Exploiting Self-Similarity for Change Detection

Giacomo Boracchi, Manuel Roveri

Abstract— **Time-series data are often characterized by a large degree of self-similarity, which arises in application domains featuring periodicity or seasonality. While self-similarity has shown to be an effective prior for modeling real data in the signal and image-processing literature, it has received much less attention in time-series literature, where only few works leveraging the self-similarity for anomaly detection have been presented. Here we introduce a novel change-detection test to detect structural changes in time series by analyzing their self-similarity. The core of the proposed solution is the definition of a change indicator to quantitatively assesses the self-similarity of the time-series data over time. In particular, the change indicator is obtained by comparing each patch to be analyzed with its most similar counterpart in a change-free training set. Experimental results on the flow measurements in the water distribution network of the Barcelona city show the effectiveness of the proposed solution.**

I. Introduction

OFTEN, signals and time series are redundant and self-similar. Namely, each patch (i.e., a segment of data) exhibits similarities to other patches in the sequence. Self-similarity is particularly evident in climate (or environmental) monitoring time series, in ECG recordings, and in water (or electricity) consumption measurements, where periodicity and seasonality result in repeated patterns. As an example, Fig. 1.a illustrates flow measurements in a District Metered Area (DMA) of the Barcelona city (Spain): the daily trend in the water demand results in a highly self-similar time series.

A plethora of algorithms in the signal and image processing literature indicates the self-similarity as an effective prior for handling real-world data. Fractal models for natural images [1] and fractal block coding [2] explicitly rely on the assumption that images exhibit similar content in different locations. Such a paradigm was further developed in texture synthesis and completion [3] and in the nonlocal-means denoising algorithm [4], which has inspired several successful noise attenuation algorithms for images [5], [6], videos [7], [8], surfaces [9], and point-clouds [10]. Exploiting the non-local self-similarity is now a widely accepted paradigm, which has been adopted in a broad range of signal and image processing applications, see [11].

Conversely, only few works in the time-series literature consider self similarity, and these mainly focus on the identification of similar [12], [13] or anomalous patches [14], [15]. In particular, the identification of anomalous patches (also called "discords" [16]) is generally formulated as an anomaly/novelty detection problem [17], [18]. Therefore, these techniques are not designed to detect *structural changes*, i.e., situations where the data-generating process shifts from *in-control* to *out-of-control* conditions, while they are instead meant to identify outliers in a data sequence.

Detecting structural changes is a primary issue in time series analysis, representing a mean to identify faults in the sensing apparatus or unforeseen evolutions of the data-generating process. Even though most of structural changes introduces patterns that are highly dissimilar to those generated in stationary conditions, self-similarity has never been used as a prior for detecting structural changes.

Most of the change-detection solutions for time series [19], [20], [21], [22] relies on predictive or approximation models estimated from a change-free training set. During the operational life, the residual, i.e., the discrepancy between observation and the output of the estimated predictive/approximation model, is monitored by means of a standard change-detection test (CDT) [23], [24], [25],[26], [27]. CDTs are statistical techniques meant for analyzing, in an online and sequential manner, independent and identical distributed (i.i.d.) realizations of a random variable. Unfortunately, a residual-based approach might not be feasible in complex real-world scenarios, where estimating accurate models is particularly difficult [21].

Here, we introduce a CDT for time series exhibiting self-similarity that does not require to estimate a predictive or approximation model of the data-generating process. Our self-similarity based CDT detects, in an on-line and sequential manner, a structural change as soon as the monitored data are no more similar to those belonging to an initial, change-free, training sequence. In particular, we compare the patch around each new sample with all the patches in the training sequence to identify the most similar one. Then, we compute the change indicator as the difference between the values at the center of these two patches. Structural changes can then be successfully detected by monitoring any variation in the statistical behavior of the change indicator by means of a CDT. To the best of our knowledge, this is the first work exploiting self-similarity to perform detection of structural changes in time series. Our self-similarity based CDT has been successfully tested on flow measurements coming from the Barcelona water distribution network, where structural changes have been artificially introduced.

This paper is organized as follows: the literature concerning anomaly detection and change-detection methods exploiting self-similarity is presented in Section II, while the self-similarity based CDT is described in Section III. Experiments are presented and discussed in Section IV, while Section V draws concluding remarks and ongoing works.

Giaocmo Boracchi and Manuel Roveri are with Dipartimento di Elettronica, Informazione e Bioingegneria (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.

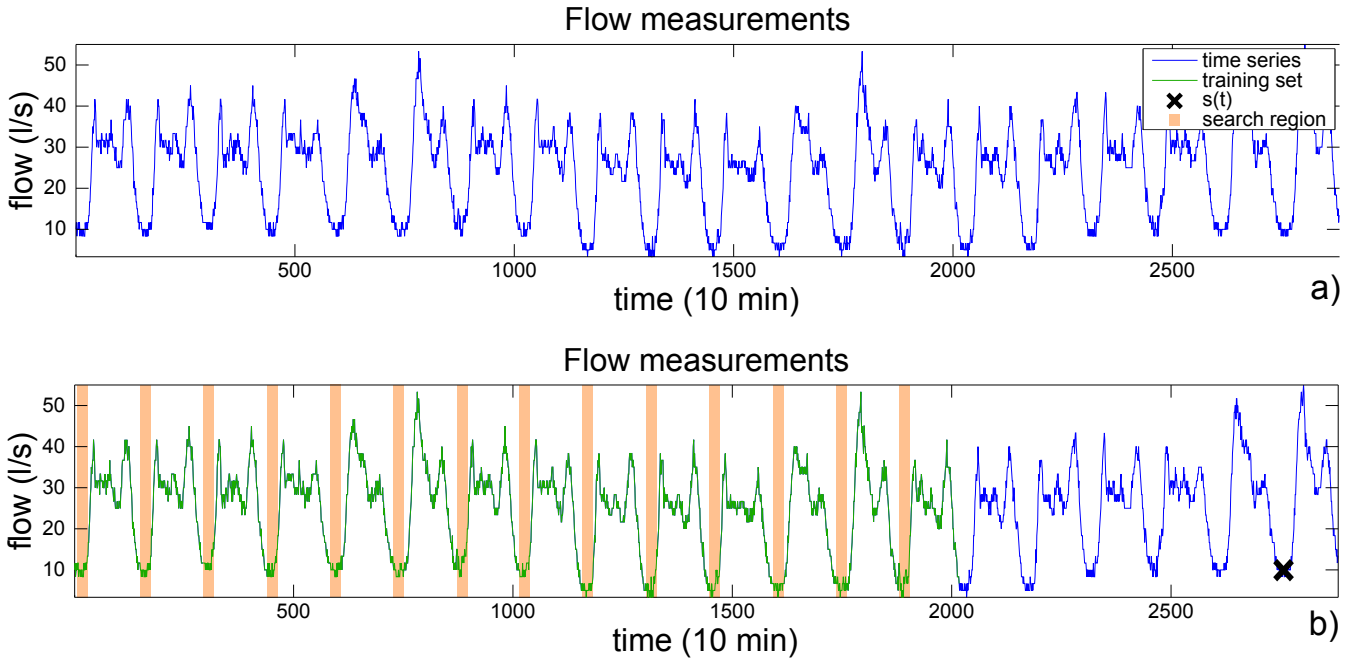

Fig. 1. a) Flow measurements in a DMA of Barcelona over 20 days: measurements are taken every 10 minutes (144 samples per day). This time series is highly self-similar: the water demand over each day follows a clear trend determined by citizens' habits. b) An example of search neighborhood $R_{t,\phi,\delta}$, where $t = 2500$, $\phi = 144$ and $\delta = 10$.

II. Related Works

To the best of our knowledge, the problem of detecting structural changes in time series through self-similarity has never been addressed in the literature, and only few works concerning anomaly detection and self-similarity in time series have been presented [15], [28], [14], [29], [30], [16].

Most of these works [15], [14], [29], [28] follow the *negative selection* approach, which is inspired by a specific mechanism of the human immune system [18]. In particular, our immune system randomly generates receptors (T-cells) by genetic rearrangements. Antigens (e.g., viruses or bacteria) are detected as soon as they bind with such detectors. To prevent that receptors bind also with cells of the human body, randomly generated T-cells undergo a censoring process: those matching with self-cells are discarded. An example of negative selection algorithm is presented in [15], where data are patch-wise encoded into binary strings. Anomaly detectors are created by randomly generating strings and removing those matching with strings in the training set. During the operational life, incoming patches are compared with all these detectors, and an anomaly is detected as soon as a match is found. The main drawback of this solution is the binary encoding of the patches, as this might dramatically reduce the information provided to the detection algorithm. A real-valued negative selection algorithm that is able to operate real-valued patches is presented in [28]. There, the initial configuration phase aims at creating a training set composed of both self (normal) patches and randomly generated patches to train a classifier (i.e., a multi-layer preceptron) devoted to negative selection.

Similarly, [14] suggests an immunity-based negative selection algorithm operating on real values, where the discords-detection mechanism exploits rules generated by genetic algorithms. This approach has been successfully applied to the detection of intrusions in computer networks. The algorithm in [29] extends [14] by replacing the crisp detection-rules mechanism with fuzzy logic. A technique to generate a set of detection fuzzy rules by relying only on normal samples is also proposed.

Remarkably, all the above mentioned works measure some form of similarity with alternative detectors. In contrast, our approach consists in exploiting self-similarity to characterize the in-control state of the data-generating process, and we detect any structural change as any break in such self-similarity. Therefore, we do not need to generate any anomaly detector for out-of-control states.

Anomaly-detection techniques for identifying faults in a simulated refrigerator system were investigated in [30], where several matching rules were considered (e.g., threshold on the euclidean distance between patches or analysis of binary encoded patches). However, the solutions proposed in [30] are not sequential, and out-of-control states yielding patches that are not clearly dissimilar to those in the training sequence are not easily detectable, even when the change persists over time. It is possible to increase the change sensitivity of such techniques by lowering the detection threshold, at the expenses of a large number of false alarms.

In contrast with solutions inspired to the immune system, [16] performs anomaly detection by computing a distance index between any patch and the most similar patch in the rest of the time series. As in [15], this distance-based anomaly-detection technique relies on the analysis of quantized data

by means of symbolic aggregate approximation. The authors adopt a heuristic procedure to reduce the computations when calculating such distance index for all the patches in the time series. Unfortunately, this technique is not sequential and cannot be used to detect structural changes.

III. Detecting Changes in Self Similarity

A. Problem Formulation

Let us denote by $S = \{s(\tau), \tau = 1, \dots, \}, s(\tau) \in \mathbb{R}$ a real-valued time series, which is generated by a process $\mathcal{S}$. We assume that, when $\mathcal{S}$ is in-control state, the time series S features self-similarity, i.e., for each patch of S it is possible to find at least another similar patch within S itself. We say that $\mathcal{S}$ presents a structural change at time T^* when it shifts into an out-of-control state that affects the self similarity of S. In particular, there is no more similarity between patches in $\{s(t), t = 1, \dots, T^* - 1\}$ and $\{s(t), t \geq T^*, \dots\}$.

We address the problem of detecting structural changes by analyzing the time series in a sequential manner. To this purpose, we assume that an initial training set $TS = \{s(\tau), \tau = 1, \dots, L\}$, generated when $\mathcal{S}$ is in-control, is provided.

B. Notation

To assess the self-similarity in a time series S, we operate in a patches-wise manner and we adopt the following notation

$$\mathbf{s}_t = \{s(t-\nu), \dots, s(t), \dots, s(t+\nu)\}, \tag{1}$$

for indicating a patch having size $2\nu + 1$ and centered in t. Since $\mathbf{s}_t \in \mathbb{R}^{2\nu+1}$, each patch corresponds to a vector, and the distance between two patches can be computed as the Euclidean distance

$$d(s_{t_1}, s_{t_2}) = \|\mathbf{s}_{t_1} - \mathbf{s}_{t_2}\|_2 = \sqrt{\sum_{i=-\nu}^{\nu} \big(s(t_1+i) - s(t_2+i)\big)^2}, \tag{2}$$

for any pair of patches centered in t_1 and t_2. We denote by $\widehat{T}$ the time instant when a structural change is detected.

C. The Proposed Solution

We tackle the problem of detecting structural changes in $\mathcal{S}$ by determining when patches in the time series S are no more similar to any patch in the training set. To this purpose we build a training set of patches

$$\mathbf{P} = \{\mathbf{s}_\tau, \tau = \nu + 1, \dots, M - \nu - 1\}, \tag{3}$$

where $M \leq L$ indicates the portion[1] of the training set TS that is used for building the training set of patches $\mathbf{P}$. We approach the detection of structural changes in $\mathcal{S}$ with the following design assumption:

$$\begin{cases} \exists\, \mathbf{s}_u \in \mathbf{P} \text{ similar to } \mathbf{s}_t, & \forall t < T^* \\ \nexists\, \mathbf{s}_u \in \mathbf{P} \text{ similar to } \mathbf{s}_t, & \forall t \geq T^* \end{cases}. \tag{4}$$

[1]Part of TS is often required to configure the CDT that is used in the next steps of the algorithm, as it will be clarified later on in Section III-E of the training set TS used for building $\mathbf{P}$.

We then compute a meaningful change indicator, denoted by x, to highlight any departure from the in-control state as (4), and we monitor any variation in the statistical behavior of x by means of a CDT. More precisely, our desiderata are that when $\mathcal{S}$ is in-control (i.e., $t < T^*$), the values of the change indicator are stationary, i.e., $\{x(t), t < T^*\}$ are i.i.d. realizations of a unique random variable. In contrast, when $\mathcal{S}$ goes out-of-control, the distribution of the change indicator should vary. This allows us to detect structural changes by means of a suitable CDT. Detecting when $\mathcal{S}$ goes out-of-control is thus reformulated as the problem of detecting changes in the probability density function (pdf) of a random variable by monitoring a sequence of its realizations over time, i.e., the change indicators $X = \{x(\tau), \tau = L+1, \dots\}$.

The proposed change indicator is computed as follows: for each time instant $t > L$, we extract $\mathbf{s}_t$, the patch around $s(t)$, and look for the most similar patch within $\mathbf{P}$, which is centered in

$$\pi(t) = \underset{\tau=\nu+1,\dots,M-\nu-1}{\operatorname{argmin}} d(\mathbf{s}_t, \mathbf{s}_\tau). \tag{5}$$

Equation (5) defines a function $\pi(\cdot) : \mathbb{N} \to \{\nu+1, \dots, M-\nu-1\}$, which maps each time instant t to the center of the patch $\mathbf{P}$ that is the most similar to $\mathbf{s}_t$. The patch in $\mathbf{P}$ that is most similar to $\mathbf{s}_t$ is from now on denoted by $\mathbf{s}_{\pi(t)}$. The change indicator then defined as

$$x(t) = s(t) - s(\pi(t)), \tag{6}$$

namely, the difference between the values at the center of the patches $\mathbf{s}_t$ and $\mathbf{s}_{\pi(t)}$, which have been associated according to (5).

To clarify why the change indicator (6) is suitable for being monitored with a CDT, let us consider the ideal situation where any patch $\mathbf{s}_t, t < T^*$ has a perfect match and $\mathbf{s}_{\pi(t)}$ differs only because of noise. In such ideal situation, the expectation of the change indicator is zero, i.e., $E[x(t)] = 0$ and the sequence $X = \{x(\tau), \tau = L, \dots, T^* - \nu - 1\}$ contains i.i.d. realization of a random variable following an unknown pdf. In contrast, any out-of-control state affecting the self similarity (4), would introduce a bias or some form of correlation or increase the dispersion of x. In practice, structural changes of $\mathcal{S}$ would modify the pdf of the change indicators $\{x(\tau), \tau \geq T^* - \nu\}$ and, therefore, a suitable CDT is able to detect when $\mathcal{S}$ goes out-of-control, by monitoring the stationarity of X.

Unfortunately, such an ideal situation is rarely met in real-world time series: patches that are identical except of noise are rare and, even when these exist, they could eventually not be paired through (5) because of noise. Out of this ideal situation we have no warranty that $\{x(\tau), \tau = L, \dots, T^* - \nu\}$ contains i.i.d. data. Nevertheless, when time series feature self-similarity, it is reasonable to monitor the change indicator (6) by means of a CDT. In fact, the similarity between two patches is well correlated with the similarity of their central pixels (also the nonlocal-means denoising algorithm [4] relies on such a modeling assumption) and therefore out-of-control states that meet our designing assumptions (4) would affect

the distribution of x. However, since the distribution of x would not be perfectly i.i.d. during the in-control state, we expect the CDT to be less effective, and in particular, to have larger false positive rates than in the ideal situation previously described.

D. Dealing with Periodic Time Series

Often, self similarity in time series is due to the periodic or cyclic nature of $\mathcal{S}$ – as in the case of water/electricity demand. This information has not been considered so far, since in (5), the most similar patch is searched among all the training patches in $\mathbf{P}$. It follows that out-of-control states corresponding to a shift in the period of $\mathcal{S}$ would not be perceived, as these do not affect the self-similarity of patches. To detect such out-of-control states in time series that exhibits a cyclic behavior, the most similar patch has to be searched within a suitable search region rather than in the whole training set of patches. In practice, if the time series is expected to have a period ϕ – which can be either known or estimated from TS – , it is possible to replace (5) by

$$\pi(t) = \underset{\mathbf{s}_\tau \in R_{t,\phi,\delta}}{\operatorname{argmin}}\, d(\mathbf{s}_t, \mathbf{s}_\tau), \tag{7}$$

where the search region $R_{t,\phi,\delta}$ is defined as

$$R_{t,\phi,\delta} = \bigcup_{i=1}^{n} \{\tau, |t_0 + i\phi - \tau| < \delta\}, \tag{8}$$

being $n = \lfloor M/\phi \rfloor$ the number of periods in the training set, $t_0 = t - \left\lfloor \frac{t}{\phi} \right\rfloor \phi$ the point corresponding to t in the first period in the training set ($\lfloor \cdot \rfloor$ denotes the floor approximation), and $\delta \in \mathbb{N}$ a user-defined parameter. Thus, the search region in (8) consists of patches centered in windows opened around points corresponding to t with respect to the period ϕ and having width 2δ. Obviously, constraining the search region from $\mathbf{P}$ to $R_{t,\phi,\delta}$ also to reduces the computational complexity of the whole algorithm. However, in periodic time series, the choice between (5) and (7) is mostly determined by the desired sensitivity of the CDT being designed, since this choice determines the out-of-control states that can be detected. Fig. 1.b illustrates an example of search region.

E. The Algorithm

We here summarize the self-similarity based CDT for time series featuring self similarity. The CDT is detailed in Algorithm 1 independently from the specific CDT used for monitoring the distribution of change indicators, since any nonparametric CDT for scalar data can be used[2]. Most of nonparametric CDTs are configured from a sequence of change indicators thus, the initial $M > 0$ samples of TS will be used to build $\mathbf{P}$ as in (3) (line 2), while the rest of TS, i.e., $\{s(\tau), \tau = M+1, \ldots, L\}$, will be used to compute the change indicators (lines 3 - 7) to configure the CDT (line 8). Other inputs required at line 1 are the patch size ν and, when the time series shows periodic trends, the period ϕ and parameter δ that determines the size of the search regions $R_{t,\phi,\delta}$.

Disregarding whether the change indicators are computed for CDT configuration (lines 3 - 7) or during the operational life (lines 11 - 14), the algorithm proceeds as follows: at time t, the the patch $\mathbf{s}_t$ around the current data $s(t)$ is extracted (line 4, line 11), then the search region $R_{t,\phi,\delta}$ is defined (line 5, line 12) and the most similar patch $\mathbf{s}_{\pi(t)}$ is extracted (line 6, line 13). Then, the change indicator $x(t)$ easily follows from (6), (line 7, line 14).

During the operational life, i.e., when $t > L$, the CDT is then applied to the sequence of change indicators $X = \{x(\tau), \tau = M+1, \ldots, t\}$, and as soon as the CDT detects a change in X, the process $\mathcal{S}$ is considered out-of-control. Note that there is an intrinsic delay of ν samples in each detection since the whole patch $\mathbf{s}_t$ around t is required for computing $x(t)$: this motivates the delay introduced at line 9. When the monitored time series does not show periodicity, lines 5 and 12 have not to be considered, and the most similar patch $\mathbf{s}_{\pi(t)}$ is defined according to (5) rather than (7).

In our implementation[3] we monitor the sequence of change indicators X by means of the ICI-based CDT described in [24], which detects both changes in mean and variance in i.i.d. realizations of a random variable. Remarkably, the ICI-based CDT can be executed with a fixed computational complexity. Thus, since searching for the most similar patch in (5) or (7) requires a fixed number of operations, the overall computational burden per input is also fixed.

```
1-  input: {s(τ), τ = 1,...,L}, ν, δ, φ, M
2-  define P from TS = {s(τ), τ = 1,...,M} as in (3),
3-  for (t = M + 1; t ≤ L; t++) do
4-  |   extract the patch s_t as in (1),
5-  |   define the search region R_{t,φ,δ} as in (8),
6-  |   compute the patch most similar to s_t in R_{t,φ,δ}, (7)
7-  |   compute the change indicator x(t) as in (6),
    end
8-  configure the CDT on {x(t), t = M + 1,...,L},
9-  wait for the next ν samples,
10- while (s(t + ν) arrives) do
11- |   extract the patch s_t as in (1),
12- |   define the search region R_{t,φ,δ} as in (8),
13- |   compute the patch most similar to s_t in R_{t,φ,δ}, (7)
14- |   compute the change indicator x(t) as in (6),
15- |   if (CDT({x(τ), τ = M,...,t}) == 1) then
16- |   |   detect a structural change in S at T̂ = t.
17- |   |   return.
    |   end
18- |   t = t + 1;
    end
```

Algorithm 1: Self-similarity based CDT for time series.

We remark that parameters of the specific CDT used at

[2]We consider nonparametric CDTs since these do not assume any apriori information about the distribution of the monitored data before or after the change.

[3]Codes of the self-similarity based CDT are available for download at `http://home.deib.polimi.it/boracchi/Projects/projects.html`

line 15 are also required as input, though not reported in Algorithm 1. In case of nonparametric CDTs that do not require any explicit configuration phase, such as [27], [31], it is possible use the whole training sequence to build $\mathbf{P}$, setting $M = L$.

IV. Experiments

A. Dataset Description

The CDT in Algorithm 1 has been tested on a dataset of 10 time series containing flow measurements coming from different DMAs of Barcelona. In each DMA the flow is measured every 10 minutes, 144 times per day. Each time series lasts 82 days. Fig. 1 shows the highly self-similar behaviour of the time series induced by the periodicity of citizens' habits. We manipulated these sequences to reproduce the following structural changes starting at day 41:

Offset: an offset o is added to the time series after T^*, i.e.

$$s(t) = s(t) + o\,, \quad t > T^*.$$

Such an additive offset models the effects of a leak in the pipes or junctions, which increases the flow measurements at each time instant. We considered an offset of both 0.5 and 0.25 of the mean flow during the in-control state. Examples to illustrate the effects of an offset on time series are reported in Fig. 3.a and Fig 4.a.

Sensor degradation: a zero-mean Gaussian random variable is added to the time series after T^*, i.e.,

$$s(t) = s(t) + \eta(t)\,, \quad t > T^*,\ \eta \sim \mathcal{N}(0, \sigma^2),$$

modeling the effects of a degradation of the flow sensor. In our experiments, the noise standard deviation σ is set to 0.5 or 0.25 the average flow during in the control state.

Source Change: two different time series are juxtaposed at time T^*

$$s(t) = s_1(t)\,, \quad t > T^*,$$

being s_1 data from a time series recorded in a different DMA. Each DMA is characterized by its own daily profile, thus, such a source change breaks the self similarity in the time series. An example of time series affected by a source change is reported in Fig. 5.a.

Stack-at: The measurements become constant after T^*, i.e.,

$$s(t) = s(T^*)\,, \quad t > T^*,$$

which is a situation that could be induced by a sensor or communication fault.

B. Considered CDTs

The self-similarity based CDT is contrasted against both a residual-based and a template-based approach. The residual-based approach consists in computing the residual between the observation $s(t)$ and the output of a nonlinear ARX predictive model f_{NARX} trained on the initial M samples of the time series [32]. The residuals are computed as

$$r(t) = s(t) - f_{\text{NARX}}(s(t), s(t-1), \ldots, s(t-k)), \qquad (9)$$

where k is the order of the autoregressive component (the eXogenous component of the NARX predictive model is not considered since we are dealing with time series).

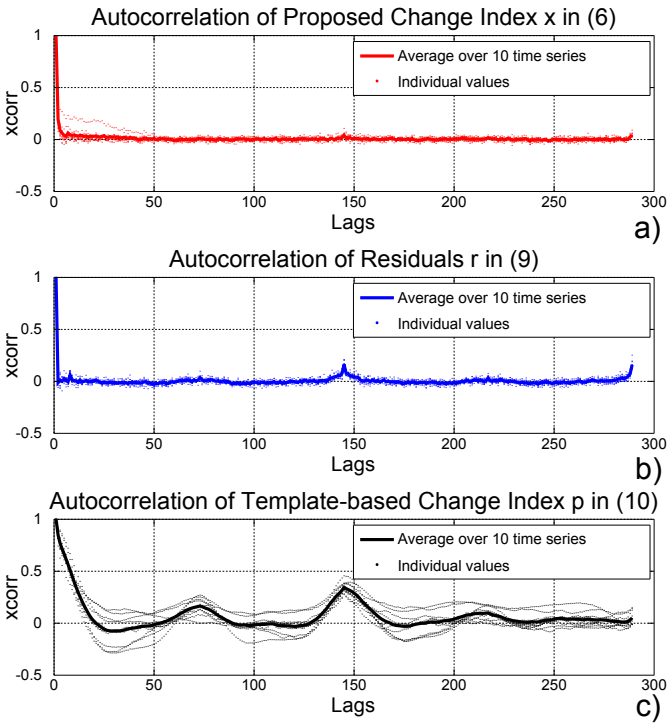

Fig. 2. Autocorrelation of the change indicator x in (6), the residuals in (9) and the template-based change indicator in (10). The autocorrelation was computed considering only change indicators before the structural change, and both the average over the 10 time series and the value on each time series have been reported. Peaks at indicates that the change indicators are not white, and this is particularly evident at lags equal to 144 (which corresponds to the period).

The template-based approach represents the most straightforward way to cope with periodic patterns corrupted by additive white noise. We estimate the average flow profile by averaging the flow measurements recorded at the same time over different days: this corresponds to the template of the daily water flow. During the operational life, the template is used as an estimate of the expected flow measurements. The template-based change indicator is defined as

$$p(t) = s(t) - \frac{1}{n}\sum_{i=1}^{n} s(t_0 + i\phi) \qquad (10)$$

being, as in (8), $n = \lfloor M/\phi \rfloor$ and $t_0 = t - \left\lfloor \frac{t}{\phi} \right\rfloor \phi$. In (10), the term $\frac{1}{n}\sum_{i=1}^{n} s(t_0 + i\phi)$ represents the value of the template estimated from the training set.

The ICI-based CDT is then applied to sequences of x in (6), r in (9) and p in (10), to fairly compare the detectability of the structural changes from these different change indicators at time t.

Configuration: The patch training set $\mathbf{P}$, the NARX predictive model, and the average daily profile have been estimated from the first two weeks of each time series, i.e., $M = 2016$. The ICI-based CDTs were then configured over

	Self-Similarity based			Residuals-based			Template-based		
	FPR	**FNR**	**DD**	**FPR**	**FNR**	**DD**	**FPR**	**FNR**	**DD**
offset 0.5	0.1	0.0	156.4	0.0	0.0	1554.0	0.2	0.0	332.0
offset 0.25	0.1	0.0	914.2	0.0	0.3	2803.4	0.2	0.0	747.0
sensor degradation 0.5	0.1	0.0	174.2	0.0	0.0	170.0	0.2	0.0	269.5
sensor degradation 0.25	0.1	0.0	336.4	0.0	0.0	288.0	0.2	0.0	652.0
source change	0.1	0.0	103.1	0.0	0.0	800.0	0.2	0.0	219.5
stack-at	0.1	0.0	169.8	0.0	0.0	160.0	0.2	0.0	534.5

TABLE I

DETECTION PERFORMANCE OF THE THREE CONSIDERED SOLUTIONS.

400 change indicators, thus $L = M + 400 = 2416$. The true change-time was $T^* = 5904$[4]. All the ICI-based CDT were configured by setting $\Gamma = 2.5$; further details about the ICI-based CDT, can be found in [24]. The change indicator x was computed by setting $\nu = 5$, $\delta = 5$ and the period $\phi = 144$, while in (9) we considered a wavelet network as nonlinear estimator of the NARX predictive model, where k has been experimentally fixed to 7.

C. Figures of Merit

The detection performance is assessed by means of the following figures of merit:

- False Positive Rate (FPR), the percentage of time series where a structural change was erroneously detected when $\mathcal{S}$ is in-control, i.e., $\widehat{T} < T^*$.
- False Negative Rate (FNR), the percentage of time series where an actual structural change in $\mathcal{S}$ was not detected.
- Detection Delay (DD), the average of $\widehat{T} - T^*$ computed on runs where $\widehat{T} \geq T^*$;

D. Discussion

The detection performance of the proposed CDT relies on the ability to correctly exploit self-similarity of the time series: namely, the sequence of change indicators has to be i.i.d., during the in-control conditions, and its statistical behavior should be modified by any structural change. To assess whether $x(\cdot)$, $p(\cdot)$ and $r(\cdot)$ are suited for monitoring in-control, self-similar processes, we compute the autocorrelation of these when $t < T^*$. The autocorrelation of the proposed change detection index x (Fig 2.a) shows that it is very close to be i.i.d.. This means that we are operating fairly close to the ideal conditions for CDTs, as discussed in Section III-C. This result justifies the use of a CDT designed to work with i.i.d. data, such as the ICI-based CDT. The analysis of the autocorrelation of $r(t)$, that is depicted in Fig. 2.b, shows a peak around lag 144. This is not surprising since it corresponds to the period of the daily-cycle. This result indicates that the NARX predictive model is not effective in capturing the time-series dynamic. Finally, the autocorrelation of $p(t)$ (Fig 2.c) shows highly correlated values over time. Even in this case, there is a peak around 144, indicating that template-based change indicator does not properly model the time series.

These results are well reflected by the detection performance of the three considered CDTs reported in Table I. Remarkably, the proposed change indicator x yields detections in the offset and source change scenarios that are far prompter than those of the residual-based and the template-based approaches. The residual-based approach provides slightly better results in the sensor degradation and stuck-at scenario. The template-based CDT is less effective than the other solutions and this is due to the high-correlation of the change indicator $p(t)$, indicating that the assumption of strict periodicity does not hold for these time series.

Figure 3-5 show three examples of considered time series together with the corresponding change indicators and detections results. The statistical behaviour of x is particularly interesting since it clearly behaves like an i.i.d. random variable before the change. It is also worth noting that after T^*, the values of x show a very different behavior, indicating that self-similarity was successfully exploited for detecting structural changes. As expected, $r(t)$ exhibits a similar behaviour, while $p(t)$ clearly shows autocorrelated values (as also highlighted with the analysis of Figure 2).

Note that each CDT is characterized by constant FPR over different application scenarios, since false positives do not depend on the specific structural change introduced. Fig 3 reports the time series where the self-similarity based CDT had a false positive detection and, even though no change was artificially introduced, it clearly emerges an anomalous behavior of the flow measurements around sample 4500.

V. CONCLUSIONS AND ONGOING WORKS

We introduced the self-similarity based CDT, a novel solution for detecting structural changes in time series by analyzing their of self-similarity. This CDT computes a change indicator that quantitatively assess the similarity between patches in the recent data and patches from the initial training set. Changes in the self-similarity of the time series are detected by monitoring such change indicator over time, by means of a non-parameteric CDT. The proposed solution revealed to be very effective in detecting structural

[4]This specific choice of T^* was made to maximize both the number and the length of the time series extracted from the Barcelona water distribution network dataset.

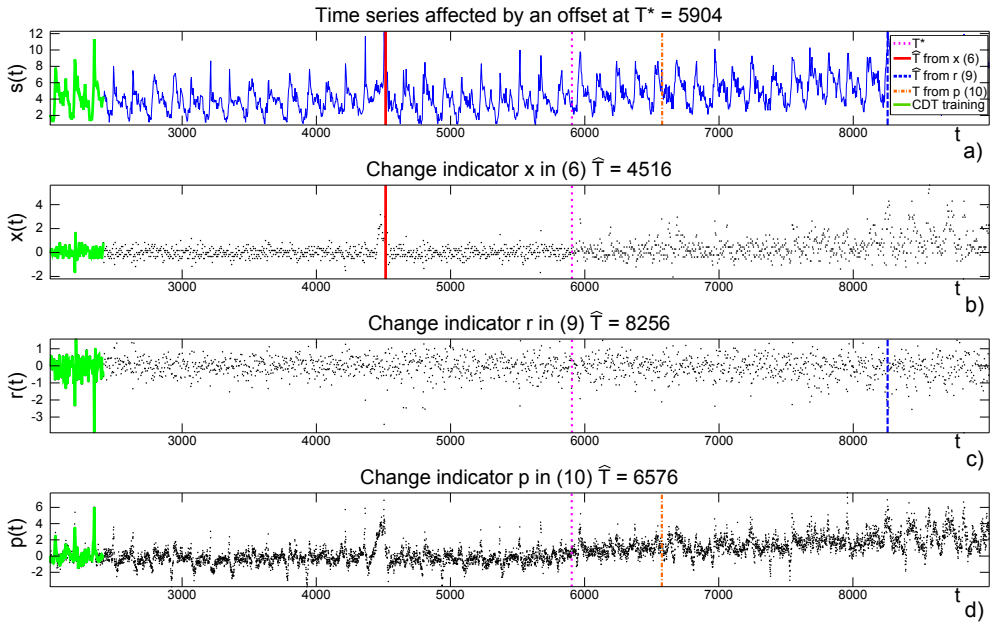

Fig. 3. An example of structural change inducing an offset of +0.25 times the average flow during in-control conditions

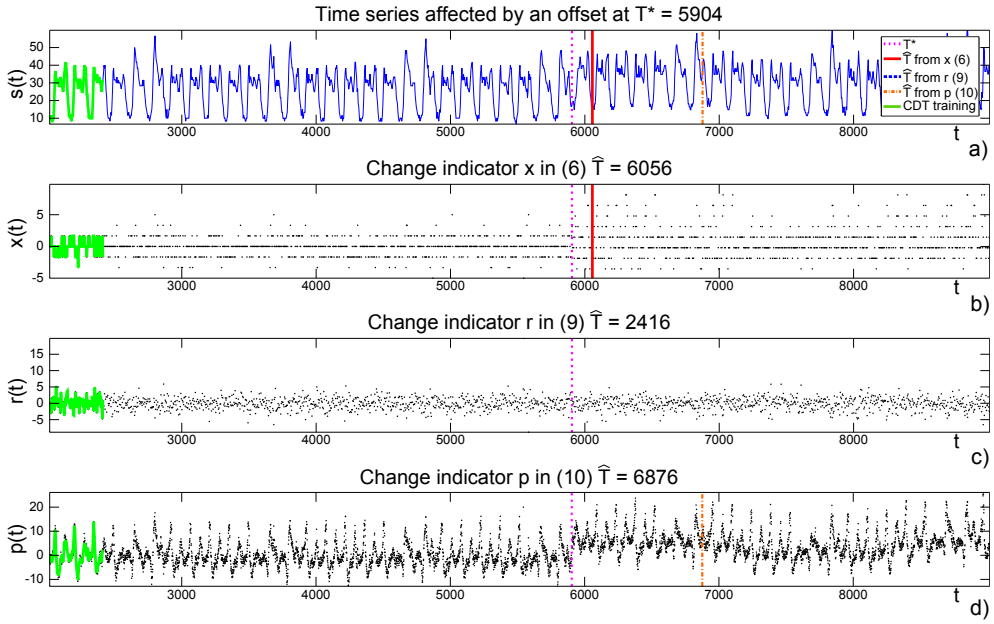

Fig. 4. An example of structural change inducing an offset of +0.5 times the average flow during in-control conditions

changes in the time series of flow measurements coming from the Barcelona water distribution network.

Ongoing works concern the investigation of different change indicators, the analysis of both the impact of the search region and the patch size on the CDT performance, as well as the use of self-similarity to perform change detection in application scenarios where time series exhibits self-similarity though not periodicity.

ACKNOWLEDGEMENT

The authors would like to thank Prof. Vicenç Puig from Universitat Politècnica de Catalunya, Barcelona, Spain, for providing us the flow dataset and for the meaningful discussions.

REFERENCES

[1] M. F. Barnsley, *Fractals everywhere*, 2nd ed. Academic Press, 1993.

[2] A. Jacquin, "Image coding based on a fractal theory of iterated contractive image transformations," *IEEE Transactions on Image Processing*, vol. 1, no. 1, pp. 18–30, 1992.

[3] A. Efros and T. Leung, "Texture synthesis by non-parametric sampling," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 2, 1999, pp. 1033–1038 vol.2.

[4] A. Buades, B. Coll, and J. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling Simulation*, vol. 4, no. 2, p. 490, 2005.

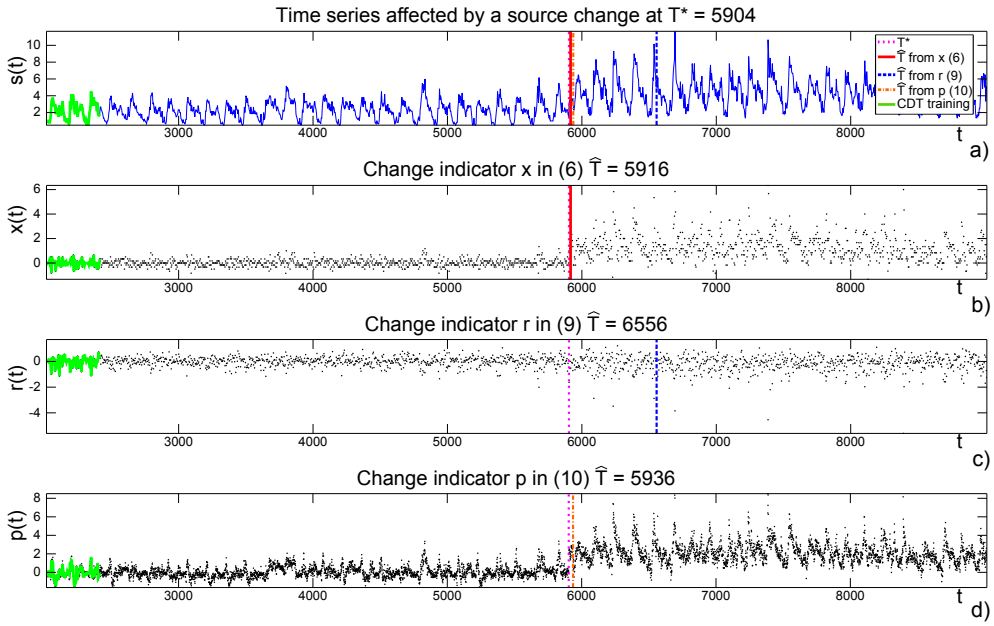

Fig. 5. An example of source change.

[5] C. Kervrann and J. Boulanger, "Optimal spatial adaptation for patch-based image denoising," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 2866–2878, October 2006.

[6] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, August 2007.

[7] A. Buades, B. Coll, and J.-M. Morel, "Nonlocal image and movie denoising," *Int. J. Comput. Vision*, vol. 76, no. 2, pp. 123–139, Feb. 2008. [Online]. Available: http://dx.doi.org/10.1007/s11263-007-0052-1

[8] M. Maggioni, G. Boracchi, A. Foi, and K. Egiazarian, "Video denoising, deblocking and enhancement through separable 4-d nonlocal spatiotemporal transforms," *IEEE Transactions on Image Processing*, vol. 21, no. 9, pp. 3952– 3966, September 2012.

[9] B. Dong, J. Ye, S. Osher, and I. Dinov, "Level Set Based Nonlocal Surface Restoration," *Multiscale Modeling & Simulation*, vol. 7, no. 2, pp. 589–598, 2008. [Online]. Available: http://scitation.aip.org/getabs/servlet/GetabsServlet?prog=normal\&id=MMSUBT000007000002000589000001\&idtype=cvips\&gifs=yes

[10] G. Rosman, A. Dubrovina, and R. Kimmel, "Patch-collaborative spectral point-cloud denoising," *Computer Graphics Forum*, pp. n/a–n/a, 2013. [Online]. Available: http://dx.doi.org/10.1111/cgf.12139

[11] A. Buades, B. Coll, and J. M. Morel, "Image denoising methods. a new nonlocal principle," *SIAM Rev.*, vol. 52, no. 1, pp. 113–147, Feb. 2010. [Online]. Available: http://dx.doi.org/10.1137/090773908

[12] B. Chiu, E. Keogh, and S. Lonardi, "Probabilistic discovery of time series motifs," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2003, pp. 493–498.

[13] T. Warren Liao, "Clustering of time series dataa survey," *Pattern Recognition*, vol. 38, no. 11, pp. 1857–1874, 2005.

[14] D. Dasgupta and F. González, "An immunity-based technique to characterize intrusions in computer networks," *Evolutionary Computation, IEEE Transactions on*, vol. 6, no. 3, pp. 281–291, 2002.

[15] S. Forrest, A. S. Perelson, L. Allen, and R. Cherukuri, "Self-nonself discrimination in a computer," in *Research in Security and Privacy, 1994. Proceedings., 1994 IEEE Computer Society Symposium on*. Ieee, 1994, pp. 202–212.

[16] E. Keogh, J. Lin, and A. Fu, "Hot sax: Efficiently finding the most unusual time series subsequence," in *Data mining, fifth IEEE international conference on*. IEEE, 2005, pp. 8–pp.

[17] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 15:1–15:58, Jul. 2009. [Online]. Available: http://doi.acm.org/10.1145/1541880.1541882

[18] M. A. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, "A review of novelty detection," *Signal Processing*, no. 0, pp. –, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S016516841300515X

[19] J. Gertler, *Fault detection and diagnosis in engineering systems*. CRC press, 1998.

[20] L. H. Chiang, R. D. Braatz, and E. L. Russell, *Fault detection and diagnosis in industrial systems*. Springer, 2001.

[21] R. Isermann, *Fault-diagnosis systems: an introduction from fault detection to fault tolerance*. Springer Verlag, 2006.

[22] F. Gustafsson, *Adaptive Filtering and Change Detection*. Wiley, Oct. 2000. [Online]. Available: http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471492876,descCd-description.html

[23] M. Basseville and I. V. Nikiforov, *Detection of abrupt changes: theory and application*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[24] C. Alippi, G. Boracchi, and M. Roveri, "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.

[25] ——, "A hierarchical, nonparametric, sequential change-detection test," in *Proceedings of IJCNN 2011*, 31 2011-aug. 5 2011, pp. 2889 –2896.

[26] A. Tartakovsky and H. Kim, "Performance of certain decentralized distributed change detection procedures," in *Information Fusion, 2006 9th International Conference on*, july 2006, pp. 1 –8.

[27] 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.

[28] F. A. González and D. Dasgupta, "Anomaly detection using real-valued negative selection," *Genetic Programming and Evolvable Machines*, vol. 4, no. 4, pp. 383–403, 2003.

[29] J. Gómez, F. González, and D. Dasgupta, "An immuno-fuzzy approach to anomaly detection," in *Fuzzy Systems, 2003. FUZZ'03. The 12th IEEE International Conference on*, vol. 2. IEEE, 2003, pp. 1219–1224.

[30] D. W. Taylor and D. W. Corne, "An investigation of the negative selection algorithm for fault detection in refrigeration systems," in *Artificial Immune Systems*. Springer, 2003, pp. 34–45.

[31] G. Ross and N. M. Adams, "Two nonparametric control charts for detecting arbitrary distribution changes," *Journal of Quality Technology*, vol. Vol 44, No. 22, pp. 102–116, 2012.

[32] L. Ljung, *System identification*. Wiley Online Library, 1999.