

# A reconfigurable and element-wise ICI-based change-detection test for streaming data

Giacomo Boracchi, Manuel Roveri

**Abstract**—Detecting changes in data-generating processes is a primary requirement for adaptive and flexible systems endowed with computational intelligence abilities. In order to maintain/improve their performance in evolving or dynamic environments, these systems have to detect any variation in the data-generating process and react and adapt to the new operating conditions. The problem of detecting changes in streams of data is generally addressed by means of Change-Detection Tests (CDTs) and, recently, a family of CDTs based on the Intersection-of-Confidence-Interval (ICI) rule has been presented. ICI-based CDTs monitor data streams by extracting Gaussian distributed features from non-overlapping data windows. The drawback of such a window-wise operational mode is a structural delay, which is particularly evident when the change magnitude is large.

We present a novel ICI-based CDT that overcomes this problem by operating in an element-wise manner thanks to a Gaussian transform of the acquired data. Such an element-wise CDT is characterized by a high change-detection ability and a reduced computational complexity, which makes it suitable for the execution on low-power embedded systems. The proposed CDT is also provided with a reconfiguration mechanism that, after any detected change, allows the CDT to be reconfigured on the new working conditions to detect further changes. A wide experimental campaign shows the effectiveness of the proposed element-wise CDT both on synthetic and real datasets.

## I. INTRODUCTION

The ability to detect variations in the statistical behavior of a data-generating process is a primary and distinguishing feature of adaptive and flexible systems for environmental monitoring, engineering and industrial applications. In fact, detecting changes in the data-generating process allows to promptly react and adapt to variations of the environment where the system operates, or to activate countermeasures when the change affects the cyber-physical system itself (e.g., a fault affecting the sensors or the electronic boards). Therefore, change detection and adaptation are necessary characteristics of any intelligent system aiming at maintaining/improving its quality of service in dynamic and evolving environments.

Detecting changes in data-generating processes becomes even more important when the systems implement computational intelligence techniques, since these often assume, either implicitly or explicitly, that the data-generating process is stationary. This assumption assures that the knowledge acquired during the initial training phase remains valid

during the operational life. Therefore, when the stationarity assumption is violated, the performance of these systems might decrease, even dramatically. As a relevant example of adaptive and flexible systems that relies on change-detection abilities, we mention active classifiers [1]–[5], where change-detection mechanisms act as triggers to refresh the classifier knowledge base. In the classification scenario, changes in the data-generating process are commonly referred to as *concept drift*.

To obtain adaptive systems that operate effectively and efficiently on data streams, the change-detection activity has to be performed 1) in a truly nonparametric manner, 2) with a reduced computational complexity and 3) on streams of data that may undergo a sequence of changes. The first requirement is motivated by the fact that often, when monitoring streaming data, no information about the data-generating process is provided and the effects of changes are unpredictable. The second requirement is due to the fact that, as for example in distributed scenarios such as wireless sensor networks or networked embedded systems, processing units are characterized by constraints on the computational power, memory availability and energy consumption. The third requirement entails that, after detecting a change, the system has to automatically reconfigure itself on the novel operating conditions, to be ready to detect any next change.

Changes on streaming data are typically detected by means of Change-Detection Tests (CDTs) that are statistical techniques able to assess, in an online and sequential manner, the stationarity of a data-generating process. The literature concerning CDTs is very wide [6]–[10] and, recently, a family of CDTs relying on the Intersection-of-Confidence Interval (ICI) rule [11], [12] as core technique has been presented [2], [13]. ICI-based CDTs are nonparametric CDTs that are particularly effective despite their reduced computational complexity and, furthermore, they can be easily reconfigured after each detection: these peculiarities make the ICI-based CDTs very suitable for addressing the change-detection task in flexible and adaptive systems.

Unfortunately, all the ICI-based CDTs operate in a window-wise manner by extracting features from non-overlapping data windows. These features, representing relevant and condensed information about the data-generating process, are modeled as independent and identically distributed (i.i.d.) realizations of a Gaussian random variable. The ICI rule is then applied on the sequence of features to detect nonstationarity in the data-generating process. On the one hand, employing large windows guarantees an increased signal-to-noise ratio of the change to be detected, because the

Giacomo Boracchi and Manuel Roveri are with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milano, Italy, email: {giacomo.boracchi, manuel.roveri}@polimi.it.

This work was supported by the FP7 EU project i-Sense, Making Sense of Nonsense, Contract No: INSFO-ICT-270428.

noise in the extracted features is better suppressed and the change magnitude becomes more evident. On the other hand, the window-wise processing induces a structural detection delay, since changes can be detected with window coarseness rather than at each single sample. Such a delay might not be acceptable when the change magnitude is large, and might represent a problem in those scenarios where the detection delay is a critical figure of merit such as the detection of a contaminant in an intelligent building [14] or the detection of variations in the topology or in the working conditions of a sensor network [15].

To address this problem, we develop an ICI-based CDT operating in an element-wise manner. Such an element-wise CDT learns a specific transform to make the data approximately Gaussian distributed. Then, the ICI rule is directly applied to the transformed data (without extracting features) to detect changes in their expected value. Such an element-wise CDT achieves lower detection delays than its windowed-counterpart, since decisions about the (possible) change are taken at every sample. In particular, the advantages provided by the element-wise CDT become more evident when short training sequences are provided for CDT configuration.

According to [13], we embed the element-wise CDT into a hierarchical, two-layered, architecture to address both the change-detection and the reconfiguration tasks. In fact, from such a hierarchical architecture, it naturally follows a reconfiguration mechanism that allows the element-wise CDT to reconfigure itself once a change has been detected, to be able to detect further changes from the new working conditions.

The paper is organized as follows: Section II details the problem statement, while Section III describes the use of the ICI rule for change-detection purposes. The proposed element-wise CDT is described in Section IV and the experimental results are presented in V. Conclusions are finally drawn in Section VI.

## II. PROBLEM STATEMENT

Let  $x(t) \in \mathbb{R}$  be the scalar data acquired at time  $t$  that is generated from a process  $\mathcal{X}$ . We assume that  $\mathcal{X}$  is a random process, generating over time a stream of i.i.d. data  $X = \{x(\tau), \tau = 1, \dots\}$ . The general change-detection model assumes that, at an unknown time instant  $T^*$ , the probability density function (pdf) of  $\mathcal{X}$  changes, i.e.,

$$x(t) \sim \begin{cases} \phi_0 & t < T^* \\ \phi_1 & t \geq T^* \end{cases}, \quad (1)$$

where  $\phi_0$  and  $\phi_1$  (with  $\phi_1 \neq \phi_0$ ) are the pdfs before and after the change, respectively. Since we pursue a nonparametric approach, we do not assume  $\phi_0$  and  $\phi_1$  to be known.

The aim of any CDT is to analyze the stream  $X$  and detect changes in the pdf of  $\mathcal{X}$  like those in (1). CDTs operate in a sequential manner and assert, at each time instant  $t$ , whether the sequence

$$X_t = \{x(\tau), \tau \leq t\} \quad (2)$$

contains a change point, i.e., if a change in  $\mathcal{X}$  has occurred. Let  $\hat{T}$  be the time instant when a change is detected. The goal of any CDT is to detect such changes with a short delay (i.e.,  $\hat{T} - T^*$ ), while reducing the false positive and negative rates, namely the percentage of detections not corresponding to an actual change in  $\mathcal{X}$  and the percentage of missed detections, respectively.

## III. CHANGE DETECTION USING THE ICI RULE

### A. The ICI Rule

The Intersection of Confidence Intervals (ICI) rule [11], [12] is a technique to define adaptive supports for polynomial regression. The ICI rule operates on observations  $z(i)$  corrupted by Gaussian white noise:

$$z(i) \sim \mathcal{N}(\mu(i), \sigma^2), \quad i \in W \quad (3)$$

where  $\mu(\cdot)$  is the noise-free signal, which is to be estimated by means of polynomial regression,  $\sigma$  is the noise standard deviation and  $W \subset \mathbb{Z}$  is a uniformly spaced sampling grid.

The ICI rule selects, for each specific  $i_0 \in W$ , a neighborhood  $U_{h+}$  within a predefined set of nested neighborhoods  $\{U_h, h \in H\}$ , where  $H$  determines the admissible sizes of these neighborhoods<sup>1</sup>. Neighborhood selection is performed by analyzing  $\{\hat{\mu}_h(i_0), h \in H\}$ , that is a set of point-wise estimates of  $\mu(i_0)$  computed via polynomial regression over the corresponding  $\{U_h, h \in H\}$ <sup>2</sup>. The variance of these polynomial estimators is  $\sigma^2/|U_h|^2$ , and therefore decreases with  $|U_h|$ , the cardinality (size) of  $U_h$ , because the number of data involved in the polynomial regression increases. In contrast, their bias is non-decreasing with respect to  $|U_h|$ , since  $\mu(\cdot)$  might not anymore be a polynomial over  $U_h$ . There is indeed a bias-variance trade off ruled by  $h$ , which in practice determines the risk of the corresponding estimates  $\{(\mu(i_0) - \hat{\mu}_h(i_0))^2, h \in H\}$ .

The ICI rule can be stated as follows. Denote with  $\sigma_h$  the standard deviation of the polynomial estimator over  $U_h$  and with  $I_h$  the confidence interval of  $\hat{\mu}_h(i_0)$  defined as

$$I_h = [\hat{\mu}_h(i_0) - \Gamma \cdot \sigma_h; \hat{\mu}_h(i_0) + \Gamma \cdot \sigma_h], \quad (4)$$

where  $\Gamma > 0$  is a tuning parameter. Then, the estimate  $\hat{\mu}_{i+}(i_0) \in \{\hat{\mu}_h(i_0), h \in H\}$  minimizing the risk corresponds to the largest neighborhood  $U_{h_{i+}}$  for which the intersection of confidence intervals

$$\mathcal{I}_{i+} = \bigcap_{i \leq i+} I_{h_i} \quad (5)$$

is not empty.

<sup>1</sup>The conditions on the neighborhood collection  $\{U_h, h \in H\}$  are formally expressed as  $i_0 \in U_h, \forall h \in H$  and  $U_{h_j} \subset U_{h_{j+1}} \subset W$  provided that  $h_j < h_{j+1}, \forall h_j, h_{j+1} \in H$ . Furthermore, the smallest considered neighborhood has to be  $\{i_0\}$ . For the sake of notation we do not report the subscript  $i_0$  in the neighborhoods  $U_h$ .

<sup>2</sup>In principle, the estimates  $\hat{\mu}_h(i_0)$  have to be computed via least square regression of a polynomial function of order  $m$  over the corresponding support. However, in practice, also different polynomial estimators, such as the Local Polynomial Approximation (LPA) [16] have been used.

### B. Change Detection on streaming using the ICI rule

The ICI rule has not been originally designed to perform change detection on streaming data but the definition of adaptive neighborhoods can be used for change-detection purposes. To more clearly describe the use of ICI rule for change detection we assume  $\mathcal{X}$  to be a Gaussian stationary process, i.e.,

$$x(t) \sim \mathcal{N}(\mu, \sigma^2). \quad (6)$$

In this scenario, the ICI rule can be used to estimate  $\mu$ , identifying the best neighborhood for 0-th order polynomial regression among the set of nested neighborhoods  $\{X_t, t = 1, \dots\}$ , defined as in (2). In stationary conditions, the bias of the polynomial regressor over  $X_t$  is always zero<sup>3</sup>, while the variance decreases as  $1/t^2$ . Therefore, in stationary conditions, the intersection of confidence intervals (5) is expected not to be empty. In contrast, an empty intersection (5) indicates that the polynomial estimator became biased: the data expectation is no more constant and, hence, a change in the expected value of  $\mathcal{X}$  occurred. This is the core mechanism of all the ICI-based CDTs.

Interestingly, when data are not Gaussian, i.e., (6) does not hold, it is possible to perform a preliminary feature-extraction to compute Gaussian-distributed features providing relevant and condensed information about  $\mathcal{X}$  over time. In addition, to guarantee the values of each feature to be independent, the ICI-based CDT presented in [2] originally adopted features extracted from non-overlapping windows of data. More specifically, in [2], the sample mean and the sample variance on non-overlapping data windows were considered (the sample mean approaches the Gaussian distribution thanks to the Central Limit Theorem, while the sample variance is made approximately Gaussian by means of an ad-hoc power-law transformation [17]).

The ICI-rule allows us to perform change detection when the expectation of features follows a polynomial  $\mu(\cdot)$ , and are not just constant as in (6). Assuming that the features follow a polynomial trend actually extends the model typical of the change-detection framework (1), including polynomial trends among stationary conditions and detecting changes affecting these trends, as in [21].

Unfortunately, feature extraction entails a window-wise processing and as such the ICI-based CDT makes decisions about the process stationarity only on each window of data, rather than at each sample. In Section IV we introduce a Gaussian transform on the data  $x(t)$  to obtain an element-wise ICI-based CDT.

## IV. THE ELEMENT-WISE ICI-BASED CDT

We illustrate the element-wise ICI-based CDT by detailing its two core functionalities: the change-detection and reconfiguration. Change-detection consists in transforming the acquired data to be approximately Gaussian distributed,

<sup>3</sup>In the change-detection scenario there is no need to specify at which sample  $\mu$  is estimated (i.e, the location  $i_0$  reported in the previous section), since  $\mu$  is expected to be constant. It is enough to specify the sequence  $X_t$  used for regression.

```

1- input:  $TS_x = \{x(t), t = 1, \dots, L\}$ ,  $\Gamma$ , reconfigure;
2- Estimate  $\lambda$  of (7) on  $TS_x$ ;
3-  $\hat{\mu}_L = \sum_{t=0}^L \mathcal{M}_\lambda(x(t))/L$ ,
    $\hat{\sigma} = \sqrt{\sum_{t=0}^L (\mathcal{M}_\lambda(x(t)) - \hat{\mu}_L)^2 / (L-1)}$ ;
4-  $\mathcal{I}_L = [\hat{\mu}_L - \Gamma \cdot \hat{\sigma} / \sqrt{L}, \hat{\mu}_L + \Gamma \cdot \hat{\sigma} / \sqrt{L}]$ ;
5-  $t = L + 1$ ;
6- while ( $x(t)$  arrives) do
7-    $y(t) = \mathcal{M}_\lambda(x(t))$ ;
8-    $\hat{\mu}_t = ((t-1) \cdot \hat{\mu}_{t-1} + y(t)) / t$ ;
9-    $\mathcal{I}_t = [\hat{\mu}_t - \Gamma \cdot \hat{\sigma} / \sqrt{t}, \hat{\mu}_t + \Gamma \cdot \hat{\sigma} / \sqrt{t}] \cap \mathcal{I}_{t-1}$ ;
10-  if ( $\mathcal{I}_t == \emptyset$ ) then
11-    A change is detected at time  $\hat{T} = t$ ;
12-    if (reconfigure == 1) then
13-      Estimate  $T_{\text{ref}}$ , using the refinement
14-      procedure, Algorithm 3 in [2];
15-       $TS_x = \{x(t), t = T_{\text{ref}}, \dots, \hat{T}\}$ ;
16-      Estimate  $\lambda$  of (7) on  $TS_x$ ;
17-       $L = \hat{T} - T_{\text{ref}} + 1$ ;
18-       $\hat{\mu}_t = \sum_{t=T_{\text{ref}}}^{\hat{T}} \mathcal{M}_\lambda(x(t))/L$ ,
19-       $\hat{\sigma} = \sqrt{\sum_{t=T_{\text{ref}}}^{\hat{T}} (\mathcal{M}_\lambda(x(t)) - \hat{\mu}_L)^2 / (L-1)}$ ;
20-       $\mathcal{I}_t = [\hat{\mu}_L - \Gamma \cdot \hat{\sigma} / \sqrt{L}, \hat{\mu}_L + \Gamma \cdot \hat{\sigma} / \sqrt{L}]$ ;
21-    else
22-      Report the detection and exit;
23-    end
24-  end
25-   $t = t + 1$ ;
26- end

```

**Algorithm 1:** The Element-wise ICI-based CDT.

and then assessing the stationarity of the transformed data by means of the ICI rule. Both the Gaussian transformation and the distribution of the transformed data in stationary conditions are learned from an initial training set  $TS_x$  containing data generated by  $\mathcal{X}$  in stationary conditions. After each detection, a new training set is automatically identified from the recent data and the element-wise ICI-based CDT can be reconfigured on the new state of the data-generating process to detect further changes. The element-wise ICI-based CDT is detailed in Algorithm 1 and described in the rest of the section.

### A. The Element-wise Change-Detection

As stated in Section II, the aim of the CDT is to sequentially analyze the stream  $X$  to detect variations in the data-generating process  $\mathcal{X}$ . In contrast with the ICI-based CDT presented in [2], which aggregates data into non-overlapping windows to extract i.i.d. Gaussian features, the proposed CDT is able to process data  $x(t)$ s in an element-wise manner thanks to a suitable transformation that makes them approximately Gaussian distributed. Most of the Gaussian

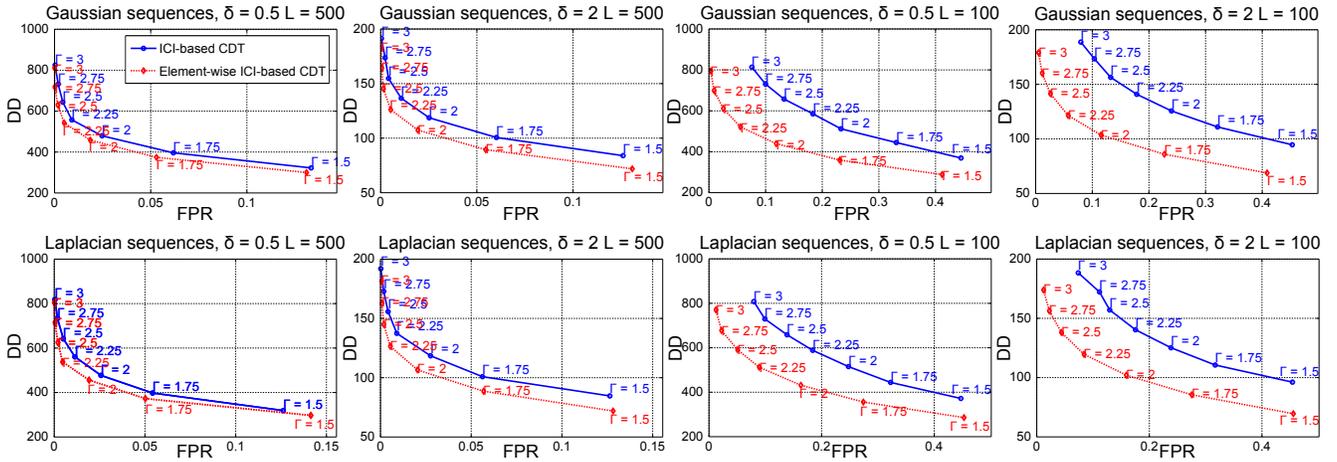


Fig. 1. Test on the *Synthetic* dataset. The plots present the  $(FPR, DD)$  achieved when  $\Gamma = \{1.5, 1.75, 2, 2.25, 2.5, 2.75, 3\}$ .

transformations in the literature are power transforms [18] and, probably, the most known transformation is the Box-Cox one [19]. The drawback of such a transform is that it requires data to be positive or bounded below by a known value. Unfortunately, these may represent a serious limitation when analyzing data streams. Manly [20] introduced an exponential transform which can be applied also to negative data, and that was proved to be effective in turning skewed unimodal distributions into nearly-symmetric Gaussian ones. The Manly transform is defined as follow

$$\mathcal{M}_\lambda(x(t)) = \begin{cases} (e^{\lambda x(t)} - 1)/\lambda; & \lambda \neq 0 \\ x(t); & \lambda = 0 \end{cases} \quad (7)$$

where the  $\lambda \in \mathbb{R}$  parameter is estimated from the training set  $TS_x$  following the maximum likelihood approach in [20].

The configuration phase of the element-wise CDT is reported in lines 1-4 of Algorithm 1. The CDT requires an initial training set  $TS_x$ , the parameter  $\Gamma$  that regulates the detection promptness in all the ICI-based CDTs, and the flag parameter *reconfigure*, which determines whether the CDT has to be automatically reconfigured after each detection (line 1). The parameter  $\lambda$  of the Manly transform is estimated from  $TS_x$  (line 2) as in [20]. The CDT configuration phase consists in computing the confidence interval for the 0-th order polynomial fit of the transformed training set (lines 3 and 4) as described in Section III-B.

During the operational life (line 7), samples  $x(t)$ s are transformed into  $y(t)$ s by means of  $\mathcal{M}_\lambda$  (7). Then, the estimate of  $\hat{\mu}_t$  from all the transformed data is computed (line 8) as well as the corresponding confidence interval. The intersection of all the confidence intervals until  $t$  is computed as in (5) (line 9). As soon as this intersection becomes empty, a change in the expected value of the transformed data (and hence in  $\mathcal{X}$ ) is detected (line 10) and the reconfiguration phase is eventually activated (lines 11 - 14).

### B. Reconfiguration after a change

Every time a change has been detected, a reconfiguration phase is activated (when *reconfigure* == 1) to re-

configure the element-wise ICI-based CDT on the new state to detect further changes. This reconfiguration phase relies on an estimate  $T_{\text{ref}}$  of the change point  $T^*$  (line 11). To compute this estimate we exploit the refinement procedure described in Algorithm 3 of [2]. This procedure consists in repeatedly invoking the ICI-based CDT on shorter sequences of (transformed) data.

Remarkably, all the samples between  $T_{\text{ref}}$  and  $\hat{T}$  can be safely associated with the new state of the data-generating process and represent a new training set  $TS_x$  to reconfigure the element-wise ICI-based CDT. Hence, a new value of the parameter  $\lambda$  (7) is estimated on  $TS_x$  (line 12) and  $\hat{\mu}_L$  and  $\hat{\sigma}$  for configuring the ICI-based CDT are computed from the new transformed observations (line 13). Finally (line 14), the confidence interval for the new estimator is computed, and the element-wise ICI-based CDT is ready to detect further changes of  $\mathcal{X}$ .

## V. EXPERIMENTS

In the experiments we contrast the proposed element-wise CDT with the ICI-based CDT in [2], which monitors the sample mean over windows of  $\nu = 20$  samples. We considered both synthetic and real-world datasets and measured the CDT performance according to the following figures of merit:

- False Positive Rate (*FPR*), the percentage of runs where a change was erroneously detected in stationary conditions, i.e.,  $\hat{T} < T^*$ .
- False Negative Rate (*FNR*), the percentage of runs where a true change in  $\mathcal{X}$  was not detected.
- Detection Delay (*DD*), the average of  $\hat{T} - T^*$  (expressed in samples) computed on runs where  $\hat{T} > T^*$ ;

### A. Datasets Description

*Synthetic Dataset* refers to data-generating process  $\mathcal{X}$  that in stationary conditions follow either a Gaussian or a Laplace distribution. Each sequence in the dataset lasts 6000

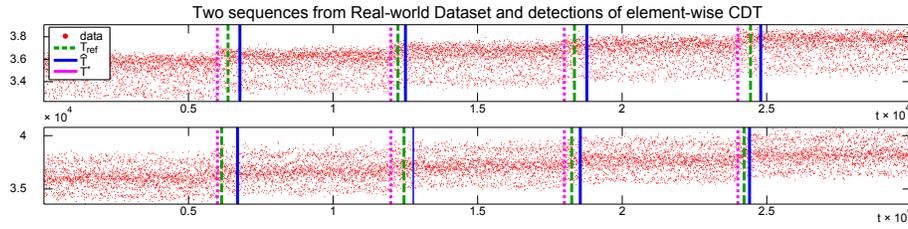


Fig. 2. Two sequences from the real-world dataset together with the true change time instant ( $T^*$ ), the detections ( $\hat{T}$ ) and the refined estimate ( $T_{\text{ref}}$ ) provided by the element-wise ICI-based CDT.

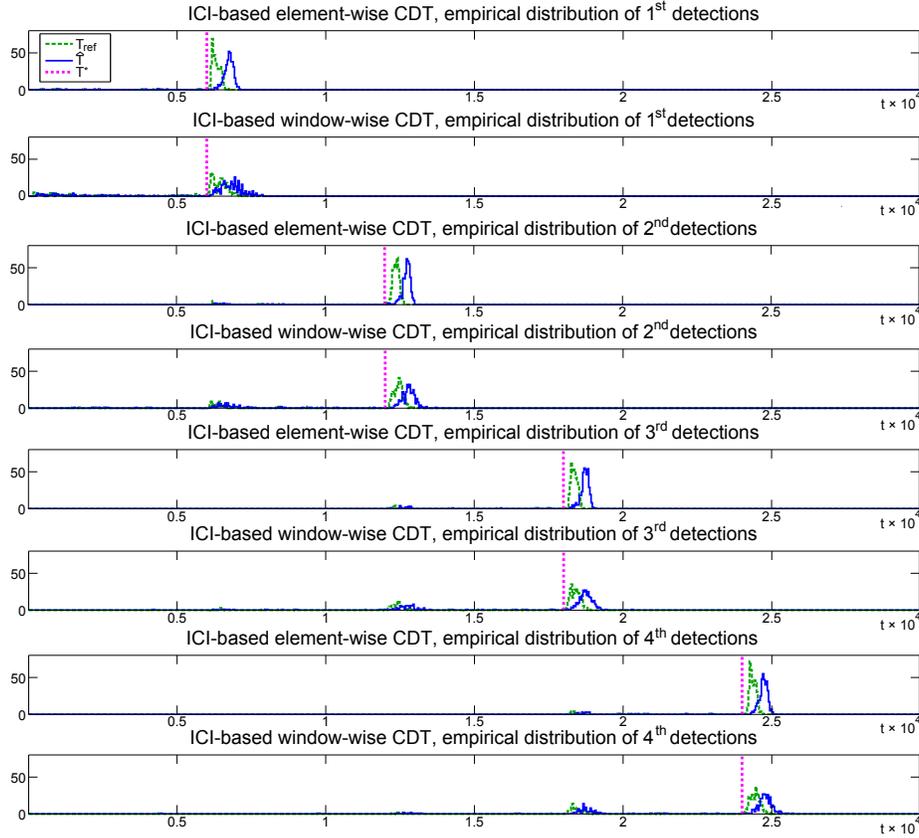


Fig. 3. Empirical distribution of the detections ( $\hat{T}$ ) and refined estimates ( $T_{\text{ref}}$ ) of the two CDTs on the real-world dataset ( $\Gamma = 2.5$  in both CDTs and  $\Gamma = 2$  in the refinement procedures). The histograms indicate that detections provided by the element-wise CDT are better localized than the ICI-based CDT; furthermore, the element-wise CDT suffers from less false positives than the ICI-based CDT.

samples and at  $T^* = 4000$ , an abrupt additive perturbation of magnitude  $\delta$  affects the mean value of  $\mathcal{X}$ , i.e.,

$$x(t) \sim \begin{cases} \phi & t < T^* = 4000 \\ \phi + \delta & t \geq T^* = 4000 \end{cases}, \quad (8)$$

being  $\phi$  the pdf of either a standard normal distribution (i.e.,  $\mathcal{N}(0, 1)$ ) or a zero-mean, unitary variance, Laplacian distribution (i.e.,  $\mathcal{L}(0, 1/\sqrt{2})$ ). We set  $\delta = 0.5$  and  $\delta = 2$  corresponding to half and twice the variance of  $\mathcal{X}$ , respectively. Figures of merit have been computed over datasets of 5000 sequences each. Gaussian sequences were considered to compare the core mechanism of the CDTs where the Gaussian transform and feature extraction operate in their ideal conditions. In this dataset, we considered two different

training sequence length  $L$  for both CDTs, i.e.,  $L = 100$  and  $L = 500$ . We used  $\Gamma$  ranging from 1.5 to 3 for both CDTs and no reconfiguration is activated after each detection (i.e.,  $reconfigure = 0$ ).

*Real-world dataset* is composed of 430 sequences of measurements acquired from photodiodes. Each sequence lasts 30000 samples and contains 5, artificially introduced, additive shifts like (8) of magnitude  $\delta = 0.5 \cdot S^2$ , being  $S^2$  the sample variance of data in stationary conditions. This dataset was prepared to test both the CDTs performance on non-Gaussian data (Fig. 2 shows two examples of data from fairly skewed and heavy tailed distribution) and their reconfiguration abilities when changes occur in a sequence.

Both CDTs have been configured with  $\Gamma = 2.5$  ( $\Gamma = 2$

in the refinement procedure) and  $reconfigure=1$ . After each detection, for both CDTs, we require a minimum training set length of 100 samples for the reconfiguration purposes.

### B. Discussion

Experimental results on the *Synthetic* datasets are reported in Fig. 1, where it is shown that, for all the considered values of  $\delta$ , the element-wise CDT outperforms the window-wise CDT both in terms of  $DD$  and  $FPR$  (in all the experiments  $FNR = 0$ ). Interestingly, the red dotted curve (element-wise ICI-based CDT) is always below the blue solid one (ICI-based CDT) in all the ( $FPR$ ,  $DD$ ) plots, indicating that the performance of the element-wise ICI-based CDT cannot be achieved by the window-wise CDT by simply adjusting  $\Gamma$ . The advantages of the element-wise CDT become more evident when the training set is small ( $L = 100$ ). In fact, operating on non-overlapping windows implies less training samples for the configuration (i.e., only  $L/\nu$  features): this becomes a critical issue when the training set is small.

The performance of both CDTs on the Laplacian and the real-world dataset are in line with those of the Gaussian distribution, meaning that the Manly transform and the feature extraction allow to correctly operate on non-Gaussian data. In particular, the histograms in Fig. 3 shows the empirical distributions, over the whole dataset, of the first 4 detections on each sequence (solid blue line) and the corresponding refined estimates  $T_{ref}$  (the dashed green line) of both the element-wise and the window-wise CDTs on the real-world dataset. In these runs, both CDTs are automatically reconfigured, after each detection, to detect next changes. Interestingly, the element-wise ICI-based CDT reveals to be more effective than the window-based one, since the detections have lower spread and are closer to  $T^*$  (the dotted magenta lines). Moreover, the element-wise CDT is characterized by a lower number of false positive detections, as the area of the histograms before  $T^*$  is lower for the element-wise than for the window-wise CDT. The values of  $T_{ref}$  computed by the element-wise CDT are also closer to  $T^*$  than those of the window-wise CDT, hence guaranteeing better reconfiguration abilities.

We experienced that in situations where the data distribution over  $T'S_x$  is multi modal or suffers from heavy outliers, the transformed data may be far from being Gaussian, and this may seriously impair the performance of the element-wise CDT. The feature extraction is then a viable option to cope with these situations, and the ICI-based CDT executed on the sample mean over non-overlapping data windows achieves satisfactory change-detection performance.

## VI. CONCLUSIONS

The paper presented a novel ICI-based CDT that, thanks to a Gaussian transformation learned during the initial training phase, makes decisions about the stationarity of the data-generating process at every new acquired sample. Such element-wise CDT overcomes the limitations of the ICI-based CDTs performing feature extraction, and represents a viable alternative to detect changes in the expected value of

an unknown data-generating process. The element-wise CDT is particularly suited for streaming data, thanks to its reduced computational complexity and its automatic reconfiguration capabilities. The advantages provided by the element-wise ICI-based CDT have been confirmed both on synthetic and real-world datasets.

## REFERENCES

- [1] C. Alippi, G. Boracchi, and M. Roveri, "Just-in-time classifiers for recurrent concepts," *Neural Networks and Learning Systems, IEEE Transactions on*, vol. 24, no. 4, pp. 620–634, April.
- [2] —, "A Just-In-Time adaptive classification system based on the Intersection of Confidence Intervals rule," *Neural Networks*, vol. 24, no. 8, pp. 791 – 800, 2011.
- [3] C. Alippi and M. Roveri, "Just-In-Time adaptive classifiers – part II: Designing the classifier," *Neural Networks, IEEE Transactions on*, vol. 19, no. 12, pp. 2053 –2064, dec. 2008.
- [4] J. Gama, P. Medas, G. Castillo, and P. Rodrigues, "Learning with drift detection," in *Advances in Artificial Intelligence—SBIA 2004*. Springer, 2004, pp. 286–295.
- [5] K. Nishida and K. Yamauchi, "Learning, detecting, understanding, and predicting concept changes," in *Proceedings of IJCNN 2009*. IEEE, 2009, pp. 2280–2287.
- [6] M. Basseville and I. V. Nikiforov, *Detection of abrupt changes: theory and application*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.
- [7] A. G. Tartakovsky, B. L. Rozovskii, R. B. Blak, and H. Kim, "Detection of intrusions in information systems by sequential change-point methods," *Statistical Methodology*, vol. 3, no. 3, pp. 252 – 293, 2006.
- [8] G. J. Ross, D. K. Tasoulis, and N. M. Adams, "Nonparametric monitoring of data streams for changes in location and scale," *Technometrics*, vol. 53, no. 4, pp. 379–389, 2011.
- [9] 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.
- [10] Y. Kawahara and M. Sugiyama, "Sequential change-point detection based on direct density-ratio estimation," *Stat. Anal. Data Min.*, vol. 5, no. 2, pp. 114–127, Apr. 2012.
- [11] A. Goldenshluger and A. Nemirovski, "On spatial adaptive estimation of nonparametric regression," *Math. Meth. Statistics*, vol. 6, pp. 135–170, 1997.
- [12] 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, pp. 223–235, 2002.
- [13] C. Alippi, G. Boracchi, and M. Roveri, "A hierarchical, nonparametric, sequential change-detection test," in *Proceedings of IJCNN 2011*, 31 2011-aug. 5 2011, pp. 2889 –2896.
- [14] T. H. Jeys, W. D. Herzog, J. D. Hybl, R. N. Czerwinski, and A. Sanchez, "Advanced trigger development," *Lincoln Laboratory Journal*, vol. 17, no. 1, pp. 29–62, 2007.
- [15] S. Brienza, D. De Guglielmo, C. Alippi, G. Anastasi, and M. Roveri, "A learning-based algorithm for optimal mac parameters setting in ieee 802.15. 4 wireless sensor networks," in *Proceedings of PE-WASUN '13*. ACM, 2013, pp. 73–80.
- [16] V. Katkovnik, V. Katkovnik, K. Egiazarian, and J. Astola, *Local Approximation Techniques in Signal and Image Processing*, ser. Press Monographs. Society of Photo Optical, 2006.
- [17] 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, pp. pp. 479–485, 1981.
- [18] R. Sakia, "The box-cox transformation technique: a review," *The statistician*, pp. 169–178, 1992.
- [19] G. E. Box and D. R. Cox, "An analysis of transformations," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 211–252, 1964.
- [20] B. Manly, "Exponential data transformations," *The Statistician*, pp. 37–42, 1976.
- [21] C. Alippi, G. Boracchi, and M. Roveri, "An effective Just-In-Time adaptive classifier for gradual concept drifts," in *Proceedings of IJCNN 2011*, 31 2011-aug. 5 2011, pp. 1675 –1682.