

A multisensor-array device for breath analysis: development of a data quality assessment framework and application in the clinical practice

Valentina Pizzillo^{1,2*}, Jolanda Palmisani^{1,2*}, Alessia Di Gilio^{1,2}, Annamaria Catino³, Alba Quarato³, Gianluigi de Gennaro^{1,2}, Sergio Ivan Lopes⁴

¹ *Department of Biosciences, Biotechnology and Environment, University of Bari Aldo Moro, 70125 Bari, Italy*

² *Apulian Regional Centre for the Breath Analysis, Istituto Tumori 'Giovanni Paolo II', 70124 Bari, Italy;*

³ *Istituto Tumori 'Giovanni Paolo II', 70124 Bari, Italy;*

⁴ *ADiT-Lab, Instituto Politécnico de Viana do Castelo, 4900-348 Viana do Castelo, Portugal*

**co-first authors*

Corresponding Author's e-mail address: v.pizzillo@phd.uniba.it

Summary:

In the present work, a multi-sensor array device for real-time monitoring of target compounds (CO₂, CO, TVOCs, NO₂, H₂S, SO₂) was developed and validated under controlled environmental conditions. To face the critical issue of data quality, a multidimensional Data Quality Assessment (DQA) framework was proposed. Outcomes showed that environmental/external conditions significantly affect sensors behavior and data reliability. The developed approach therefore enabled the identification of reliable acquisition conditions. The validation of the system for diagnostic purposes on a population of malignant pleural mesothelioma patients is ongoing.

Keywords: Breath analysis, Multi-sensor array system, Data Quality Assessment, early diagnosis, malignant pleural mesothelioma

Background, Motivation and Objective

Breath analysis has emerged as a promising non-invasive approach for the early diagnosis and monitoring of chronic and oncological diseases. The chemical characterization of exhaled breath, combined with the real-time detection of selected compounds, enables the identification of metabolic alterations associated with pathological conditions [1,2]. Multisensor array systems have been widely investigated as effective tools for this purpose. However, their performance is strongly influenced by environmental conditions and acquisition-related factors, which may affect baseline stability, sensor response, and measurement reliability. In addition, temporal misalignment and inter-sensor variability complicate the interpretation of multisensor data. Although several studies address data quality in sensor systems [3,4], it is often treated as a preprocessing step rather than as part of system characterization. The system considered in this study is designed for breath acquisition from both healthy subjects and patients affected by oncological diseases (such

as lung cancer and malignant pleural mesothelioma).

In this context, this work presents a multi-sensor array device for the real-time monitoring of target compounds (CO₂, CO, TVOCs, NO₂, H₂S, SO₂) in human breath. To face the well-known critical issue of data quality provided by these kinds of systems, a quality-aware workflow for multisensor breath analysis integrating preprocessing for multi-rate synchronization, CO₂-based breath/background segmentation and a multidimensional Data Quality Assessment (DQA) framework was developed and applied to background data. Besides exploring the data quality across multiple complementary dimensions, the present work aims to explore the potentialities of the multi-sensor array device in the clinical practice for diagnosis and prediction purposes. Preliminary results coming from the use of the system on a population of patients affected by malignant pleural mesothelioma (MPM) are also discussed.

Description of the New Method or System

In this work, a multisensor-array system for breath analysis was developed, integrating commercial sensor technologies for the real-time monitoring of selected organic and inorganic compounds within the ppb–ppm concentration range. The system consists of a multi-sensor array for the high temporal resolution monitoring of: 1. TVOCs concentration - photoionization detectors (PIDs); 2. CO₂ concentration and temperature monitoring - infrared (IR) sensor; 3. H₂S, SO₂, CO, and NO concentration monitoring along with relative humidity determination - electrochemical sensors. The system, visualized in Figure 1, operates in continuous acquisition mode, enabling simultaneous multisensor measurements.

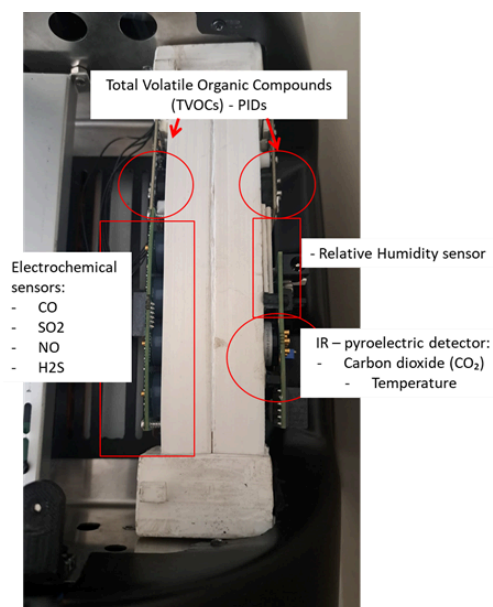


Fig. 1. Overview of the multisensor array chamber.

Sensor data acquired at different sampling rates were temporally aligned within a common time reference. An automatic segmentation procedure was developed to separate breath-related responses from environmental background using CO₂ dynamics as a physiological reference (capnogram). Breath and background data were processed separately, and the background component was used to evaluate system behaviour independently of respiratory signals. Breath data acquired from subjects are also presented in this work, providing preliminary observations, while the main focus remains on background characterization and data quality assessment. System performance was quantified through indicators including baseline stability, signal

variability, drift, and inter-sensor correlation. Temporal consistency was evaluated through autocorrelation and timestamp analysis. Experimental measurements were performed under different environmental conditions (e.g., occupancy and ventilation variations), enabling a systematic characterization of sensor behaviour and sensitivity to environmental perturbations. Once exploring the data quality across multiple complementary dimensions and identifying the reliable acquisition conditions, the breath analysis system is currently undergoing validation in collaboration with the medical team of Istituto Tumori 'Giovanni Paolo II' on individuals affected by MPM for diagnosis and prediction purposes.

Results

The obtained outcomes demonstrated that environmental/external conditions may significantly influence sensor responses and baseline stability. More specifically, variations in occupancy and ventilation conditions (rate and mode) may lead to increased baseline variability and higher inter-sensor correlation. Drift analysis highlighted the time-dependent changes in sensor behaviour, more pronounced under perturbed environmental conditions. These findings therefore emphasize the importance of background characterization for reliable multisensor measurements and confirm the effectiveness of the proposed approach in identifying stable acquisition conditions. The use of the system on a population of MPM patients provided promising results, although preliminary.

References

- [1] IA. G. Buma et al. Lung cancer detection by electronic nose analysis of exhaled breath: a multicentre prospective external validation study, *Annals of Oncology* 36, 786–795 (2025)
- [2] E. A. Chapman, et al. A breath test for malignant mesothelioma using an electronic nose, *European Respiratory Journal* 40, 448–454 (2012)
- [3] E. Hosseinzadeh et al. Data quality assessment in healthcare, dimensions, methods and tools: a systematic review, *BMC Medical Informatics and Decision Making* 25, 296 (2025)
- [4] I. Hayward et al. Low-Cost Air Quality Sensors: Biases, Corrections and Challenges in Their Comparability, *Atmosphere* 15, 1523 (2024)