



HAL
open science

The dimensionality of MOX sensors in Air quality assessment

Luiz Miranda, Nathalie Redon, Caroline Duc, Marie Verrielle, Jugurta Montalvão,
Bernadette Dorizzi, Mossaab Hariz, Dijana Petrovska, Jerome Boudy

► To cite this version:

Luiz Miranda, Nathalie Redon, Caroline Duc, Marie Verrielle, Jugurta Montalvão, et al.. The dimensionality of MOX sensors in Air quality assessment. Eurosenors 2025, Sep 2025, Wroclaw, Poland, Poland. 2025, <10.5162/EUROSENSORS2025/MP44>. <hal-05300871>

HAL Id: hal-05300871

<https://hal.science/hal-05300871v1>

Submitted on 7 Oct 2025

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 CC BY 4.0 - Attribution - International License

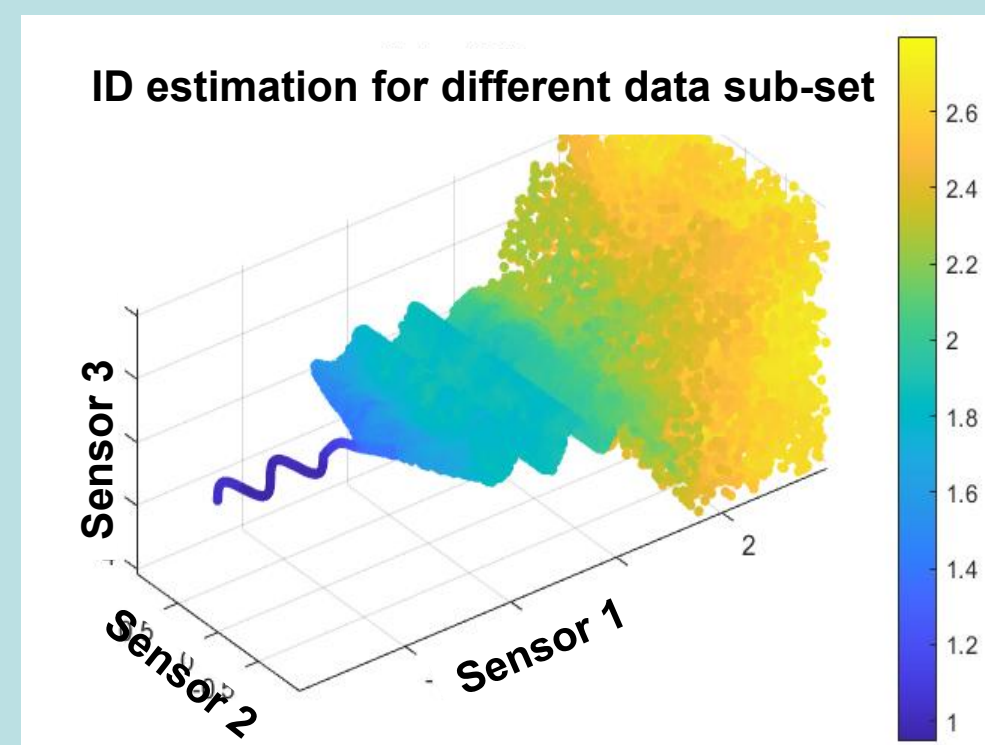
Context

- Metal-oxide (**MOX**) gas sensors are widely used in various gas sensing despite their low selectivity
- Solution: use **sensor arrays** to benefit from an enhanced response
- Applications often use **more than 10 sensors**, but in **air quality monitoring only 2 to 4 sensors** are generally used [1]

→ **Why do small MOX sensor arrays work effectively for air quality monitoring?**

What is Intrinsic Dimensionality (ID) ?

The **Dataset dimensionality** represents the number of parameters measured i.e. the apparent dimensionality, **D**. The **Intrinsic dimensionality (ID), d**, often lower than **D**, measures degrees of freedom and “perceived dimensionality”.



ID Estimation Methods mainly used:

- **PCA**: Number of components explaining >90% variance. Tends to overestimate due to linearity.
- **MLE** [3]: Finds dimension **d** that best fits **k**-nearest neighbor distance distributions (**k=20**).
- **GP** [4]: Slope of log C(r) vs log r, where C(r) counts pairs within distance r. Multiple stable regions possible.

Results and discussion

- Estimators show **higher ID for Dataset 1 than Dataset 2**, reflecting complexity differences in sensor count and measured phenomena.
- Relative response (**Rs**) captures **temporal information**, yielding **higher ID than average response** values.
- **PCA overestimated ID** except for **Rs in Dataset 1**. Further investigation needed.
- Air quality applications target **few specific gases** (like Dataset 2), resulting in **low complexity** requiring fewer sensors, explaining why 2-4 sensors suffice.

Methodology

Study the **dimensionality (ID)** of two datasets with different complexities :

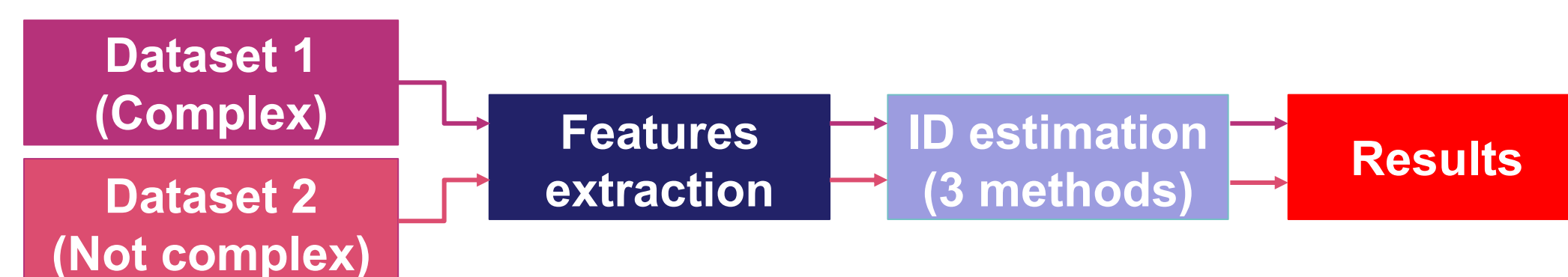
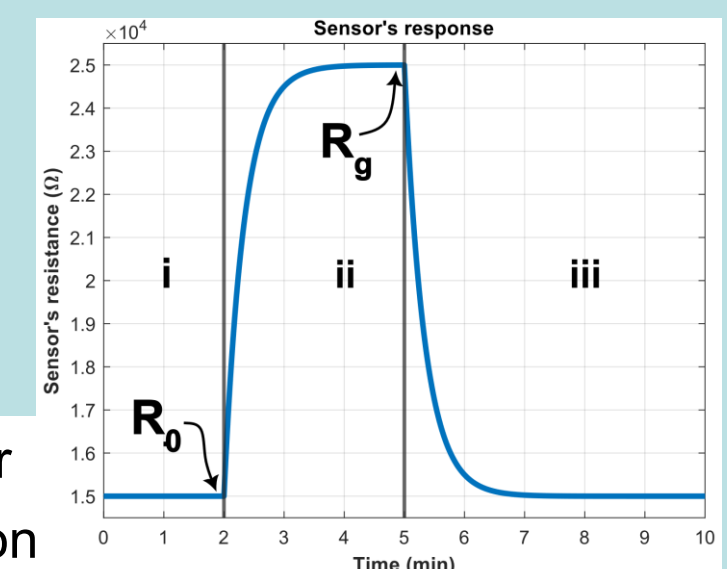
- **Dataset 1 [22 MOX-sensors]** is from an indoor activity detection experiment [2].
- **Dataset 2 [4 MOX-sensors]** is from a VOC metrological qualification set;

• **Features** from datasets:

- **Average** sensor measurement;
- Normalized relative response **R_s**

$$R_s = \frac{R_g - R_0}{R_0}$$

- **i and iii**: no activity / clean air
- **ii**: activity / target gas injection



Sensor reference	Present in		Target pollutant (according to datasheet)
	Dataset 1	Dataset 2	
CCS 801	X	X	C ₂ H ₅ OH + CO + C ₇ H ₈
COM 31	X		CO
GM 102B	X		NO ₂
GM 302B	X		Alcohol
GM 402B	X		C ₃ H ₈ + CH ₄
GM 502B	X		VOCs
GM 602B	X		H ₂ S
GM 702B	X		CO
GM 802B	X		NH ₃
H2SM 31	X		H ₂ S
MICS 2714	X		NO ₂
MICS 5524	X	X	CO + C ₂ H ₅ OH + H ₂ + NH ₃ + CH ₄
MICS 5914	X	X	NH ₃ + C ₂ H ₅ OH + H ₂ + C ₃ H ₈ + C ₄ H ₁₀
SB 5100	X		H ₂ S
SBAQ 106	X		VOCs
TGS 2442	X		CO
TGS 2600	X		H ₂ + C ₂ H ₅ OH + C ₄ H ₁₀
TGS 2602	X		VOCs (toluene + ethanol)
TGS 2603	X		C ₂ H ₅ OH + C ₃ H ₈ N + CH ₄ S + H ₂ S + H ₂
TGS 2620	X		VOCs
TGS 8100	X	X	C ₂ H ₅ OH + H ₂ + CO + C ₄ H ₁₀
VOCM 31	X		VOCs

Estimator	Dataset 1 (D = 21)		Dataset 2 (D = 4)	
	Average	Rs	Average	Rs
PCA (Linear)	6	5	3	4
MLE (Non-linear)	4,4	6,5	1,9	2,8
GP (Non-linear)	3,2	3.5 or 5.8	1.4 or 2.4	2,5

Conclusions

- **Few diverse sensors** capture essential air quality info within constraints.
- Relative response yields higher ID than average response, indicating **temporal information** improves gas detection.

References

- [1] Z. Yuan et al., Adv. Eng. Mater. 26, 2400782 (2024), DOI: [10.1002/adem.202400782](https://doi.org/10.1002/adem.202400782)
- [2] Miranda et al. (2025), Labelled dataset, DOI: [10.5281/zenodo.15462760](https://doi.org/10.5281/zenodo.15462760)
- [3] E. Levina and P. Bickel, NIPS 17 (2004).
- [4] P. Grassberger and I. Procaccia, Phys. Rev. Lett. 50, 346 (1983), DOI: [10.1103/PhysRevLett.50.346](https://doi.org/10.1103/PhysRevLett.50.346)

Further info:



Partners



UNIVERSIDADE
FEDERAL DE
SERGIPE



AMUSENS



Funded by the European Union under grant agreement no. 101130159. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.