



**HAL**  
open science

## **Intrinsic dimension estimation as a tool to sensor selection for an indoor air quality multisensory system**

Luiz Miranda, Caroline Duc, Nathalie Redon, Marie Verrielle, Bernadette Dorizzi, Jérôme Boudy, Jugurta Montalvão

► **To cite this version:**

Luiz Miranda, Caroline Duc, Nathalie Redon, Marie Verrielle, Bernadette Dorizzi, et al.. Intrinsic dimension estimation as a tool to sensor selection for an indoor air quality multisensory system. Healthy Buildings Europe - 2021, Jun 2021, Oslo / Online, Norway. hal-04638695

**HAL Id: hal-04638695**

**<https://hal.science/hal-04638695v1>**

Submitted on 10 Jul 2024

**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.

---

# Intrinsic dimension estimation as a tool to sensor selection for an indoor air quality multisensory system

Luiz MIRANDA\*<sup>1</sup>, Caroline DUC<sup>1</sup>, Nathalie REDON<sup>1</sup>, Marie VERRIELE<sup>1</sup>, Bernadette DORIZZI<sup>2</sup>, Jérôme BOUDY<sup>2</sup>, Jugurta MONTALVÃO<sup>3</sup>

<sup>1</sup> IMT Lille-Douai, Douai, France

<sup>2</sup> Samovar, Télécom SudParis, Institut polytechnique de Paris, France

<sup>3</sup> Universidade Federal de Sergipe, São Cristóvão, Brazil

\* *Corresponding author: luiz.miranda@imt-lille-douai.fr*

---

## SUMMARY

To estimate the minimum number of sensors necessary in an air quality monitor capable to discriminate hazardous volatile organic compounds (VOCs) an intrinsic dimension (ID) estimator was used. Considering the sensors as dimensions, ID estimators often find a smaller number of dimensions than the original number of sensors, indicating that some sensors provide redundant information. An ID estimator was applied to the signals collected by 45 commercial multi-sensor indoor air monitors (Rubix PODs) using 16 different sensors in 6 different situations in which 6 different gases were injected in the room which the PODs were in. The resulting estimation was 3 dimensional, meaning that, for this experiment, 3 specific sensors would be enough to identify each of the 6 gases injected. This experiment, however, is very simple compared to what multisensory systems face during day-to-day use. More experiments in realistic scenarios should provide a better understanding of which sensors are really needed.

## INTRODUCTION

One of the current trends on the study of Indoor Air Quality (IAQ) is to continuously monitor Indoor Air Pollution (IAP) in different possible environments, such as offices or homes (Tran, Park & Lee, 2020). Much research is being done in developing systems that are capable of detecting the target pollutants of IAQ while keeping them affordable (Saini, Dutta & Marques, 2020). The main cost of these systems are the sensors, and, as highly selective sensors have a higher price, the selected sensors tend to be cheaper but non-selective (Saini, Dutta & Marques, 2020b). The use of low-cost sensors creates the need to employ several different sensors to achieve a good degree of selectiveness and versatility (Feng et al., 2019). However, these sensors are selected often by the information provided by the manufacturer, which can be unreliable (Rackes, Ben-David & Waring, 2018), or by availability. This work proposes a tool to aid in the selection of sensors to be used in IAQ monitoring

systems by estimating the Intrinsic Dimension (ID) of the sensors readings.

## Intrinsic Dimension of sensors readings

Consider an IAQ monitoring system with number of sensors equals to  $D$ . When processing signals in this type of system, the number of sensors corresponds to its dimensionality (Hines, Llobet & Gardner, 1999), if the ID of these signals is estimated, the result approximates how many sensors provided significative information for the task they were exposed to (Camastra & Staiano, 2016; Bennet, 1969). As the sensors used in IAQ monitoring systems are, in general, non-selective and based on the same detection mechanism, it is very likely that two or more sensors provide the same kind of information or at least show some level of redundancy. Then, it is not unreasonable to expect that the ID estimated for these kinds of systems is a value  $d < D$ .

There are some methods of estimating ID, as it can be seen in (Camastra & Staiano, 2016), but the method selected to be used in this work was the one proposed by Grasberger & Procaccia (1986) with the implementation suggested in (Montalvão, Canuto & Miranda, 2020). This method was selected due to its versatility and facility to implement and understand. This technique was then applied to the results of a set of experiments to determine the minimum number of sensors necessary to identify certain types of gas, this experiment is described in the next section.

## The experiment

45 commercial multi-sensor indoor air monitors (Rubix PODs) were grouped in a 12,5 m<sup>2</sup> room in which different gases were injected during several experiments. Each POD contained at least 16 different sensors to measure various sources of IAP and ambient parameters (temperature, humidity, etc.). The experiment that generated the dataset used in this work consisted in 6 different days in which each day a different gas was injected in the room at a designated concentration and for a given period of time. Then the gas was ejected from the room via an exhaust pipe. The PODs registered readings from the room during all experiments and saved the data for later processing.

An example of one of these readings is presented in Figure 1.

The gases injected were volatile organic compounds (VOCs) representative of IAP: acetaldehyde (283 ppb), acetone (128 ppb), ethanol (284 ppb), formaldehyde (400 ppb), limonene (880 ppb) and a mixture of 6 VOCs (acetaldehyde at 204 ppb, acetone at 108 ppb, ethanol at 125 ppb, formaldehyde at 64 ppb, limonene at 131 ppb and toluene at 113 ppb). These compounds have been selected for their ubiquity in indoor environments and/or for their sanitary impact and/or to represent typical indoor pollution activities.

After the experiment the resulting dataset was arranged in a way that made the six days of the experiment became a single large experiment, that is, the data was chained together. This made possible to evaluate the resulting dataset as a gas separation task. The data from all the sensors were considered, and rescaled so its readings are between 0 and 1, this was done so the different scale of measurements did not affect the ID estimation. In Figure 1 it can be seen an example of the resulting process.

## Results and Discussion

With the dataset pre-processed the ID estimation method was applied to it, resulting in a 3-dimensional ID. This meant that three ideal sources of signal would be enough to generate the signal produced by all 16 sensors present in the PODs during the experiments.

To determine what sensors influenced the most the ID estimation the following procedure was done: (i) one sensor would be removed from the dataset; (ii) the ID would be estimated again; (iii) if the resulting ID was lower, the sensor would be re-introduced to the dataset, if not, the sensor would be kept out of the dataset for the next steps. These steps three were repeated until no more sensors could be removed.

After doing this procedure, 4 sensors remained in the dataset, all of them were metal oxide semiconductor (MOX) gas sensors. This means that the essential information for the proposed task (gas separation) is contained in these 4 sensors. It is important to notice that these sensors do not yield independent signals, but they form the best subset of available signals that still yields the ID estimated with all sensors. Moreover, to have a visual representation of how the gases could be separated, the readings of the remaining sensors were plotted using each sensor as an axis.

Due to the 3-dimensional limitation of human vision only three sensors could be represented in each plot, so four different plots representing all possible combinations of the 4 sensors were made, the plot using sensors MOX1, MOX3 and MOX4 was the one that showed a higher degree of separation between gases, that is, groups of gases are farther apart as it can be seen in Figure 2. Because of that, those three sensors were considered sufficient to separate the different gases injected in this experiment.

## Conclusions

The use of Intrinsic Dimension (ID) estimation was proposed as a tool to aid the selection of sensors for a given task (in this work, a gas separation task) in Indoor Air Quality (IAQ). Its use was illustrated by experiments using a commercial IAQ monitoring system in a 12,5m<sup>2</sup> room. The experiments showed that, by using only 3 of the 16 sensors, it is possible to separate different IAP gases in an experimental setup. On the other hand, to be able to do this task in a situation that is more representative of the day-to-day use of IAQ monitors, more experiments are need along with deeper study of the behavior of the sensors in the presence of IAP.

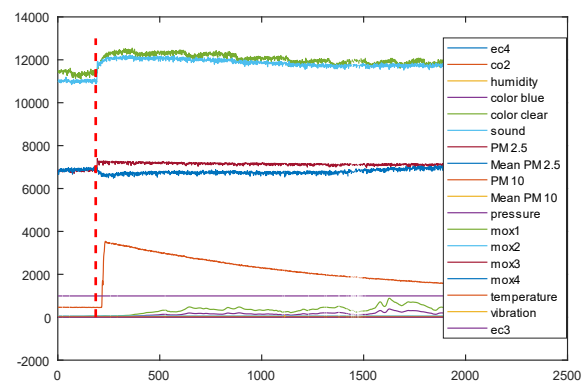


Figure 1. Raw measurements of one injection of acetaldehyde at 283 ppb. The vertical red dashed line corresponds to the instant of the injection. Measurements dominated by mox1, mox2, mox3, mox4 and co2.

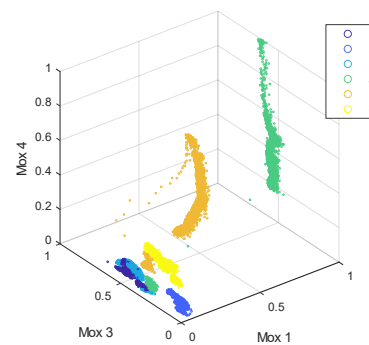


Figure 2. Tridimensional plot of sensors readings. Different colors are different gases.

## ACKNOWLEDGMENTS

Authors acknowledge the financial support from IMT.

## REFERENCES

Bennett, Robert. "The intrinsic dimensionality of signal collections." IEEE Transactions on Information Theory 15.5 (1969): 517-525.

- Camastra, Francesco, and Antonino Staiano. "Intrinsic dimension estimation: Advances and open problems." *Information Sciences* 328 (2016): 26-41.
- Feng, Shaobin, et al. "Review on smart gas sensing technology." *Sensors* 19.17 (2019): 3760.
- Hines, E. L., E. Llobet, and J. W. Gardner. "Electronic noses: a review of signal processing techniques." *IEE Proceedings-Circuits, Devices and Systems* 146.6 (1999): 297-310.
- Montalvão, Jugurta, Jânio Canuto, and Luiz Miranda. "Bias-Compensated Estimator for Intrinsic Dimension and Differential Entropy." *Journal of Communication and Information Systems* 35.1 (2020): 300-310.
- Rackes, Adams, Tom Ben-David, and Michael S. Waring. "Sensor networks for routine indoor air quality monitoring in buildings: Impacts of placement, accuracy, and number of sensors." *Science and Technology for the Built Environment* 24.2 (2018): 188-197.
- Saini, Jagriti, Maitreyee Dutta, and Gonçalo Marques. "A comprehensive review on indoor air quality monitoring systems for enhanced public health." *Sustainable Environment Research* 30.1 (2020): 6.
- Saini, Jagriti, Maitreyee Dutta, and Gonçalo Marques. "Indoor air quality monitoring systems based on Internet of things: A systematic review." *International journal of environmental research and public health* 17.14 (2020): 4942.
- Tran, Vinh Van, Duckshin Park, and Young-Chul Lee. "Indoor air pollution, related human diseases, and recent trends in the control and improvement of indoor air quality." *International journal of environmental research and public health* 17.8 (2020): 2927.