A Hierarchical, Nonparametric, Sequential Change-Detection Test

Cesare Alippi, Giacomo Boracchi and Manuel Roveri

Abstract – **Design of applications working in nonstationary environments requires the ability to detect and anticipate possible behavioral changes affecting the system under investigation. In this direction, the literature provides several tests aiming at assessing the stationarity of a data generating process; of particular interest are nonparametric sequential change-point detection tests that do not require any a-priori information regarding both process and change. Moreover, such tests can be made automatic through an on-line inspection of sequences of data, hence making them particularly interesting to address real applications. Following this approach, we suggest a novel two-level hierarchical change-detection test designed to detect possible occurrences of changes by observing incoming measurements. This hierarchical solution significantly reduces the number of false positives at the expenses of a negligible increase of false negatives and detection delays. Experiments show the effectiveness of the proposed approach both on synthetic dataset and measurements from real applications.**

I. Introduction

Despite the fact that change detection tests, i.e., methods used to assess the stationary hypothesis of a data generating mechanism, have been studied for several decades, research has never been interrupted because of their relevance in real-world applications (e.g., fault detection, quality analysis of products, clinical trials,…). Change-detection tests (CDT) can be grouped into three main families [1]: statistical hypothesis tests, sequential hypothesis tests, and change-point detection tests.

Statistical hypothesis tests work by accepting/rejecting the stationarity hypothesis of a process by analyzing a finite set of samples and building up an estimator for confuting/accepting such hypothesis. In particular, these tests either require to verify the validity of an assumption abut the pdf (e.g., the Z-test and t-test [2]) or partition the available samples into two disjoint subsets for evaluating statistical discrepancies between them (e.g., the Mann-Whitney U test [3] or the Kolmogorov-Smirnov test [4]). Unfortunately, as stated in [5], the partitioning strategy used on the available samples affects the change-detection performance; at the same time, partitioning becomes a critical issue in on-line classification systems. In fact, the need of a continuous inspection of data streams makes difficult to propose a good partitioning even when considering sliding windows, whose size definition would require a-priori knowledge of the change dynamics.

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. This research has been funded by the European Commission's 7th Framework Program, under grant Agreement INSFO-ICT-270428 (iSense).

Differently from statistical hypothesis tests, where the number of considered samples is fixed, in sequential hypothesis tests the number of samples depends on available data, thus it is a random variable. These tests (e.g., the Sequential Probability Ratio Test [6], [7] and the Repeated Significance Test [8], [9]) sequentially analyze acquired samples one by one, until the decision to accept or refuse the "no-change" hypothesis can be taken with a given level of confidence. Sequential hypothesis tests are theoretically well grounded but suffer from the need to make a decision about the null hypothesis when a given confidence level is reached. This is not a drawback in general, as often it is required to take a decision about a null hypothesis (e.g., in medical applications one has to decide if a certain drug is or is not effective). However, this can be an issue in quality analysis where the aim of the change-detection test is to identify when the process under monitoring changes its statistical properties.

Sequential change-point detection tests [1] monitor the process by looking at indications for the occurrence of a change (i.e., the null hypothesis is rejected) without the need to make a decision as soon as an acceptable confidence level is granted. The concept of *state of statistical control* has been introduced by Shewhart [10] to model the statistical behavior of selected characteristics of the process at a specific time instant. Any change from the control state, which indicates a variation in the pdf of the process under monitoring, is detected by means of predefined thresholds on the monitored characteristics. To increase the robustness of the Shewhart chart (i.e., reduce the false positives), several moving average mechanisms have been proposed (e.g., [11]). A different approach is suggested by the CUSUM test [12], [13], which exploits the log-likelihood ratio between the pdfs of the null and the alternative hypotheses. This test guarantees an high detection ability but requires knowledge of the pdf before and after the change, thus belonging to the class of parametric change-detection tests. Assuming that the pdf is given is often impractical in real-world applications, and nonparameteric extensions of the CUSUM test have been presented in the literature (e.g., see [14], [15], [16]). In particular, [14] suggests a distribution-free CUSUM test aiming at detecting abrupt changes in the median of a distribution, [15] presents a nonparametric CUSUM-based test for the detection of changes in the mean value, and [16] proposes a self-configuring and nonparametric CUSUM test following a computational-intelligence approach. Differently, [17] presents a nonparametric sequential change-point detection test that uses the Intersection of Confidence Intervals (ICI) rule to monitor the evolution of

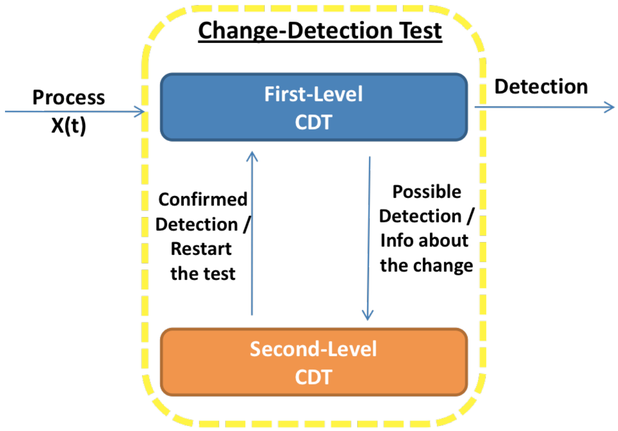

Fig. 1. The architecture of the proposed hierarchical, change-detection test.

the data generating process.

This work proposes a novel approach in the field of sequential change-point detection by suggesting a two-level hierarchical change-detection test. The first-level consists in a change-detection test that activates, when a change is detected, a second-level change-detection test to confirm (or not) the suspected change. If the change is confirmed, the change-detection mechanism signals a detection, otherwise the first-level detection output is considered to be a false positive and the test (eventually retrained) restarts to assess further changes. It should be noted that the proposed change-detection test differs from the mere execution of two subsequent tests. In fact, by referring to Fig. 1, the first- and the second-level change-detection tests are able to cooperate by exchanging information/evaluations about the detected change to improve on-line the detection abilities.

Experiments show that the suggested change-detection-test outperforms state of the art tests by significantly reducing false positives and guaranteeing a low false negatives and detection delays.

The paper is organized as follows: Section II introduces the problem statement, while the general approach of the suggested change-detection test is presented in Section III. Section IV specifies the general approach by describing a hierarchical ICI-based change-detection test. Experimental results are shown in Section V.

II. Problem Statement

In stationary conditions, the process under monitoring $X: \mathbb{N} \rightarrow \mathbb{R}$ generates independent and identically distributed (i.i.d.) observations over time t, extracted from an unknown probability density function (pdf). Let $O_T = \{X(t), t = 1, \dots, T\}$ be the sequence of observations up to time T, and let us assume that (at least) the first T_0 observations come from the process X in a stationary state. Let O_{T_0} be the training set of the change-detection test, i.e., a set of observations generated in stationary conditions used to configure the test parameters. The suggested approach does not require the pdfs of the process before and after the change, which remain unknown. Moreover, the pdf after the change may be time independent (abrupt changes) or evolve with time (drifts).

The goal of the change detection test is to identify the time instant T^* where the process X changes its statistical properties. A good test has to reduce the detection delays (i.e., the time needed to detect the change) and minimize both false positives and negatives.

III. A Hierarchical Change-Detection Test: The General Approach

According to the hierarchical approach illustrated in Fig. 1, the process is steadily monitored by a first-level CDT. Whenever this test detects a change, it activates a second-level test to validate its proposal. If the detection is confirmed, then that is the outcome of the hierarchical test and a change is detected in the process. On the contrary, if the detection is not confirmed, the first-level test is reconfigured to keep on monitoring the forthcoming data. The first-level CDT has to guarantee prompt detection abilities, even at cost of a relatively high false-positive rate, while the second-level acts only when triggered by a change detected at the first-level, to confirm or confute the first-level detection.

The hierarchical CDT is summarized in Algorithm 1. More in detail, the hierarchical CDT starts by configuring the first-level CDT on the training sequence O_{T_0} (line 1). Then, during the operational life (line 2), each new received observation (line 3) is analyzed by the first-level CDT, which assesses stationarity for the data generating process. If the first-level test identifies a change in the incoming observations (line 4), the data are further analyzed to provide the estimate T_{ref} of the change time instant T^* (line 5). The second-level CDT exploits T_{ref} to assess differences in the statistical behavior of the process under monitoring before and after T_{ref} (line 6). Whenever the second-level CDT states that samples generated before and after T_{ref} comes from two different probability density functions, the detection raised by the first-level CDT is confirmed (line 7): thus the hierarchical test reveals a change in the data-generating process and the hierarchical test is reconfigured on the new state (as done, for example, in [18] and [19]). On the contrary, when the second-level CDT does not detect a variation (line 8) then the first-level test is said to have a false positive. Hence, the first-level CDT is stopped and reconfigured to improve its performance (while keeping O_{T_0} as training sequence).

The interaction between the two levels can be further improved, if needed, by modifying the parameters of the first-level CDT, thus reducing the probability of having a false positive in forthcoming data (e.g., by increasing a threshold).

We now separately illustrate the requirements of the first-level CDT and the second-level one.

A. First-Level Change-Detection Test

The first-level requires an on-line CDT to promptly detect changes in the process under monitoring, possibly without relying on a-priori information about the process distribution

ALGORITHM I:
HIERARCHICAL CDT(THE APPROACH)

1. Configure the first-level CDT on O_{T_0}
2. **while** (1){
3. **input** receive a new observation
4. **if** (first-level CDT detects a nonstationary behavior){
5. Estimate T_{ref}
6. **if** (second-level test detects a different statistical behavior before and after T_{ref}) {
7. The first-level detection at t is validated}
8. **else**{
9. The first-level detection is discarded and first-level CDT is reconfigured.
10. Restart the first-level CDT}
11. }
12. }

before and after the change. Since this test has to be executed on-line, its computational complexity might be a critical issue if the code is running on an embedded system. These requirements led us to consider nonparameteric sequential change-point detection tests (see Section I), which generally are nonparameteric versions of traditional parametric sequential change-point detection tests. As an additional requirement, the first-level test should provide, together with each detection, an estimate T_{ref} of the change time-instant T^*.

B. Second- Level Change-Detection Test

The second-level CDT aims at validating detections raised by the first-level, as such this does not operate on-line, and it is executed only when a first-level detection occurs. In particular, we rely on the value T_{ref} provided by the first-level CDT to partition the available observations into two disjoint subsequences (aiming at representing observations before and after the suspected change T_{ref}) and then we apply a statistical hypothesis tests for comparing the two subsequences, assessing possible variations in the data-generating process. Hypothesis tests are theoretically well founded and, in their nonparametric versions, do not require a priori information about the monitored process or nature of the change. The literature about statistical hypothesis tests is rich (e.g., [2]-[4], and [20]): we point out that the choice of the particular hypothesis test to be used at the second-level is strictly related test at the first-level. In fact, the second-level test must be able to validate a hypothesis on (at least) the same statistical quantities monitored by the first-level test (e.g., the pdf, the mean, the median, the variance). For example, if the first-level CDT assesses variations in the mean of the monitored process, the second-level hypothesis test should compare the means (or possibly the distributions) of the process that generated the two aforementioned observation subsequences. In this case, a second-level test that compares the higher-order moments of these two populations (and not their means) is useless, as this would not be able to validate the outputs of the first-level.

IV. A HIERARCHICAL ICI-BASED CHANGE-DETECTION TEST

This section presents a hierarchical CDT, which relies on the ICI-based CDT [17] at the first-level, and on a multivariate hypothesis test exploiting the Hotelling's T-square statistic [20] for the second one. The ICI-based CDT monitors the stationarity of an unknown data-generating process by means of features (e.g. the sample mean and variance computed on disjoint subsequences of observations), as detailed in [17]. Features are suitably defined by means of ad-hoc transformations (when needed) to guarantee Gaussian distribution of their values. Then, the Intersection of Confidence Intervals (ICI) rule is applied to assess the stationary of the features values. Differently from other nonparametric sequential change-point detection tests, the ICI-based CDT is endowed with a *refinement procedure* [19] that, once a change is detected, provides (by relying on repeated executions of the test on shorter observation subsequences) an accurate estimate T_{ref} of the time instant T^*.

The second-level multivariate hypothesis test exploits T_{ref} and the feature values provided by the ICI-based CDT to conclude whether a null hypothesis on the mean value of Gaussian multivariate random variable can be rejected or not. Every time a change is detected, the null hypothesis consists in assuming that the means of the features values before and after T_{ref} are the same (i.e., their statistical difference is null), to validate the detection at the first-level.

As stated in [19], the ICI-based CDT is particularly effective in detecting changes (low false positives and negatives) and it requires a low-computational complexity, but it suffers from an increase of the detection delays when T^* increases. Lower detection delays could be obtained by reducing the test parameter Γ (Γ_{ref} in the refinement procedure), at the expenses of higher false positive rate. The second-level hypothesis test aims at mitigating the occurrence of false positives, while maintaining the first-level change-detection test promptness. Thus, for a given false positive rate, the second-level CDT allows to use lower values of Γ, hence providing prompter detections. The suggested hierarchical ICI-based CDT is detailed in Algorithm 2.

More in detail, during the training phase (line 1) the test computes the feature values and their confidence intervals on O_{T_0}. For each new sample $X(t)$ (line 3), the ICI-based CDT assesses the stationary of the data-generating process (it processes disjoint observation subsequences). Every time a change is detected (line 4), the refinement procedure [19] is activated to estimate T_{ref} (line 5). Then, the feature values in $[0, t]$ are partitioned into two subsequence, as those belonging to $[0, T_{\text{ref}}]$ and $[T_{\text{ref}}, t]$. Let F_0 and F_1 be the mean of ICI-based CDT features on $[0, T_{\text{ref}}]$ and $[T_{\text{ref}}, t]$, respectively. The covariance matrix Σ of the features is computed (line 6) by pooling the covariances [20] estimated before and after the change (since in stationary conditions X is i.i.d., we can assume, as a null hypothesis, that the covariances are the same). Then, the Hotelling's T-square statistic is used to assess if the null hypothesis "$F_0 - F_1$

ALGORITHM II:
ICI-BASED HIERARCHICAL CDT

1. Configure the ICI CDT on O_{T_0}
2. **while** (1){
3.   **input** receive a new observation $X(t)$
4.   **if** (ICI CDT(subsequence containing $X(t)$) detects a nonstationary behavior){
5.   Run the refinement procedure to estimate T_{ref}
6.   Compute F_0 and F_1, the mean features in $[0, T_{\text{ref}}]$ and $[T_{\text{ref}}, t]$. Compute the features covariance Σ.
7.   **if** (multivariate hypothesis test rejects the null hypothesis "$F_0 - F_1$ equals 0") {
8.   The first-level detection is validated: the hierarchical CDT detects a change in t}
9.   **else**{
10.   Restart the ICI CDT .}
11.   }
13. }

equals 0" can be rejected according to a defined significance level α (line 7).

When the multivariate hypothesis test validates the first-level detection (i.e., the null hypothesis is rejected), the hierarchical CDT reveals a change in the subsequences containing $X(t)$ (line 8). On the contrary, when the multivariate hypothesis test does not detect a variation between the two subsequences, the ICI-based CDT is newly configured from the original training set O_{T_0}, and restarted on $X(t+1)$ (line 10).

A. Discussion

The multivariate hypothesis test, which aims at reducing the false positive rate of the ICI-based CDT, relies on a partitioning of the observations in two subsequences generated by the process in stationary, and in the (suspected) nonstationary conditions. The more accurate the estimate of T_{ref}, the more effective the multivariate hypothesis test is. In other words, an inaccurate estimate T_{ref} of T^* might induce either, when $T_{\text{ref}} < T^*$ the presence of feature values computed on samples coming from stationary conditions in F_1 or, when $T_{\text{ref}} > T^*$, the presence of feature values computed on samples coming from nonstationary ones in F_0. In both cases, the multivariate hypothesis test might not be effective in detecting a change in the process, hence, resulting in a false negative of the hypothesis test. This effect is more evident for small perturbations (e.g., an increase of the process mean about 10% of the process standard deviation) where the power of the multivariate hypothesis test is limited even by errors in estimating T_{ref}. On the contrary, in case of higher magnitude perturbations, an accurate estimate of T_{ref} is not so critical since variations are more easily detectable.

We emphasize that other partitioning strategies are possible: for example a more conservative choice would be to compute F_0 from the original training set O_{T_0}, without considering features in $[T_0, T_{\text{ref}}]$. However, we experienced that this latter choice is not successful for validating detections induced by small perturbations.

V. EXPERIMENTS

To validate the effectiveness of the suggested change-detection test we consider both synthetically generated datasets (application D1) and measurements recorded from photodiodes (application D2). We compare the performance of the proposed hierarchical CDT with the ICI-based CDT [17], the CUSUM test [13], and the NP-CUSUM test [15]. Three indexes are used to assess the performances 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.
- Mean Delay (MD): it measures the time delay in detecting a change. It is therefore an estimate of $\mathrm{E}[\hat{T} - T^*]$, being $\hat{T}$ the change-detection outcome, and E[] the mathematical expectation.

Application D1 - refers to a simple mono-dimensional process ruled by a Gaussian pdf (with $\mu = 0$, $\sigma = 1$). The process lasts 60000 samples. We considered two kinds of perturbations having intensity of $+\delta$ affecting the mean value at $T^* = 30000$: an abrupt perturbation (i.e., the mean becomes $\mu + \delta$ in T^*), and a drift (i.e. the mean increases linearly from $T^* = 30000$ to achieve $\mu + \delta$ at the end of the sequence). For each kind of perturbations we considered the following intensity values $\delta \in \{0.1\sigma, 0.5\sigma, \sigma, 2\sigma\}$.

Application D2 - refers to a dataset composed of 250 sequences of light measurements acquired from photodiodes. Dataset of abrupt changes is composed of observation sequences that are similar to the synthetic ones: they last 60000 samples and have been selected to provide a perturbation affecting the mean at sample 30000 in the $\delta \in (0.1\sigma, 2\sigma)$ range (being δ the sample standard deviation of estimated from observations in [0,30000]). However, the sample distribution can be far from being Gaussian, as shown in Fig.4-6. Drift dataset is composed of shorter sequences: they last about 6000 samples and they have been manually aligned to guarantee that their change-points lie at sample 2050. Examples of such sequences are shown in Fig 7, 8.

The proposed hierarchical CDT, the ICI-based CDT, and the NP-CUSUM tests have been configured with the first 400 samples of each dataset as a training sequence. The CUSUM has been configured assuming the pdf of X as known both before and after the change: for this reason it has not been used on the drift datasets in D1 and for application D2. In both application D1 and D2 we experimentally fixed $\Gamma = \Gamma_{\text{ref}} = 2$ in the hierarchical CDT, and we considered two different configurations of the parameter α in the second-level hypothesis test: $= 0.05$, and

TABLE I
SIMULATION RESULTS FOR THE CONSIDERED DATASETS

			CUSUM	ICI CDT	Hierarchical CDT		NP -CUSUM
					α=0.05	α=0.1	
D1	Abrupt δ = 0.1σ	FP (%)	0	35.2	4.4	10	21.2
		FN (%)	0	0	10	6	14.8
		MD (sample)	3687.4	6822.6	7135.3	7089.6	10067.8
	Abrupt δ = 0.5σ	FP (%)	0	35.2	4.4	1	21.2
		FN (%)	0	0	0	0	0
		MD (sample)	153.7	1206.1	1041.0	1070.3	1052.5
	Abrupt δ = 1σ	FP (%)	0	35.2	4.4	10	21.2
		FN (%)	0	0	0	0	0
		MD (sample)	39.6	606.0	525.6	540.8	494.8
	Abrupt δ = 2σ	FP (%)	0	35.2	4.4	10	21.2
		FN (%)	0	0	0	0	0
		MD (sample)	10.6	311.7	275.0	282.0	241.9
	Drift δ = 0.1σ	FP (%)	NA	35.2	4.4	10	21.2
		FN (%)	NA	2.8	38.8	30.0	42.4
		MD (sample)	NA	20655.8	22227.1	22097.5	20999.4
	Drift δ = 0.5σ	FP (%)	NA	35.2	4.4	10	21.2
		FN (%)	NA	0	2.7	1.8	0
		MD (sample)	NA	8876.7	9591.9	9267.7	10144.6
	Drift δ = 1σ	FP (%)	NA	35.2	4.4	10	21.2
		FN (%)	NA	0	0	0	0
		MD (sample)	NA	6099.0	6287.4	6227.3	6622.4
	Drift δ = 2σ	FP (%)	NA	35.2	4.4	10	21.2
		FN (%)	NA	0	0	0	0
		MD (sample)	NA	4246.8	4123.8	4138.6	4377.4
D2	Abrupt	FP (%)	NA	35.2	5.6	10	26.8
		FN (%)	NA	0	0	0	2
		MD (sample)	NA	614.4	510.9	521.7	2763.7
	Drift	FP (%)	NA	8.4	2	2.4	3.6
		FN (%)	NA	0	0	0	0
		MD (sample)	NA	151.1	150.8	150.8	93.9

$\alpha = 0.1$. To ease the comparison we set $\Gamma = 2$ also in the ICI-based CDT. The threshold of the CUSUM test has been experimentally fixed to 20. We tuned the NP-CUSUM to guarantee a mean delay comparable with the one provided by the hierarchical CDT, and we had to use two different configurations for each application scenario, since the data ranges are significantly different (in application D1, $C = 0.05$ and $h = 500$, while in application D2, $C = 20$ and $h = 25$, being C and h specified as in [15]). On the contrary, the ICI-based CDT and the hierarchical CDT use the same configurations in both applications.

Table I shows the comparison among the considered change detection tests: performance values are averaged over 250 runs. The label "NA" denotes a "not applicable" situation as the CUSUM cannot be properly used when the process distribution is unknown (i.e., in those situations where, after the change, the process undergoes a drift or in application D2). As far as the abrupt sequences of application D1 are concerned, the proposed hierarchical CDT provides a lower false positive rate than the ICI-based CDT, and the NP-CUSUM. As expected, the hierarchical CDT shows lower FP when $\alpha = 0.05$ and $\alpha = 0.1$. On the contrary, in this latter configuration it provides lower FN and MD. In both configurations, the hierarchical CDT shows higher FNs than the ICI-based CDT: such increase in FNs is particularly evident for very low perturbations ($\delta = 0.1\sigma$) and disappears for higher magnitude perturbations. We emphasize that this drawback is well compensated by the meaningful reduction in the false positives. Note that, although the hierarchical CDT employs the ICI-based CDT at its first-level, it may provide lower MDs than the sole ICI-based CDT, as this may have false detections that are instead filtered out by the second-level.

The suggested hierarchical CDT outperforms the NP-CUSUM in terms of FPs, FNs, and MDs. This is particularly relevant since the NP-CUSUM detects variations only in the

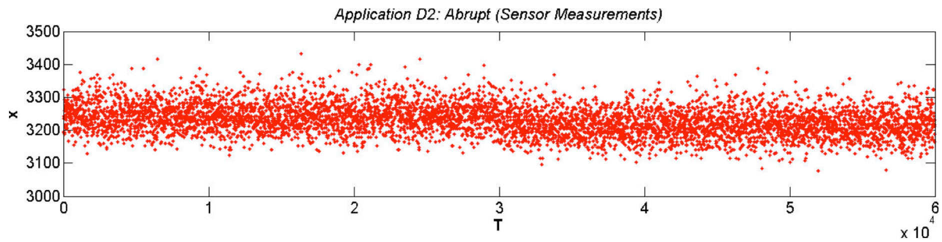

Fig. 2. An example of abrupt change sequence (Application D2).

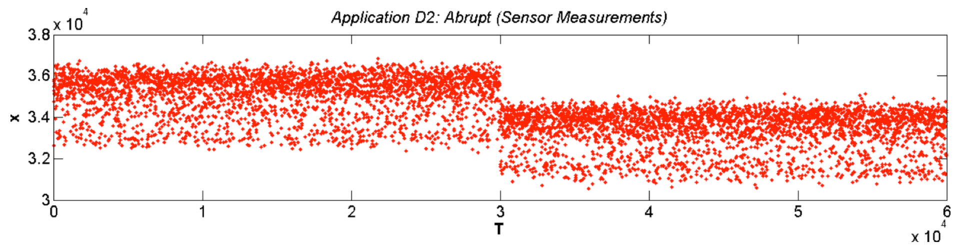

Fig. 3 . An example of abrupt change sequence, generated by an asymmetric process (Application D2).

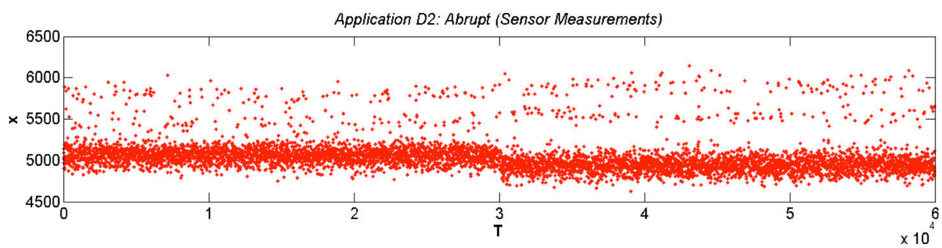

Fig. 4. An example of abrupt change sequence, generated by a multimodal process (Application D2).

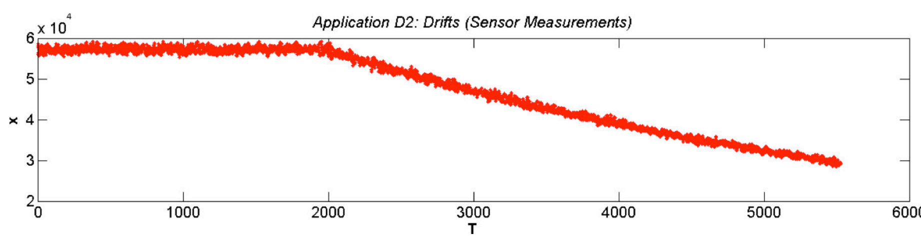

Fig. 5. An example of concept drift almost linear (Application D2).

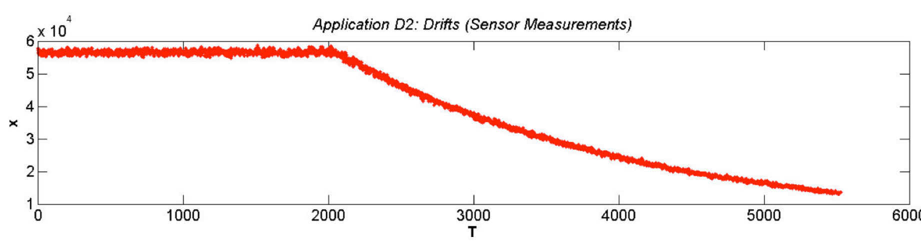

Fig. 6. An example of concept drift (Application D2).

mean value of the process under monitoring, while the suggested hierarchical test allows for assessing changes both in mean and in variance of the data generation process (since it relies by on ICI-based CDT and the multivariate hypothesis test on its feature values). The CUSUM test provides the best performance thanks to the knowledge of the pdfs both before and after the change. Obviously, as δ increases, all the considered tests decrease both FNs and MDs. FPs remain constant in D1 since, to ease the comparison, the 250 datasets on which the different perturbations have been applied are the same for all the experiments.

Experiments on the drift datasets in application D1 are in line with the ones on the abrupt case. We remark that both the MDs of the considered tests, as well as the FNs (when $\delta = 0.01$ and $\delta = 0.05$) increase w.r.t. the abrupt datasets since the change affecting the mean is smooth and less easy to detect. The performance of the hierarchical CDT and the

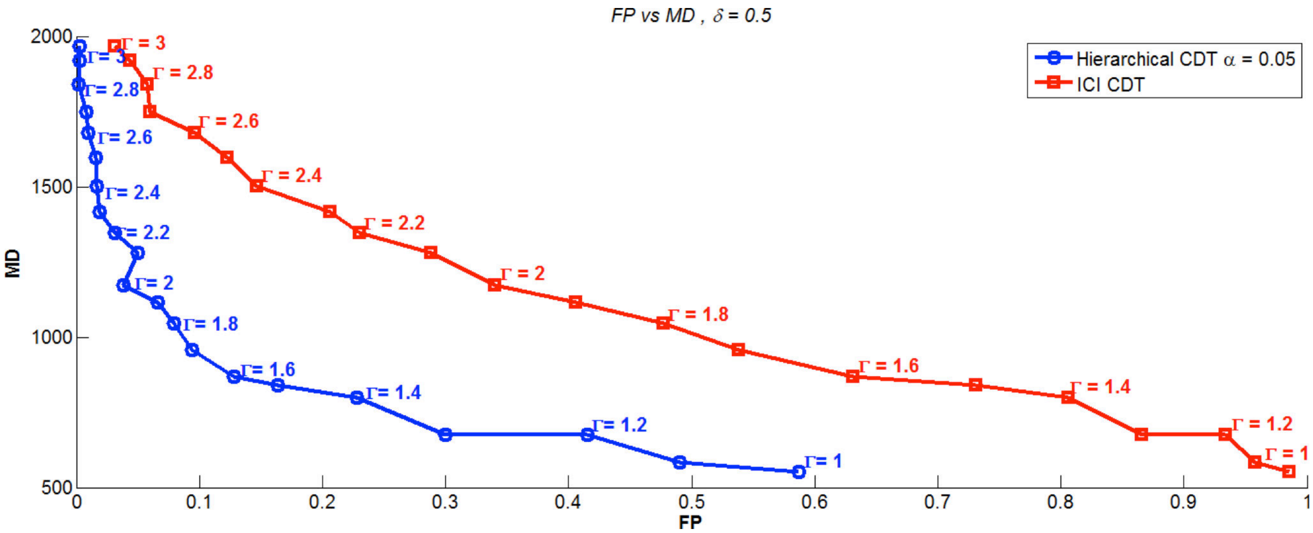

Fig. 7. Mean Delay (MD) w.r.t. False Positives (FP) for the hierarchical CDT (with α=0.05) and the ICI CDT test for a perturbation of δ = 0.5σ

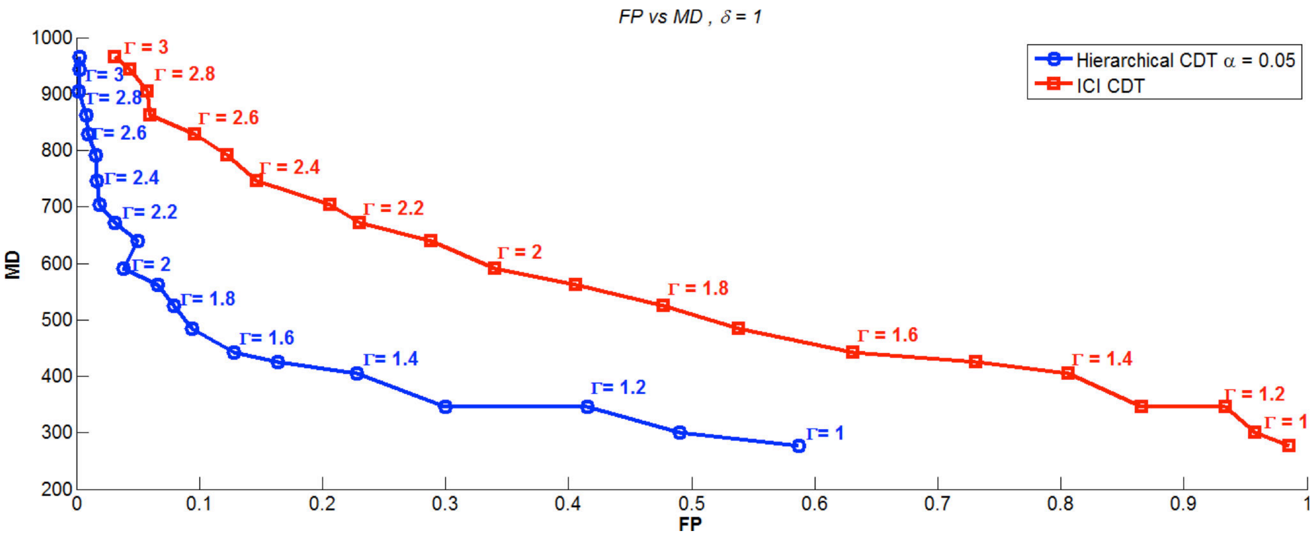

Fig. 8. Mean Delay (MD) w.r.t. False Positives (FP) for the hierarchical CDT (with α=0.05) and the ICI CDT test for a perturbation of δ = 1σ.

ICI-based CDT on application D2 show that both tests can effectively cope with data having non-Gaussian distributions. In fact, the performance of these tests in the abrupt datasets coincides with those of application D1: this means that the features computed by the ICI-based CDT (which are also used in the multivariate hypothesis test at the second CDT level) well approaches the Gaussian distribution. We emphasize that both the ICI-based CDT and the hierarchical CDT have been executed using the same configuration as in application D1, while the NP-CUSUM had to be reconfigured since observations in D2 have a different range.

We performed also more detailed comparison to assess the advantages of the proposed hierarchical test w.r.t. the ICI-based CDT, by analyzing how the performance varies w.r.t. the parameter Γ. We considered Γ ranging from 1 to 3 and fixed $\Gamma = \Gamma_{\mathrm{ref}}$ and $\alpha = 0.05$ for the hierarchical test. We executed 1000 runs of ICI-based CDT and of the hierarchical CDT on datasets of application D1-abrupt with $\delta = 0.5$ and $\delta = 1$. The relationship between false positives and mean delays when Γ varies is illustrated in Fig. 2 and 3. These results are particularly interesting for the following reason: as stated in Section IV.A, by increasing the value of Γ one can reduce the false positives of the ICI-based CDT at the expenses of increasing the detection delays. The hierarchical CDT outperforms the ICI-based CDT since it guarantees low false positives at values of Γ significantly smaller than the ones of ICI-based CDT. In other words, given a fixed percentage of false positives allowed by the application, the proposed hierarchical CDT guarantees a significantly reduced MD, or, similarly, at equal values of MD (for any change δ), the hierarchical CDT guarantees lower false positives than the ICI-based CDT.

VI. Conclusions

This paper suggests a novel hierarchical approach in the field of sequential change-point detection tests. The proposed change-detection test relies on a two-level CDT composed by a first-level that exploits an on-line change-detection test aiming at providing prompt detections, and a second-level, where a different test (e.g., an hypothesis test) validates the detections raised by the first-level. Experiments performed on both synthetically generated datasets and sequences of photodiodes measurements show that the suggested hierarchical change-detection test is able to significantly reduce false positives while negligibly increasing the amount of false negatives and detection delays.

REFERENCES

[1] T.L. Lai, "Sequential analysis: some classical problems and new challenges," in *Statistica Sinica*, vol. 11, n. 2, pp. 303--350, 2001.

[2] A.M. Mood, F.A.Graybill, and D.C. Boes, *Introduction to the theory of statistics* 3rd ed, New York: McGraw--Hill, 1974.

[3] H.B. Mann and D.R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," in *The Annals of Mathematical Statistics*, vol. 18, n. 1, pp. 50-60, 1947.

[4] F.J. Massey "The Kolmogorov-Smirnov test for goodness of fit", in *Journal of the American Statistical Association*, vol. 46, n. 253, pp. 68-78, 1951, JSTOR

[5] S. Muthukrishnan, E. van den Berg, and Y. Wu, "Sequential Change Detection on Data Streams," in *Proc. of IEEE Data Mining Workshops,*pp.551-550, 28-31 2007

[6] A. Wald, "Sequential Tests of Statistical Hypotheses," in *Annals of Mathematical Statistics*, vol. 16, n. 2, pp. 117–186, 1945.

[7] A.Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," in *The Annals of Mathematical Statistic.* vol 19, pp 326-339.

[8] P. Armitage, *Sequential Medical Trials*, 2nd ed. Oxford, U.K.: Blackwell, 1975.

[9] H. R. Lerche, "An optimal property of the repeated significance test," in *Proc. Nat. Acad. Sci. USA*, no. 83, pp. 1546–1548, 1986.

[10] W. Shewhart, *Economic control of quality of manufactured product* New York: D. Van Nostrand Company. 1931

[11] S.W, Roberts, "Control chart tests based on geometric moving averages," in *Technometrics*, vol. 42, n. 1, pp. 97-101, 1959.

[12] Page, E. S. "Continuous inspection schemes," in *Biometrika* vol. 41, pp. 100-114, 1954

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

[14] C. A. McGilchrist and K. D. Woodyer, "Note on a Distribution-Free CUSUM Technique," in *Technometrics*, vol. 17, no. 3,pp. 321-325, 1975

[15] A. G. Tartakovsky, B. L. Rozovskii, R. B. Blazek, and H. Kim, "Detection of intrusions in information systems by sequential change-point methods," in *Statistical Methodology*, vol. 3, no. 3, 2006, pp. 252-293.

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

[17] C. Alippi, G. Boracchi, and M. Roveri, "Change Detection Tests Using the ICI rule", in *Proc. of IJCNN*, pp. 1-7, 2010.

[18] C. Alippi, G. Boracchi, and M. Roveri, "Just in time classifiers: Managing the slow drift case," in *Proc of IJCNN,* 2009, pp.114-120.

[19] C. Alippi, G. Boracchi, and M. Roveri, "Adaptive Classifiers with ICI-based Adaptive Knowledge Base Management," in *Proc of ICANN* 2010,Lecture Notes on Computer Science. Springer Berlin / Heidelberg, vol 6353, pp. 458-467.

[20] J.A. Johnson and D.W. Wichern *Applied multivariate statistical analysis* 4th ed. Prentice Hall, 1998