



**HAL**  
open science

# Investigation of indoor air quality in a low energy high school building combining micro gas sensors and unsupervised learning

Benjamin Hanoune, Alexandre Caron, Nathalie Redon, Patrice Coddeville

## ► To cite this version:

Benjamin Hanoune, Alexandre Caron, Nathalie Redon, Patrice Coddeville. Investigation of indoor air quality in a low energy high school building combining micro gas sensors and unsupervised learning. IOP Conference Series: Materials Science and Engineering, 2019, IOP Conference Series: Materials Science and Engineering, 609, pp.042027. 10.1088/1757-899x/609/4/042027 . hal-02375462

**HAL Id: hal-02375462**

**<https://hal.univ-lille.fr/hal-02375462>**

Submitted on 5 Jan 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

PAPER • OPEN ACCESS

## Investigation of indoor air quality in a low energy high school building combining micro gas sensors and unsupervised learning

To cite this article: B Hanoune *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **609** 042027

View the [article online](#) for updates and enhancements.

# Investigation of indoor air quality in a low energy high school building combining micro gas sensors and unsupervised learning

**B Hanoune<sup>1,\*</sup>, A Caron<sup>1,2</sup>, N Redon<sup>2</sup>, P Coddeville<sup>2</sup>**

<sup>1</sup> Univ. Lille, CNRS, UMR 8522 – PC2A – Physicochimie des Processus de Combustion et de l'Atmosphère, F-59000 Lille, France

<sup>2</sup> IMT Lille Douai, Univ. Lille, SAGE Dept., F-59000 Lille, France

\* benjamin.hanoune@univ-lille.fr

**Abstract.** Because of their size and price, miniature gas sensors are good candidates for long-term, large-scale, continuous monitoring of the air quality in confined environments, even in the presence of occupants. In spite of their still somewhat limited metrological performances, these tools are able to provide relevant information on the pollutants spatial and temporal evolution. They can therefore be used to identify pollution sources and automatically control ventilation and remediation systems, provided they are associated with adequate data treatment procedures. In the present study, four sensors for the detection of CO<sub>2</sub>, NO, NO<sub>2</sub> and O<sub>3</sub> have been deployed, without previous calibration, inside a classroom of a low energy high school building, together with standard analytical instruments. The data are analyzed with a procedure based on the bisecting K-means algorithm. This unsupervised classification allows the identification of similar measurements, which can be merged into clusters. An excellent agreement has been found between the classification results provided by the analyzers and by the sensors, even if the latter were not calibrated before deployment. These results validate the data treatment methodology proposed in this work, and demonstrate the potential of using commercial micro sensors in real conditions.

## 1. Introduction

Air quality in confined environments has been recognized as a major health problem worldwide [1]. Yet our knowledge and understanding of these environments and of the chemical mechanisms leading to high pollutants concentrations indoors is still limited. There is therefore a need for large-scale continuous measurements of the indoor air quality (IAQ) for research purposes, as well as for regulatory and operational purposes, such as the mandatory or voluntary IAQ assessment in buildings, or the triggering and control of a ventilation and/or air treatment system. There is also an increasing awareness about IAQ, resulting in individuals or constituted bodies demanding information about the air they breathe, and willing to participate in the acquisition of the data.

While the conventional analytical techniques, such as gas analyzers or chromatographs, used for laboratory research, or for regulatory monitoring of the air quality outdoors, are well suited for research oriented measurement campaigns, they are not adapted for real-time monitoring of occupied indoor environments. Indeed these often bulky instruments generate nuisances, such as noise or vibrations, which would prevent normal occupancy or activities in the room under investigation. In addition, they are highly technical instruments, with a response sometimes delayed. Finally yet importantly, they are too expensive to be deployed massively in many places.

In recent years, there has been a considerable development of gas and PM microsensors [2], that emerge as promising tools for indoor air quality monitoring [3]. Several studies have investigated their performances and limitations in real or realistic environments [4]. Still, metrological issues remain,



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

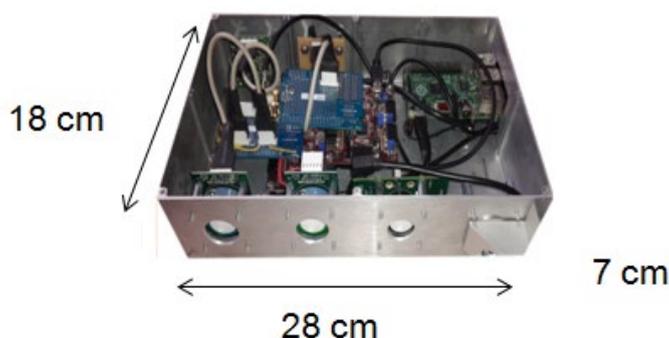
such as cross-sensitivity for some compounds, dependence of the signal with humidity or temperature, drift with time of the sensor response [5]. It is therefore necessary to develop adequate procedures to analyze the data from the sensors.

The goal of the present work is to develop IAQ analysis procedures based exclusively on the parameters measured by sensors located indoors, taking into account possible sensor calibration issues. We present here the results of an approach based on unsupervised clustering [6]. Unsupervised classification refers to algorithms that require no training set (blind partitioning), no a priori knowledge of the structure of the dataset, and automatically define the different classes. This approach is therefore adapted to our goal. The unsupervised classification method selected here is based on the K-means algorithm.

## 2. Materials and methods

### 2.1. Instruments and measurement conditions

The experimental results analyzed here were acquired during an intensive measurement campaign in February and March 2015 in a junior high school in northern France. This campaign was part of the MERMAID study (2012-2015), which aimed to investigate the specificities, if any, of air quality within low energy consumption public buildings [7]. Along with all the conventional or advanced instruments deployed, we also installed miniature sensors nodes (figure 1) developed in the lab, based on commercial electrochemical sensors (Alphasense, B4 series) for the quantification of NO, NO<sub>2</sub>, and O<sub>3</sub>, and a NDIR CO<sub>2</sub> sensor (Alphasense IRC-A1), together with non-selective metal oxide sensors. Data collection, storage and transmission were achieved with a Raspberry Pi B+ and an Arduino board. The performances and limitations of these sensor boxes for the investigation of indoor air quality events have been previously described [3]. The sensor boxes were installed at different places within the room under investigation, on the floor, at mid-height, representative of the exposure of the schoolchildren, at the air inlet and outlet on the ceiling.



**Figure 1.** Sensor box used during the MERMAID campaign.

### 2.2. Experimental datasets

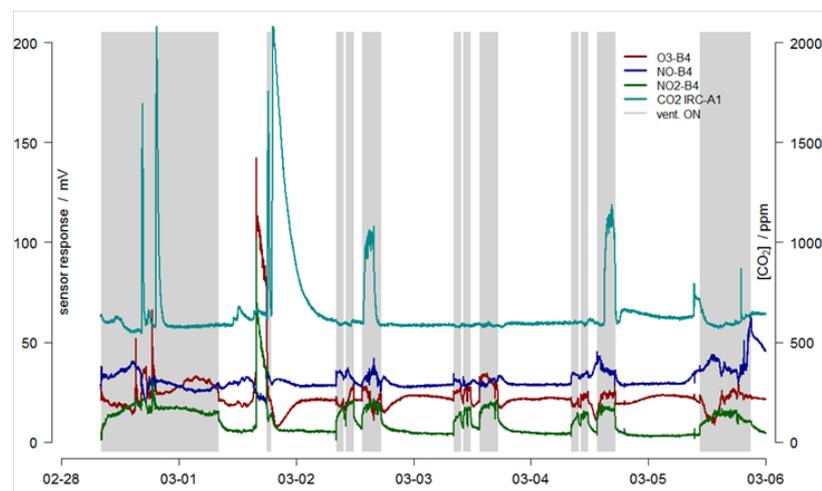
In the present paper, we focus only on the analysis of the signals from the selective sensors: NO, NO<sub>2</sub>, O<sub>3</sub> and CO<sub>2</sub>, during the intensive campaign from February 28 to March 6, 2015, averaged over 1 minute (total of 8160 datapoints per sensor). In the absence of specific sources, such as combustion devices, NO<sub>x</sub> and ozone are mostly pollutants coming from outdoors. CO<sub>2</sub> indoors mostly comes from the respiration of the occupants, with possible additional intake from outdoors. The CO<sub>2</sub> sensor directly outputs the concentration, while the electrochemical sensors output is a voltage. The sensors were not calibrated before their installation in the room under investigation. The same species were also simultaneously measured with calibrated online gas analysers (Thermo Scientific 42i and 49i for NO<sub>x</sub> and O<sub>3</sub>) and with a Testo 480 probe for CO<sub>2</sub>.

### 2.3. K-means classification

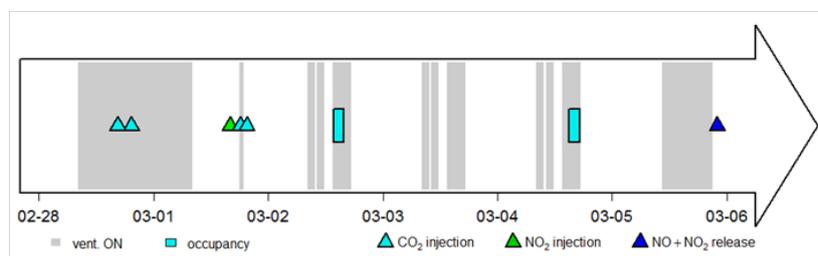
The analysis of the data is performed using the bisecting K-means procedure. It is an iterative centroid-based partitioning algorithm, widely used in data mining [6], thanks to its simplicity, ease of implementation and use, speed of convergence, and ability to process datasets with missing values. This algorithm, devised decades ago, presents surprisingly good performances, which are still not fully understood. The K-means algorithm has been applied here to the raw output of the sensors, except for the CO<sub>2</sub> concentration for which the logarithm (base 10) has been considered, because of the wide range of values. In addition to the data, the only necessary input from the user is the number of clusters into which the data must be split.

### 3. Results and discussion

Figure 2 presents the time series of the sensors signals, together with the ventilation status of the room. The different air quality events that took place are shown in figure 3. These include CO<sub>2</sub> injections to determine the air exchange rate, with (February 28) and without (March 1<sup>st</sup>) ventilation, two periods of occupancy (March 2 and 4), one injection of NO<sub>2</sub> on March 1<sup>st</sup>, and an involuntary NO and NO<sub>2</sub> release on March 5.

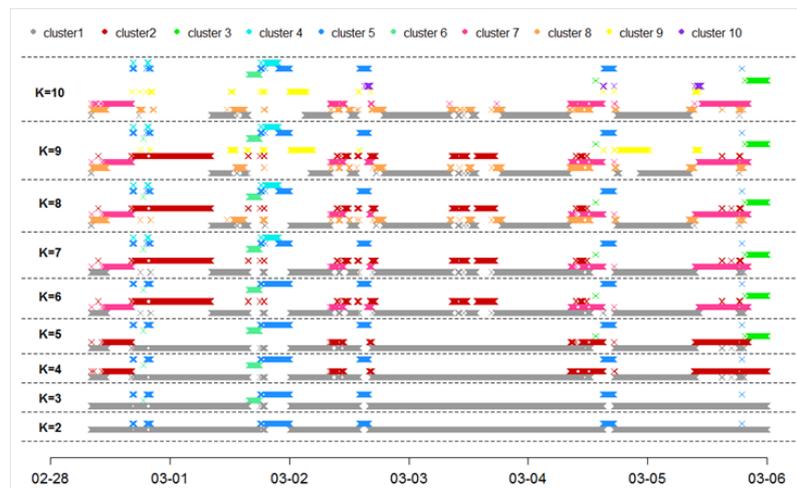


**Figure 2.** Time series of the sensors signals during the MERMAID measurements. Grey areas indicate the periods with active ventilation.



**Figure 3.** Specific events in the room during the measurements.

The output of the clustering procedure is shown on figure 4, for an increasing predetermined number of clusters. When the number of clusters increases up to 7, these clusters can be related to the events taking place in the room. Additional clusters cannot be related to specific air quality events, so we will restrict the discussion to the case with 7 clusters which seems to be optimal in our experiment.



**Figure 4.** Evolution of the clustering of the sensors signals.

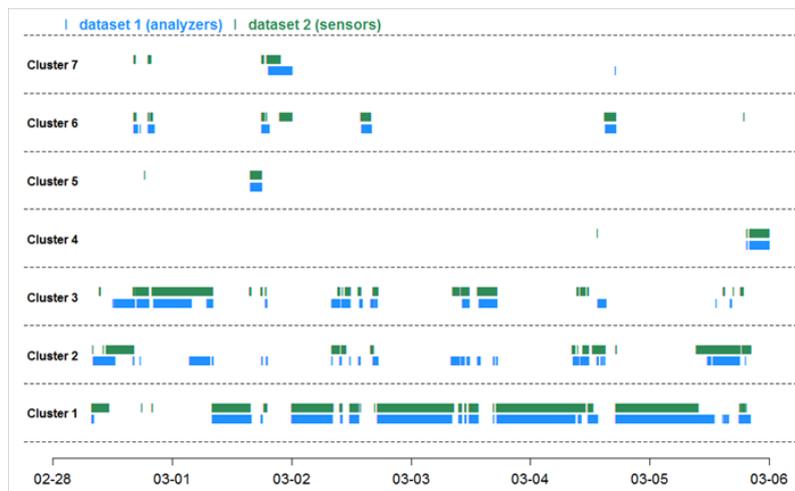
The centers (mean voltage or concentration) of the clusters are summarized in table 1. Cluster 1 groups the majority of the data points (4394 observations), and is characterized with low levels of pollutants. This cluster corresponds to the periods with no ventilation, and therefore no intake of outdoor pollutants. Cluster 2 (1246 observations) has a higher response from the NO and NO<sub>2</sub> sensors, while cluster 3 (1515 observations) has in addition a higher response of the O<sub>3</sub> sensor. These two clusters correspond to periods when the outdoor pollutant contribution is significant, that is when the ventilation is active. Cluster 4 (225 observations) appears at the end of the measurement period, with a significant increase in the NO sensor response, and corresponds to the NO and NO<sub>2</sub> spill. Cluster 5 (129 observations) corresponds to the maximal value of the O<sub>3</sub> and NO<sub>2</sub> sensor, and is related to the NO<sub>2</sub> injection in the room (the high cross-response of these two gases has been previously reported). Clusters 6 (208 observations) and 7 (443 observations), with significant concentrations of CO<sub>2</sub>, match with the CO<sub>2</sub> injections in the room to determine the air exchange rate, or with the presence of occupants in the room, with cluster 6 being more particularly linked to the ventilation OFF periods.

**Table 1.** Summary of the clusters for the partitioning into 7 classes.

Cluster	Event	n. obs.	O <sub>3</sub> / mV	NO / mV	NO <sub>2</sub> / mV	CO <sub>2</sub> / ppm
<b>C1</b>	Ventilation OFF	4394	21.20	29.54	5.80	614
<b>C2</b>	Ventilation ON	1246	19.19	37.20	13.90	601
<b>C3</b>	Ventilation ON	1515	28.32	28.82	16.53	592
<b>C4</b>	NO+NO <sub>2</sub>	225	22.13	51.49	7.34	642
<b>C5</b>	NO <sub>2</sub>	129	95.63	22.23	49.28	650
<b>C6</b>	CO <sub>2</sub> (Vent. OFF)	208	14.22	31.09	9.39	1658
<b>C7</b>	CO <sub>2</sub> (Vent. ON)	443	21.26	32.48	13.65	1035

These almost straightforward assignments show by themselves the interest of the K-means partitioning to analyze the data from the sensors. The only necessary user input resides in the determination of the optimal number of classes. This must be done considering the stability of the clusters when increasing their number, and, if available, using contextual information, in the present case the daily log of the events in the room, to assign the events.

A further argument is given by the comparison with the results obtained from the measurements by the gas analyzers, also leading to an empirical optimal value of 7 classes. The comparison between the two datasets is shown in figure 5.



**Figure 5.** Comparison of the classes obtained from the measurements by the analyzers (dataset 1) and from the signals of the sensors (dataset 2).

Some of the classes overlap strongly, such as cluster 1 (ventilation off, 91.1% overlap), cluster 4 (NO and NO<sub>2</sub> spill, 97.8%), cluster 5 (NO<sub>2</sub> injection, 98.4%). The two clusters corresponding to high CO<sub>2</sub> concentrations tend to be rather mixed between the two datasets, and so do the two clusters corresponding to the basis case with the ventilation active. It is probable that taking into account the signal from other sensors, especially if they can discriminate between pollutants from outdoor and from indoor origin, would help achieve a better differentiation.

#### 4. Conclusions

This work demonstrates that classification methods applied to the signals from sensor arrays, even when the sensors are not calibrated beforehand, can be used to analyze the air quality inside real environments. Several issues remain open, and in particular the determination of the optimal number of classes to be used. On one hand, it should be possible to apply mathematical procedures and criteria to determine this number, in particular considering the stability of the classes when increasing their total number. But on another hand, the interpretation of the classes in terms of physical events requires a specific expertise in indoor air and building sciences, and contextual information to supplement the sensors data, such as the ventilation status or the outdoor pollution levels. Another approach to overcome this problem would be to construct a database or set of classes for typical indoor air quality events, by applying the sensors and the associated classification methodology we propose here, to other buildings or in laboratory conditions. This would result in a database of characteristic air quality event and associated classes. This database could then be used as the basis of a supervised classification model, which would not require the expertise of a user. This would therefore make the management of buildings more easy and efficient, for instance by coupling this analysis with the control of the ventilation or filtration system.

#### Acknowledgments

The authors thankfully acknowledge the financial support from the University of Lille (BQR 2015). This work is a contribution to the CaPPA project (Chemical and Physical Properties of the Atmosphere), funded by the French National Research Agency (ANR) through the PIA (Programme d'Investissement d'Avenir) under contract ANR-11-LABX-005-01, and a contribution to the CPER research project CLIMIBIO, with financial support from the French Ministère de l'Enseignement Supérieur et de la Recherche, the Hauts de France Region and the European Funds for Regional Economical Development. The MERMAID project was financially supported through the PRIMEQUAL program of the French Environment and Energy Management Agency ADEME (Agence De l'Environnement et de la Maîtrise de l'Énergie).

**References**

- [1] Tham KW 2001 *Energ. Buildings* **130** 637-50
- [2] Aleixandre M and Gerboles M 2012 *Chem. Eng. Trans.* **30** 169-74
- [3] Kumar P, Skouloudis AN, Bell M, Viana M, Carotta MC, Biskos G and Morawska L 2016 *Sci. Total Environ.* **560-561** 150-9
- [4] Caron A, Redon N, Thevenet F, Hanoune B and Coddeville P 2016 *Build. Environ.* **107** 19-28
- [5] Lewis AC, Lee JD, Edwards PM, Shaw MD, Evans MJ, Moller SJ, Smith KR, Buckley JW, Ellis M, Gillot SR and White A 2016 *Faraday Discuss.* **189** 85-103 *Environ.* **123** 517-28
- [6] Verma M, Srivastava M, Chack N, Diswar AK and Gupta N 2012 *Int. J. Eng. Res. Appl.* **2** 1379-84
- [7] Verrielle M, Schoemaeker C, Hanoune B, Leclerc N, Germain S, Gaudion V, Locoge N 2016 *Indoor Air* **26** 702-13