1 = \{\hat{M}_1, \hat{C}_1\}$ follows by considering a $100(1-\alpha)\%$ confidence interval for the sample mean and the covariance matrix of $\varphi_Y(t)$. In particular, in analogy with (2) and (3), we have

$$\hat{M}_1 = \hat{M}_0 \pm \gamma \sqrt{\frac{n}{N}} Z_{\alpha/2} \sqrt{\mathrm{diag}(\hat{C}_0)} \tag{8}$$

$$\hat{C}_{1,\min} = \hat{C}_0 - \gamma \left(\frac{\frac{N}{n} - 1}{\chi^2_{\alpha/2; \frac{N}{n}-1}} - 1 \right) \hat{C}_0$$

$$\hat{C}_{1,\max} = \hat{C}_0 - \gamma \left(\frac{\frac{N}{n} - 1}{\chi^2_{1+\alpha/2; \frac{N}{n}-1}} - 1 \right) \hat{C}_0 \tag{9}$$

where, again, γ references the sensitivity parameter of the test and $\sqrt{\mathrm{diag}(\hat{C})}$ is the vector containing the square root of diagonal elements of $\hat{C}$. Finally, we compute $R(t)$, $m(t)$, and $g(t)$ on X. The maximum value of $g(t)$ in the configuration sequence is used as threshold value $h = \max_{1 \le \tau \le N/n} g(\tau)$; the extended CUSUM test can now be applied also to multidimensional problems.

A. *Geometrical Displacement of* Θ^1 *in the Hypothesis Space*

Effectiveness of the computational intelligence-based CUSUM (CI-CUSUM) test requires selection of the most appropriate alternative hypotheses Θ^1's. Because each Θ^1 corresponds to a different test, we have to tradeoff computational complexity of the test phase with its performance (generation of false positive and negative in change detection). In fact, while generation of the alternative hypotheses Θ^1 in the extended CUSUM can provide four possible alternative hypotheses, the multidimensional nature of $\varphi_Y(t)$ would generate an exponential number of alternative hypotheses in the dimension of $\varphi_Y(t)$. This NP-hard computational problem is less critical than it might appear as experienced in real applications where rarely the number of features exceeds five (inducing 32 alternative hypotheses) thanks to PCA. Moreover, because features $\varphi_Y(t)$ have been designed to generate positive increments in response to changes in the process, only alternative hypotheses in (8) and (9) inducing a positive increment of $g(t)$ are of interest.

Which alternative hypothesis should the designer consider in a large hypothesis space? If *a priori* information about the possible change in the process is available, i.e., the trajectory of $\varphi_Y(t)$ induced by the change in the feature space is known, *ad hoc* tests can be designed. Conversely, when such an information is not available—which is generally the case—four sets of configurations can be suggested as candidates to the alternative hypothesis. By defining false positive and false negative as the times, a test detects a change in the sequence when the change is not there and it does not detect a change when the change is there, and by denoting the size of the hypothesis vector Θ with $\lambda = |\Theta|$, the increase of the ith parameter in $\Theta = \{\theta_1, \theta_2, \ldots, \theta_i, \ldots, \theta_\lambda\}$ computed according to (8) and (9) with $\Delta\theta_i > 0$, and the null hypothesis with $\Theta^0 = \{\theta_1^0, \theta_2^0, \ldots, \theta_i^0, \ldots, \theta_\lambda^0\}$, the suggested four sets of configurations for Θ^1 are as follows.

- *Configuration #1* is composed by a single alternative hypothesis representing the situation where all components in $\varphi_Y(t)$ increase $\Theta^1 = \{\theta_1^0 + \Delta\theta_1, \theta_2^0 + \Delta\theta_2, \ldots, \theta_i^0 + \Delta\theta_i, \ldots, \theta_\lambda^0 + \Delta\theta_\lambda\}$. For its nature, the solution guarantees the lowest false positive rate (among the suggested configurations), but it might suffer from the presence of a high false negative one. The computational complexity of the test configuration does not depend on λ.
- *Configuration #2* contains configuration #1 and all test configurations characterized by a null increment only in a generic component of the Θ^1 vector. The complexity of the test solution is $O(\lambda + 1)$, linear in λ. This configuration improves over the previous one by reducing the false negative rate.
- *Configuration #3* starts from the null hypothesis and increases only one component at time. The result is a set of λ alternative hypotheses. This configuration guarantees the lowest false negative rate compared to the proposed configurations but it might suffers from false positives due to the curse of dimensionality.
- *Configuration #4* considers all possible combinations $(\theta_i + \Delta\theta_i)$ of the λ parameters of Θ^1. Even if this approach is intuitively justified by the need to consider all possible variations, the exponential computational complexity $O(2^\lambda/2)$ makes this solution intractable for large λ's.

Configuration #2 is an appealing solution for most of applications. We experimentally found it provides a convenient compromise between change detection performance and computational complexity. The designers can select from configuration #1 to address low false positive rates or configuration #3 for low false negative ones. Fig. 1 shows the four candidate configurations in the $\lambda = 3$ case.

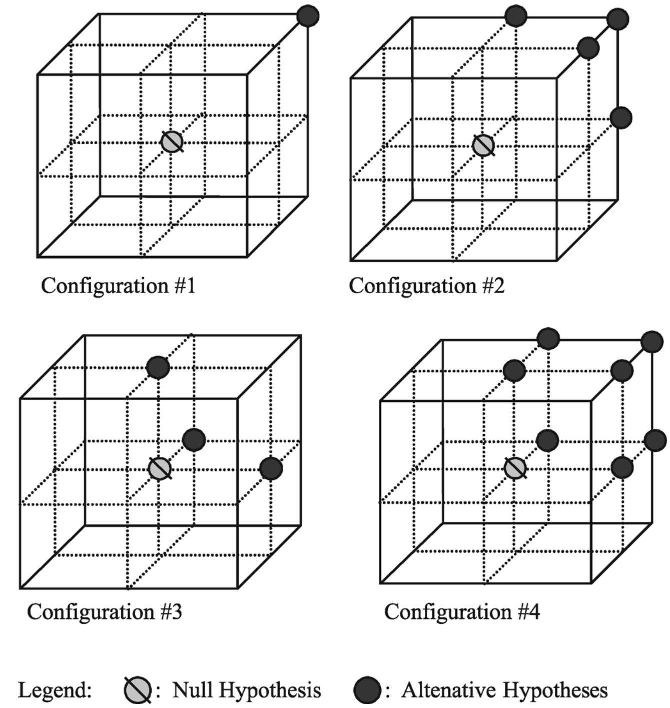

Fig. 1. Four CI-CUSUM configuration, $\lambda = 3$.

IV. Experimental Section: Assessing the Change Detection Tests

To validate the effectiveness of the suggested nonstationary detection tests, we considered six applications, two of which (D1 and D3) are simulated and four (D2, D4, D5, and D6) are coming from real experiments. Obtained results are compared and contrasted with the ones provided by traditional CUSUM and Mann–Kendall whenever such methods can be applied (CUSUM requires the pdf of the process generating the data while Mann–Kendall can only be applied to monodimensional signals).

- *Application D1* refers to a simple monodimensional process ruled by a known Gaussian pdf. The process lasts 5000 samples; a nonspectral perturbation affecting the mean value (which changes from $\mu_0 = 100$ to $\mu_1 = 105$, 5% perturbation, $\sigma = 3$) is injected after 2500 samples. The CUSUM test was configured by using the available *a priori* information $\Theta^0 = \{\mu_0\}$ and $\Theta^1 = \{\mu_1\}$.
- *Application D2* coincides with the SATIMAGE benchmark [39], i.e., classification of Landsat Multispectral Scanner images in seven classes (6435 samples of 36 features). A feedforward neural network classifier was considered as suggested in [40] (single hidden layer of six neurons and one output neuron). The change affects the neural network weights at sample 4000.
- *Application D3* refers to self-assembled monolayer (SAM) [41], [42] gas sensors. The model of the sensor resistance (to be considered unknown to all change detection tests) is in stationary conditions

$$R = R^0 \left[1 + K \frac{\sum_i \alpha_i p_i}{1 + \sum_i \beta_i p_i} \right]$$

  where α_i and β_i are real numbers, p_i is the partial pressure of the ith gas, $K > 0$ is a physics constant, and R^0 is the sensor resistance measured in a reference gas. The application considers a set of five SAM sensors, two features (the sensor measurement and its derivative) extracted from each signal generating a ten-dimensional feature vector to be inspected for variations (sensors differ only for the production process). The parameter affected by the variation is R^0 for each sensor; 10 000 samples coming from 150 different simulated acquisitions (pseudorandom binary signals, i.e., steps with random amplitudes and time duration) were considered; changes started at sample 5000.
- *Application D4* refers to a monitoring application; signals come from a set of photodiodes receiving X-rays which, in similar absorbing conditions, show a group behavior. Here, the goal is to recognize nonstationary behaviors of the photodiode X-ray source system in subsequent several days acquisitions by inspecting 16 photodiodes (11 680 samples per experiment).
- *Data Set D5* refers to the physiological data benchmark suggested in [43]. The data set contains two measurements [the respiratory signal (RPS) and the electrooculography signal (EOG)] retrieved from naps of 15 healthy people. According to neurophysiologists, the afternoon nap represents a "switch into a different mode" for the neuronal activity, i.e., the cerebral activity commutes from a stationary condition into a new stationary state. The aim of the experiment is to detect the state change by exploiting the respiratory and the electroencephalogram signals (from 9650 to 19 060 samples per experiment).
- *Data Set D6* refers to the Tennessee Eastman benchmark addressing the control of a complete chemical plant [44]. The data set [45], which includes 41 observation variables, consists of 22 different simulation runs (one fault-free case and 21 fault-affected cases). Each run addresses 48 hours with the fault affecting the chemical plant induced at hour 8.

Whenever the process was known, i.e., in applications D1 and D2, both abrupt and drift changes were considered. For abrupt changes, a multiplicative perturbation model was applied, which requires that the generic parameter P is affected by a perturbation δ changing its value from P^0 to P^1: $P^1 = P^0(1+\delta)$. Drifts have been modeled as a linear evolution of parameter P from P^0 to $P^1 = P^0(1+\delta)$ at the end of the experiment. The level of significance for Mann–Kendall was fixed at 99.9%; $n = 20$ and $\gamma = \sqrt{N/n}$ for the extended CUSUM and the CI-CUSUM tests. The chosen γ allows the designer to consider a 99.9% confidence interval on the mean value. The extended CUSUM test and the CI-CUSUM test were configured using $N = 1000$ samples. The CI-CUSUM test was designed by considering configuration #2; results of other configurations (#1, #3, and #4) for the considered D1–D6 applications are presented (and discussed) at the end of this section.

Results given in the following have been averaged over 150 runs, with δ uniformly extracted in each experiment from the $[-2, 2]$ interval.

The following four indices are suggested to assess the performance of the tests:

- false positive index (FP): it counts the times a test detects a change in the sequence when it is not there;
- false negative index (FN): it counts the times a test does not detect a change when it is there;

TABLE I
SIMULATION RESULTS FOR THE CONSIDERED DATA SETS

			CUSUM	Mann-Kendall	Extended CUSUM	CI-CUSUM
D1	**Abrupt**	**FP (%)**	0	2.6	0	**0**
		FN (%)	0	0	3.33	**0**
		RCS (sample)	17.15	60.7	63.6	**385.9**
		CT (s)	0.08	230.4	0.10	**3.5**
D2	**Abrupt**	**FP (%)**	Na	0	0	**0**
		FN (%)	Na	3.3	11.3	**1.9**
		RCS (sample)	Na	204.4	155.1	**536.7**
		CT (s)	Na	97.7	0.10	**4.8**
	Drift	**FP (%)**	Na	0	0	**0**
		FN (%)	Na	3.3	13.3	**2.6**
		RCS (sample)	Na	372.6	428.6	**512.7**
		CT (s)	Na	93.6	0.1	**7.79**
D3	**Abrupt**	**FP (%)**	Na	Na	Na	**1.33**
		FN (%)	Na	Na	Na	**6.71**
		RCS (sample)	Na	Na	Na	**380.5**
		CT (s)	Na	Na	Na	**276.4**
	Drift	**FP (%)**	Na	Na	Na	**1.3**
		FN (%)	Na	Na	Na	**7.3**
		RCS (sample)	Na	Na	Na	**1780.2**
		CT (s)	Na	Na	Na	**276.5**
D4		**FP (%)**	Na	9.8	0	**0**
		FN (%)	Na	5.7	82.3	**4.2**
		RCS (sample)	Na	1061.7	1319.8	**952.1**
		CT (s)	Na	130.6	0.07	**12.1**
D5	**EOG**	**FP (%)**	Na	2.02	0.1	**17**
		FN (%)	Na	72.9	74.7	**7.6**
		RCS (sample)	Na	Na	Na	Na
		CT (s)	Na	2832.1	0.13	**56.3**
	RSP	**FP (%)**	Na	11.3	0.2	**19**
		FN (%)	Na	24.3	76.6	**5.2**
		RCS (sample)	Na	Na	Na	Na
		CT (s)	Na	2730.4	0.13	**49.2**
D6		**FP (%)**	Na	Na	Na	**0**
		FN (%)	Na	Na	Na	**13.6**
		RCS (sample)	Na	Na	Na	**1129.9**
		CT (s)	Na	Na	Na	**190.5**

- recognition capability speed index (RCS): it measures the detection promptness by considering the time delay in detecting the change;
- computational time index (CT): it provides the execution time needed to perform the test (reference platform: Intel Centrino 1.7 GHz, 1Gb RAM, Windows XP, all unnecessary processes aborted).

Results are given in Table I, where Na denotes a "not applicable" situation, in the sense that the test cannot be run either for lack of *a priori* information (CUSUM) or for the presence of multidimensional signals (Mann–Kendall).

The extended CUSUM and the CI-CUSUM tests provide performance in line with that given by the traditional CUSUM in application D1 but without requiring any *a priori* information. Mann–Kendall test provides comparable results but it is computationally more expensive than tests based on CUSUM.

In application D2, the CI-test performs slightly better than the Mann–Kendall as far as FP and FN are concerned. It is less prompt in detecting changes but its computational complexity is significantly lower. Conversely, the extended CUSUM is characterized by a low complexity. It is very fast in detecting abrupt changes but it is also less accurate. In drifts (which are generally more difficult to identify than abrupt changes), the CI-CUSUM test is very effective with low FP and FN and a reasonable computational load.

TABLE II
CHANGE DETECTION ERROR (PERCENT) FOR DATA SET D5

	EOG	RSP
Mann-Kendall	74.9	35.6
Extended CUSUM	74.9	76.8
CI-CUSUM	24.9	24.6
Literature	Na	30.3

TABLE III
CHANGE DETECTION PERFORMANCE FOR DATA SET D6

	CI-CUSUM	Literature
FP (%)	0	12.1
FN (%)	13.6	16.1
Detection delay (min)	677.5	143.4
Execution time (s)	190.5	Na

TABLE IV
SIMULATION RESULTS FOR THE DIFFERENT CONFIGURATIONS OF CI-CUSUM

			CI-CUSUM -Conf. #1-	CI-CUSUM -Conf. #2-	CI-CUSUM -Conf. #3-	CI-CUSUM -Conf. #4-
D1	**Abrupt**	**FP (%)**	1.33	**0**	0.66	0.66
		FN (%)	0	**0**	0	0
		RCS (sample)	230.4	**385.9**	285.1	152.9
		CT (s)	6.8	7	7.0	7.1
D2	**Abrupt**	**FP (%)**	0	**0**	0	0
		FN (%)	1.9	**1.9**	0.66	0.66
		RCS (sample)	323.9	**536.7**	283.4	578.2
		CT (s)	4.8	**4.8**	5.4	5.8
	Drift	**FP (%)**	0	**0**	0	0
		FN (%)	2.6	**2.6**	0.66	4.66
		RCS (sample)	228.4	**512.7**	402.7	175.5
		CT (s)	7.1	**7.79**	9.6	15.6
D3	**Abrupt**	**FP (%)**	1.33	**1.33**	9.8	2.6
		FN (%)	7.33	**6.71**	3.2	5.3
		RCS (sample)	341.8	**380.5**	200	481.7
		CT (s)	273.5	**276.4**	315.9	349.7
	Drift	**FP (%)**	1.7	**1.3**	14.7	1.9
		FN (%)	8.9	**7.3**	2.2	9.5
		RCS (sample)	1785.6	**1780.2**	1196.1	1562.4
		CT (s)	270.5	**276.5**	285.9	295.4
D4		**FP (%)**	0	**0**	0	0
		FN (%)	4.1	**4.2**	2	4.1
		RCS (sample)	852.1	**952.1**	750.3	832.5
		CT (s)	15.1	**12.1**	14.3	16.1
D5	**EOG**	**FP (%)**	17.6	**17**	17.4	17.7
		FN (%)	7.2	**7.6**	7.4	7.1
		RCS (sample)	Na	**Na**	Na	Na
		CT (s)	51.7	**52.3**	53.1	55.5
	RSP	**FP (%)**	19.8	**19**	19.8	19.9
		FN (%)	4.8	**5.2**	3.9	4.7
		RCS (sample)	Na	**Na**	Na	Na
		CT (s)	48.0	**49.2**	52	52.3
D6		**FP (%)**	0	**0**	33.1	33.1
		FN (%)	33.1	**13.6**	0	0
		RCS (sample)	1451.1	**1129.9**	1108.7	1119.1
		CT (s)	180.1	**190.5**	195.2	201.4

Application D3 shows that the CI-CUSUM test is particularly effective in detectingdrift changes with contained FP and FN ratios, yet it performs well in correspondence with abrupt changes. It is expected that detection performance should improve by enlarging the temporal horizon onto which tests are configured. The comparison between the CI-CUSUM test and the others cannot be done here because CI-CUSUM is the only test that can be applied to multidimensional signals.

For data set D4, the CI-CUSUM test provides both the highest detection accuracy and the quickest recognition ability. The Mann–Kendall test achieves satisfactory results but its computational complexity is very high. The extended CUSUM guarantees the lowest computational complexity but is unsatisfactory as far as the detection accuracy is concerned.

For data set D5, the results provided by CI-CUSUM are better than the ones proposed in [43] in case of change detection based on the respiratory data (while Mann–Kendall results are in line with them). When considering EOG signals, no comparison can be done because the results are not provided in [43]. A detailed comparison is presented in Table II: when applicable, the CI-CUSUM improves over literature results.

Results provided by CI-CUSUM in case of data set D6 (and summarized in Table III) are particularly interesting. The CI-CUSUM test improves accuracy of results present in the literature [1] at the expenses of a delay in detecting the fault without requiring supervised data for training (differently from methods suggested in the literature which require supervised data).

On benchmarks D1–D6, we tested the four alternative hypothesis configurations suggested for the CI-CUSUM test. As presented in Table IV, configuration #2 (the one used in Table I) guarantees the best balancing between FP and FN. Moreover, its computational complexity is acceptable and in line with the one provided by the best case (configuration #1).

As suggested in Section III, configuration #1 guarantees the lowest false positive rate but suffers from the presence of high false negative rate. Configuration #3 provides the best performance in terms of RCS and FN but does not guarantee false positive rates comparable with the ones of configurations #1 and #2. Moreover, the computational time of configuration #3 is higher than the computational time of configurations #1 and #2. Configuration #4 provides low false negative rates but it has the highest computational time.

As a final note, we suggest to consider configuration #2 and CI-CUSUM as the most suitable change detection test for identifying a possible evolution of a process and design consequent adaptive classification systems.

V. Conclusion

This paper presents a novel approach to the change detection problem, aspect relevant in the neural network community when designing classification systems able to track changing environments. In contrast to the literature, where simple features are considered for detecting changes, we suggest the use of a set of features for detecting trends and drifts. A novel test is suggested, the CI-CUSUM test, which somehow inherits the detection ability of the extended CUSUM test procedure and a computational intelligence philosophy, here also suggested to generate a pdf-free extension of the traditional CUSUM. The CI-CUSUM test performs well in detecting abrupt changes in nonstationarity (e.g., due to faults), yet it is particularly sensitive and effective in detecting small and smooth drifts, situations where traditional change detection tests show their weakness.

References

[1] L. H. Chiang, E. L. Russel, and R. D. Braatz, *Fault Detection and Diagnosis in Industrial Systems*. New York: Springer-Verlag, 2001.

[2] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[3] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

[4] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, "Review of process fault diagnosis—Parts I, II, III," *Comput. Chem. Eng.*, vol. 27, no. 3, pp. 293–346, 2003.

[5] P. Frank, "Fault diagnosis in dynamic systems using analytical knowledge-based redundancy—A survey and some new results," *Automatica*, vol. 26, no. 3, pp. 459–474, 1990.

[6] M. A. Kramer and F. E. Finch, "Developemnent and classification of expert systems for chemical process fault diagnosis," *Robot. Comput.-Integr. Manuf.*, vol. 4, pp. 437–446, 1988.

[7] M. Y. Chow, R. N. Sharpe, and J. C. Hung, "On the application and design of artificial neural networks for motor fault detection. Parts I, II," *IEEE Trans. Ind. Electron.*, vol. 40, no. 2, pp. 181–196, Apr. 1993.

[8] C. Alippi and M. Roveri, "Just-in-time adaptive classifiers," Politecnico of Milan, Dipartimento di Elettronica e Informazione, Milan, Italy, Tech. Rep., 2007.

[9] A. Kuh, "Performance of analog neural networks subject to drifting targets and noise," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jul. 10–16, 1999, vol. 4, pp. 2406–2409.

[10] A. Kuh, "Comparison of tracking algorithms for single layer threshold networks in the presence of random drift," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 640–649, Mar. 1997.

[11] G. A. Carpenter, S. Grossberg, and J. H. Reynolds, "A fuzzy artmap nonparametric probability estimator for nonstationary pattern recognition problems," *IEEE Trans. Neural Netw.*, vol. 6, no. 6, pp. 1330–1336, Nov. 1995.

[12] L. Rutkowski, "Adaptive probabilistic neural networks for pattern classification in time-varying environment," *IEEE Trans. Neural Netw.*, vol. 15, no. 4, pp. 811–827, Jul. 2004.

[13] R. K. Aggarwal, Q. Y. Xuan, A. T. Johns, F. Li, and A. Bennett, "A novel approach to fault diagnosis in multicircuit transmission lines using fuzzy ARTmap neural networks," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 1214–1221, Sep. 1999.

[14] A. Ilin, H. Valpola, and E. Oja, "Nonlinear dynamical factor analysis for state change detection," *IEEE Trans. Neural Netw.*, vol. 15, no. 3, pp. 559–575, May 2004.

[15] M. R. Azimi-Sadjadi, D. Yao, A. A. Jamshidi, and G. J. Dobeck, "Underwater target classification in changing environments using an adaptive feature mapping," *IEEE Trans. Neural Netw.*, vol. 13, no. 5, pp. 1099–1111, Sep. 2002.

[16] J. Wang, M. R. Azimi-Sadjadi, and D. Reinke, "A temporally adaptive classifier for multispectral imagery," *IEEE Trans. Neural Netw.*, vol. 15, no. 1, pp. 159–165, Jan. 2004.

[17] S. C. Tan and C. P. Lim, "Condition monitoring and fault prediction via an adaptive neural network," in *Proc. TENCON*, Sep. 2000, vol. 1, pp. 13–17.

[18] P. Robertson and J. M. Brady, "Adaptive image analysis for aerial surveillance," *IEEE Intell. Syst. Appl.*, vol. 14, no. 3, pp. 30–36, May–Jun. 1999.

[19] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York: Academic, 1972.

[20] R. L. Iman and W. J. Conover, *A Modern Approach to Statistics*. New York: Wiley, 1983.

[21] I. W. Burr, *Statistical Quality Control Methods*. New York: Dekker, 1976.

[22] G. A. F. Seber, *Linear Regression Analysis*. New York: Wiley, 1976.

[23] R. Gnanadesikan, *Methods for Statistical Data Analysis of Multivariate Observations*. New York: Wiley, 1977.

[24] R. J. Herrnstein, D. H. Loveland, and C. Cable, "Natural concepts in pigeons," *J. Exp. Psychol.: Animal Behavior Process.*, vol. 2, pp. 285–302, 1976.

[25] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics*, vol. 1, pp. 80–83, 1945.

[26] M. Kendall, *Rank Correlation Methods*, 4th ed. London, U.K.: Griffin, 1975.

[27] B. F. J. Manly and D. I. MacKenzie, "A cumulative sum type of method for environmental monitoring," *Environmetrics*, vol. 11, pp. 151–166, 2000.

[28] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, pp. 100–115, 1954.

[29] G. Kitigawa and W. Gersch, "A smoothness prior time-varying AR coefficient modeling of nonstationary covariance time series," *IEEE Trans. Autom. Control*, vol. AC-30, no. 1, pp. 48–56, Jan. 1985.

[30] I. V. Nikiforov and I. N. Tikhonov, "Application of change detection theory to seismic signal processing," in *Detection of Abrupt Changes in Signals and Dynamical Systems*, ser. Lecture Notes in Control and Information Sciences 77. New York: Springer-Verlag, pp. 355–373.

[31] K. Kumamaru, S. Sagara, and T. Sodersrom, "Some statistical methods for fault diagnosis for dynamical systems," in *Fault Diagnosis in Dynamic Systems—Theory and Application*, ser. International Series in Systems and Control Engineering, R. Patton, P. Frank, and R. Clark, Eds. London, U.K.: Prentice-Hall, 1989, pp. 439–476.

[32] M. Basseville, A. Benveniste, G. Moustakides, and A. Rougee, "Detection and diagnosis of changes in the eigenstructure of nonstationary multivariable systems," *Automatica*, vol. 23, no. 3, pp. 479–489, 1987.

[33] J. Kohlmorgen, "Tracking and visualization of changes in high-dimensional non-parametric distributions," in *Proc. IEEE XIV Workshop Mach. Learn. Signal Process.*, 2004, pp. 203–212.

[34] J. Kohlmorgen, "On optimal segmentation of sequential data," in *Proc. IEEE Workshop Neural Netw. Signal Process.*, 2003, pp. 449–458.

[35] R. R. Fritzemeier, H. T. Nagle, and C. F. Hawkins, "Fundamentals of testability—A tutorial," *IEEE Trans. Ind. Electron.*, vol. 36, no. 2, pp. 117–128, May 1989.

[36] H. T. Nagle, S. C. Roy, C. F. Hawkins, M. G. McNamer, and R. R. Fritzemeier, "Design for testability and built-in self test: A review," *IEEE Trans. Ind. Electron.*, vol. 36, no. 2, pp. 129–140, May 1989.

[37] C. Alippi, M. Catelani, A. Fort, and M. Mugnaini, "Automated selection of test frequencies for fault diagnosis in analog electronic circuits," *IEEE Trans. Instrum. Meas.*, vol. 54, no. 3, pp. 1033–1044, Jun. 2005.

[38] F. Desobry, M. Davy, and C. Doncarli, "An online kernel change detection algorithm," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2961–2974, Aug. 2005.

[39] C. Blake and C. Merz, "UCI Machine Learning Databases Repository," Dept. Inf. Comput. Sci., Univ. California—Irvine, Irvine, CA [Online]. Available: http://www.ics.uci.edu/~mlearn/databases/statlog/

[40] D. Michie, D. J. Spiegelhalter, and C. C. Taylor, *Machine Learning, Neural and Statistical Classification*. Upper Saddle River, NJ: Ellis Horwood, 1994.

[41] M. Bollano, R. Piagge, A. Charai, and D. Narducci, "Experimental evidence and computational analysis of the electronic density modulation induced by gaseous molecules at Si(001) surfaces upon self-assembling organism monolayer," *Appl. Surface Sci.*, vol. 175–176, pp. 379–385, 2001.

[42] D. Narducci, P. Bernardiello, and M. Oldani, "Investigation of gas-surface interactions at self-assembled silicon surfaces acting as gas sensor," *Appl. Surface Sci.*, vol. 212–213, pp. 491–496, 2003.

[43] J. Kohlmorgen, K.-R. Müller, J. Rittweger, and K. Pawelzik, "Identification of nonstationary dynamics in physiological recordings," *Biol. Cybern.*, vol. 83, no. 1, pp. 73–84, 2000.

[44] J. J. Downs and E. F. Vogel, "A plant-wide industrial process control problem," *Comput. Chem. Eng.*, vol. 17, no. 3, pp. 245–255, 1993.

[45] Tennessee Eastman Dataset [Online]. Available: http://brahms.scs.uiuc.edu

Cesare Alippi (SM'94–F'06) received the Dr.Ing. degree in electronic engineering (*summa cum laude*) and the Ph.D. degree in computer engineering, both from Politecnico di Milano, Milano, Italy, in 1990 and 1995, respectively.

He has completed research work in computer sciences at the University College London, London, U.K., and the Massachusetts Institute of Technology, Cambridge. Currently, he is a Full Professor of Information Processing Systems at the Politecnico di Milano. His research interests include application-level analysis and synthesis methodologies for embedded systems, computational intelligence, and wireless sensor networks. His research results have been published in more that 120 technical papers in international journals and conference proceedings.

Dr. Alippi is the Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS and the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT.

Manuel Roveri received the Dr.Eng. degree in computer science engineering from the Politecnico di Milano, Milano, Italy, in June 2003, the M.S. degree in computer science from the University of Illinois at Chicago, Chicago, in December 2003, and the Ph.D. degree in computer engineering from Politecnico di Milano, in May 2007.

His research interests include machine learning, computational intelligence, adaptive algorithms, and wireless sensor networks.