A cognitive monitoring system for detecting and isolating contaminants and faults in intelligent buildings

Giacomo Boracchi, Michalis Michaelides, Manuel Roveri

Abstract—**Intelligent buildings are typically endowed with sensing devices that are able to measure the concentration of specific contaminants in relevant zones. The collected measurements are subsequently processed by intelligent algorithms in order to enable the prompt detection and isolation of contaminant sources inside the building. Unfortunately, in real-world conditions, these sensing devices may suffer from faults affecting the sensors or the embedded electronics. Such faults, generally result in perturbed or missed data in the acquired data-stream, that can induce false alarms (or possibly missed alarms) and compromise the contaminant detection and isolation ability. This paper proposes a three-layer cognitive monitoring system for the detection and isolation of both contaminants and sensor faults in intelligent buildings. The first two layers are designed for the prompt detection of small variations in the concentration of a specific contaminant, while reducing the possible occurrence of false alarms. At the third layer, a cognitive mechanism employing a propagation model for the contaminant, which is based on the airflows between the building zones, allows to isolate the source zone and discriminate between sensor faults and the presence of a contaminant source. The proposed method is validated using a realistic 14-zone building scenario.**

I. Introduction

Advances in sensing devices and embedded systems are transforming our homes and work environments into *intelligent buildings*. These systems have the ability to adapt and control the building environment in order to save energy and create more comfortable, healthy and safe living conditions for their occupants [1], [2]. The safety of the occupants is directly associated with the Indoor Air Quality which can be easily compromised by an accident (e.g., CO leakage from a faulty furnace) or a terrorist attack. In particular, recent terrorist events and alarms to potential hostile attacks with airborne Chemical and Biological Agents (CBA) have created a crucial and world-wide concern for building and occupant safety, e.g. see [3]. Most CBAs are highly poisonous and a small amount of CBA can cause morbidity and mortality. Under these safety-critical conditions, data collected in real-time from sensors monitoring the concentration of a CBA can be used to alert the occupants and determine appropriate solutions like indicating safe spaces, or isolating and cleaning contaminated

Giacomo Boracchi and Manuel Roveri are with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan, Italy, 20133 IT e-mail: {giacomo.boracchi,manuel.roveri}@polimi.it.

Michalis Michaelides is with the Department of Electrical Engineering, Computer Engineering and Informatics, Cyprus University of Technology, 30 Archbishop Kyprianos Str., CY-3036 Lemesos, Cyprus (e-mail: {michalis.michaelides}@cut.ac.cy).

areas. Therefore, the prompt detection and accurate isolation of contaminant sources is an essential task in the design of intelligent buildings.

The development of near real-time biological and chemical agent sensors (see [3], [4] and references therein) has recently enabled effective protective measures against such CBA threats. These measures can be *low-disruption actions* in response to the sensors' readings, e.g., automatically modifying the operational mode of a building's Heating Ventilation and Air Conditioning (HVAC) system, or *high-disruptive actions*, e.g., the building's complete evacuation. Highlights of this new sensor technology include inexpensive, moderately sensitive remote and point-trigger sensors and rapid identifiers which can be exploited for quickly covering wide areas. At their current state, however, these trigger sensors can only support low-disruption actions because they can suffer from high false positive rates [3] (i.e., false alarms induced by noise causing the incorrect detection of the contaminant). Note that frequent false positives can make the protection system useless because of the reluctance of the occupants to cooperate with the required protective actions (cry-wolf effect).

In the real world, data gathered by sensors are corrupted by noise. In addition, sensors and the embedded electronics can suffer from a wide range of faults. These faults can induce incorrect or missed data in the acquired data-stream, which can heavily affect the contaminant detection and isolation abilities of the monitoring system (inducing either missed or false alarms). Moreover, the correct operation of the monitoring system relies on the ability to discriminate between sensor faults and the real presence of a contaminant source.

In this paper we propose a novel cognitive monitoring system that can be combined with real-time point-trigger sensors for performing contaminant and sensor-fault diagnosis, namely detection and isolation, in intelligent buildings. The proposed system reduces the occurrence of false alarms (and of consequent disruptive actions) by means of a three-layered architecture for detection and isolation and employs cognitive mechanisms to discriminate sensor faults from the presence of a real contaminant source in the building. At the first layer, we rely on Change-Detection Tests (CDTs) for promptly detecting small amounts of the monitored contaminant. CDTs are statistical techniques able to sequentially analyse the gathered contaminant measurements to inspect for changes in their distribution. The second layer performs a validation of the changes detected at the first layer by means of Change-Point Methods (CPMs) to reduce the possible occurrence of

false alarms. These CPMs are statistical hypothesis tests that, operating on a fixed length sequence, are able to confirm (or not) the presence of a change-point in the sequence (by also providing an estimate of the change-point position in the data sequence). Finally, the third (cognitive) layer aims at isolating the zone where the contaminant source has been inserted or the sensor fault has occurred. In addition, it identifies whether the detection can be associated to the true presence of a contaminant source or to a fault affecting a sensor. This is achieved by taking into account the expected propagation of the contaminant within the intelligent building by considering the building topology and the direction of the airflows.

In summary, the proposed three-layer cognitive monitoring system constitutes the main contribution of this work and offers the following advantages:

- change-detection tests at the first layer that are able to promptly detect the presence of a specific contaminant, even at small concentrations;
- change-point methods at the second layer that validate each detection raised by the first layer to reduce false alarms;
- a cognitive layer on top of the architecture able to isolate the zone of the building in which either the contaminant has been released or the sensor fault has occurred and discriminate between the two.

The paper extends our preliminary results on contaminant detection in smart buildings [5] by introducing the cognitive layer, which analyses in an aggregated manner sensor measurements in different zones of the building, to discriminate between sensor faults and contaminants, as well as isolate the zone in which the contaminant source has been inserted (or the sensor fault has occurred).

The paper is organized as follows: Section II presents the related literature and Section III formulates the addressed problem. The proposed cognitive monitoring system is detailed in Section IV, while experiments are presented in V and, finally, conclusions are drawn in Section VI.

II. RELATED LITERATURE

A. Fault and Contaminant Diagnosis in Buildings

During the last four decades, a lot of important results have been achieved in the area of fault diagnosis for a variety of engineering systems. In a recent two-part survey [6], [7], fault-diagnosis methods and their applications are comprehensively reviewed and classified as model-based, signal-based, knowledge-based, hybrid (which combine at least two methods) and active (where a suitably designed input signal is injected in order to enhance detectability of potential faults). In another recent survey [8], the various fault diagnosis methods are reviewed from a data-driven perspective based on the type of data and how these are processed. The proposed cognitive monitoring system can be considered a hybrid method as it combines signal-based methods with a qualitative model of the building airflows. In fact, the statistical tools employed at the first and second layers (CDTs and CPMs, respectively) are signal-based methods that analyze the stream of contaminant measures in the various building zones, while the qualitative,

physics-based model of the building airflows is used in the cognitive layer for differentiating between the presence of a contaminant event and a sensor fault. With respect to fault diagnosis in building scenarios, a two-part survey [9], [10] reviews methods for fault detection, diagnostics and prognostics with particular emphasis on Heating, Ventilation, Air-Conditioning and Refrigeration (HVAC & R) systems.

The problem of contaminant event monitoring (or contaminant diagnosis) has received considerable attention in the literature over the last decade. A detailed report on the literature related to the inverse tracking of pollutants in both groundwater and air fields is presented in [11]. Two methods that have been successfully applied to isolate contaminant sources in building environments are the Bayesian updating method [12] and the adjoint probability method [13]. However, both methods require some form of prior knowledge about the considered scenario, either in the form of a constructed scenario database before the event or concerning one of the source characteristics during the event (location or the time of contaminant release). More recently, the state space method [14] has been proposed for contaminant event monitoring. In the developed multi-zone formulation, the presence of a contaminant source is modeled as a fault in the process, which enables the contaminant detection and isolation by means of advanced fault-diagnosis tools. This method is appropriate for situations where a model of the building airflows is available and bounds on the modeling uncertainty can be calculated. In principle, this method can completely eliminate the presence of false alarms in the system; the main challenge involved is the design of tight, adaptive thresholds for bounding the modeling uncertainty. We emphasize that the use of conservative thresholds can lead to missed detections, especially under low signal-to-noise conditions. Note that this, as well as other fault diagnosis methods involving state observers and Kalman Filters are not directly comparable to this work as they are model-based, requiring the existence of a state-space or input-output model of the system.

Differently, [4] suggests the design of contaminant detection systems based on the Scalar Trigger Algorithm (STA). In such systems, a detection threshold is dynamically adapted to compensate for the effects of measurement noise in order to guarantee a pre-specified false alarm probability when applied to a fixed sequence of data. Of course, reducing false alarms can be done at the expense of increasing detection delay and false negative rate (i.e., the percentage of missed alarms when the contaminant is present), especially in situations characterized by high noise or low contaminant concentration.

Compared to existing methods in the literature, the solution here proposed does not require a detailed model of the building airflows, bounds on modeling uncertainty or any prior information about the source characteristics (onset time, location and generation rate). It employs a layered architecture to reduce the presence of false alarms and, at the same time, it ensures the detection of even small quantities of contaminants inside the building thanks to the sequential monitoring technique adopted. At the cognitive layer, the proposed system constructs a conceptual model based on the wind direction and the topological relations between the different zones of the building to

isolate the source zone (without requiring a priori knowledge of the actual values of the building airflows) and discriminate between sensor faults and the presence of a contaminant source in the monitored area. To the best of our knowledge, this is the first method proposed in the literature that jointly addresses the problem of contaminant and sensor-fault diagnosis (detection and isolation) in intelligent buildings. It is also worth pointing out that the algorithms and results of this paper, although custom-designed for the specific application of intelligent buildings, may find applicability to many other domains. In fact, the issue of differentiating between process and sensor faults in a system remains an important challenge for the fault diagnosis community.

B. Contaminant Transport Simulation Software

The evaluation and testing of contaminant monitoring systems require the release of contaminants in the building interior which is often prohibited due to safety regulations or difficult to implement in practice. Due to the difficulty in obtaining real datasets, researchers often resort to simulation tools. For the indoor air and contaminant transport simulation there are two main modeling approaches: Computational Fluid Dynamics (CFD) and multi-zone modeling. On the one hand, CFD modeling involves the numerical solution of the conservation equations of mass, momentum, energy and species concentrations by dividing the space into a finite number of discrete cells and then by using an iterative procedure to achieve a converged solution. This approach can provide the spatial distributions and temporal evolutions of air pressure, velocity, temperature, humidity, contaminants, and turbulence intensity. However, the degree of accuracy comes at the expense of high computational cost. On the other hand, multi-zone models provide a computational efficient solution by representing a building as a network of well-mixed zones. Temperature, humidity, air velocity and pollutant concentration are assumed uniform within each zone. Different zones are connected by discrete flow paths such as doors, windows, wall cracks, ducts and hallways. In this multi-zone model, a zone may correspond to an entire room or to a part of it. The model predicts the flow parameters based on mass conservation and component interaction. The most popular multi-zone simulation programs are COMIS [15] by Lawrence Berkeley National Laboratory (LBNL) and CONTAM [16] by the US National Institute of Standards and Technology (NIST). In this paper, we rely on the Matlab-CONTAM Toolbox described in [17] for creating the datasets used for the evaluation of our proposed algorithms in a realistic 14-zone building scenario.

In particular, we consider the Holmes house as a building case study [18], whose layout is shown in Figure 1(a). Such a well-known case study comprises of 14 zones: a garage (Z1), a storage room (Z2), a utility room (Z3), a living room (Z4), a kitchen (Z5), two bathrooms (Z6 and Z13), a corridor (Z8), three bedrooms (Z7, Z9 and Z14) and three closets (Z10, Z11 and Z12); as well as 30 leakage path openings corresponding to windows and doors (P1–P30).

III. THE PROBLEM FORMULATION

In this section we formulate the addressed problem, introduce the models we employ and state our assumptions. We conclude with the problem statement.

A. Sensor Measurement and Fault Model

Let us consider an intelligent building composed of N zones. Each zone is equipped with a sensor measuring the concentration of a specific contaminant. Let $m_i(t)$ (with $i = 1, \ldots, N$) denote the measurement provided by the sensor of the i-th zone at time t, which can be modelled as

$$c_i(t) = x_i(t) + \Delta_i(t), \tag{1}$$
$$m_i(t) = \phi(c_i(t)) + \eta_i(t), \tag{2}$$

where c_i (1) describes the true contaminant concentration inside the i-th zone, while (2) models the output of the corresponding i-th sensor that can be affected by noise and possibly faults. More specifically, $c_i(t)$ represents the true concentration of the contaminant at time t, which is the sum of the natural concentration $x_i(t)$, and the additional concentration $\Delta_i(t)$ of the anomalous source (i.e., what needs to be detected by the sensing system). Note that $x_i(t)$ can be zero when no contaminant is naturally present in the i-th zone (e.g., for toxic gases or CO), or constant (e.g. for CO_2 under steady state airflow conditions) or may follow a dynamic behaviour like $x_i(t) = f(x_i(t-1), x_i(t-2), ...)$ in the more general case. In this paper, we assume $x_i(t) = 0$, but the proposed monitoring system can be straightforwardly applied to scenarios where $x_i(t) = \zeta > 0$. The more general case where the contaminant follows a dynamic behavior requires some prediction mechanism like the one described in [19].

Next, in (2), $\phi(\cdot)$ represents the (possibly nonlinear) fault function and $\eta_i(t)$ is the independent and identically distributed (i.i.d.) random noise that affects the measurement of the sensor in the i-th zone at time t. The fault function $\phi(\cdot)$ is characterized by a time profile of the fault and a fault signature. With respect to the time profile, the fault can be classified based on the time duration (as permanent, transient or intermittent) and the evolution mode of appearance and/or disappearance (as abrupt or incipient). Furthermore, based on the signature, we can have different types of faults including offset, drift or precision degradation. In our experiments we consider permanent, abrupt and offset sensor-faults. In other words, the fault function is given by the following equation,

$$\phi(c_i(t)) = \begin{cases} c_i(t), & t < \tau \\ c_i(t) + \delta, & t \geq \tau \end{cases} \tag{3}$$

where τ represents the onset time of the fault while $\delta > 0$ is the offset amount.

Let us denote by τ also the time instant when a contaminant source is inserted into the building and let i^* (with $1 \leq i^* \leq N$) be the zone where this source is inserted, i.e., the *source zone*. Let τ_i be the time instant in which the contaminant concentration appears in the i-th zone, i.e.,

$$\begin{cases} \Delta_i(t) = 0, & t < \tau_i \\ \Delta_i(t) > 0, & t \geq \tau_i \end{cases}. \tag{4}$$

Note that, because of propagation delays, $\tau_i > \tau$ when $i \neq i^*$.

B. Contaminant-Propagation Model

The contaminant propagation through the building zones depends on a number of factors affecting the internal airflows including (i) the building structure (e.g., the interconnections of the various zones through doors and openings), (ii) environmental conditions (e.g., temperature, wind direction and velocity), (iii) HVAC operational mode (or any other type of fan causing a forced flow).

In particular, the direction of the wind influences the direction of the flows among the different zones within the building, i.e., zone i is connected to zone j if there exists an airflow from i to j. Note that this connection is not symmetric. Interestingly, wind is not the only source of flows between zones since such a physical phenomenon could also be induced by the thermal gradient (i.e., the difference in temperature between two zones).

The relationships among the measurements in different zones of the building at a given time instant, which are induced by the airflows, are represented by the *flow matrix*

$$Z(t) = \begin{bmatrix} z_{1,1}(t) & \cdots & z_{1,N}(t) \\ \vdots & \ddots & \vdots \\ z_{N,1}(t) & \cdots & z_{N,N}(t) \end{bmatrix}, \qquad (5)$$

where $z_{i,j}(t)$ equals to 1 if there is flow at time t from the j-th zone to the i-th zone and 0 otherwise; while $z_{i,i}(t) = 0, i = 1, \ldots, N$.

We assume that updated estimates of $Z(t)$ are regularly provided, as these depend on the wind direction and velocity, the HVAC operating mode, and the opening status of the doors and windows: information that is typically available in smart buildings. In general, the flow matrix $Z(t)$ as specified in (5) is time-dependent. However, under steady-state airflow conditions, the time dependence can be dropped (i.e., $Z(t) = Z$) as the flows remain constant between the various building zones. In what follows, we will be assuming steady-state airflow conditions. This simplifying assumption is well justified under forced-ventilation, where the flows are dictated by the HVAC and no big variations are expected in the flows.

Under natural-ventilation conditions, the airflows can often change as they depend on the wind direction and velocity. Even for this case, however, it is possible to assume the system is in a near-steady-state condition for a short time, and thus break-up the monitoring problem in short time intervals and apply the proposed solution in each of these intervals. It is also worth pointing out that in the proposed formulation we do not use the actual values of the flows, but only the existence of a flow between pairs of zones. So, as long as the flow directions remain the same, the proposed solution is robust to non steady-state conditions. When the flow directions do change (e.g. when a door opens), the flow matrix has to be re-estimated inside the cognitive layer of the monitoring system.

As an example, consider the Holmes house case study (presented in Section II-B) with wind coming from the North, as shown in Figure 1(b). Note that the induced airflows between the various zones are portrayed with green lines with
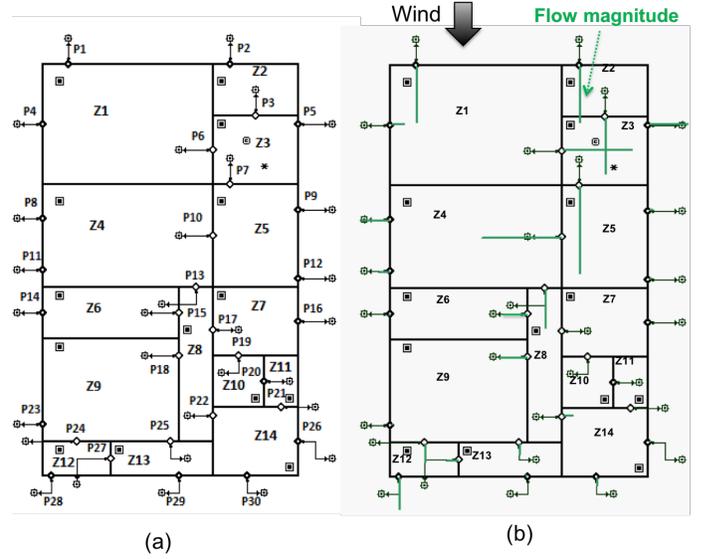

Fig. 1. The reference intelligent building case study. (a) The Holmes house. It comprises of 14 zones (Z1-Z14) and 30 leakage path openings (P1-P30); (b) The induced airflows (i.e., the green lines) between zones of the house when wind blows from North to South (the length of the lines is proportional to the flow magnitude).

	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10	Z11	Z12	Z13	Z14
Z1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Z2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Z3	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Z4	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Z5	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Z6	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Z7	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Z8	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Z9	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Z10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Z11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Z12	0	0	0	0	0	0	0	0	1	0	0	0	1	0
Z13	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Z14	0	0	0	0	0	0	0	1	0	0	0	0	0	0

Fig. 2. The flow matrix Z of the reference building case study shown in Figure 1. Here, we are assuming steady airflow conditions and time dependence has been dropped.

the length of the lines indicating the flow magnitude. The corresponding flow matrix Z for the specific example is shown in Figure 2, where flow conditions are assumed to be steady (hence, time dependence has been dropped).

C. Problem Statement

At time τ, either a single contaminant source or a single sensor fault is introduced in the building zone i^*. The aim of the proposed monitoring system is to promptly detect and accurately diagnose the specific event based on the sensor measurements. This involves three distinct phases: (i) *detection*

1- *Each unit:* Train the ICI-based CDT from $m_i(t), t = 1, \ldots, L$;
 while *(1)* **do**
2-   *Each unit*: acquire $m_i(t)$;
3-   *Each unit*: run the ICI-based CDT on $m_i(t)$
4-   *At unit i:*
5-   **if** $CDT(m_i(t)) == 1$ **then**
6- $\widehat{i} = i$;
7- $\widehat{T} = t$;
8-   **if** $CPM(\{m_i(t), t = 1, \ldots, \widehat{T}\}) == 1$ **then**
9-   Run Cognitive Layer:
10-   Compute A_D and B_D as described in Section IV-C;
11-   Compute W_{A_D} and W_{B_D} with Eq. (12) and (15) ;
12-   **if** $CPM(W_{A_D}) == 1$ **then**
13-   Compute i^0 with Eq. (14);
14-   Detection = "Contaminant";
    else
15- $i^0 = \widehat{i}$;
16-   **if** $CPM(W_{B_D}) == 1$ **then**
17-   Detection = "Contaminant";
     else
18-   Detection = "Fault";
     end
    end
   end
  end
 end

Algorithm 1: The proposed cognitive monitoring system for intelligent buildings.

(a binary yes/no decision about whether the event has occurred), (ii) *isolation* (locate the zone i^* where the contaminant source or the fault has been introduced) and (iii) *identification* (determine whether the alert was due to a contaminant source or a sensor fault). The detection/diagnosis of multiple or simultaneous events is out of the scope of this work and part of our future research.

IV. THE PROPOSED COGNITIVE MONITORING SYSTEM FOR INTELLIGENT BUILDINGS

The proposed cognitive monitoring system, which is summarized in Figure 3, is characterized by a hierarchical architecture composed of the following three layers:

1) the *change-detection layer*, which involves a sequential and on-line CDT running at each of the N sensors of the intelligent building, is responsible for monitoring the concentration of the contaminant. The goal of this layer is to guarantee the prompt detection of any anomalous concentration of the contaminant by monitoring the statistical behaviour of $\{m_i(t), 1 \leq i \leq N\}$. This layer relies on measurements coming from a single sensor, hence, it can be directly executed at the sensor level, provided that there are enough computational resources.
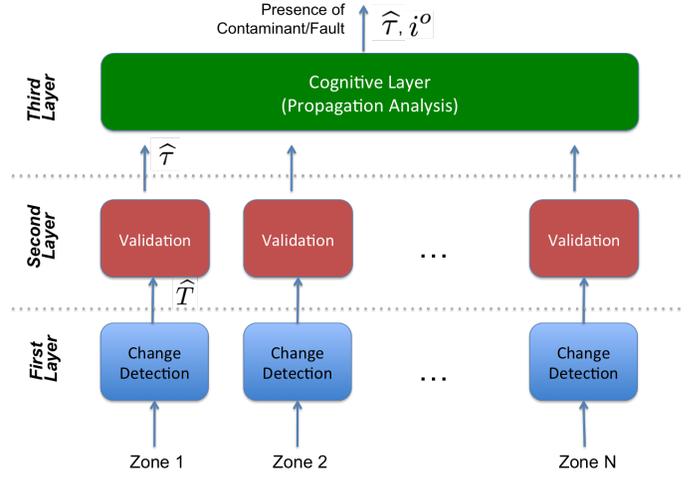

Fig. 3. The three-layer hierarchical architecture of the proposed cognitive monitoring system.

The zone where the CDT detects a change is referred to as the *detection zone* and denoted by $\widehat{i}$.

2) The *validation layer* is meant to reduce the false positives raised by the change-detection layer and, to this purpose, it employs a CPM. The CPM analyses the concentration measurements acquired at the detection zone to confirm whether any detection raised by the CDT at the first layer actually corresponds to a change in the statistical behaviour of the acquired measurements. Since this operation relies on measurements coming from a single sensor, it can be also directly executed at the sensor level, as long as enough computational resources are available.

3) The *cognitive layer* addresses two relevant tasks: first, isolating the zone where either the contaminant has been released (which in principle might not be $\widehat{i}$) or the sensor fault has occurred and, second, determining whether the validated detection is due to a contaminant source, or to a fault affecting the sensing apparatus. The core idea is that a contaminant would propagate through the building zones according to the flow matrix Z, while a fault would not. To perform such analysis, the cognitive layer builds two propagation trees to arrange the zones according to the expected contaminant propagation. The measurements from the relevant zones of the propagation trees are then aggregated, and analyzed through a CPM to determine if the contaminant has propagated therein. Since this layer involves the measurements coming from different zones of the building, it has to be executed in a centralized manner.

In practice, the first two layers implement in each sensor a *hierarchical* CDT [20] that has been customized for contaminant-detection purposes. The third layer, instead, has been specifically designed for addressing diagnosis and identification issues in a smart building environment. The three layers are detailed in what follows, while the algorithm of the proposed cognitive monitoring system is detailed in Algorithm 1.

A. The Change-Detection Layer

The aim of this layer is to provide prompt detection of changes in the statistical behaviour of sensor measurements. To achieve this goal, we opted for CDTs that are statistical techniques able to operate on data-streams to detect perturbations in the statistical distribution of the data.

CDTs in the literature are divided into parametric and nonparametric ones: the former [21], [22], like the CUSUM test [23], assume the distribution of the data before and after the change to be known. In contrast, nonparametric CDTs [24]–[28] operate without knowing the distribution of the observations either before or after the change. In real world applications, non-parametric CDTs are often preferred, mainly because of the unpredictability of the change. Sequential CDTs have been originally developed as rigorous statistical tools for quality-control applications [29]. Since then, CDTs have been used in a plethora of applications, including intrusion detection in computer networks [25] and fault detection in sensing/control apparatus [5], as the primary tools for detecting a change. Other relevant application scenarios are the environmental monitoring scenario, where CDTs can be used to detect evolutions in the micro-acoustic recordings to monitor a rock face [30], or in critical infrastructure monitoring, where CDTs can be used to detect leaks in a water distribution network [31].

Among the several sequential and on-line CDTs present in the literature, we focused on *the ICI-based CDT* [24]. A distinguishing feature of this CDT is the use of the Intersection of Confidence Intervals (ICI) rule, which is an adaptation technique for defining supports for polynomial regression. The ICI-based CDT is characterized by low computational complexity and this is a very important feature for analyzing streaming data. In addition, the ICI-based CDT operates without any a-priori information about the distribution of the data-generating process (either before or after the change).

In our previous work [5], we considered an ICI-based CDT monitoring two features extracted from the measurement stream: the sample mean and the sample variance computed on non-overlapping windows of ν observations. In this work, we replace the sample mean on non-overlapping data windows, used in [24], with the Manly transform [32], which yields approximately Gaussian distributed output and is defined as follows,

$$y_i(t) = \begin{cases} \left(e^{\lambda m_i(t)} - 1\right)/\lambda; & \lambda \neq 0 \\ m_i(t); & \lambda = 0 \end{cases}, \tag{6}$$

where $\lambda \in \mathbb{R}$ is a transform parameter. The Manly transform allows us to perform element-wise monitoring, thus possibly reducing the detection delay of relevant changes with respect to the window-wise operational modality. This is particularly relevant in critical application scenarios, like contaminant detection, where the detection delay has to be reduced as much as possible. Other transformations yielding output that are approximatively Gaussian distributed (e.g., the Box-Cox transform) could be considered as well; further details about element-wise ICI-based CDTs can be found in [33].

Besides the element-wise monitoring of the expected value of $y_i(t)$, the ICI-based CDT monitors the variance of $m_i(t)$ [5], by means of a power-law transformation of the sample variance computed on nonoverlapping windows. The transformed sample variance $V_i(s)$ for the s-th subsequence is defined as

$$V_i(s) = \left(\frac{S_i(s)}{\nu - 1}\right)^{\gamma}, \text{where} \tag{7}$$

$$S_i(s) = \sum_{t=(s-1)\nu+1}^{\nu s} \left(m_i(t) - M_i(s)\right)^2 \text{ and}$$

$$M_i(s) = \frac{1}{\nu} \sum_{t=(s-1)\nu+1}^{\nu s} m_i(t).$$

In these equations, $M_i(s)$ and $S_i(s)$ are the sample mean and the sample variance of the contaminant measurements in the s-th data window containing ν observations, respectively, while $\gamma > 0$ is the parameter of the power-law transformation which yields values of $V_i(s)$ that are approximatively Gaussian distributed [34].

The stationarity of $y_i(t)$ and $V_i(s)$ is then monitored over time through the ICI rule [35], [36] as described in [24]: changes in the feature distribution would indicate changes in the measurement streams that could be associated to either a contaminant source or a sensor fault. The detection phase of the ICI-based CDT is reported at line 3 of Algorithm 1, where $CDT(m_i(t))$ is equal to 1 when the ICI-based CDT detects a change in either in $y_i(t)$ or $V_i(s)$ (computed from the window containing $m_i(t)$), and 0 otherwise. In what follows, we denote by $\hat{i}$ the detection zone, i.e., the first zone where the ICI-based CDT detects a change in the stream of sensor measurements, and we denote by $\widehat{T}$ the time instant when this change has been detected.

As mentioned above, the ICI-based CDT does not require any a priori information about the data distribution but, for each sensor to be monitored, it requires an initial training sequence $\{m_i(t), t = 1, \ldots, L\}$ of L measurements acquired assuming the i-th sensor is fault-free and no contaminant sources are present. A first portion of the training sequence is used to estimate the CDT parameters (line 1 of Algorithm 1), in particular, λ of the Manly transform (6) and the power-law transform parameter γ in (7). The λ parameter is estimated via the maximum likelihood approach described in [32], while $\gamma = 1 - (\kappa_1 \kappa_3)/3\kappa_2^2$ with κ_j representing the j-th cumulant of $m_i(t)$ defined in [34]. The remaining part of the training sequence is used to compute the sample mean and the standard deviations of the feature values, which are used by the ICI-based CDT to detect changes. The ICI-based CDT is also characterized by a user-defined parameter Γ, which regulates the responsiveness of the CDT.

B. The Validation Layer

Every time the first layer raises a detection, it activates the validation layer which assesses whether this could be associated to a false positive of the CDT, and in this case prevents the activation of unnecessary emergency procedures.

To achieve this goal we resort on a CPM [37], namely a hypothesis test aiming at determining whether all the data within the analyzed sequence are i.i.d. samples (thus generated from the same distribution), or if there exists a change point, namely a point that separates these data in two sequences that have been generated by two different distributions.

Thus, CPMs operate on a fixed-length sequence and when they validate the change, they also estimate the position of the change point in the sequence. Like CDTs, CPMs can be also divided into parametric and non-parametric ones, depending on the assumptions made on the considered measurement sequence. Recently, optimized and approximated implementations of CPMs have been proposed to operate on data-streams [38], [39], and an ensemble of CPMs [40] was proposed to handle residuals from approximating models which might exhibit some form of correlation in stationary conditions. Estimating the position of a change point is of paramount importance in several application domains [41], [42], [43], but CPMs have never been applied to contaminant and sensor-fault diagnosis in intelligent buildings.

In more detail, the CPM is applied to a sequence of measurements coming from the detection zone $\widehat{i}$ before the detection time $\widehat{T}$. We denote this sequence as $\mathcal{C}_{\widehat{i}} = \{m_{\widehat{i}}(t), t = 1, \ldots, \widehat{T}\}$[1], while the CPM on this sequence is executed on line 8 of Algorithm 1. The CPM operates as follows (to simplify the notation we omit $\widehat{i}$ where not necessary): for each time instant $1 \leq u \leq \widehat{T}$, the set $\mathcal{C}_{\widehat{i}}$ is split into two parts,

$$
\begin{aligned}
A_u &= \{m_{\widehat{i}}(t), \ t = 1, \ldots, u\}, \\
B_u &= \{m_{\widehat{i}}(t), \ t = u+1, \ldots, \widehat{T}\},
\end{aligned}
$$

and a specific test statistic $\mathcal{T}$ is used to compute

$$
\mathcal{T}_u = \mathcal{T}(A_u, B_u),
$$

which is used to determine whether the set A_u and B_u contain data following different distributions.

The values of $\mathcal{T}_u$ are computed for all the samples $1 \leq u \leq \widehat{T}$, yielding $\{\mathcal{T}_u, u = 1, \ldots, \widehat{T}\}$. Let $\mathcal{T}_M$ be the maximum value of the statistic $\mathcal{T}$ over all the possible splits, i.e.,

$$
\mathcal{T}_M = \max_{u=1,\ldots,\widehat{T}} (\mathcal{T}_u - 1) \tag{8}
$$

and let $\widehat{\tau}$ be the sample for which $\mathcal{T}$ is maximum, i.e.,

$$
\widehat{\tau} = \underset{u=1,\ldots,\widehat{T}}{\operatorname{argmax}} (\mathcal{T}_u - 1).
$$

The value of $\mathcal{T}_M$ is then compared with a predefined threshold $h_{\widehat{T},\alpha_v}$, which is a function of the statistic $\mathcal{T}$, the cardinality $\widehat{T}$ of $\mathcal{C}_{\widehat{i}}$ and a given confidence level α_v that sets the percentage of type I errors (i.e., false positives) of the CPM operating at the validation layer. When $\mathcal{T}_M$ is larger than $h_{\widehat{T},\alpha_v}$, there is enough statistical evidence for the CPM to confirm the presence of a change in the measurements. On the other hand, when the test statistic does not exceed the threshold, there is not enough statistical evidence for claiming that the sequence $\mathcal{C}_{\widehat{i}}$ contains a change point. Hence (line 8 of Algorithm 1), the

---

[1]Note that other memory-efficient solutions could also be considered, e.g., by considering the training set and a buffer over the latest samples.

outcome of the CPM can be defined by the following binary variable:

$$
\mathrm{CPM}(\{m_i(t), t = 1, \ldots, \widehat{T}\}) = \begin{cases} 1 & \text{if } \mathcal{T}_M \geq h_{\widehat{T},\alpha_v} \\ 0, & \text{if } \mathcal{T}_M < h_{\widehat{T},\alpha_v} \end{cases}. \tag{9}
$$

When the outcome of the CPM is equal to 1, the detection raised by the CDT is validated and $\widehat{\tau}$ is identified as the time instant when the contaminant (or the fault) first appeared in the $\widehat{i}$-th zone. Otherwise, the detection is not validated and the CDT at the change-detection layer is newly configured from the training sequence $\{m_{\widehat{i}}(t), \ t = 1, \ldots, L\}$.

Choosing the right test statistic $\mathcal{T}$ is crucial to effectively validate changes detected by the change-detection layer. In our case, among the wide range of test statistics in the literature (e.g., see [37] and reference therein), we opted for a non-parametric statistic, hence not requiring any a priori information about the distribution of measurements to be analyzed. In particular, we focused on the Lepage test statistic [44], which corresponds to the sum of the Mann-Whitney [45] and the Mood [46] statistics. The Mann-Whitney statistic is able to detect changes in the location and the Mood statistic detects changes in the scale of a distribution. Thus, for all practical purposes, the Lepage statistic can be used to detect changes in the mean and variance of data belonging to $\mathcal{C}_{\widehat{i}}$, without any a priori assumption on the measurement distribution.

One of the most critical aspects when dealing with CPMs is the computation of the threshold $h_{\widehat{T},\alpha_v}$. In fact, the statistic of the CPM is indeed the maximum of all the test statistics (8), and it is typically not possible to analytically compute values of threshold that would yield a controlled amount of type-I errors (this is particularly true for statistics like the Lepage one which are not easily tractable by calculations). So, in practice, thresholds $h_{\widehat{T},\alpha_v}$ are numerically computed by means of Monte-Carlo simulations, as described in [37], [38], by generating sequences having different length. In this regard it's worth mentioning that, since the test statistic is nonparametric, it is possible to compute these thresholds using i.i.d. data generated from any arbitrary distribution and, of course, these do not depend on the change amount. We also emphasize that α_v, which sets the probability of type-I errors (i.e. false positives), is a user-defined parameter and can be set the same for all the sensors.

The solution presented in [5] computes, after each detection at the first layer, the p-values of both the Mann-Whitney and the Mood statistic to determine whether the detection was most likely due to an anomalous concentration of a contaminant (when the p-value of the Mann-Whitney test was lower than the one of the Mood) or to a sensor degradation fault (when the p-value of the Mood test was lower than that of the Mann-Whitney). This strategy, which was originally suggested in [38], was meant for a simplified observation and fault scenario like the one in [5]. Unfortunately, this simplified approach might not be effective when more general changes affect the distribution of measurements. In the more general model described by (1) and (2), the presence of contaminant sources and sensor faults have to be discriminated by analysing the contaminant propagation by means of the cognitive mechanisms described in what follows.

C. The Cognitive Layer

Once the detection raised at the CDT at the first layer has been confirmed by the CPM at the second layer (line 8 of Algorithm 1), the cognitive layer is activated (line 9 of Algorithm 1) to isolate the zone within the building in which either the contaminant has been released or the fault has occurred (*isolation* phase) and then distinguish between the two (*identification* phase). It is worth noting that the analysis at the cognitive layer is performed at the building level, by aggregating information coming from sensors placed in selected zones of the building.

1) Propagation Trees: The isolation and identification phases rely on the analysis of the contaminant concentration in those zones of the building that induce flow to, or receive flow from the detection zone $\widehat{i}$, according to the flow matrix Z, defined in Section III. To inspect the contaminant concentration in zones providing/receiving flow to/from $\widehat{i}$, we organize the zones of the building in two different trees: an *isolation tree* and an *identification tree*. These two trees, which represent the core ingredients of the cognitive layer, form the basis of the proposed cognitive monitoring system.

Without loss of generality, we are here assuming that $\hat{i}$ and $\hat{T}$ are unique, namely there are no simultaneous detections in different zones. The proposed method could be easily extended to manage simultaneous detections, by activating multiple executions of the three layers (detection-validation-cognitive), provided that the corresponding propagation and isolation trees do not overlap (possibly by relying on an ad-hoc mechanism for tree-generation able to force such a non-overlapping property).

The *isolation tree* $\mathcal{A} = <\mathcal{V}_{\mathcal{A}}, \mathcal{E}_{\mathcal{A}}>$ is constructed by considering the set of zones providing flow to the detection zone $\widehat{i}$, where $\mathcal{V}_{\mathcal{A}}$ is the set of vertices of $\mathcal{A}$ and $\mathcal{E}_{\mathcal{A}}$ is the set of edges connecting $\mathcal{V}_{\mathcal{A}}$. The isolation tree is constructed as follows: the root is the zone $\widehat{i}$ and the vertices of the first level represent those zones directly providing flow to the zone $\widehat{i}$ according to the flow matrix Z, i.e., the j-th zone is a zone of the first level if $z_{\widehat{i},j} = 1$. Next, the vertices of the second level represent those zones that, according to the flow matrix Z, induce positive flow to the vertices of the first level, and the construction proceeds iteratively until all the reachable zones are included. An example of an isolation tree for the considered building case study is portrayed in Figure 4, where the detection zone is $\widehat{i} = 3$.

Similarly, the expected propagation path from the detection zone $\widehat{i}$ to the rest of the building is analysed by constructing the *identification tree* $\mathcal{B} = <\mathcal{V}_{\mathcal{B}}, \mathcal{E}_{\mathcal{B}}>$, where $\mathcal{V}_{\mathcal{B}}$ is the set of vertices of $\mathcal{B}$ and $\mathcal{E}_{\mathcal{B}}$ is the set of edges connecting $\mathcal{V}_{\mathcal{B}}$. The identification tree is defined as follows: the root is the zone $\widehat{i}$; the vertices of the first level are those zones that directly receive a flow from the zone $\widehat{i}$ according to the flow matrix Z, i.e., the j-th zone is a zone of the first level if $z_{j,\widehat{i}} = 1$. Similarly, the vertices of the second level represent the zones that are directly connected to the vertices of the first level and the construction proceeds iteratively until all the reachable zones are included. An example of an identification tree for the considered case study is portrayed in Figure 4, where the detection zone is $\widehat{i} = 3$.

We emphasize that, while the isolation tree $\mathcal{A}$ indicates whether or how the contaminant has propagated before its detection in $\widehat{i}$ (in this sense $\mathcal{A}$ provides a possible view of the "incoming" contaminant to $\widehat{i}$), the identification tree $\mathcal{B}$ provides a description of the expected propagation of the contaminant within the building from $\widehat{i}$ to the other zones (in this sense, $\mathcal{B}$ provides a view of the "outgoing" contaminant from the detection zone).

For both the isolation and identification tree, we fix a maximum depth D of the tree and, to ease the description, we consider the same value of D for both trees but, in practice, we could consider different maximum depths. The effect of different values of D on the proposed cognitive monitoring system will be evaluated in Section V. We also comment that, in principle, D could be adapted based on the airflows and the amount of contaminant measured in the detection zone: the larger the flows and the measured amount, the larger the D used (since the contaminant is expected to quickly propagate to the rest of the building). On the other hand, small values of D, thus small propagation trees, would be preferable when dealing with multiple sources simultaneously.

When creating the propagation trees we assume that flows in Z do not to form cycles but, when this assumption does not hold, we can always use dummy, duplicate nodes in order to break-up the cycles, so as to be able to apply the proposed method.

Given D, we define A_D and B_D representing the sets of pairs (zone, depth) corresponding to vertices of $\mathcal{A}$ and $\mathcal{B}$ respectively, having distance from the root larger than zero and smaller than $D + 1$. More formally, the sets A_D and B_D are defined as follows (line 10 of Algorithm 1):

$$A_D = \{(i, d_i) \,|\, i \in \mathcal{V}_{\mathcal{A}}, d_i \in \{1, \ldots, D\}\} \qquad (10)$$

and

$$B_D = \{(i, d_i) \,|\, i \in \mathcal{V}_{\mathcal{B}}, d_i \in \{1, \ldots, D\}\} \qquad (11)$$

where i identifies the zone and d_i its distance from the root in the tree. In the example demonstrated in Figure 4 and using $D = 3$, $A_3 = \{(1,1), (2,1)\}$ and $B_3 = \{(5,1), (4,2), (8,3)\}$.

2) Isolation and Identification Phases: We approach the isolation and identification phases by assessing the presence of contaminant within the zones of A_D and B_D, respectively. The core statistical tool for assessing the contaminant in these zones is again a CPM, whose goal here is to detect variations even in those zones where the change-detection layer was not able to detect a contaminant. This is reasonable since the CPM, which operates in an off-line manner on a fixed-length measurement sequence, is characterized more powerful than the on-line change-detection mechanism operating at the change-detection layer. This means than the CPM might be able to detect changes even when the corresponding detection layer was not.

To effectively assess the contaminant propagation within the building we aggregate the sequence of measurements acquired from all the zones of A_D (and similarly of B_D) and apply a single CPM on the *aggregated measurement sequence*.
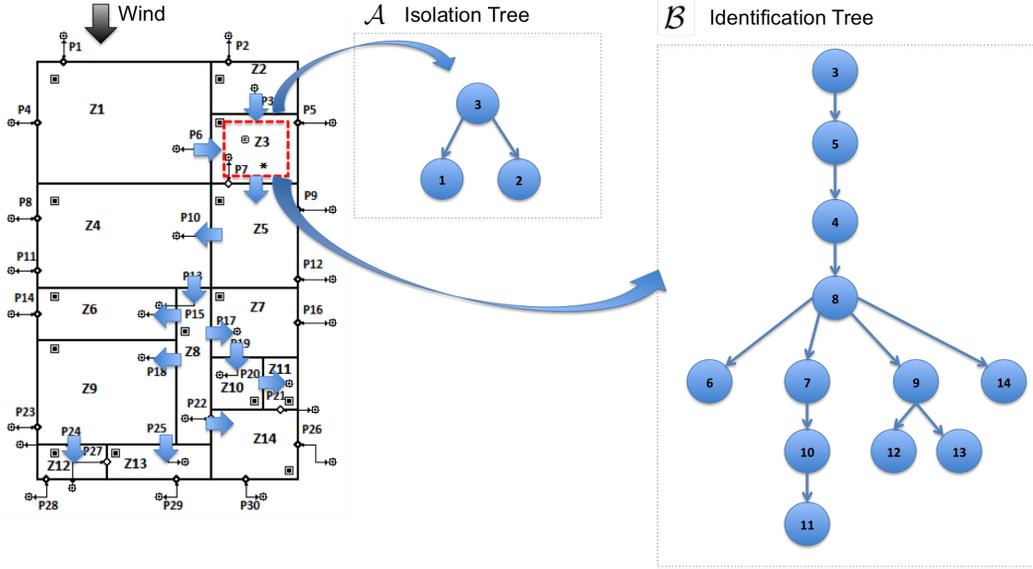
Fig. 4. The Holmes house layout and the corresponding isolation tree $\mathcal{A}$ and identification tree $\mathcal{B}$ when the detection zone is $\widehat{i} = 3$. The directions of the airflows between the various building zones are indicated with thick blue arrows on the building layout.

(i) Isolation Phase: The measurement sequences from all the zones in A_D are aggregated into a weighted-average measurement sequence:

$$W_{A_D}(t) = \sum_{(j,d_j) \in A_D} w_j^{A_D} m_j(t), \quad t \in \{1, \ldots, \widehat{T}\}, \quad (12)$$

where the aggregation weights $\{w_j^{A_D}\}$ are set inversely proportional to the distance from the detection zone $\widehat{i}$ along A_D, namely,

$$w_j^{A_D} = \frac{1/d_j}{\sum_{(j,d_j) \in A_D} 1/d_j}, \qquad \forall (j,d_j) \in A_D, \quad (13)$$

where j is the zone index and d_j its distance from the root zone in the tree A_D (line 11 of Algorithm 1). Setting weights inversely proportional to the distance from $\widehat{i}$ follows from the implicit assumption that the detection zone is close (possibly equal) to the source zone i^*, thus measurements from zones closer to $\widehat{i}$ yield a larger contribution in the aggregated measurements. Of course, other solutions where weights decrease with the distance from the detection zone could be considered as well. Note that the weights in (13) are normalized.

A CPM based on the Mann-Whitney test statistic (described in Section IV-B) is here considered since we are interested in detecting changes in the average contaminant concentration. The CPM is applied with a confidence level α_c to the aggregated measurement sequence W_{A_D}, i.e., CPM(W_{A_D}) at line 12 of Algorithm 1: when the outcome of the CPM is equal to 1, the presence of the contaminant is validated in the isolation tree meaning that the detection in $\widehat{i}$ is most likely the result of incoming flow from a zone of A_D. Consequently, the estimated source zone i^o becomes the zone in A_D characterized by the

largest amount of contaminant acquired from $t = 1$ to $t = \widehat{T}$ (line 13 of Algorithm 1), i.e.,

$$i^o = \operatorname*{argmax}_{(j,d_j) \in A_D} \sum_{t=1}^{\widehat{T}} m_j(t). \quad (14)$$

Note that, in this case the detection in $\widehat{i}$ can be safely associated to the presence of a contaminant (line 14 of Algorithm 1), hence there is no need to execute the subsequent identification phase. On the contrary, when the outcome of the CPM is equal to 0, no contaminant is detected in A_D, hence the estimated source zone i^o becomes $\widehat{i}$ and the subsequent identification phase must be executed to distinguish between the presence of a contaminant (whose source zone is $\widehat{i}$) and a possible sensor fault.

(ii) Identification Phase: The measurement sequences from all the zones in B_D are aggregated into a weighted average measurement sequence

$$W_{B_D}(t) = \sum_{(j,d_j) \in B_D} w_j^{B_D} m_j(t), \quad t \in \{1, \ldots, \widehat{T}\}, \quad (15)$$

where the aggregation weights $\{w_j^{B_D}\}$ are defined as in (13), by replacing A_D with B_D (line 11 of Algorithm 1). The CPM based on the Mann-Whitney test statistic is then applied, with a confidence level α_c, to the aggregated measurements W_{B_D} as in line 16 of Algorithm 1.

When $CPM(W_{B_D}) = 1$, the contaminant propagation is validated and the detection in zone $\widehat{i}$ is associated to the *real presence of a contaminant* in the building (line 17 of Algorithm 1). On the contrary, when $CPM(W_{B_D}) = 0$, no propagation of contaminant has been detected in B_D, hence corroborating the idea of having a fault affecting the sensor in zone $\widehat{i}$ (line 18 of Algorithm 1). We also take into consideration the extreme cases where $B_D = \emptyset$, a situation that occurs when the source zone is not connected to any other building zone. In

such a case, the cognitive layer cannot make any propagation analysis and the most conservative decision should be made, i.e., the change is considered to be a real contaminant.

Summing up, a CPM is used both for the isolation and the identification phases. Their confidence levels, α_c and α_v, which regulate the percentage of false positives, could be equal. We emphasize that, here at the cognitive layer, thresholds refer to the Mann-Whitney statistic, while the ones used in the validation phase referred to the Lepage statistic.

We also emphasize that, once the contaminant propagation has been confirmed by the cognitive layer, the proposed cognitive monitoring system could be temporarily suspended in those zones belonging to the propagation tree of $\widehat{i}$ to avoid consecutive activations/detections due to the propagation of the contaminant through the building. In this case, small values of D are preferable so as to reduce the number of zones in which the contaminant detection is suspended during the cognitive analysis.

3) Other Solutions: Except from the proposed solution, a variety of other simpler solutions using CPMs could be also considered for detecting contaminant variations in A_D or B_D. We detail two such solutions below.

(i) Nearest Neighbour: The most straightforward solution would be to analyse the sequence of measurements only in the nearest zones (one hop away) of the detection zone $\widehat{i}$ and to detect the contaminant propagation by means of CPMs applied to measurements coming from these zones. In particular, for this solution a CPM is separately applied to the sequences of measurements in all the zones at the first level of either the isolation or the identification tree. This solution, however, is not able to exploit the full contaminant propagation path through the zones of the building, hence reducing its ability to identify it.

(ii) K over N analysis: Another straightforward solution consists in executing CPMs independently in all the zones of either A_D or B_D, and in assessing that the contaminant has propagated as soon as K of these CPMs reveal a change. This solution allows to fully exploit the natural propagation of the contaminant within the building but it has two main drawbacks. First, it is very difficult to set K, since it actually depends on the contaminant propagation which is a-priori unknown; for this reason, the most conservative choice of $K = 1$ is often made [47]. Second, in situations where the detection at the first layer is due to a fault and there is no contaminant propagation, there is a substantial risk of having false positives by running several CPMs in parallel (the risk increases with the number of considered zones) leading to the erroneous identification of faults as contaminants.

V. EXPERIMENTAL RESULTS

The aim of this section is to evaluate the proposed solution on a wide experimental campaign encompassing both contaminant sources and sensor fault scenarios.

A. The considered scenarios

As mentioned in Section II-A, in our analysis we considered datasets generated by the Matlab-CONTAM toolbox [17]

Scenario	Type	Characteristics
S1	Contaminant	Source at constant emission rate 50 g/h
S2	Contaminant	Source with step change every 2 hours between 25 and 50 g/h
S3	Contaminant	Source at constant emission rate 33 g/h
S4	Contaminant	Source with step change every 2 hours between 13 and 33 g/h
S5	Fault	Abrupt permanent additive fault of magnitude $\delta = 1$
S6	Fault	Abrupt permanent additive fault of magnitude $\delta = 0.5$

TABLE I
THE SIX CONSIDERED SCENARIOS.

referring to the Holmes house case study. More specifically, through the Matlab-CONTAM toolbox, we generated datasets according to 6 different scenarios: 4 scenarios with actual contaminant sources of different emission rates and 2 scenarios of sensor faults. For the sensor faults, we are here assuming permanent abrupt additive faults whose effect is to add a permanent bias δ to the acquired measurements. Note that in the considered scenarios the contaminant source and the sensor fault do not occur at the same time. The six considered scenarios are detailed in Table I. In all these scenarios, the simulation time is 48 hours while the sampling time is 1 sample per minute: thus, each sensor acquires 2880 measurements.

It is assumed that natural ventilation is the dominant cause of air flows in the building with wind coming from the north $(0°)$ at a speed of 10 m/s. All the openings (doors or windows) are assumed to be in the fully open position. We assume that at time $\tau = 25$ hours, a contaminant source or a sensor fault is activated in the utility room (Z3), i.e., $i^* = 3$, as shown in Figure 4. There is one sensor in each zone able to record the concentration of the contaminant at regular intervals at its own location but the sensor measurements are corrupted by noise following the model in (2). Noise is assumed to be i.i.d. following a Gaussian distribution $\mathcal{N}(0, \sigma^2)$. In our experiments we considered five different values of σ, i.e., $\sigma = \{1, 1.5, 2, 2.5, 3\}$.

Note that, to ease the comparison in all the displayed scenarios, we report results for a fixed source zone, i.e. $i^* = 3$, and wind direction, i.e., 0°. However, additional experiments have been performed with different configurations of source zones and wind directions obtaining similar results.

B. Figures of merits

The proposed cognitive monitoring system for contaminant and sensor fault diagnosis has been evaluated from two different perspectives: (i) detection and (ii) isolation/identification. For this purpose we defined two different sets of figures of merit.

Regarding detection, the following three figures of merit have been considered:

- False Positive Rate (FPR), the percentage of experiments in which the change has been detected before τ;
- False Negative Rate (FNR), the percentage of experiments in which the change has not been detected;

- Detection Delay DD (in samples), that is the average value of $\widehat{T} - \tau$;

With respect to isolation/identification, we considered the following two figures of merit:

- ϵ_{iso}, the percentage of experiments in which the source zone has not been correctly recognized;
- ϵ_{id}, the percentage of experiments in which the validated change has not been correctly identified (i.e., the presence of a contaminant is associated to a sensor fault or vice versa).

We computed the above figures of merit over 250 runs for each scenario.

C. The Configuration of the Proposed Cognitive Monitoring System

The proposed cognitive monitoring system has been configured as follows. The ICI-based CDT at the change-detection layer is trained on the first $L = 400$ samples, while the parameter Γ that regulates the CDT responsiveness was set to $\Gamma = 2$. At the validation layer, the confidence α_v of the CPM based on the Lepage statistics has been fixed at 0.05. This configuration yielded successful detection performance when used in a hierarchical CDTs based on ICI-based CDT, as in [20]. Differently, at the cognitive layer, we considered two different values for the maximum depth D, i.e., $D = 3$ and $D = 5$, for both the isolation and identification trees. We also considered two different confidence values for the CPM used in the cognitive layer (the one based on the Mann-Whitney statistic), i.e., $\alpha_c = 0.05$ and $\alpha_c = 0.01$.

D. Alternative solutions

The detection ability of the proposed cognitive monitoring system is compared against that of the Scalar Trigger Algorithm (STA) described in [4]. As mentioned in II-A, STA does not require a priori information about the considered scenario and is dynamically able to deal with noisy measurements. In order to operate, STA requires to set several parameters related to the sizes of three windows obtained over the acquired measurements (i.e., the *background*, the *guard* and the *present* window) and a threshold. To allow a fair comparison, the STA parameters have been experimentally fixed to guarantee FPRs in line with those provided by the proposed solution, i.e., the size of the three windows mentioned above were set to 400, 20 and 100, respectively, while the threshold was set to 4.5.

We should stress at this point that STA is based on thresholding mechanisms, while the proposed solution does not use a threshold but relies instead on a sequential statistical analysis of the datastreams (using ICI-based CDTs). This allows the detection of *any* variations in the contaminant concentration (even small ones).

The isolation performance of the proposed cognitive monitoring system has been compared with the straightforward solution where the detection zone $\widehat{i}$ is considered as the *estimated source zone*. This "naive" solution is supported by the idea that the first zone detecting a change (either sensor fault or contaminant presence) can be reasonably assumed to be the source zone. However, as described below, this choice might not provide satisfactory performance, in particular when sensors are affected by heavy noise and the contaminant concentration is low.

Finally, the identification ability of the proposed cognitive monitoring system has been compared with the nearest neighbour (NN) approach and the K over N analysis (K/N) described in Section IV-C with $K = 1$ (as suggested in [47]).

E. Discussion

1) Detection: The comparison between the detection ability of the proposed cognitive monitoring system and STA is detailed in Table II. Several comments arise.

First, the proposed solution is very effective and yields prompt detection of both contaminants (scenarios S1 to S4) and sensor faults (scenarios S5 and S6). Results in Table II show that the proposed solution is able to keep under control the false positives (in all the configurations of the noise the FPR is 0%), while providing a very good detection performance both in terms of false negatives and detection delay. At the same time, STA achieves similar FPR (since this is the criteria we used for tuning it), but at the expense of very high FNRs leading to an ineffective detection system. Note that even though STA provides lower DDs, this advantage is negligible since most of the changes are not detected.

Second, as expected, both the FNR and DD increase with σ. This is due to the fact that the signal-to-noise ratio (SNR)

Scenario	σ	Proposed Solution			STA		
		FPR	FNR	DD	FPR	FNR	DD
S1	1.0	0.000	0.000	191.4	0.02	0.95	105.38
	1.5	0.000	0.000	279.8	0.00	0.98	244.40
	2.0	0.000	0.000	378.0	0.00	0.99	89.00
	2.5	0.000	0.000	489.2	0.00	0.98	234.33
	3.0	0.000	0.008	590.9	0.01	0.99	259.00
S2	1.0	0.000	0.000	262.3	0.02	0.96	195.43
	1.5	0.000	0.000	401.5	0.00	0.98	322.75
	2.0	0.000	0.000	538.3	0.00	1.00	74.00
	2.5	0.000	0.020	702.2	0.00	0.98	288.33
	3.0	0.000	0.096	840.1	0.01	0.98	493.50
S3	1.0	0.000	0.000	281.2	0.02	0.96	156.20
	1.5	0.000	0.000	429.6	0.00	0.98	357.00
	2.0	0.000	0.004	597.9	0.00	1.00	74.00
	2.5	0.000	0.076	790.9	0.00	0.99	336.00
	3.0	0.000	0.228	950.0	0.01	0.99	259.00
S4	1.0	0.000	0.000	401.5	0.02	0.98	216.00
	1.5	0.000	0.000	622.0	0.00	0.98	357.00
	2.0	0.000	0.108	845.0	0.00	1.00	74.00
	2.5	0.000	0.472	1005.3	0.00	0.99	336.00
	3.0	0.000	0.712	1078.8	0.01	0.99	259.00
S5	1.0	0.000	0.000	188.3	0.02	0.96	54.14
	1.5	0.000	0.000	301.4	0.00	0.98	282.75
	2.0	0.000	0.000	424.6	0.00	0.99	89.00
	2.5	0.000	0.012	571.3	0.00	0.98	234.33
	3.0	0.000	0.052	700.7	0.01	0.99	259.00
S6	1.0	0.000	0.000	421.4	0.02	0.98	201.50
	1.5	0.000	0.096	693.5	0.00	0.98	357.00
	2.0	0.000	0.296	924.7	0.00	1.00	74.00
	2.5	0.000	0.716	1029.6	0.00	0.99	336.00
	3.0	0.000	0.856	1151.1	0.01	0.99	259.00

TABLE II

DETECTION: COMPARISON BETWEEN THE PROPOSED SOLUTION AND STA ON THE CONSIDERED SCENARIOS.

decreases as σ increases. In the scenarios affected by contaminant (i.e., S1 - S4), the SNR is defined as the ratio between $\Delta_{i^*}^2$ and σ^2. If we estimate Δ_{i^*} in (1) as the average of contaminant measurements after τ_{i^*} in zone i^*, we obtain 1.19, 0.87, 0.78 and 0.57 for scenarios S1, S2, S3 and S4, respectively. Then, the resulting SNR for scenario S1 ranges from 1.41 to 0.15 (depending on the value of σ), while for scenario S4 the SNR ranges from 0.32 to 0.03. This justifies the performance loss of the proposed solution in scenario S4 when $\sigma > 2$. Similarly, for the sensor fault scenarios (i.e., scenarios S5 and S6) the SNR can be measured as the ratio between δ^2 and σ^2. Hence, for scenarios S5 and S6, the SNR ranges from 1 to 0.11 and from 0.25 to 0.03, respectively. Thus, by looking at the results in Table II we could conclude that, in order to guarantee a FNR less than 0.1, the SNR should be larger than 0.01 in both the contaminant and sensor-fault scenarios. Examples of the acquired measurements in the source zone for scenarios S1 and S5 can be found in Figure 5. Note from this figure how difficult the problem becomes under low SNR conditions, making the detection task impossible with the naked eye and without using sophisticated detection solutions as the one proposed in this paper. Another interesting thing to observe from the figure is that by just looking at the source zone alone, it is impossible to tell whether the increased concentration is due to a contaminant source (S1) or a sensor fault (S4), which motivates the proposed cognitive identification solution.

2) Isolation: The isolation ability of the proposed cognitive monitoring system compared to the one of the *naive* solution are shown in Table III - Columns 3 and 4.

As expected, the naive solution does not provide satisfactory performance, in particular when σ is large and the contaminant concentration is low. In these situations, the source zone i^* might not be the first one to detect the presence of the contaminant and often the first detection of contaminant might be provided by different zones receiving flow from i^*. Differently, the proposed cognitive isolation phase (described in Section IV-C) is able to better deal with measurement noise by relying on the analysis of the isolation tree and the propagation of contaminant up to $\widehat{i}$. Consequently, in scenarios S5 and S6 where there is no contaminant propagation (these are the sensor-fault scenarios) the proposed and the naive solution achieve similar results.

This corroborates the idea that the proposed analysis of the expected propagation of the contaminant up to the detection zone improves the isolation performance of the monitoring system.

3) Identification: The last three columns of Table III show the comparison among the proposed, the NN and the K over N solutions. Results are particularly interesting and two main comments arise.

First, for the first three contaminant scenarios (i.e., scenarios S1, S2 and S3) the identification ability of the proposed solution is quite impressive. In fact, for almost all the σ values considered, the value of ϵ_{id} does not exceed 0.01 indicating that the proposed solution is extremely good in recognizing the presence of a contaminant. In scenario S4 the results slightly worsen (recall this is scenario characterized by the lowest SNR), while maintaining values of ϵ_{id} smaller than 0.08. Similarly, the proposed solution is able to effectively identify the presence of faulty sensors. In fact, in Scenarios S5 and S6 the values of ϵ_{id} are generally less than 0.06 meaning that in most of the cases the change is correctly associated to a fault affecting the sensor.

Second, the proposed solution provides better performance than the two alternative solutions. As expected, the NN solution is not able to fully exploit the contaminant propagation through all the zones of the building since it analyses only the zones at the first level of the $\mathcal{B}_D$. This results in high values of ϵ_{id} especially in cases of low contaminant concentration and high values of σ (e.g., Scenario S4). Differently, the K *over* N *analysis* is able to fully exploit the natural propagation of the contaminant within the building since it analyses all the zones of $\mathcal{B}_D$. This results in very low values of ϵ_{id} in all the contaminant scenarios (S1-S4), but at the expense of the identification ability in the sensor fault scenarios (S5 and S6) where the values of ϵ_{id} are considerably higher than the proposed solution.

In addition, we have tested the robustness of the proposed solution by varying three of the key parameters involved: D, α_c and L. The experimental results with $D = 5$ presented in Table IV, are close to the ones with $D = 3$ presented in Table III. This is not surprising since the proposed solution is able to take into account the distance of each considered zone from the detection zone thanks to the weighting mechanisms described in (12) and (15). Similarly, results with $\alpha_c = 0.01$ presented in Table V, are in line with those of $\alpha_c = 0.05$ presented in Table III, meaning that the proposed solution is effective using both configurations. Finally, with respect to the training sequence length L, as expected, our results indicate that smaller values of L would result in a slight increase of both FPR and FNR, while larger ones would allow to reduce the FNR. These results have not been included in the paper due to space limitations, but can be provided upon request.

4) Propagation Trees: As mentioned in Section IV-C, the isolation and identification trees represent the core of the cognitive isolation and identification phases of the proposed monitoring system. The usefulness and effectiveness of these trees can be evaluated by comparing the experimental results in Tables III and VI. While the first table has been extensively commented above, the second one shows the isolation and identification ability of the considered solutions without the isolation and identification trees (all the zones of the house are considered regardless of the distance from the detection zone). By comparing Tables III and VI, the advantages provided by the isolation and identification trees become meaningful and relevant for both the isolation and the identification phase. As expected, the values of ϵ_{iso} and ϵ_{id} are larger in Table VI meaning that exploiting the expected propagation path from and to the detection zone is truly able to improve the isolation and identification performance. It is interesting to note from Table VI that without using the propagation trees, the identification performance of the proposed solution is only affected for contaminant source scenarios (S1-S4), while the performance of the other solutions (NN and K/N) becomes worse for the sensor fault scenarios (S5-S6).
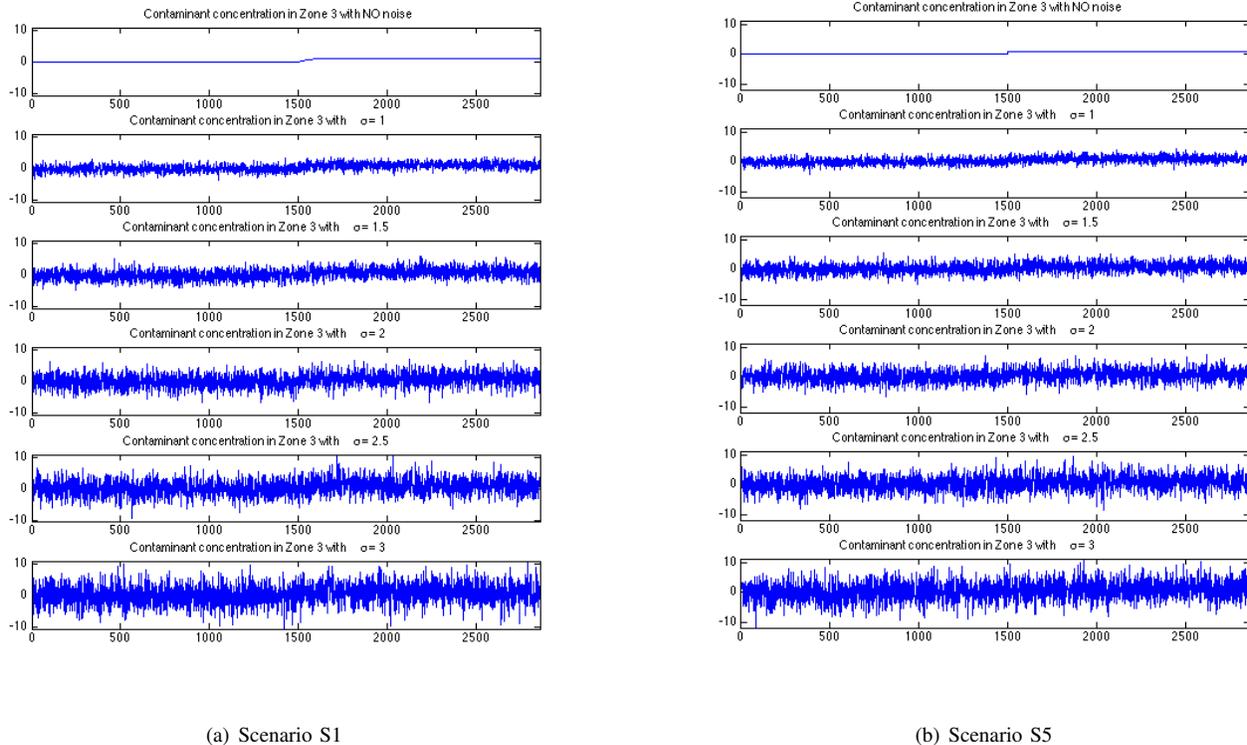
(a) Scenario S1

(b) Scenario S5

Fig. 5. Examples of acquired measurements in the source zone without noise and with different levels of σ in Scenarios S1 and S5.

VI. CONCLUSIONS

The prompt detection and isolation of possible contaminants is a key issue in intelligent buildings to ensure the quality of life and safety of the occupants. This paper introduces a novel cognitive monitoring system able to promptly detect variations in the contaminant concentration, isolate the source zone and identify whether the change can be safely associated to the presence of a contaminant source or to a fault affecting a sensor, which could easily occur in systems operating in real-world conditions. The proposed cognitive solution relies on a three-layer hierarchical architecture (the change-detection, the validation and the cognitive layer) encompassing theoretically-grounded statistical techniques and the ability to exploit the expected propagation path of the contaminant within the building for discriminating between contaminant sources and sensor faults. It is worth pointing out that the proposed cognitive monitoring system can be operated on-line with very little a priori information. The proposed solution was demonstrated to be particularly effective in a realistic building case study encompassing scenarios of both contaminant sources as well as sensor faults.

In the future, we plan to investigate more complex scenarios involving building case studies with a large number of zones and a limited number of sensors. Moreover, we will consider scenarios involving multiple contaminant sources or multiple sensor faults or even the simultaneous presence of both. For dealing with these more complex scenarios, we plan to enhance our cognitive solution by considering other relevant information for constructing the propagation trees, like the estimated detection times of the contaminant in the various building zones and the sequence of detection times.

REFERENCES

[1] J. Braun, "Intelligent building systems-past, present, and future," in *Proc. of American Control Conference (ACC'07)*. IEEE, 2007, pp. 4374–4381.

[2] X. Lu, D. Clements-Croome, and M. Viljanen, "Past, present and future of mathematical models for buildings," *Intelligent Buildings International*, vol. 1, pp. 23–38(16), 2009.

[3] D. Cousins and S. D. Campbell, "Protecting buildings against airborne contamination," *Lincoln Laboratory Journal*, vol. 17, no. 1, 2007.

[4] 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.

[5] G. Boracchi, M. Michaelides, and M. Roveri, "A cognitive monitoring system for contaminant detection in intelligent buildings," in *International Joint Conference on Neural Networks (IJCNN 2014)*, 2014.

[6] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques-Part I: Fault diagnosis with model-based and signal-based approaches," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 6, pp. 3757–3767, June 2015.

[7] ——, "A survey of fault diagnosis and fault-tolerant techniques-Part II: Fault diagnosis with knowledge-based and hybrid/active approaches," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 6, pp. 3768–3774, June 2015.

[8] X. Dai and Z. Gao, "From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 4, pp. 2226–2238, November 2013.

Scenario	σ	ϵ_{iso} Proposed	Naive	ϵ_{id} Proposed	NN	K/N
S1	1.0	0.00	0.02	0.01	0.01	0.01
	1.5	0.01	0.08	0.00	0.00	0.00
	2.0	0.02	0.13	0.00	0.00	0.00
	2.5	0.07	0.27	0.01	0.00	0.00
	3.0	0.12	0.31	0.00	0.01	0.00
S2	1.0	0.01	0.04	0.00	0.00	0.00
	1.5	0.05	0.18	0.01	0.01	0.01
	2.0	0.04	0.24	0.00	0.00	0.00
	2.5	0.14	0.32	0.01	0.03	0.01
	3.0	0.25	0.46	0.02	0.10	0.02
S3	1.0	0.01	0.08	0.00	0.01	0.01
	1.5	0.06	0.21	0.01	0.01	0.01
	2.0	0.08	0.27	0.01	0.00	0.00
	2.5	0.23	0.43	0.01	0.04	0.01
	3.0	0.34	0.55	0.03	0.16	0.04
S4	1.0	0.04	0.16	0.01	0.02	0.01
	1.5	0.13	0.35	0.01	0.02	0.02
	2.0	0.19	0.38	0.01	0.12	0.01
	2.5	0.41	0.60	0.05	0.16	0.07
	3.0	0.44	0.58	0.08	0.33	0.11
S5	1.0	0.00	0.00	0.05	0.04	0.15
	1.5	0.00	0.00	0.04	0.04	0.13
	2.0	0.00	0.00	0.05	0.05	0.14
	2.5	0.00	0.00	0.04	0.04	0.13
	3.0	0.01	0.01	0.06	0.06	0.14
S6	1.0	0.01	0.01	0.05	0.04	0.17
	1.5	0.01	0.01	0.05	0.04	0.14
	2.0	0.01	0.02	0.08	0.07	0.22
	2.5	0.08	0.08	0.06	0.03	0.14
	3.0	0.15	0.15	0.10	0.02	0.12

TABLE III

ISOLATION AND IDENTIFICATION: PERFORMANCE EVALUATION WITH $D = 3$ AND $\alpha_c = 0.05$

Scenario	σ	ϵ_{iso} Proposed	Naive	ϵ_{id} Proposed	NN	K/N
S1	1.0	0.00	0.02	0.01	0.01	0.00
	1.5	0.01	0.08	0.01	0.00	0.00
	2.0	0.02	0.13	0.01	0.00	0.00
	2.5	0.07	0.27	0.01	0.00	0.00
	3.0	0.12	0.31	0.00	0.01	0.00
S2	1.0	0.00	0.04	0.00	0.00	0.00
	1.5	0.05	0.18	0.01	0.01	0.01
	2.0	0.04	0.24	0.01	0.00	0.00
	2.5	0.14	0.32	0.01	0.03	0.01
	3.0	0.25	0.46	0.02	0.10	0.02
S3	1.0	0.00	0.08	0.00	0.00	0.00
	1.5	0.06	0.21	0.01	0.01	0.01
	2.0	0.08	0.27	0.01	0.00	0.00
	2.5	0.22	0.43	0.02	0.04	0.01
	3.0	0.34	0.55	0.04	0.16	0.03
S4	1.0	0.04	0.16	0.02	0.01	0.01
	1.5	0.12	0.35	0.01	0.01	0.01
	2.0	0.19	0.38	0.02	0.12	0.01
	2.5	0.40	0.60	0.06	0.16	0.05
	3.0	0.44	0.58	0.10	0.33	0.08
S5	1.0	0.00	0.00	0.06	0.04	0.40
	1.5	0.00	0.00	0.07	0.04	0.43
	2.0	0.00	0.00	0.03	0.05	0.43
	2.5	0.00	0.00	0.04	0.04	0.41
	3.0	0.01	0.01	0.03	0.06	0.45
S6	1.0	0.01	0.01	0.06	0.04	0.40
	1.5	0.01	0.01	0.06	0.04	0.38
	2.0	0.01	0.02	0.08	0.07	0.45
	2.5	0.08	0.08	0.05	0.03	0.39
	3.0	0.15	0.15	0.07	0.02	0.22

TABLE IV

ISOLATION AND IDENTIFICATION: PERFORMANCE EVALUATION WITH $D = 5$ AND $\alpha_c = 0.05$

[9] S. Katipamula and M. R. Brambley, "Review article: Methods for fault detection, diagnostics, and prognostics for building systems-A review, Part I," *HVAC&R Research*, vol. 11, no. 1, pp. 3–25, January 2005.

[10] ——, "Review article: Methods for fault detection, diagnostics, and prognostics for building systems-A review, Part II," *HVAC&R Research*, vol. 11, no. 2, pp. 169–187, April 2005.

[11] X. Liu and Z. Zhai, "Inverse modeling methods for indoor airborne pollutant tracking: literature review and fundamentals," *Indoor Air*, vol. 17, no. 6, pp. 419–438, 2007.

[12] M. Sohn, R. Sextro, A. Gadgil, and J. Daisey, "Responding to sudden pollutant releases in office buildings: 1. Framework and analysis tools," *Indoor Air*, vol. 13, no. 3, pp. 267–276, 2003.

[13] X. Liu and Z. Zhai, "Prompt tracking of indoor airborne contaminant source location with probability-based inverse multi-zone modeling," *Building and Environment*, vol. 44, no. 6, pp. 1135–1143, 2009.

[14] M. Michaelides, V. Reppa, M. Christodoulou, C. Panayiotou, and M. Polycarpou, "Contaminant event monitoring in multi-zone buildings using the state-space method," *Building and Environment*, vol. 71, pp. 140–152, 2014.

[15] H. Feustel, "COMIS–an international multizone air-flow and contaminant transport model," *Energy and Buildings*, vol. 30, no. 1, pp. 3–18, 1999.

[16] G. Walton and W. Dols, "CONTAM 2.4 user guide and program documentation," *National Institute of Standards and Technology, NISTIR*, vol. 7251, 2005.

[17] M. P. Michaelides, D. G. Eliades, M. Christodoulou, M. Kyriakou, C. Panayiotou, and M. Polycarpou, "A Matlab-CONTAM toolbox for contaminant event monitoring in intelligent buildings," in *Artificial Intelligence Applications and Innovations*. Springer, 2013, pp. 605–614.

[18] L. Wang, W. Dols, and Q. Chen, "Using CFD capabilities of CONTAM 3.0 for simulating airflow and contaminant transport in and around buildings," *HVAC&R Research*, vol. 16, no. 6, pp. 749–763, 2010.

[19] G. Boracchi, M. Michaelides, and M. Roveri, "Detecting contaminants in smart buildings by exploiting temporal and spatial correlation," in *Intelligent Embedded Systems (IES), 2015 IEEE Symposium on*, December 2015, pp. 1–8.

[20] C. Alippi, G. Boracchi, and M. Roveri, "Hierarchical change-detection tests," *IEEE Transactions on Neural Networks and Learning Systems*, vol. PP, no. 99, pp. 1–13, 2016.

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

[22] A. Polunchenko and A. Tartakovsky, "State-of-the-art in sequential change-point detection," *Methodology and Computing in Applied Probability*, vol. 14, no. 3, pp. 649–684, 2012. [Online]. Available: http://dx.doi.org/10.1007/s11009-011-9256-5

[23] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1, pp. 100–115, 1954. [Online]. Available: http://biomet.oxfordjournals.org/content/41/1-2/100.extract

[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. G. Tartakovsky, B. L. Rozovskii, R. B. Blaek, and H. Kim, "Detection of intrusions in information systems by sequential change-point methods," *Statistical Methodology*, vol. 3, no. 3, pp. 252 – 293, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1572312705000493

[26] 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.

[27] Y. Kawahara and M. Sugiyama, "Sequential change-point detection based on direct density-ratio estimation," *Statistical Analysis and Data Mining*, vol. 5, no. 2, pp. 114–127, 2012. [Online]. Available: http://dx.doi.org/10.1002/sam.10124

[28] C. Alippi, G. Boracchi, and M. Roveri, "A hierarchical, nonparametric, sequential change-detection test," in *International Joint Conference on Neural Networks (IJCNN 2011)*, 31 2011-aug. 5 2011, pp. 2889 –2896.

[29] W. A. Shewart, *Economic control of Quality of Manufactured Product*. New York: Van Nostrand Reinhold Co., 1931.

[30] C. Alippi, G. Boracchi, and B. Wohlberg, "Change detection in streams of signals with sparse representations," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, May 2014, pp. 5252–5256.

Scenario	σ	ϵ_{iso} Proposed	Naive	ϵ_{id} Proposed	NN	K/N
S1	1.0	0.00	0.02	0.01	0.01	0.01
	1.5	0.01	0.08	0.00	0.00	0.00
	2.0	0.02	0.13	0.00	0.00	0.00
	2.5	0.07	0.27	0.01	0.00	0.00
	3.0	0.15	0.31	0.01	0.01	0.00
S2	1.0	0.01	0.04	0.00	0.00	0.00
	1.5	0.05	0.18	0.01	0.01	0.01
	2.0	0.06	0.24	0.00	0.00	0.00
	2.5	0.18	0.32	0.02	0.03	0.01
	3.0	0.29	0.46	0.02	0.10	0.02
S3	1.0	0.01	0.08	0.00	0.01	0.01
	1.5	0.06	0.21	0.01	0.01	0.01
	2.0	0.10	0.27	0.01	0.00	0.00
	2.5	0.28	0.43	0.03	0.10	0.03
	3.0	0.34	0.55	0.03	0.16	0.04
S4	1.0	0.04	0.16	0.01	0.02	0.01
	1.5	0.17	0.35	0.01	0.02	0.02
	2.0	0.24	0.38	0.01	0.13	0.01
	2.5	0.47	0.60	0.08	0.20	0.08
	3.0	0.47	0.58	0.11	0.35	0.13
S5	1.0	0.00	0.00	0.05	0.04	0.15
	1.5	0.00	0.00	0.04	0.04	0.13
	2.0	0.00	0.00	0.05	0.05	0.14
	2.5	0.00	0.00	0.04	0.04	0.13
	3.0	0.01	0.01	0.05	0.06	0.14
S6	1.0	0.01	0.01	0.05	0.04	0.17
	1.5	0.01	0.02	0.05	0.05	0.14
	2.0	0.01	0.02	0.08	0.07	0.21
	2.5	0.08	0.08	0.06	0.03	0.13
	3.0	0.14	0.14	0.10	0.02	0.12

TABLE V

ISOLATION AND IDENTIFICATION: PERFORMANCE EVALUATION WITH $D = 3$ AND $\alpha_c = 0.01$

Scenario	σ	ϵ_{iso} Proposed	Naive	ϵ_{id} Proposed	NN	K/N
S1	1.0	0.04	0.02	0.29	0.01	0.00
	1.5	0.12	0.08	0.19	0.00	0.00
	2.0	0.19	0.13	0.19	0.00	0.00
	2.5	0.27	0.27	0.16	0.00	0.00
	3.0	0.31	0.31	0.16	0.00	0.00
S2	1.0	0.06	0.04	0.29	0.00	0.00
	1.5	0.23	0.18	0.16	0.00	0.00
	2.0	0.26	0.24	0.13	0.00	0.00
	2.5	0.32	0.32	0.16	0.02	0.00
	3.0	0.37	0.46	0.13	0.04	0.00
S3	1.0	0.12	0.08	0.21	0.00	0.00
	1.5	0.24	0.21	0.18	0.00	0.00
	2.0	0.32	0.27	0.14	0.00	0.00
	2.5	0.38	0.43	0.15	0.02	0.00
	3.0	0.41	0.55	0.14	0.07	0.01
S4	1.0	0.20	0.16	0.17	0.00	0.00
	1.5	0.32	0.35	0.16	0.01	0.00
	2.0	0.39	0.38	0.15	0.05	0.00
	2.5	0.51	0.60	0.19	0.09	0.04
	3.0	0.56	0.58	0.22	0.13	0.03
S5	1.0	0.00	0.00	0.06	0.13	0.47
	1.5	0.00	0.00	0.04	0.16	0.53
	2.0	0.00	0.00	0.04	0.13	0.50
	2.5	0.00	0.00	0.04	0.16	0.52
	3.0	0.01	0.01	0.07	0.21	0.55
S6	1.0	0.01	0.01	0.04	0.14	0.52
	1.5	0.01	0.01	0.04	0.20	0.53
	2.0	0.01	0.02	0.06	0.20	0.55
	2.5	0.08	0.08	0.03	0.13	0.55
	3.0	0.15	0.15	0.05	0.15	0.44

TABLE VI

ISOLATION AND IDENTIFICATION: PERFORMANCE EVALUATION WITHOUT USING PROPAGATION TREES

[31] G. Boracchi and M. Roveri, "Exploiting self-similarity for change detection," in *International Joint Conference on Neural Networks (IJCNN 2014)*, 2014.

[32] B. Manly, "Exponential data transformations," *The Statistician*, pp. 37–42, 1976.

[33] G. Boracchi and M. Roveri, "A reconfigurable and element-wise ici-based change-detection test for streaming data," in *2014 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA 2014)*, 2014.

[34] 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.

[35] A. Goldenshluger and A. Nemirovski, "On spatial adaptive estimation of nonparametric regression," *Math. Meth. Statistics*, vol. 6, pp. 135–170, 1997.

[36] 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.

[37] D. M. Hawkins, P. Qiu, and C. W. Kang, "The changepoint model for statistical process control," *Journal of Quality Technology*, vol. Vol. 35, No. 4, pp. 355–366, 2003.

[38] 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.

[39] 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.

[40] C. Alippi, G. Boracchi, and M. Roveri, "Ensembles of change-point methods to estimate the change point in residual sequences," *Soft Computing*, vol. 17, no. 11, pp. 1971–1981, 2013. [Online]. Available: http://dx.doi.org/10.1007/s00500-013-1130-7

[41] J. Bai, "Estimation of a change point in multiple regression models," *The Review of Economics and Statistics*, vol. 79, no. 4, pp. 551–563, November 1997.

[42] J. Reeves, J. Chen, X. L. Wang, R. Lund, and L. QiQi, "A review and comparison of changepoint detection techniques for climate data," *Journal of Applied Meteorology and Climatology*, vol. 46, no. 6, pp. 900–915, 2007. [Online]. Available: http://journals.ametsoc.org/doi/abs/10.1175/JAM2493.1

[43] C. Alippi, G. Boracchi, V. Puig, and M. Roveri, "A hierarchy of change-point methods for estimating the time instant of leakages in water distribution networks," in *Proceedings of LEAPS, the 1st Workshop on Learning stratEgies and dAta Processing in nonStationary environments, in 9th AIAI Conference*, September 2013, pp. 1–10.

[44] Y. Lepage, "A combination of Wilcoxon's and Ansari-Bradley's statistics," *Biometrika*, vol. Vol. 58, No. 1, pp. 213–217, April 1974.

[45] H. B. Mann and D. R. Whitney, "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other," *The Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60, 1947. [Online]. Available: http://dx.doi.org/10.2307/2236101

[46] A. M. Mood, "On the asymptotic efficiency of certain nonparametric two-sample tests," *The Annals of Mathematical Statistics*, vol. Vol. 25, No. 3, pp. 514–522, September 1954.

[47] A. G. Tartakovsky and V. V. Veeravalli, "Quickest change detection in distributed sensor systems," in *Proceedings of the 6th International Conference on Information Fusion*, 2003, pp. 756–763.