Change Detection Tests Using the ICI rule

Cesare Alippi, Giacomo Boracchi and Manuel Roveri

***Abstract* - Designing tests able to effectively detect changes in the stationarity of a process generating data is a challenging problem, in particular when the process is unknown, and the only information available has to be extracted from a set of observations. This work proposes a novel approach for detecting changes in a process generating data whose distribution is unknown. Peculiarity of the approach is the use of the Intersection of Confidence Intervals (ICI) rule to monitor the process evolution. A change detection test derived from this approach is also presented. Experimental results show that the proposed test outperforms state-of-the art solutions, both in terms of efficiency and effectiveness, in particular when a reduced test configuration set is available.**

I. INTRODUCTION

RELIABLE systems designed for processing data collected from real-world phenomena (e.g. for classification, sensing and monitoring purposes), have to handle the occurrence of unpredictable events causing changes in the data generating process. Change detection, i.e., verifying whether drifts, abrupt changes and smooth deviations have affected the data generating mechanism, is of relevant interest for two main reasons. First, changes could be induced by faults, malfunctioning and ageing effects in the hardware involved and need to be discovered as soon as possible. Second, any reliable adaptive system (such as a classifier) has to promptly adjust to the new operating conditions, in order to preserve performance [1], [2].

There is a wide literature concerning change detection tests [3] – [16], which are often employed in industrial, environmental, and medical applications [3].

The most common change detection techniques are parametric and non-parametric statistical tests [3], [4]. Parametric techniques [5], [6] require availability (or an estimate) of the probability density function of the process generating the data before and after the change and/or information about the nature of the drift itself. Conversely, non-parametric tests do not require strong a priori information [6], [7] and are thus more flexible in dealing with a large class of applications. The Mann-Whitney U test for independent samples [8] (which relies on the possibility to rank two independent samples of observations) is a non-parametric test originally designed for verifying whether two independent samples come from the same distribution or not. Unfortunately, the test was not designed to work in sequential analysis and focuses on detecting a single point-change. Differently, the Mann-Kendall [9] (designed for environmental sciences) and the CUmulative SUM (CUSUM) [10], [11] (developed in the system control community to detect structural changes) non-parametric 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 [12], [13], changes in mechanical systems [14], [14]. Unfortunately, these tests 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 must be identified by exploiting a priori information or through a trial-and-error approach.

A different approach is presented in [16] where two automatic tests for detecting non stationarity phenomena, which differ for detection accuracy and computational complexity, are suggested. These tests are particularly interesting since they propose a parameter configuration phase that allows for automatically configuring the parameters of the tests (i.e., they do neither require a priori information nor assumptions about the process generating the data).

This work introduces a novel approach in designing non parametric change detection tests that have to operate without any a priori information about the process generating data, as in [16]. Changes are detected by extracting features first and apply then the Intersection of Confidence Interval (ICI) rule [17]. More specifically, the feature extraction step is meant to produce Gaussian distributed features from observations while the ICI rule is used to assess the stationarity of the features, and hence that of the data generating process.

A simple and computationally-light change detection test has been designed and applied in the experimental session. Experiments shows that the test outperforms state of the art solutions working under the same assumptions, guaranteeing high performance in detecting small changes and slow drifts also when short training sequences are used to configure the test.

The rest of the paper is organized as follows: Section II introduces the general framework while the formulation of the change detection methodology is presented in Section III. The design of a change detection test is presented in Section IV, whose experimental results are shown in Section V.

Cesare Alippi, Giacomo Boracchi and Manuel Roveri are with Politecnico di Milano, Dipartimento di Elettronica e Informazione, Milano, Italy e-mail: cesare.alippi@polimi.it, giacomo.boracchi@polimi.it, manuel.roveri@polimi.it

II. PRELIMINARIES

A. Problem Statement

Let $X: \mathbb{N} \to \mathbb{R}^d$ be a stochastic process generating (at least initially) independent and identically distributed (i.i.d.) d dimensional real data (observations or measurements) according to an unknown *pdf*. Let $O_T = \{X(t),\ t = 1, \dots, T\}$ be the sequence of observations measured up to time T, and assume that at least the first $T_0 < T$ observations come from the process X in a stationary state. As such, $X(t)$ is a d-dimensional vector of measurements at time t. Define $TS = O_{T_0}$ the set of initial supervised observations that constitute the training sequence of the change detection test.

The goal of a change detection test is to determine the time instant $t = T^*$ in which the process X changes its statistical properties. Since this work is meant for application scenarios where no *a priori* information is available (i.e., we are not requiring the knowledge of the distribution of the process generating the data), the change detection test solely relies on features extracted from observations. In particular, it separately analyzes each feature and detects a change in X as soon as a change in such a feature is perceived by the ICI rule. In the following, we present at first the ICI rule and the core elements of the test and then, in Section III we give a detailed description of the test.

B. The ICI Rule

The ICI rule [17],[18] has been proposed as a method for developing adaptive estimates for regression of functions from noisy observations. Although the ICI rule has been originally introduced for processing 1D signal, it has been proved to be particularly effective in image processing applications, e.g. [19] – [23].

The ICI rule operates, combined with a polynomial regression technique, on sequences of noisy data $z \in \mathbb{R}$ extracted with a given sampling frequency from a Gaussian distribution of mean $\mu(t)$ and standard deviation σ

$$z(t) \sim N\left(\mu(t), \sigma^2\right) \quad t \in W, \tag{1}$$

where t represents the time instant and W the uniformly spaced sampling grid. For ease of understanding, data are here assumed to be scalar, as the ICI rule can be applied to data of arbitrary dimensions in a component-wise manner.

For each $t \in W$, the ICI rule adaptively identifies an optimal neighborhood $U_{i^+}(t)$ and substitutes (regularizes) $z(t)$ with an estimate $\hat{\mu}(t)$ of $\mu(t)$, obtained by least squared error polynomial fit of order m of the noisy data belonging to the optimal neighborhood $\{z(t),\ t \in U_{i^+}(t)\}$. The neighborhood $U_{i^+}(t)$ is adaptively selected from a set of L nested neighborhoods $\{U_i(t)\ \ i = 1, \dots, L\}$ (obtained by scaling a reference neighborhood $\tilde{U}(t)$ centered at t). Let $\{\hat{\mu}_i(t),\ i = 1, \dots, L\}$ be the sequence of estimates $\hat{\mu}(t)$ evaluated over $U_i(t)$, and $\{\sigma_i, i = 1, \dots, L\}$ be the associated standard deviations. Note that since σ is constant in (1), σ_i depends only on $U_i(t)$, σ and m.

The ICI rule can be stated as follows. Let $\mathcal{I}_i$ be the confidence interval of $\hat{\mu}_i(t)$

$$\mathcal{I}_i = \left[\hat{\mu}_i(t) - \Gamma\sigma_i(t)\,;\, \hat{\mu}_i(t) + \Gamma\sigma_i(t)\right], \tag{2}$$

where $\Gamma > 0$ is a parameter configured at design time. The ICI rule selects the adaptive neighborhood $U_{i^+}(t)$ as the one corresponding to the largest index j for which $\bigcap_{i=1}^{j} \mathcal{I}_i$ is not empty. The ICI selected neighborhood satisfy optimality criteria which are expressed in [17]: basically the estimate corresponding to the neighborhood U_{i^+} approaches the minimum least squared error estimate of $\mu(t)$.

C. Features

The proposed test relies on a set of functions $\{F^k\ k = 1, \dots, N\}$, that transform the observations $A \subset O_T$ into values $F^k\ (A) \in \mathbb{R}^d$, which are called features. When the features are distributed as in (1), we can use U_{i^+}, the ICI-selected interval as a valuable tool for change detection. As such we assume the features distributed as in (1): in Section IV we present two examples of features satisfying this assumption.

III. CHANGE DETECTION USING THE ICI RULE IN THE FEATURES DOMAIN

Stationarity in the process generating the data is monitored by using each feature $\{F^k\ k = 1, \dots, N\}$ separately: a change is detected in X when at least one of the features shows a change. As such, we can focus on a single feature and devise change detection tests using the ICI rule on a single feature output sequence.

Since the features are distributed as in (1), we can use the ICI rule to select adaptive neighborhoods for performing polynomial estimates of their values in feature domain. However, we are not interested in the estimates provided by the ICI rule themselves; rather, in their neighborhoods. More precisely, we cast the ICI rule in a particular scenario where it is used to select the adaptive neighborhood for estimating the expected feature value at $t = 1$ only (i.e. $\mu(1)$).

The ICI rule selects the adaptive neighborhood among intervals having the leftmost extreme at $t = 1$, i.e. $[1, T]$, $T > 0$ (refer to Fig. 1a). Thus, the set of neighborhoods considered for analyzing observations at time T is $\{[1, T_0], \dots, [1, T]\}$. Note that the confidence intervals related to the observations that belong to the training set TS do not need to be computed, since the observations have been generated by X in the initial stationary state.

Finally, the estimates associated to each neighborhood are obtained by the 0^{th} order polynomial fit of feature values: this choice reflects the intuitive idea that stationary processes provide constant feature values. In this framework, any ideal neighborhood selected by the ICI rule contains a set of features that must be considered constant in a stationarity scenario and have been generated from the same Gaussian distribution. Contrarily, feature values belonging to different ICI-selected neighborhoods should be considered as values generated by different Gaussian distributions: Fig. 1b illustrates an example of adaptive scale selection using the ICI rule.

Let $\{F^k(t) \quad t = 1,\ldots,T_0\}$ be the sequence of values assumed by the feature F^k within the training set; then

$$\hat{\mu}^k_{T_0} = \sum_{t=1}^{T_0} \frac{F^k(t)}{T_0} \quad \text{and} \qquad (3)$$
$$\hat{\sigma}^k_{T_0} = \sqrt{\sum_{t=1}^{T_0} \frac{\left(F^k(t) - \hat{\mu}^k_{T_0}\right)^2}{T_0 - 1}},$$

represent an estimate of the expected value of the k^{th} feature and its standard deviation, when X is in the initial, stationary state. Note that (3) defines the first confidence interval $\mathcal{I}^k_{T_0}$ according to (2), and thus the test configuration consists of computing $\mathcal{I}^k_{T_0}, \hat{\mu}^k_{T_0}$, and $\hat{\sigma}^k_{T_0}$.

During the operational phase, for each $T > T_0$, the test computes the 0^{th} order polynomial fit of the features belonging to the interval $[0,T]$. Then, the ICI rule is used to assess whether these values have been generated from the same Gaussian distribution or not. Hence, the ICI rule acts as a nonstationarity test determining, by means of the features, if X is constant within the time interval $[0,T]$, and thus stationary. The estimate coming from the 0^{th} order polynomial fit and its standard deviation can be efficiently computed by means of the following iterative formulas:

$$\hat{\mu}^k_T = \frac{(T-1)\mu^k{}_T + F^k(T)}{T} \quad \text{and} \qquad (4)$$
$$\sigma^k_T = \frac{\sigma^k_{T_0}}{\sqrt{T}}, \quad \forall T > T_0.$$

Finally, the ICI rule reveals a change in stationarity in X when the $\bigcap_{t=T_0}^{T} \mathcal{I}^k_t$, $T > T_0$ becomes an empty set.

IV. A Change Detection Test

In the following we present a change detection test derived from the methodology detailed in Section III.

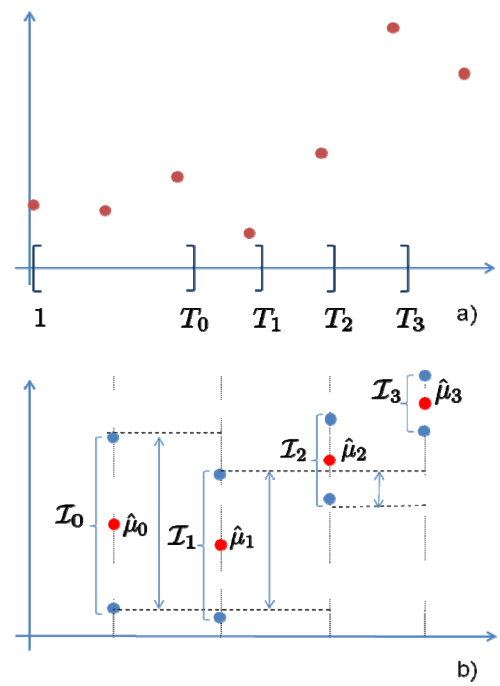

Fig. 1. An illustrative example of the ICI rule in the settings used for change detection: **a)** features and the set of intervals $\{[1,T_0],[1,T_1],[1,T_2],[1,T_3]\}$ **b)** the corresponding polynomial estimates (0^{th} order) and the confidence intervals. The ICI rule selects as ideal interval, $[1,T_2]$ thus $i^+ = 2$. The brackets in **b)** represents the confidence intervals, while the arrows represents their intersection.

A. The Considered Features

As per the extended CUSUM and the CI-CUSUM change detection tests [16] we segment the set of observations $X(t)$ in disjoint subsequences of length ν:

$$Y_s = \{X(t), (\nu-1)s \le t < \nu s\}, \qquad (5)$$

where $s \in \mathbb{N}$ is the subsequence index. The proposed test exploits features that are evaluated only on these subsequences: it follows that changes in X are detected as events occurring within a subsequence, and thus with an accuracy $\pm\nu/2$. To simplify the notation used, we introduce the following abbreviation: $F^k(s) = F^k(Y_s), \quad k = 1,\ldots,n.$ The polynomial fit of 0^{th} order and the ICI rule are then applied to the set of subsequences indexes $\{[1,S_0],\ldots,[1,T/\nu]\}$, being $S_0 = T_0/\nu$. The proposed test relies on two features only, computed on each subsequence.

The first feature is the sample mean M evaluated in each subsequence Y_s:

$$M(s) = \frac{1}{\nu} \sum_{t=(\nu-1)s}^{\nu s} X(t). \qquad (6)$$

Since $X(t)$ s are i.i.d. (when X is stationary), the Central Limit Theorem guarantees that $M(s)$ approaches a Gaussian distribution (and thus these values are distributed as in (1)).

The second feature is a quadratic deviation, evaluated in each subsequence Y_s:

$$S(s)=\sum_{t=(\nu-1)s}^{\nu s}\left(X(t)-M(s)\right)^2. \quad (7)$$

Note that $S(s)/(\nu-1)$ is the sample variance of Y_s. Unfortunately, $S(s)$ is not Gaussian distributed when the samples $X(t)$ are drawn from an arbitrary distribution. Thus, we adopt the power-law transform presented in [24] to obtain a derived feature, which is related to the variance of X and is drawn from a Gaussian distribution. This Gaussian approximation holds when the sample variance is computed from i.i.d. samples; the transform has the form

$$\mathcal{T}\left(S(s)\right)=\left(\frac{S(s)}{\nu-1}\right)^{h_0}, \quad (8)$$

where the exponent h_0 is expressed as a function of the cumulants of X, which in turn can be estimated from the observations (refer to [24] for details about this transform). The cumulants of X and hence h_0, can be computed from the observation in the training set TS. The second feature is defined as $V(s)=\mathcal{T}\left(S(s)\right)$.

B. Algorithm Details

The following scheme describes the algorithmic steps constituting the proposed change detection test. Lines 1-11 represent the test configuration phase, where the parameters $\hat{\mu}^M_{S_0},\hat{\sigma}^M_{S_0},\hat{\mu}^V_{S_0}$ and $\hat{\sigma}^V_{S_0}$ of feature distributions are estimated. The operational phase of the test is represented by the loop at lines 14-21: the test iterates every time a subsequence of ν observations are collected.

ALGORITHM I :
ICI CHANGE DETECTION TEST

1. Compute $\left\{M(s), s=1,\ldots,S_0\right\}$,
2. $\hat{\mu}^M_{S_0}=\sum_{s=1}^{S_0}\frac{M(s)}{S_0}$.
3. $\hat{\sigma}^M_{S_0}=\sqrt{\sum_{s=1}^{S_0}\frac{\left(M(s)-\hat{\mu}^M_{S_0}\right)^2}{S_0-1}}$.
4. Define $\mathcal{I}^M_{S_0}=\left[\hat{\mu}^M_{S_0}-\hat{\sigma}^M_{S_0}\ ;\ \hat{\mu}^M_{S_0}+\hat{\sigma}^M_{S_0}\right]$.
5. Compute the first six cumulants of X from TS.
6. Compute h_0 as described in [24], and define $\mathcal{T}$.
7. Compute $\left\{S(s), s=1,\ldots,S_0\right\}$.
8. Compute $\left\{V(s)=\mathcal{T}\left(S(s)\right), s=1,\ldots,S_0\right\}$
9. $\hat{\mu}^V_{S_0}=\sum_{s=1}^{S_0}\frac{V(s)}{S_0}$.
10. $\hat{\sigma}^V_{S_0}=\sqrt{\sum_{s=1}^{S_0}\frac{\left(V(s)-\hat{\mu}^V_{S_0}\right)^2}{S_0-1}}$.
11. Define $\mathcal{I}^V_{S_0}=\left[\hat{\mu}^V_{S_0}-\hat{\sigma}^V_{S_0}\ ;\ \hat{\mu}^V_{S_0}+\hat{\sigma}^V_{S_0}\right]$.
12. Set $s=S_0$
13. **while** $\left(\mathcal{I}^M_s\neq\emptyset\ \&\&\ \mathcal{I}^V_s\neq\emptyset\right)$ {
14.   Set $s=S_0+1$
15.   Wait for ν observations, until $Y(s)$ is populated
16.   Compute $M(s)$ and $V(s)$.
17. $\hat{\mu}^M_s=\frac{(s-1)\cdot\hat{\mu}^M_{s-1}+M(s)}{s}$
18. $\hat{\sigma}^M_s=\frac{\hat{\sigma}^M_{S_0}}{\sqrt{s}}$
19. $\hat{\mu}^V_s=\frac{(s-1)\cdot\hat{\mu}^V_{s-1}+V(s)}{s}$
20. $\hat{\sigma}^V_s=\frac{\hat{\sigma}^V_{S_0}}{\sqrt{s}}$
21. $\mathcal{I}^M_s=\left[\hat{\mu}^M_s-\hat{\sigma}^M_s\ ;\ \hat{\mu}^M_s+\hat{\sigma}^M_s\right]\cap\mathcal{I}^M_{s-1}$
22. $\mathcal{I}^V_s=\left[\hat{\mu}^V_s-\hat{\sigma}^V_s\ ;\ \hat{\mu}^V_s+\hat{\sigma}^V_s\right]\cap\mathcal{I}^V_{s-1}$
23. }
24. Detect a change in $\left[(s-1)\nu\ ,\ s\nu\right]$

V. EXPERIMENTS

To validate the effectiveness of the suggested non-stationary detection test we considered three applications, two of which (D1, D3) are simulated, and one (D2) with observations collected from real experiments. The results of tests presented in Section IV.B (in what follows ICI test) are compared and contrasted with the ones provided by the CUSUM, the traditional Mann-Kendall test, and by the CI-CUSUM test [16]. Four indexes 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 there it is not.
- False negative index (FN): it counts the times a test does not detect a change when there it is.
- Recognition capability speed index (RCS): it measures the detection promptness by considering the time delay in detecting a change.
- Computational time index (CT): it provides the execution time needed to perform the test (reference platform: Intel Xeon CPU 2.33 GHz)

Application D1 - refers to a simple mono-dimensional process ruled by a known Gaussian pdf. The process lasts 6000 samples. We considered two kinds of perturbations injected after 4000 samples: a perturbation affecting the mean value (which changes from $\mu_0=100$ to $\mu_1=105$, $\sigma=3$) and a perturbation affecting the standard deviation (from $\sigma_0=3$ to $\sigma_1=5$,). In addition, we modeled both an abrupt (i.e., the parameter changes suddenly at t=4000) and a smooth changes (the parameters changes slowly from t=4000 to achieve the new value at the end of the experiment) for the perturbations on mean and the variance. Thus, for application D1, we considered 4 kinds of changes: abrupt change on the mean value, drift change on the mean value, abrupt change on the standard deviation, drift change on the standard deviation. The CUSUM test was configured using the pdfs of X (for abrupt changes only).

Application D2 - coincides with the SATIMAGE benchmark [25], i.e., classification of Landsat Multi-Spectral Scanner images in seven classes (6435 samples of 36 features). A Feed-Forward Neural Network classifier [26] was considered as suggested in [27] (single hidden layer of 6 neurons and 1 output neuron). The change affects the neural network weights at sample 4000. For each experiment, we randomly partition the dataset into training set and validation set.

Application D3 - refers to Self-Assembled-Monolayers (SAM) [28][29] gas sensors. The model of the sensor resistance (to be considered unknown to all change detection tests) is, in stationary conditions,

$$R = R^0 \left[1 + K \frac{\sum_{i=1}^{5} \alpha_i p_i}{1 + \sum_{i=1}^{5} \beta_i p_i} \right] \quad (9)$$

where $\alpha_i, \beta_i \in \mathbb{R}$, p_i is the partial pressure of the i^{th} 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 10-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; 12000 samples coming from 150 different simulated acquisitions (pseudo random binary signals, i.e., steps with random amplitudes and time duration) were considered; changes started at sample 6000.

Besides application D1, where the perturbation is fixed, we considered a multiplicative perturbation model both for application D2 and D3. In case of abrupt changes 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 dataset. The perturbation δ has been uniformly extracted in each experiment from the interval $[-2, 2]$. The level of significance for Mann-Kendall [9] was fixed at 99.9%; $n = 20$ and $\gamma = \sqrt{T_0 / \nu}$ for the CI-CUSUM [10][11] tests. The ICI test has been configured with $\Gamma = 2$. The experiments for the ICI test and the CI-CUSUM [16] test have been configured using 2000 training samples, i.e., $T_0 = 2000$ (6th and 8th column in Table I).

Results given in Table I have been averaged over 150 runs. The symbol *Na* denotes a "not applicable" situation, in the sense that the test cannot be run either for lack of a priori information (for the CUSUM when the change is a drift) or for the presence of multidimensional signals (Mann-Kendall).

As far as application D1-Abrupt Mean is concerned, ICI and CI-CUSUM test provide the same detection accuracy of the traditional CUSUM test in application D1 but without requiring a-priori information about the pdfs before and after the change. In particular, ICI is faster in detecting changes and less computationally demanding than CI-CUSUM (i.e. ICI has lower RCSs and CTs). Mann-Kendall provides comparable results but it is computationally very expensive. Similarly, in case of the D1-Drift Mean, the ICI test provides results comparable with CI-CUSUM but guarantees lower RCSs and CTs.

Simulation results on application D1-Abrupt Variance and D1-Drift Variance are in line with those of changes on the process mean. The ICI test provides the same results of the CUSUM test but does not require any a-priori information about the nature of the changes. Moreover, the ICI test performs better than the CI-CUSUM test and it is prompter in detecting changes. Even in this case, ICI guarantees CTs lower than CI-CUSUM. Mann-Kendall is not able to detect changes in the variance since it has been designed for detecting trends affecting the mean.

Also in application D2, the ICI test outperforms CI-CUSUM and Mann-Kendall in terms of FPs and FNs and guarantees a prompter detection of changes (lowest RCSs) both in the abrupt and in the drift case. Moreover, the computational times (CT) of ICI test are significantly lower. Application D3 shows that ICI test provides higher detection accuracies than CI-CUSUM (lower FPs and FNs) both in the abrupt and in the drift case. Moreover, ICI test slightly outperforms CI-CUSUM as far as the recognition speed is concerned and heavily reduces the computational complexity.

Summarizing, the ICI test is particularly effective in detecting changes, guaranteeing reduced FPs and prompt recognitions. Moreover, its computational complexity is significantly lower than that of other tests. In all these simulations we set the ICI parameter $\Gamma = 2$; higher values of Γ would reduce the probability of FPs but at the expenses of larger RCSs (and possibly more FNs).

Following the spirit of [16], where the joint use of non-stationary detection tests and adaptive knowledge mechanisms allows the adaptive classifiers for reacting to changes in non-stationary conditions by tracking the evolution of the data generating process, we designed an additional set of experiments with a shorter training sequence. In this manner we simulate most of real working conditions in which the adaptive classifier has to quickly adapt to the unknown process evolution (i.e., little supervised information is available after the change).

Simulation results for the considered datasets with the ICI and CI-CUSUM test configured using $T_0 = 500$ samples are presented in the 7th and 9th column of Table I. These

TABLE I
SIMULATION RESULTS FOR THE CONSIDERED DATASETS

			CUSUM	Mann-Kendall	CI-CUSUM		ICI test	
					T_0=2000	T_0=500	T_0=2000	T_0=500
D1	Abrupt Mean	FP (%)	0	7.3	0	7.7	**0**	**5.5**
		FN (%)	0	0	0	0	**0**	**0**
		RCS (sample)	11.4	94.9	386.1	345.0	**149.5**	**140.5**
		CT (s)	0.5	1044.0	6.9	6.9	**0.12**	**0.1**
	Drift Mean	FP (%)	Na	8	0	8.0	**0**	**5.9**
		FN (%)	Na	0	0.3	0	**0**	**0**
		RCS (sample)	Na	590.0	1110.5	832.9	**793.2**	**764.2**
		CT (s)	Na	1046.9	7.1	4.5	**0.1**	**0.2**
	Abrupt Variance	FP (%)	0	10	0	8.0	**0**	**5.9**
		FN (%)	0	90	2.0	0	**0**	**0**
		RCS (sample)	39.5	Na	642.2	437.9	**300.3**	**280.9**
		CT (s)	0.5	1037.5	9.2	6.61	**0.1**	**0.1**
	Drift Variance	FP (%)	Na	10	0	9.4	**0**	**5.8**
		FN (%)	Na	90	0	0	**0**	**0**
		RCS (sample)	Na	Na	1029.1	765.8	**630.8**	**597.6**
		CT (s)	Na	1050.3	8.8	7.4	**0.13**	**0.2**
D2	Abrupt	FP (%)	Na	6.0	0	26.6	**0.0**	**7.8**
		FN (%)	Na	41.8	12.0	7.3	**6.0**	**2.6**
		RCS (sample)	Na	1003.0	574.7	487.1	**196.1**	**229.1**
		CT (s)	Na	51.7	2.3	2.0	**0.05**	**0.07**
	Drift	FP (%)	Na	6.1	0	25.3	**0**	**7.3**
		FN (%)	Na	58.5	22.6	9.3	**10.7**	**8.6**
		RCS (sample)	Na	1718.1	1304.1	996.5	**831**	**811.4**
		CT (s)	Na	50.6	2.3	1.5	**0.05**	**0.05**
D3	Abrupt	FP (%)	Na	Na	0.3	60.6	**1.3**	**8.6**
		FN (%)	Na	Na	5.3	2.6	**1.3**	**2.0**
		RCS (sample)	Na	Na	384.5	438.1	**361.7**	**295.4**
		CT (s)	Na	Na	81.4	55.1	**0.1**	**0.5**
	Drift	FP (%)	Na	Na	0.7	61.6	**0.6**	**14**
		FN (%)	Na	Na	12	1.3	**4.6**	**4**
		RCS (sample)	Na	Na	1911.3	1924.1	**1890.1**	**1843.1**
		CT (s)	Na	Na	49.2	73.4	**0.1**	**0.1**

simulation results are particularly interesting and show that the ICI test, although it slightly decreases the detection accuracy, generally maintains effective detection ability when shorter training sequences are available. On the contrary, the detection accuracy of the CI-CUSUM test severely decreases as the training sequence get shorter (this is evident by considering the high number of FPs).

RCSs and CTs of ICI and CI-CUSUM tests are in line with those of the $T_0 = 2000$ case.

VI. CONCLUSIONS

The paper presents a methodology to design a novel class of change detection tests. The proposed approach exploits the well theoretically-founded Intersection of Confidence Interval (ICI) rule applied to features extracted from observations, to promptly detect non-stationary changes in the data generating process.

A particular test derived from this methodology has been also presented: simulation results on synthetic and real datasets show that it provides high detection accuracies, promptness in detecting abrupt changes as well as small drift, and a reduced computational complexity. The proposed test preserves the detection abilities even in case of reduced training sets, and thus it represents a particularly appealing candidate for being embedded in adaptive systems working in non-stationary environments.

REFERENCES

[1] C. Alippi and M. Roveri, "Just-in-Time Adaptive Classifiers--Part II: Designing the classifier," *Neural Networks, IEEE Transactions on*, vol. 19, no. 11, pp. 2053-2064, December 2008

[2] C. Alippi, G. Boracchi and M. Roveri, "Just in time classifiers: managing the slow drift case". In *Proceedings of the 2009 international Joint Conference on Neural Networks* (Atlanta, Georgia, USA, June 14 - 19, 2009), 1537-1543, 2009.

[3] L.H. Chiang, E.L. Russel and R.D. Braatz, *Fault Detection and Diagnosis in Industrial Systems*. Springer, 2001.

[4] M. Basseville, and I.V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Prentice-Hall, Inc. 1993.

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

[6] I. W. Burr, *Statistical quality control methods*, New York: Dekker. - XII, 1976.

[7] R. Gnanadesikan, *Methods for statistical data analysis of multivariate observations*, John Wiley and Sons: New York, 1977

[8] R.J. Herrnstein, D.H. Loveland and C. Cable, "Natural concepts in pigeons", *Journal of Experimental Psychology: Animal Behavior Processes*, vol. 2, pp. 285-302, 1976.

[9] M. Kendall, *Rank correlation methods*, Ch. Griffin & Co., London, 4ht edn, 1975.

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

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

[12] G. Kitigawa and W. Gersch. "A smoothness prior time-varying AR coefficient modeling of nonstationary covariance time series", *Automatic Control, IEEE Transactions on*, vol. 30, no.1, pp.48-56, 1985.

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

[14] K. Kumamaru, S. Sagara and T. Sodersrom."Some statistical methods for fault diagnosis for dynamical systems", *Fault Diagnosis in Dynamic Systems - Theory and Application (R. Patton, P. Frank, R. Clark, eds.). International Series in Systems and Control Engineering*, Prentice Hall International, London, UK, pp.439-476, 1989.

[15] 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.

[16] C. Alippi and M. Roveri, "Just-in-Time Adaptive Classifiers--Part I: Detecting Nonstationary Changes," *Neural Networks, IEEE Transactions on*, vol. 19, no. 7, pp. 1145-1153, July 2008.

[17] A. Goldenshluger and A. Nemirovski, "On spatial adaptive estimation of nonparametric regression", *Mathematical Methods of Statistics*, vol 6 (1997), 135-170.

[18] V. Katkovnik, "A new method for varying adaptive bandwidth selection", *Signal Processing, IEEE Transactions on,* vol. 47, no. 9, pp. 2567-2571, 1999.

[19] V. Katkovnik, K. Egiazarian and J. Astola, "A Spatially Adaptive Nonparametric Regression Image Deblurring", *Image Processing, IEEE Transactions on,* vol. 14, no. 10, pp. 1469-1478, Oct. 2005.

[20] A. Foi, V. Katkovnik and K. Egiazarian, "Pointwise Shape-Adaptive DCT for High-Quality Denoising and Deblocking of Grayscale and Color Images", *Image Processing, IEEE Transactions on*, vol. 16, no. 5, pp. 1395-1411, May 2007.

[21] V. Katkovnik, K. Egiazarian and J. Astola, "Adaptive window size image de-noising based on intersection of confidence intervals (ICI) rule", *J. of Math. Imaging and Vision*, vol. 16, no. 3, pp. 223-235, 2002.

[22] A. Foi, V. Katkovnik, K. Egiazarian, and J. Astola, "Inverse halftoning based on the anisotropic LPA-ICI deconvolution", *Proc. Int. TICSP Workshop Spectral Meth. Multirate Signal Process., SMMSP 2004*, Vienna, pp. 49-56, 2004.

[23] V. Katkovnik, K. Egiazarian, and J. Astola, "Local Approximation Techniques in Signal and Image Processing", *SPIE Press, Monograph* Vol. PM157, September 2006.

[24] G. S. Mudholkar and M. C. Trivedi "A Gaussian Approximation to the Distribution of the Sample Variance for Nonnormal Populations", *Journal of the American Statistical Association,* Vol. 76, No. 374 (Jun., 1981), pp. 479-485

[25] C. Blake and C. Merz. Uci machine learning databases repository, university of california-irvine,department of information and computer science. [Online]. Available: *http://www.ics.uci.edu/~mlearn/databases/statlog/*

[26] Matworks MATLAB Neural Network Toolbox V6.0 (R2008a), http://www.mathworks.com/

[27] D. Michie, D. J. Spiegelhalter and C. C. Taylor. *Machine Learning, Neural and Statistical Classification*, Ellis Horwood, 1994

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

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