Pollution activity detection based on metal-oxide gas sensors and Intrinsic Dimensionality estimation for Indoor Air Quality applications

Luiz Miranda[1,2], Caroline Duc[2], Marie Verriele[2], Nathalie Redon[2], Bernadette Dorizzi[3], Jugurta Montalvão[4], Jérôme Boudy[3]
[1]ADEME - l'Agence de la transition écologique, 49000 Angers, France
[2]IMT Nord Europe, Institut Mines-Télécom, Univ. Lille, Centre for Energy and Environment, F-59000 Lille, France
[3] Samovar, Télécom SudParis, Institut Polytechnique de Paris, 91120 Palaiseau, France
[4]Universidade Federal de Sergipe, Departamento de Engenharia Elétrica, São Cristóvão - SE, 49100-000, Brazil
luiz.miranda@imt-nord-europe.fr

Abstract – The monitoring of indoor air quality (IAQ) is essential to prevent its potential effect on our health. One of the main sources of pollutants is the daily occupant activities, such as the use of cleaning products and cooking. To avoid that the levels of indoor pollution induced by these sources get harmful, and to aware people of their occurrence and impact, indoor air pollution events should be identified through the use of gas sensors. This work proposes a method of detection of indoor pollution activities through the estimation of the intrinsic dimensionality on windows of observation of metal-oxide gas (MOX) sensors. This approach was tested on a dataset extracted from 2 months of experiments in which typical household activities were performed in a 17m² room using 21 unique models of MOX sensors. Results show that this method is capable of detecting activities in the environment but more understanding is needed to fully characterize it.

Keywords: Indoor activity detection, metal-oxide gas sensors, Intrinsic Dimensionality, Indoor Air Quality.

I. INTRODUCTION

To monitor indoor air quality (IAQ) is to know how good is the air that we breath during most of our everyday life. Considering commutes to work, office and home periods, we spend at least 80% of our time indoors [1]. Daily activities such as cooking, housekeeping, smoking, use of candle and incense, are significant sources of pollutants and especially of volatile organic compounds (VOC) [2] and particulate matter (PM) [3] and health issues relative to poor IAQ, gathered as sick building syndrome, have been extensively reported [4].

To avoid that pollutant levels become harmful, it is essential to detect when events responsible for high emissions happen. The task of detecting these emission events in the context of IAQ is often called just “Event Detection” [5]–[7], but in this work, as pollution events are often related to activities performed on the indoor environment, we are calling it Activity Detection.

This task is usually performed by setting a threshold on the sensors’ measurements that signalizes the beginning of a pollution event when measurements go above it, and the end of a pollution event when measurements go below it. A few studies, however, try to go further by performing, for example, frequency analysis [8], support vector machines [7] or statistical analysis [2] on the sensors’ measurements to detect indoor activities.

In this work, we propose the use of the intrinsic dimensionality estimation of metal-oxide based gas (hereby called MOX) sensors’ measurements as a metric to detect when indoor activities, such as cleaning the floor or cooking, are being performed, as this method is not affected by the long-term drift of the sensors. In section II of this paper, we provide a short discussion on the choice of sensors. Section III provides the theoretical background as well as the method of detection. The experiment on which this method was tested is presented in section IV, followed by its results and discussions in section V and, finally, we present our conclusions in section VI.

II. WHY MOX?

Indoor activity detection can be achieved using a large variety of sensors, such as cameras, microphones or binary sensors. These technologies, despite providing good results in detection, pose a few problems. Cameras and microphones raise the concern about privacy when these sensors are spread through a residence. Moreover, when using cameras, there is also the need of high computational power to process the signal they provide [9].

Binary sensors such as passive infrared motion detectors, magnetic reed switches, ultrasonic and float switches are frequently designed to detect presence. They can also provide good results on activity detection, as seen in [6], but they do not allow any detection of the gaseous emissions resulting from the daily pollution activity, (as well as cameras and microphones) which is essential information in the context of IAQ.

This leads us to gas sensors, which have been used successfully for both activity detection and recognition [10], [11]. Amongst this type of sensor, MOX sensors are the best candidates, as they are the cheapest and smallest gas sensing technology currently available and, despite their typical lack of selectivity, they can detect and quantify individual pollutants emissions when used in clusters [12].

These sensors, however, often require the application of sophisticated signal processing techniques to their measurements, such as K-NNs [11] or neural networks [12], to be able to extract useful information. In this work we propose as a useful and efficient tool the estimation of the temporal Intrinsic Dimensionality of the sensors' readings.

III. INTRINSIC DIMENSIONALITY AND ACTIVITY DETECTION

When dealing with sensors, we consider that a system containing D sensors has a dimensionality of D which is sometimes called the extrinsic dimensionality. On the other hand, the Intrinsic Dimension (ID) of a multidimensional system has been described initially by [13] as *"the number of free parameters required in a hypothetical signal generator capable of producing a close approximation to each signal in the collection"* and later by [14] as *"the minimum number of free variables needed to represent the data without information loss"*.

This concept can also be understood as the "real" dimensionality of the system. In other words, if when plotting the data provided by the system using each sensor as an axis, the result is a line, the ID of this system is 1. Analogously, if the data results in a plane, the ID is 2, even if the number of sensors is greater than 2.

In this work, we use the ID estimation method proposed by Grassberger and Procaccia [15] with the implementation suggested in [16]. To understand how this method computes ID, consider a cube of side *r* placed on the space formed by the sensors' readings, that is, the measurement of each sensor will represent one axis (dimension) of this space, as the sub-figures on the right part of Figure 2. By counting the number of samples contained inside this cube, we obtain a value *C*, a number that represents a simplification of the correlation integral presented in [15]. By varying the size *r* of the cube inside a certain range and by counting the respective number of samples *C(r)* inside of it, we obtain a relation of proportionality between *C(r)* and r^d, that is $C(r) \propto r^d$, in which d is the intrinsic dimensionality of the dataset. By calculating the logarithm of this relation, we get $\log C(r) \approx d \log r + h$, in which h is the differential entropy of the dataset [16]. By plotting the relation *log(C(r))* x *log(r)* as *r* varies, we get a curve in which the slope corresponds to the estimation of ID as can be seen in Figure 1.

One of the preprocessing steps of this method is to normalize the sensors' data to avoid over-influence of sensors that provide large measurements compared to others. By doing that, when no excitation is being presented to the sensors, their measurements are dominated by the random noise, which will be amplified by the normalization step and occupy the whole space provided by the D sensors, that is, the resulting ID is D. On the other hand, when the sensors are presented to an excitation, in the form of an indoor activity, their measurements will be dominated by their transient response, which will reflect as a low dimensional structure on the space provided by the sensors. This phenomenon is illustrated on Figure 2.

By sliding a window of observation of duration T through the signal, we can extract the ID of the observations in these time intervals and study if there was activity, or not, by measuring the minima of the resulting IDs. We call the analysis of ID through time as temporal ID, and we consider that there is activity on the observed T long segment of the signal when the resulting temporal ID is below a certain threshold.

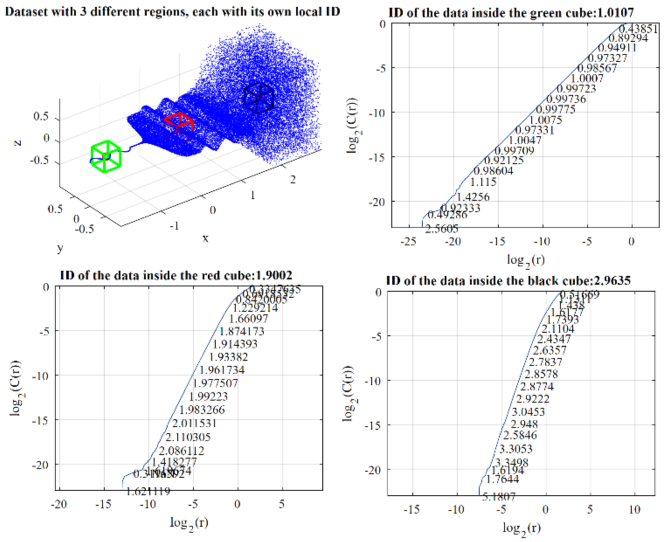

Figure 1. Illustration on how the ID estimation method works. On the top left, a dataset with three distinct intrinsic dimensions, 1D near the green cube, 2D near the red cube and 3D around the black cube. The remaining images refer to the *log(C(r))* x *log(r)* plot as *r* varies in the three different regions of the image on the top left, the numbers along the plot represent the slope, equivalent to the ID in this region. Top right is the resulting graph for the area around the green cube. Botton left, represents the area around the red cube. And bottom right, represents the are around the black cube.

But how is this different from applying a threshold directly on the sensors' measurements? There are a few reasons to use the approach, the first one is that this method is not sensitive to absolute values, this means that the known problem of long-term drift of MOX sensors [17] does not affect this type of measurement. Another reason is that this method is easy to adapt if a new sensor is needed to be added to or suppressed from the system.

However, this method may require computational resources that are not suitable for less powerful microcontrollers, making it difficult to embed it. This comes as a consequence of the approach chosen to estimate ID, using the correlation dimension, that is known to need large amounts of samples to avoid biased estimations [16].

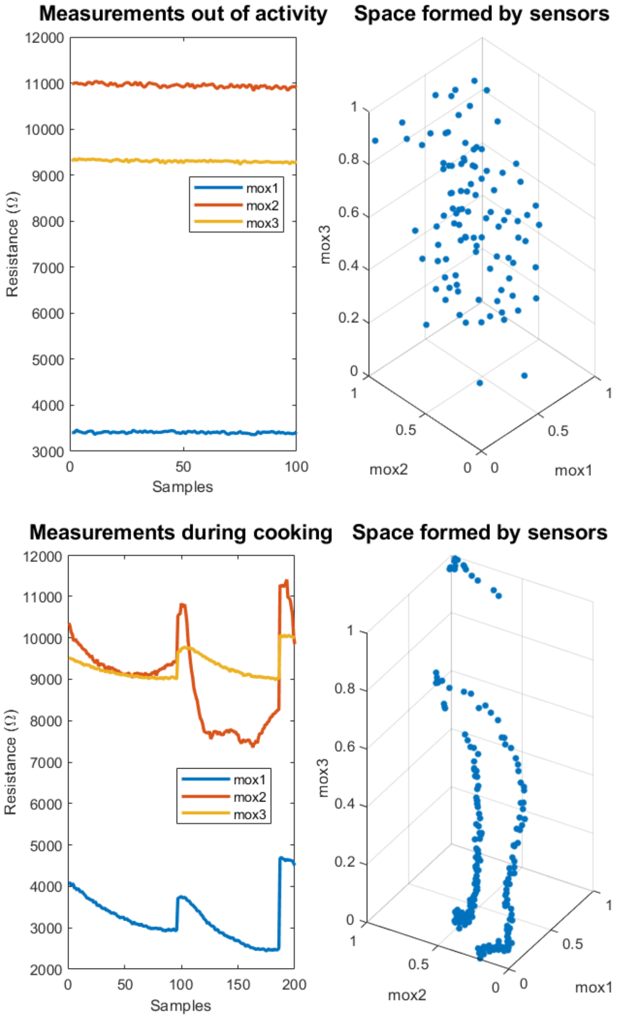
Measurements out of activity
Space formed by sensors
Measurements during cooking
Space formed by sensors

Figure 2. Measurements out of activity (top) and during activity (bottom). On the left, we have the raw measurements of three MOX sensors, that when shown on the scatter plot on the right, either occupy most of the available space (when there is no activity being performed, top) or form a low dimensional structure (a line in this case, when cooking is being performed).

IV. THE EXPERIMENT

This method was put to test on a database extracted from an indoor activity experiment. This dataset was collected in a 17 m² room, in which 26 unique models of MOX sensors were distributed between 6 unique IAQ monitoring modules. All modules had at least one copy performing measurements along with it. However, due to missing measurements and module failure we used, in fact, 21 unique models of MOX in 4 unique IAQ monitoring modules. One module with 4 sensors, one module with 6 sensors and two modules with 10 sensors each.

The activities performed were of 10 types: (i) Static activity (human present in the room seated at a desk); (ii) Cooking/eating (oil frying and eating an egg); (iii) humid activity (towel drying, no one in the room); (iv) Vacuum cleaning; (v) Floor cleaning using bleach and detergent; (vi) Low intensity exercise; (vii) Scented candle use; (viii) Aerate the room by opening a window; (ix) Emissions from the use of hot glue; and (x) Emissions from woodcutting and wood drilling. The activities were chosen to simulate normal household activities as best as possible. The activities had a suggested duration but, as to simulate a real-world scenario, variation on the duration was allowed within acceptable limits.

In total, 276 activities were performed between November and December of 2022, and between February and March of 2023, totaling 26 days of presence in the room. Two people were responsible to perform the activities and each performed about half of them. All the performed activities had their start and end time, recorded in a spreadsheet annexed to the dataset.

V. RESULTS

The dataset had a sampling period of 10 seconds, therefore, to allow for a large enough number of samples and reduce the ID estimation bias to acceptable levels [16], we set the window size, T, to be equal to 1 hour, resulting in 360 samples. This sample size is large enough to estimate ID of dimensionality up to 5 within acceptable limits of bias on the estimation, and the duration is reasonable for typical indoor activities. More details about the calculation of the minimum number of samples and the estimation bias can be found in [16]. The scanning of the signal was done with superposition of windows, that is, the window moved with steps smaller than its size. In our case, the step size was lower bound to 50 samples or 500 seconds, because of processing time.

After scanning the dataset and calculating the temporal ID, we found that during activities, this value was in average 2 when performing activities and 4.1 when not. There was, however superposition between their distributions, being that the standard deviation in activity was 0.9 and outside activity was 1.3. Knowing that, we decided to set a threshold of detection of activity based on an analysis of False Acceptance Rate (FAR) and False Rejection Rate (FRR) by finding the Equal Error Rate (EER). The result of this analysis was that the detection of activity provided an EER of 27.2% when the threshold of detection of temporal ID was equal to 3.07.

We also performed a comparison with a well referenced work by Fang et al. [6], in which they used a normalized standard deviation (NSTD) approach to detect pollution events. This approach was also based on a scanning window through the sensors' signals to detect peaks in pollutant presence. In our implementation of their method, we used the same parameters as in ours method, window of size T equal to 1 hour and window superposition of 50 samples or 500 seconds. For their work, they considered that a pollution event was happening when the resulting NSTD was above 0.27. After implementing their method and applying it to our dataset we found a EER of 27.2 % when using a threshold of 0.25. The region of convergence (ROC) plot for both methods can be seen in Figure 3.

The results obtained using our are equivalent to the ones obtained using Fang's [6]. This result indicates that our method is as comparable with the literature and if used in datasets such as the one from [6] we should expect to obtain similar results. Both methods use the sudden change of sensor measurements to detect activity so it is not surprising that they show similar results. However, this method was capable of detecting activities not listed in the dataset, such as montage of measuring equipment and high intensity exercise performed outside of campaign schedule. These activities, if listed, could have improved the EER of this method, but as we have no way to accurately point out when they were executed, we can't include them in the performance calculations. A more in-depth study is needed to fully quantify the performance of this method.

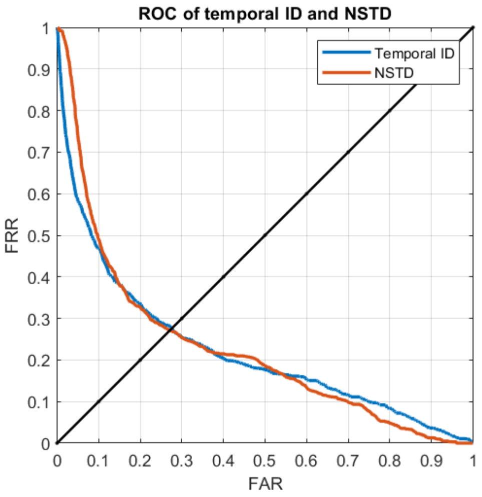

Figure 3. Region of convergence (ROC) plot of both methods tested, Temporal ID and NSTD. Both methods achieve an EER of 27.2 %.

VI. CONCLUSION AND PERSPECTIVES

A method of detection of household activities was proposed based on the use of the intrinsic dimensionality (ID) estimation of windows of observation, here called temporal ID, of metal-oxide gas sensors (MOX sensors). The method was tested on a dataset containing data from household activities performed in a 17 m² room between 2022 and 2023 extracted with 21 different MOX and the time in which activities were performed is known. We determined that activity was being detected when the temporal ID was below a certain threshold, which we established by measuring the equal error rate of classification (EER). Our results indicated a threshold of 3.07 and an EER of 27.2%. We compared this result with a well referenced method, normalized standard deviation (NSTD) **[6]**, and obtained a equivalent result EER of 27.2 % when the threshold was 0.25 (NSTD).

This result shows that the method is capable of activity detection and comparable with the literature, however, a more in-depth study is needed to both fully quantify its performance, as well as better develop its applicability, as it is possible to use certain characteristics of the household environment to improve the ability to detect activities, such as a priori assumptions of duration of activity and knowledge of sensor response.

ACKNOWLEDGMENT

Authors thank ADEME and the French region of Hauts-de-France for the provided PhD grant to Luiz Miranda. We would also like to thank Ellona for providing us with the multi-sensor platforms used to extract the dataset, in this work, without the need of financial compensation or limitations in regard to publication of results.

REFERENCES

[1] M. Śmiełowska, M. Marć, and B. Zabiegała, "Indoor air quality in public utility environments—a review," *Environ. Sci. Pollut. Res.*, vol. 24, no. 12, Art. no. 12, Apr. 2017, doi: 10.1007/s11356-017-8567-7.

[2] C. Rösch, T. Kohajda, S. Röder, M. von Bergen, and U. Schlink, "Relationship between sources and patterns of VOCs in indoor air," *Atmospheric Pollut. Res.*, vol. 5, no. 1, pp. 129–137, Jan. 2014, doi: 10.5094/APR.2014.016.

[3] L. Tofful, S. Canepari, T. Sargolini, and C. Perrino, "Indoor air quality in a domestic environment: Combined contribution of indoor and outdoor PM sources," *Build. Environ.*, vol. 202, p. 108050, Sep. 2021, doi: 10.1016/j.buildenv.2021.108050.

[4] P. Burge, "Sick building syndrome," *Occup. Environ. Med.*, vol. 61, no. 2, pp. 185–190, Feb. 2004, doi: 10.1136/oem.2003.008813.

[5] A. Caron, N. Redon, F. Thevenet, B. Hanoune, and P. Coddeville, "Performances and limitations of electronic gas sensors to investigate an indoor air quality event," *Build. Environ.*, vol. 107, pp. 19–28, Oct. 2016, doi: 10.1016/j.buildenv.2016.07.006.

[6] B. Fang, Q. Xu, T. Park, and M. Zhang, "AirSense: an intelligent home-based sensing system for indoor air quality analytics," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Heidelberg Germany: ACM, Sep. 2016, pp. 109–119. doi: 10.1145/2971648.2971720.

[7] J. Fonollosa, I. Rodriguez-Lujan, A. V. Shevade, M. L. Homer, M. A. Ryan, and R. Huerta, "Human activity monitoring using gas sensor arrays," *Sens. Actuators B Chem.*, vol. 199, pp. 398–402, Aug. 2014, doi: 10.1016/j.snb.2014.03.102.

[8] T. Oyabu, T. Onodera, H. Kimura, and T. Katsube, "Characteristics of tin oxide gas sensors for human vital signs during sleep," *Sens. Mater.*, vol. 12, no. 1, pp. 25–34, 2000.

[9] A. Szczurek, M. Maciejewska, and T. Pietrucha, "Occupancy Detection using Gas Sensors," presented at the 6th International Conference on Sensor Networks, 2017, pp. 99–107. Accessed: Mar. 29, 2023. [Online]. Available: https://www.scitepress.org/Link.aspx?doi=10.5220/0006207100990107

[10] M. Maciejewska, A. Szczurek, and A. Dolega, "AP3.4 - Recognition of indoor events using gas sensors," in *Proceedings IMCS 2018*, Vienna, Austria: AMA Service GmbH, Von-Münchhausen-Str. 49, 31515 Wunstorf, Germany, 2018, pp. 386–387. doi: 10.5162/IMCS2018/AP3.4.

[11] A. Caron, N. Redon, P. Coddeville, and B. Hanoune, “Identification of indoor air quality events using a K-means clustering analysis of gas sensors data,” *Sens. Actuators B Chem.*, vol. 297, p. 126709, Oct. 2019, doi: 10.1016/j.snb.2019.126709.

[12] J. He, L. Xu, P. Wang, and Q. Wang, “A high precise E-nose for daily indoor air quality monitoring in living environment,” *Integration*, vol. 58, pp. 286–294, Jun. 2017, doi: 10.1016/j.vlsi.2016.12.010.

[13] R. S. Bennet, “REPRESENTATION AND ANALYSIS OF SIGNALS PART XXI. THE INTRINSIC DIMENSIONALITY OF SIGNAL COLLECTIONS.” Accessed: Mar. 29, 2023. [Online]. Available: https://apps.dtic.mil/sti/citations/AD0475844

[14] F. Camastra and A. Staiano, “Intrinsic dimension estimation: Advances and open problems,” *Inf. Sci.*, vol. 328, pp. 26–41, Jan. 2016, doi: 10.1016/j.ins.2015.08.029.

[15] P. Grassberger and I. Procaccia, “Characterization of Strange Attractors,” *Phys. Rev. Lett.*, vol. 50, no. 5, pp. 346–349, Jan. 1983, doi: 10.1103/PhysRevLett.50.346.

[16] J. Montalvão, J. Canuto, and L. Miranda, “Bias-Compensated Estimator for Intrinsic Dimension and Differential Entropy,” *J. Commun. Inf. Syst.*, vol. 35, no. 1, Art. no. 1, Dec. 2020, doi: 10.14209/jcis.2020.30.

[17] N. Dennler, S. Rastogi, J. Fonollosa, A. van Schaik, and M. Schmuker, “Drift in a popular metal oxide sensor dataset reveals limitations for gas classification benchmarks,” *Sens. Actuators B Chem.*, vol. 361, p. 131668, Jun. 2022, doi: 10.1016/j.snb.2022.131668.