A Cognitive Monitoring System for Contaminant Detection in Intelligent Buildings

Giacomo Boracchi, Michalis Michaelides and Manuel Roveri

***Abstract*—Intelligent buildings are equipped with sensing systems able to measure the contaminant concentration in the different building zones for safety purposes. The aim of these systems is to promptly detect the presence of a contaminant so that appropriate actions can be taken to ensure the safety of the people. At the same time, these sensing systems, which operate in real-world conditions, suffer from noise and sensor degradation faults. Both noise and faults can induce false alarms (resulting in unnecessary disruptive actions such as building evacuation) or missed alarms (when the presence of a contaminant is not detected).**

This paper proposes a novel cognitive monitoring system for performing contaminant detection in intelligent buildings with real-time point-trigger sensors. The proposed system reduces the occurrence of false alarms by means of a three-layered architecture, which employs cognitive mechanisms to validate possible detections and discriminate between the presence of a real contaminant source and a degradation fault affecting the sensors of the sensing system. In addition, the proposed system is able to isolate the building zone containing the contaminant source (or the faulty sensor) and estimate the onset time of the release (or the fault).

I. Introduction

Nowadays, buildings are becoming increasingly intelligent by incorporating distributed sensing devices and computer technology to adapt and control the indoor environment in order to save energy and create more comfortable, healthy and safe living conditions for their occupants [1]–[3]. The safety of the occupants is directly associated with the Indoor Air Quality, which can be easily compromised by an accident (i.e., CO leakage from a faulty furnace) or a terrorist attack with airborne chemical and biological agents [4]. Under these safety-critical conditions, real-time collected data from sensors that monitor the contaminant concentration can be used to alert the occupants and determine appropriate solutions like indicating safe spaces, or isolating and cleaning contaminated areas. Therefore, the accurate and prompt detection and isolation of contaminant events is an essential task of the intelligent building design.

G. Boracchi and M. 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).

M. 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).

This work was supported by the European Commission 7th Framework Program, under grant INSFO-ICT-270428 (iSense).

Recently, the development of effective, near real-time biological and chemical agent sensors (see [4], [5] and references therein) allows new and highly effective protective measures. These measures can be low-disruption actions in response to the sensors' readings, like automatically modifying the operational mode of a building's Heating Ventilation and Air Conditioning (HVAC) system, or high-disruptive actions like 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 covering wide areas quickly. At their current state, however, these trigger sensors can only support low-disruption actions because they can suffer from high false positive rates [4] (i.e., false alarms induced by 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).

The problem of contaminant event monitoring has received considerable attention in the literature over the last decade. A detailed report on related literature on the inverse tracking of pollutants in both groundwater and air fields can be found in [6]. Some highlights of the proposed methods for contaminant source isolation in indoor building environments include the Bayesian updating method [7], the adjoint probability method [8] and the state space method in [9]. All of these methods, however, require some form of prior knowledge, either in the form of a model of the building and the resulting airflows, or through a constructed scenario database before the event. Differently, [5] 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 noise in order to guarantee a pre-specified false alarm probability. Of course, this is done at the expense of increased false negatives (i.e., missed alarms when the contaminant is present) and large delays in detection, especially in situations of high noise or low contaminant concentration.

Apart from the effects of noise, contaminant detection sensors can also suffer from degradation faults. These types of faults are quite common in real-world sensing systems working in harsh environmental conditions, and can be generally attributed to ageing or thermal drift.

Note that, unless promptly detected and isolated by the monitoring system, sensor degradation faults can increase the occurrence of false positives or negatives, hence making the detection system ineffective.

In this paper, we propose a novel cognitive monitoring system for performing contaminant detection in intelligent buildings with real-time point-trigger sensors. The proposed system reduces the occurrence of false alarms (and of consequent disruptions) by means of a layered architecture and employs cognitive mechanisms to discriminate sensor degradation faults from the presence of a real contaminant source in the building envelope. In addition, the proposed system is able to isolate the building zone containing the contaminant source (or the faulty sensor) and estimate the onset time. To achieve these goals, the proposed monitoring system is organized into a three-layer hierarchical architecture. At the first layer, we rely on Change-Detection Tests (CDTs) based on the Intersection of Confidence Interval (ICI) rule, aiming at providing the prompt detection of even small variations in the concentration of a specific contaminant. The second layer performs a validation of the changes detected at the first layer to reduce the possible occurrence of false alarms and identify possible sensor degradation faults. In the case of a confirmed contaminant source or fault, the second layer is also able to estimate the onset time. Finally, the third layer, based on the information received from the second layer, is able to isolate the zone within the building which contains the contaminant source or the faulty sensor. This information, together with the estimate of when the contaminant release/fault started, can be then utilized by the system operator in order to take the appropriate actions to ensure the safety of the people. A main advantage of the cognitive monitoring system proposed in this paper is that it does not require a model or any other prior information; this makes it appropriate for real-world applications, where little information is commonly available.

The paper is organized as follows. Section II describes the problem statement. The proposed cognitive monitoring system for intelligent buildings is detailed in Section III, while the experimental results are presented in IV. Finally, conclusions and future directions of the work are drawn in Section V.

II. The problem statement

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 of the contaminant provided by the sensor of the i-th zone at time t, which can be modeled as

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

where $c_i(t)$ is the natural concentration of the contaminant at time t, $\Delta_i(t)$ is the amount of contaminant produced by a source at time t, i.e., the anomalous source that needs to be detected by the sensing system, and $\eta_i(t)$ is the zero-mean independent and identically distributed (i.i.d.) sensor noise at time t. The natural concentration of a contaminant in the i-th zone can be zero (i.e., $c_i(t) = 0$), when no contaminant is naturally present in the i-th zone (e.g., toxic gases), or constant (i.e., $c_i(t) = \mu_i, \mu_i > 0$), or may follow a dynamic behavior in the more general case (i.e., $c_i(t) = f(c_i(t-1), c_i(t-2), ...)$. In this paper, we assume that $c_i(t)$ is either constant or zero.

Let i^* (with $1 \leq i^* \leq N$) denote the zone of the building where the contaminant source is inserted (i.e., the *source zone*) and T^* the time of the release. Also, let T_i^* be the time instant when the contaminant first appears in the i-th zone, i.e.,

$$\begin{cases} \Delta_i(t) = 0, & t < T_i^* \\ \Delta_i(t) > 0, & t \geq T_i^* \end{cases} \tag{2}$$

Due to the propagation delays in reaching the various building zones, $T_i^* > T^*$ when $i \neq i^*$. From Eq. (2), it follows that the amount of contaminant present in a specific zone (where $\Delta_i(t) > 0$) can be attributed to two main reasons. First, the i-th zone is the source zone, and in this case $T_{i^*}^* = T^*$. Second, the i-th zone is not the source zone but contaminant has flowed through the building from the source zone to the i-th zone. The propagation of a contaminant in various 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).

We assume that sensor degradation faults occur abruptly and in the absence of any contaminant sources (i.e., $\Delta_i(t) = 0, t \geq 0$). Thus, a sensor degradation at time T^0 is modeled as

$$m_i(t) = \begin{cases} c_i(t) + \delta\eta_i(t), & t \geq T^0 \\ c_i(t) + \eta_i(t), & t < T^0 \end{cases}, \tag{3}$$

where $\delta > 1$ is the perturbation magnitude.

The proposed monitoring system aims at detecting as soon as possible the abnormal presence of a contaminant and estimating the release zone and onset time, while distinguishing between the presence of a contaminant and a degradation fault affecting a sensor.

III. The proposed cognitive monitoring system for intelligent buildings

The proposed cognitive monitoring system, which is summarized in Fig. 1, is characterized by a hierarchical architecture composed of the following three layers:

INPUT at Unit i: $\{m_i(t), t = 1, \ldots, L\}$;
Train the ICI-based CDT from $\{m_i(t), t = 1, \ldots, L\}$;
while *(1)* **do**
  Acquire $m_i(t)$;
  $[d_i(t)]$ =ICI-based CDT($m_i(t)$);
  if $(d_i(t) = 1)$ **then**
    $[H_i, \tau] = CPM\big(\{m_i(j), j = 1, \ldots, t\}\big)$;
    if $(H_i = 1)$ **then**
      $\widehat{i} = i$;
      $\widehat{T} = t$;
      $A_\tau = \{m_i(j), j = 1, \ldots, \tau\}$;
      $B_\tau = \{m_i(j), j = \tau + 1, \ldots, \widehat{T}\}$;
      p_{MW} =Mann-Whitney(A_τ, B_τ);
      p_{MO} =Mood(A_τ, B_τ);
      if $(p_{MW} \leq p_{MO})$ **then**
        Detection of contaminant in zone $\widehat{i}$
        and the time of detection is τ;
      else
        Fault affecting sensor of zone $\widehat{i}$;
        the fault started at time τ;
      end
    end
  end
end

Algorithm 1: The first two layers of the proposed cognitive monitoring system for intelligent buildings executed at the i-th Unit.

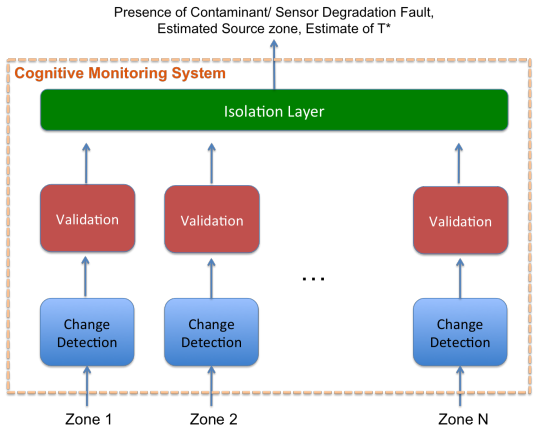

Fig. 1. The three-layer hierarchical architecture of the proposed cognitive monitoring system.

1) the *change-detection layer*, which is composed of a set of CDTs running at the N sensors of the intelligent building, is responsible for monitoring in an online manner the concentration of the contaminant. The goal of this layer is to guarantee the prompt detection of any anomalous concentration. This layer relies on measurements coming from a single sensor, hence, it can be executed locally directly at the sensor level (when enough computational power is available);
2) the *validation layer* aims at reducing the false positives raised by the change-detection layer by validating (or not) any detected change. In addition, this layer is able to distinguish between the real presence of the contaminant and a sensor degradation fault. This layer also relies on measurements coming from a single sensor, hence, it can be executed locally directly at the sensor level (when enough computational power is available);
3) the *isolation layer* is responsible for identifying the building zone containing the contaminant source or the faulty sensor. This layer analyses the detection reports of all the sensors within the building coming from the second layer; therefore, it is executed in a centralized manner.

The three layers are detailed in the next three subsections, while the algorithm of the proposed cognitive monitoring system for the first two layers is detailed in Alg. 1.

A. The change-detection layer

Among the wide range of sequential CDTs [10]–[14] available in the literature, we focused on the ICI-based CDT [15][1]. The peculiarity of this CDT is the use of the ICI rule, that is a technique to define adaptive supports for regularizing data through polynomial regression. The ICI-based CDT is nonparameteric in the sense that it does not assume to know the probability density function of the data-generating process either in stationary conditions or after the change. Furthermore, it is characterised by a reduced computational complexity and this is very important when analysing online streaming data. Basically, the ICI-based CDT running at the i-th sensor operates by assessing the stationarity of i.i.d. and Gaussian distributed features (in the stationary case) extracted from the datastream m_i. These features, which represent condensed information about m_i, are then processed according to the ICI-rule [16] to detect changes in their expectation. The features considered by the ICI-based CDT presented in [15] are the sample mean and variance computed on disjoint subsequences of ν observations.

In particular, at the i-th sensor, the first feature, i.e., the sample mean M_i, that is here computed on the s-th subsequence

$$M_i(s) = \frac{1}{\nu} \sum_{t=(s-1)\nu+1}^{\nu s} m_i(t), \tag{4}$$

approaches the Gaussian distribution thanks to the central limit theorem, while the second feature is a power-law

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

transformation of the sample variance, i.e.,

$$V_i(s) = \left(\frac{S_i(s)}{\nu-1}\right)^{h_0}, \text{being}$$
$$S_i(s) = \sum_{t=(s-1)\nu+1}^{\nu s} \big(m_i(t) - M_i(s)\big)^2.$$

This power-law transform yields values of V approximately Gaussian distributed; its exponent h_0 is computed as $h_0 = 1 - (\kappa_1\kappa_3)/3\kappa_2^2$, where κ_i is the i-th cumulant of the distribution of the sample variance.

Let $\{m_i(t), t = 1, \ldots, L\}$ be the training sequence for configuring the CDT at the i-th sensor, where the L samples have been acquired in absence of contaminant sources or faults. The CDT configuration consists of computing the parameter h_0 and the sample mean and standard deviation of the features (4) and (5). Further details on the configuration of ICI-based CDTs can be found in [11].

As a result, the ICI-based CDT$(m_i(t))$ in Alg. 1 monitors both $M_i(s)$ and $V_i(s)$ and, hence, it is able to detect any change affecting the first two moments of the unknown distribution characterizing $m_i(t)$.

Let $\widehat{i}$ be the zone where the first ICI-based CDT detects a change in the sensor measurements and let $\widehat{T}$ be the time instant in which this change has been detected.

B. The Validation Layer

Every time a detection has been raised by the change-detection layer, the validation layer is activated to reduce the occurrence of false positive detections and to prevent the activation of unnecessary emergency procedures. To achieve this goal, we consider change-point methods (CPMs) [17] that are statistical tests able to assess whether a given data-sequence contains (or not) a change point. Interestingly, CPMs do not require any training sequence since they can be directly applied to the sequence.

In more detail, a CPM is applied to a sequence of measurements coming from the zone where the contaminant has been detected, i.e., $\mathcal{C}_{\widehat{i}} = \{m_{\widehat{i}}(t), 1 \le t \le \widehat{T}\}$. To simplify the notation in what follows we omit the index $\widehat{i}$. The CPM operates as follows: for each time instant $1 \le x \le \widehat{T}$, $\mathcal{C}_{\widehat{i}}$ is split into two parts,

$$A_x = \{m_{\widehat{i}}(t),\ t = 1, \ldots, x\},$$
$$B_x = \{m_{\widehat{i}}(t),\ t = x+1, \ldots, \widehat{T}\}.$$

A specific test statistic $\mathcal{T}$ is then computed at x as

$$\mathcal{T}_x = \mathcal{T}(A_x, B_x),$$

to assess the difference between A_x and B_x.

The values of $\mathcal{T}_x$ are computed for all the samples $1 \le x \le \widehat{T}$, yielding $\{\mathcal{T}_x, x = 1, \ldots, \widehat{T}\}$. Let $\mathcal{T}_{\text{M}}$ be the maximum value of the statistic $\mathcal{T}$ over all the samples, i.e.,

$$\mathcal{T}_{\text{M}} = \max_{x=1,\ldots,\widehat{T}} (\mathcal{T}_x) \tag{5}$$

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

$$\tau = \operatorname*{argmax}_{x=1,\ldots,\widehat{T}} (\mathcal{T}_x).$$

The value of $\mathcal{T}_{\text{M}}$ is then compared with a predefined threshold $h_{\widehat{T},\alpha}$, which is function of the statistic $\mathcal{T}$, the cardinality $\widehat{T}$ of $C_{\widehat{i}}$ and a given confidence level α that sets the percentage of type I errors (i.e., false positives) of the CPM. If $\mathcal{T}_{\text{M}}$ is larger than $h_{\widehat{T},\alpha}$, there is enough statistical evidence for the CPM to confirm the presence of a change in $\mathcal{C}_{\widehat{i}}$. On the other hand, when the test statistic does not exceed the threshold, the CPM is not able to detect the change point. Hence, the outcome of the CPM can be defined by the following binary variable:

$$H_{\widehat{i}} = \begin{cases} 1 & \text{if } \mathcal{T}_{\text{M},i} \ge h_{\widehat{T},\alpha} \\ 0, & \text{if } \mathcal{T}_{\text{M},i} < h_{\widehat{T},\alpha} \end{cases}. \tag{6}$$

In the case that $H_{\widehat{i}} = 1$, the detection is validated and τ is identified as the time instant when the contaminant (or the fault) first appeared in the $\widehat{i}$-th building zone. Otherwise, when $H_{\widehat{i}} = 0$, the detection is not validated and the CDT at the detection layer is newly configured from the training sequence $\{m_{\widehat{i}}(t),\ t = 1, \ldots, L\}$.

Among the wide range of test statistics in the literature, we consider the following three nonparametric statistics: the Mann-Whitney [18], the Mood [19] and the Lepage [20]. These statistics can be used to detect, without any a priori assumption on the data distribution, changes in the location, in the scale and in both location and scale, respectively. We expect an anomalous contaminant source, which changes the expectation m_i, to be detected as a change in location, while sensor degradation, which affects the variance of m_i, to be detected as a change in the scale. One of the most critical aspects when dealing with CPMs is the definition of thresholds $h_{\widehat{T},\alpha}$. Since the analytical derivation of the threshold is in general difficult to obtain, in practice, it is often numerically computed by means of Montecarlo simulations as in [13], [17].

Every time a change has been validated by the CPM on the Lepage statistic, a cognitive analysis is activated to determine if it is due to a contaminant source or a sensor degradation fault. In practice, we determine whether the detected variation is associated to a change either in the location or in the scale of m_i. This is a valuable information because, according to (1) and (3), $\Delta_i(t)$ induces only changes in the expected value of $m_i(t)$, while sensor degradation faults affect only the variance of $m_i(t)$.

For this analysis we pursue the approach described in [13], which separately runs the Mann-Whitney and the Mood hypothesis test on the partitioning of $\mathcal{C}_{\widehat{i}}$ maximizing the Lepage test statistic, i.e., A_τ and B_τ, and compares the p-values of the two hypothesis tests. When the p-value of the Mann-Whitney test is lower than that of Mood statistics, then there is a stronger statistical evidence for a change in

the mean of the observations. In this case we can safely exclude the sensor degradation fault, since the p-values indicate a variation in the expected value of m_i. On the other hand, when the p-value of the Mood test is lower than that of the Mann-Whitney, we associate the detection to a change in the variance of m_i and, in turn, this can be safely associated to a sensor degradation fault.

Note that, in principle, the feature detecting the change in the ICI-based CDT (either M_i or V_i) could also be used to discriminate sensor degradation faults from the presence of a contaminant source. However, the analysis on the p-value is straightforward, since the values of Mann-Whitney and Moods statistics are computed when running the Lepage CPM. Furthermore, the p-values of these hypothesis tests represent confidence indicators which are not available when decisions are made by looking at the features of the ICI-based CDT.

C. The Isolation Layer

The aim of this layer is to identify the zone in which either the contaminant has been released or a degradation fault has affected the sensor. Here, the logic is rather simple: (i) When a contaminant event is detected and validated by the second layer, the first zone where the change has been detected and validated is considered to be the estimated source zone. (ii) When a sensor degradation fault is detected and validated by the second layer, then a sensor degradation fault is isolated in the respective zone which reported the detection. In other words, once a change has been validated with (6), $\widehat{i}$ becomes the zone where the contaminant source (or the fault) is isolated. At this stage, we should point out that more sophisticated solutions could be considered, e.g., by exploiting the topology of the building or the detection time instants in the different zones. We plan to investigate these issues in our future research.

IV. Experimental Results

In order to evaluate the effectiveness of the proposed solution we performed a wide experimental campaign on scenarios including both the real presence of a contaminant and sensor degradation faults. These scenarios have been generated with the Matlab-CONTAM toolbox for contaminant event monitoring in intelligent buildings described in [21]. The Holmes House [22] depicted in Figure 2 has been considered as the reference intelligent building to generate such scenarios.

A. Description of the Considered Scenarios

In particular, the following four scenarios have been considered:

- **S1**: Single contaminant source at a constant rate of 100 g/h;

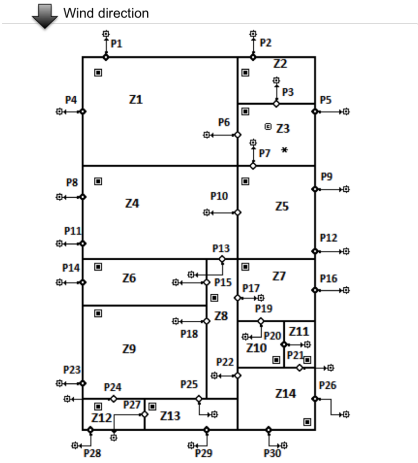

Fig. 2. The Holmes House.

- **S2**: Single contaminant source with a variable generation rate, alternating every 2 hours between the values of 50 and 100 g/h;
- **S3**: A sensor degradation fault modelled as in Eq. (3) with magnitude perturbation $\delta = 1.33$;
- **S4**: Similar to scenario S3 but the perturbation magnitude δ is set to 1.5.

In all the considered scenarios, all the sensors acquire 1 measurement per minute. The simulation time is 48 hours, yielding 2880 samples for each sequence. Both the release of contaminant and faults starts after 25 hours, i.e., $T^* = T^0 = 1500$ samples. Both the contaminant source and the sensor faults have been placed in zone 3 (i.e., $i^* = 3$), while the wind direction is 0° as shown in Figure 2. Noise $\eta_i(t)$ is assumed to be i.i.d. following a Gaussian Distibution $\mathcal{N}(0, \sigma^2)$. In our experiments we considered five different values of σ, i.e., $\sigma = \{1, 1.5, 2, 2.5, 3\}$. An example dataset of Scenario S1 with $\sigma = 2$ is shown in Figure 3.

The proposed solutions has been configured on the first $L = 400$ samples. Similarly to [15], the parameter Γ of the ICI-based CDT has been fixed to 2 (see [11] for details about this parameter). For the comparison, we also implemented the STA algorithm presented in [5], with parameters 400, 20 and 100, corresponding to the Background window, the Guard window and the Present window, respectively. In addition, to ease the comparison, the parameter γ_{STA} of STA has been fixed to 4.5 to guarantee false positive rates similar to those provided by the proposed solution.

Simulation results are averaged over 500 runs.

B. Figures of Merit

To evaluate the proposed approach we considered the following six figures of merit:

- False Positive Rate (FPR), the percentage of experiments in which the change has been detected before

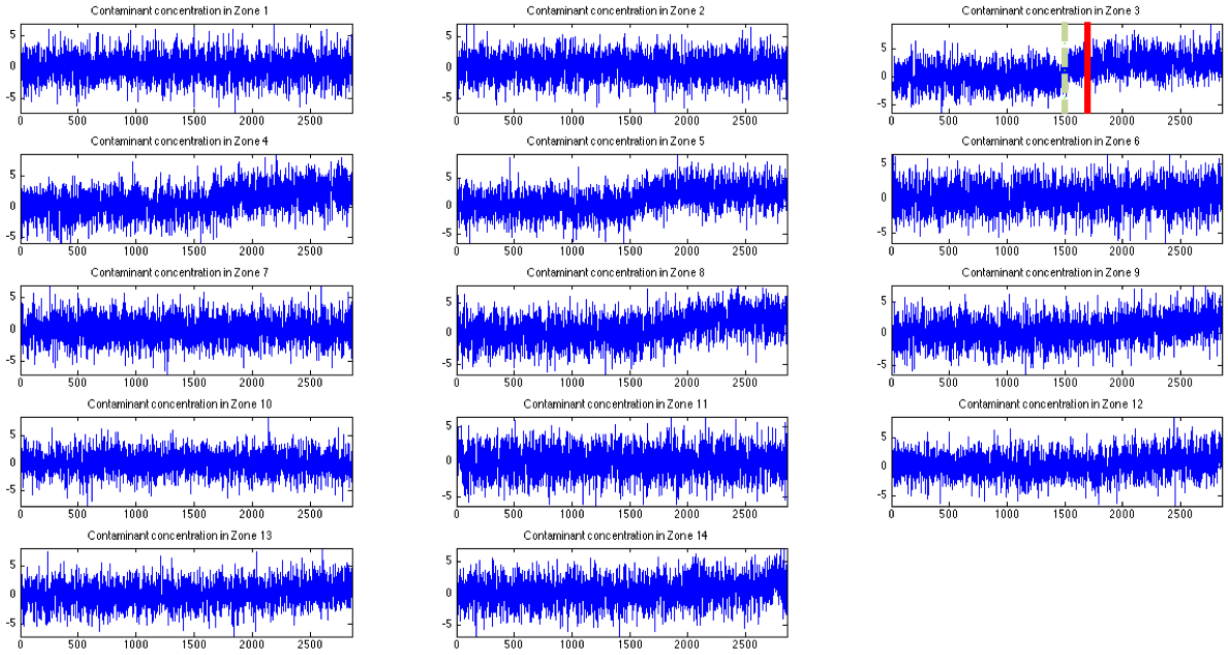

Fig. 3. An example of Scenario S1 with $\sigma = 2$. The dotted green line refers to the time instant the contaminant is released in the source zone (i.e., $t = 1500$). The red line identifies the detection time-instant provided by the proposed solution.

T^*;

- 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 $\widehat{T} - T^*$;
- Refinement Delay RD (in samples), that is the average value of $\tau_{\widehat{i}} - T^*$;
- ϵ_{iso}, the percentage of experiments in which the zone containing the contaminant source (S1, S2) or the sensor fault (S3, S4) has not been correctly isolated;
- ρ_C, the percentage of experiments in which the change has been correctly recognized; i.e. as a contaminant source for S1, S2 and as a sensor degradation fault for S3, S4.

C. Analysis of the results

Table I shows the comparison between the detection ability of the proposed solution and that of STA. Several comments arise.

First, the proposed solution is characterized by very low values of FPR and FNR. This is of paramount importance in real-life working conditions to reduce unnecessary emergency reactions as well as provide prompt detection to variations in m_i. On the contrary, to guarantee FPR similar to that achieved by the proposed solution, STA requires a high value of γ_{STA} (as described above γ_{STA} has been fixed to 4.5) that induces high FNR. This behavior is quite common in threshold-based solutions (such as STA): high values of the threshold guarantee low FPRs but at the expense of increased FNR.

Scen.	σ	Proposed Solution FPR	Proposed Solution FNR	Proposed Solution DD	STA FPR	STA FNR	STA DD
S1	1.0	0.00	0.00	103.33	0.00	0.43	83.78
	1.5	0.00	0.00	132.26	0.00	0.90	114.58
	2.0	0.00	0.00	165.80	0.01	0.95	181.35
	2.5	0.00	0.00	196.23	0.00	0.96	308.25
	3.0	0.00	0.00	227.80	0.00	0.98	117.33
S2	1.0	0.00	0.00	153.25	0.01	0.78	175.34
	1.5	0.00	0.00	191.90	0.00	0.92	213.21
	2.0	0.00	0.00	227.68	0.00	0.97	451.21
	2.5	0.00	0.00	276.10	0.01	0.97	444.18
	3.0	0.01	0.00	319.24	0.00	0.98	172.40
S3	1.0	0.00	0.00	400.78	0.01	0.92	138.35
	1.5	0.00	0.01	405.25	0.00	0.93	139.54
	2.0	0.01	0.00	398.93	0.00	0.92	146.45
	2.5	0.00	0.01	391.62	0.00	0.93	118.32
	3.0	0.00	0.00	415.19	0.00	0.93	108.74
S4	1.0	0.00	0.00	245.81	0.00	0.78	108.36
	1.5	0.00	0.00	241.88	0.00	0.81	119.55
	2.0	0.00	0.00	239.64	0.00	0.78	107.13
	2.5	0.00	0.00	243.85	0.01	0.80	85.39
	3.0	0.00	0.00	247.37	0.01	0.79	98.21

TABLE I
COMPARISON BETWEEN THE PROPOSED SOLUTION AND STA ON THE CONSIDERED SCENARIOS.

Second, as expected, the DD of the ICI-based CDT increases with σ. This is due to the fact that the ratio between the magnitude of the change and the standard deviation of the noise reduces with σ, hence requiring more samples before detecting the change. As expected, even the performance of STA decreases with σ. Even in this case,

Scen.	σ	Proposed Solution RD (Median of τ)	ϵ_{iso}	ρ_C
S1	1.0	0.239 (1521.00)	0.062	0.979
	1.5	5.213 (1522.00)	0.064	0.994
	2.0	19.392 (1526.00)	0.108	0.993
	2.5	15.620 (1526.00)	0.138	0.995
	3.0	9.915 (1526.00)	0.146	0.995
S2	1.0	28.166 (1531.00)	0.070	0.970
	1.5	43.146 (1545.00)	0.092	0.987
	2.0	52.389 (1562.00)	0.132	0.984
	2.5	53.234 (1567.00)	0.234	0.974
	3.0	48.344 (1562.00)	0.272	0.995
S3	1.0	-4.615 (1512.00)	0.080	0.993
	1.5	6.232 (1514.50)	0.076	0.996
	2.0	13.522 (1516.00)	0.066	0.994
	2.5	3.647 (1519.00)	0.068	0.998
	3.0	16.979 (1514.00)	0.074	0.985
S4	1.0	-15.544 (1500.00)	0.044	0.994
	1.5	-13.626 (1500.00)	0.048	0.989
	2.0	-11.408 (1501.00)	0.048	0.996
	2.5	-11.211 (1500.00)	0.064	0.998
	3.0	-9.697 (1500.00)	0.060	0.994

TABLE II
SIMULATION RESULTS ABOUT THE FAULT ISOLATION AND IDENTIFICATION OF THE PROPOSED SOLUTION. AMONG THE PARENTHESES WE DETAIL THE MEDIAN VALUE OF $\tau_{\widehat{i}}$.

the threshold-based solutions suffer from increases of the measurement noise.

Third, the proposed solution is able to guarantee very good results on the scenarios both with real contaminant (i.e., S1 and S2) and with faults (i.e., S3 and S4). Interestingly, the DDs of S3 are larger than those of S4 and this is reasonable since the magnitude of the fault is larger in S4 than S3.

The fault isolation and identification results of the proposed solution are detailed in Table II. Several interesting comments can be made.

First, the value of RD is low in all the considered scenarios meaning that the estimate of the time instant the change started (either release of contaminant or fault) provided by the validation layer is good. Interestingly, Scenarios S1 and S2 are generally characterized by overestimated values of τ and this is evident by looking at the value within the parenthesis that measures the median of τ over all the different runs. The reason of this behavior can be associated to the fact that the CPM identifies the point that maximizes the difference between two partitions (see Eq. (5)): in case of incipient changes of $m_i(t)$ (as in the case of real presence of contaminant), the point that maximizes the change is not at the beginning of the change but after (hence $\tau_{\widehat{i}}$ typically overestimates T^*). Interestingly, this effect is not present in scenarios S3 and S4 where the variance changes abruptly (and this is evident by looking at the median values of τ). It is also worth noting that changes in Scenario S4 are easier to be detected, and this explains why Scenario S4 is generally characterized by median values of τ that are very close to the time instant the fault occurred (i.e., $t = 1500$).

Second, the validation layer is particularly effective in distinguishing between the real presence of contaminant and a sensor degradation fault affecting a sensor and this is particularly evident by looking at the values of ρ_C for all the considered scenarios. This proves the effectiveness of using the Mann-Whitney and Mood hypothesis tests to discriminate between changes in the expected value or in the variance of m_i.

Third, the simple isolation phase suggested in Section III-C is effective in isolating the zone within the building in which either the contaminant is released or the fault occurred. Interestingly, errors in the isolation phase can be either associated to false positives (i.e., a false detection can occur in any zone of the intelligent building) or, in the case of a real contaminant, to the fact that the change is detected earlier in other zones of the building than the source zone. These are the reasons why ϵ_{iso} is generally larger in S1 and S2 than in S3 and S4 and why it increases with σ.

In addition, Figure 4 shows the histogram of the detection time-instant in all the zones of the building provided by the proposed solution for the case of Scenario S1 and $\sigma = 2$. As expected, the earliest detection occurs in the source zone (i.e., zone 3). This corroborates the idea of the isolation phase in which the zone where the first detection occurs is considered to be the source zone. Interestingly, the zones that received flow from zone 3 (i.e., zones 5, 4, 8, etc..) are characterized by detection slightly delayed with respect to zone 3. It is worth noting that zones 1 and 2 do not receive flow from 3 according to the wind direction: this is the reason why there is no detection in these zones (i.e., within the histogram all the detection times are set to 0). Differently, zones 6, 7, 10 and 11 are characterized by longer propagation delays of the contaminant (because of the topology of the building and the wind direction) and this is the reason why no detection has occurred in the considered time horizon.

V. CONCLUSIONS

This paper describes a novel cognitive monitoring system for contaminant detection in intelligent buildings. The proposed system reduces false positives and false negatives, while guaranteeing the prompt detection of a contaminant and the ability to discriminate it from sensor degradation faults. In addition, information about the zone and the time instant where the contaminant/fault has started are provided to support emergency response and guarantee people safety.

The next step of this work is to exploit the topology of the building and the detection reports raised by the sensors to improve the cognitive capabilities of the isolation and identification phases. This will allow us to deal with faults affecting the expected value of m_i, as well as more complex situations involving multiple sources of contaminant or multiple sensor faults.

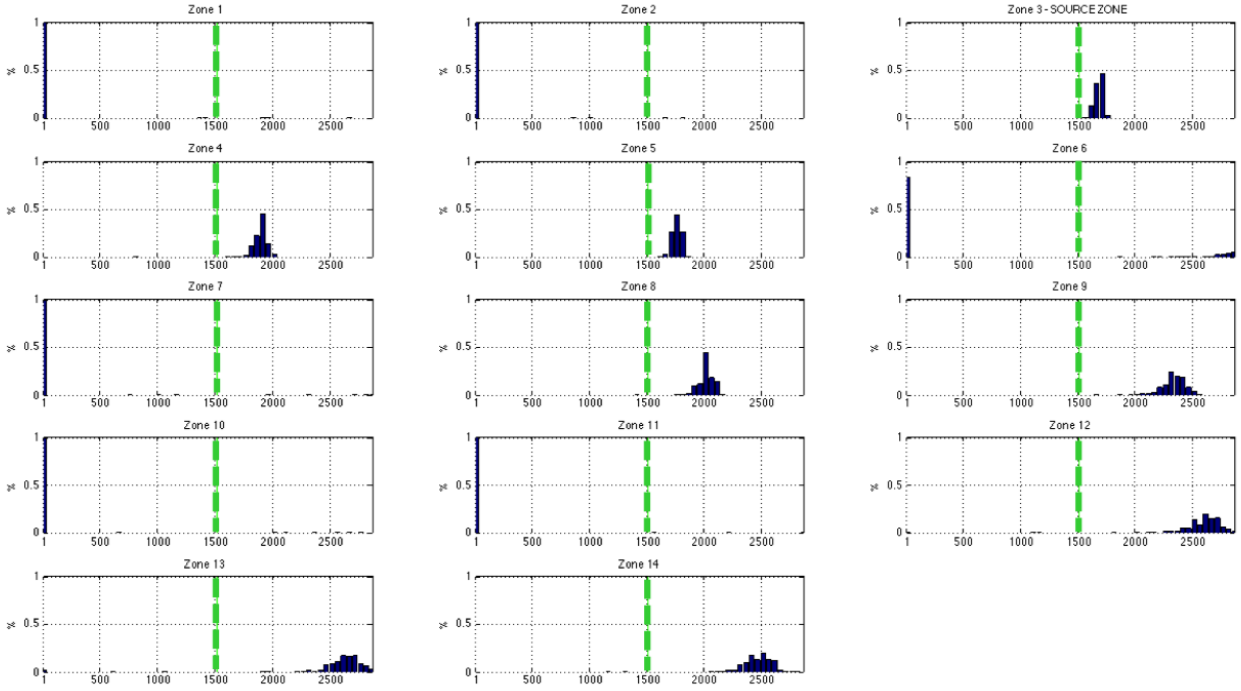

Fig. 4. Histograms of the detections of all the sensors. The source zone is 3, the wind direction is 0^o and $\sigma = 2$. The green dotted line represents the time-instant in which the contaminant is inserted in the source zone.

REFERENCES

[1] J. Braun, "Intelligent building systems-past, present, and future," in *Proc. of Am. Control Conf. (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] S. Katipamula and M. Brambley, "Methods for fault detection, diagnostics, and prognostics for building systems-A review, part I," *HVAC&R Research*, vol. 11, no. 1, pp. 3–25, 2005.

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

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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, no. 0, pp. 140 – 152, 2014.

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

[11] 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.

[12] A. G. Tartakovsky, B. L. Rozovskii, R. B. Blaek, and H. Kim, "Detection of intrusions in information systems by sequential change-point methods," *Stat. Meth.*, vol. 3, no. 3, pp. 252 – 293, 2006.

[13] 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.

[14] 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.

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

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

[17] 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.

[18] 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

[19] 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.

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

[21] 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.

[22] 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.