Advances in Science and Technology Research Journal

Advances in Science and Technology Research Journal, 2026, 20(10), 424–434
https://doi.org/10.12913/22998624/224423
ISSN 2299-8624, License CC-BY 4.0

Received: 2026.05.14
Accepted: 2026.07.15
Published: 2026.09.01

Estimation of ethylbenzene, cumene and limonene concentrations using gas sensor arrays and machine learning techniques

Magda Wlazło[1*], Agnieszka Malec[2], Dariusz Majerek[1], Anton Trnik[3], Jarosław Zubrzycki[1], Zbigniew Suchorab[2]

[1] Faculty of Mathematics and Information Technology, Lublin University of Technology, ul. Nadbystrzycka 38, Lublin 20-618, Poland
[2] Faculty of Environmental Engineering and Energy, Lublin University of Technology, ul. Nadbystrzycka 38, Lublin 20-618, Poland
[3] Constantine the Philosopher University in Nitra, Trieda Andreja Hlinku 1, Chrenová, 949 74 Nitra, Slovakia
* Corresponding author's e-mail: m.wlazlo@pollub.pl

ABSTRACT

The objective of this study is to estimate the concentration of selected microbial volatile organic compounds (mVOCs), namely ethylbenzene, cumene, and limonene, using multi-sensor gas arrays combined with statistical regression and machine learning methods. Two sensor array configurations were employed: metal oxide semiconductor (MOS) sensors and electrochemical (EC) sensors. Measurements were conducted under both field conditions, in buildings with varying degrees of microbial contamination, and laboratory conditions using fungal cultures. The results demonstrate that sensor arrays, combined with models such as support vector machines (SVM) and multi-layer perceptron (MLP), can effectively estimate concentrations of mVOCs with different physicochemical properties and concentration ranges. The findings confirm the potential of low-cost sensor systems as an alternative to conventional chromatographic techniques for indoor air quality monitoring.

Keywords: microbial volatile organic compounds (mVOCs), MOS gas sensors, indoor air quality.

INTRODUCTION

Indoor air quality (IAQ) plays a crucial role in maintaining human health and well-being. Poor air conditions in buildings are often associated with microbiological contamination, particularly the presence of fungi, which can contribute to the development of symptoms commonly referred to as sick building syndrome (SBS). In many cases, the negative impact of fungal contamination is not only related to the presence of spores or mycotoxins, but also to the emission of volatile metabolites into the indoor environment [19,21,22].

Fungi release a wide range of volatile organic compounds as a result of their metabolic activity. These substances, commonly referred to as microbial volatile organic compounds (mVOCs), represent a diverse group of low-molecular-weight chemicals that can be detected in indoor air [1–4]. Due to their origin and physicochemical properties, mVOCs are increasingly recognized as useful indicators of biological processes occurring in contaminated environments [5]. Their presence may signal active microbial growth even before visible signs of mold appear.

Among the various mVOCs identified in indoor environments, certain compounds are frequently associated with fungal activity and are therefore considered potential markers of contamination. These include aromatic hydrocarbons

such as ethylbenzene and cumene, as well as terpenes such as limonene. In addition, other groups of compounds, including alcohols (e.g., 1-octen-3-ol), ketones (e.g., 2-heptanone), and sulfur-containing compounds, are also commonly reported as products of fungal metabolism [2,6,7]. The occurrence of such substances, particularly in characteristic combinations, may indicate microbiological degradation processes taking place within building materials or indoor air.

The detection and quantification of mVOCs are therefore important for the assessment of indoor air quality and the identification of potential health risks. The presence of these compounds may serve as an early warning signal of microbiological contamination and can support preventive actions aimed at improving environmental conditions in buildings. However, reliable identification of mVOCs typically requires advanced analytical techniques.

Gas chromatography coupled with mass spectrometry (GC/MS) is widely regarded as a reference method for the analysis of volatile compounds due to its high sensitivity and selectivity [8,9]. This technique enables precise identification and quantification of individual substances, including those present at low concentration levels. Despite its advantages, GC/MS is not suitable for continuous or real-time monitoring, as it involves complex sample preparation, requires specialized equipment, and is associated with relatively high operational costs.

In response to these limitations, increasing attention has been given to alternative approaches based on multi-sensor gas arrays, often referred to as electronic nose systems. These systems consist of multiple sensors with partial selectivity, which respond to different components of gas mixtures and generate characteristic signal patterns. When combined with appropriate data processing techniques, such as statistical regression and machine learning algorithms, sensor arrays can be used not only for qualitative classification but also for quantitative estimation of selected compounds.

Among the available sensor technologies, metal oxide semiconductor (MOS) [10–13] sensors are characterized by high sensitivity to a broad range of volatile compounds, whereas electrochemical (EC) sensors offer greater selectivity and stability under varying environmental conditions [10,11,14,15]. The combination of these sensor types provides a complementary approach that can enhance the detection and estimation of compounds with different physicochemical properties and concentration ranges.

In this context, the present study focuses on the estimation of selected microbial volatile organic compounds – ethylbenzene, cumene, and limonene – using multi-sensor gas arrays combined with regression and machine learning models [16–18]. These compounds were selected due to their differing chemical characteristics and their relevance as indicators of fungal activity. By linking the presence of characteristic mVOCs with sensor responses, the study aims to evaluate the potential of sensor-based systems as a rapid and cost-effective tool for the indirect detection of microbiological contamination in indoor environments.

While numerous studies have investigated the application of gas sensor arrays for VOC classification, fewer studies have focused on quantitative estimation of specific microbial volatile organic compounds under real indoor conditions. Therefore, the objective of this study was not only to compare different predictive modeling approaches but also to evaluate their applicability for quantitative mVOC estimation using field measurements calibrated against reference GC/MS analyses.

MATERIAL AND METHODS

Materials

Examined objects

The study was conducted under field conditions in a set of indoor environments characterized by varying levels of microbiological contamination. A total of 15 rooms (marked r1-r15) were selected for analysis, representing different degrees of fungal presence – from reference environments with no visible contamination to spaces with clear signs of mold growth. The diversity of the investigated environments made it possible to analyze a wide range of concentrations of microbial volatile organic compounds (mVOCs) and to assess the performance of sensor-based estimation methods under realistic conditions.

The sampling locations were selected on the basis of visual inspection and microbiological evaluation, with particular attention paid to visible fungal contamination, moisture conditions, and the functional characteristics of the buildings. The investigated objects consisted primarily of residential buildings, supplemented by several

public utility facilities. Such diversification of the analyzed environments enabled the inclusion of spaces representing different stages of microbiological degradation, which was important for assessing the effectiveness of indirect gas-sensing methods in indoor air quality evaluation.

Gas chromatography measurement

Reference measurements of selected mVOC concentrations were carried out using gas chromatography coupled with mass spectrometry (GC/MS). The analyses were performed using solid-phase microextraction (SPME), which enabled the preconcentration of volatile compounds present in indoor air. For this purpose, Supelco CAR SPME fibers with a coating thickness of 75 µm were applied. During sampling, the fibers were exposed for 60 minutes in the central part of each investigated room, ensuring unrestricted air access from all directions and allowing effective adsorption of volatile metabolites emitted by microorganisms.

After sampling, the fibers were transported to the laboratory under cooled conditions (4 °C) and subjected to thermal desorption in the GC/MS injection port. Chromatographic analyses were performed using a TraceUltra gas chromatograph coupled with a PolarisQ mass spectrometer (Thermo Scientific), equipped with an Equity-5MS capillary column manufactured by Supelco (30 m length, 0.25 mm internal diameter, 0.25 µm film thickness). Helium with a purity of 99.999% was used as the carrier gas at a constant flow rate of 1.2 mL/min.

The chromatographic data were analyzed using Xcalibur 1.4 software together with the NIST 08 and Wiley 8 spectral libraries, which enabled identification and quantitative evaluation of selected compounds, including ethylbenzene, cumene, and limonene.

Gas sensor array measurement

In parallel with the chromatographic analyses, measurements were carried out using a multi-sensor gas array system functioning as an electronic nose. The applied measurement setup was based on a metal oxide semiconductor (MOS) sensor matrix composed of 17 gas sensors, designated from s1 to s17. All sensors were manufactured by Figaro Engineering Inc. and selected according to their sensitivity toward volatile organic and inorganic compounds potentially associated with microbiological contamination in indoor environments.

The sensor matrix incorporated sensors sensitive to general air pollutants, alcohols, hydrocarbons, methane, ammonia, hydrogen sulfide, and other volatile compounds commonly identified as microbial metabolites. The configuration included the following sensor types: TGS-2600-B00 (s1), TGS-2602-B00 (s2), TGS-2610-D00 (s3), TGS-2611-E00 (s4), TGS-2620-C00 (s5), TGS-4161 (s6), TGS-2444 (s7), TGS-2442-B02 (s8), TGS-800 (s9), TGS-825-A00 (s10 and s11), TGS-813-A00 (s12), TGS-821 (s13), TGS-823-A00 (s14), TGS-812 (s15), TGS-830 (s16), and TGS-832-A00 (s17).

The sensors were mounted inside the measurement chamber in a compact matrix configuration ensuring uniform exposure to the analyzed gas mixture and stable operating conditions during measurements. The arrangement of the sensors within the chamber enabled simultaneous acquisition of multidimensional response patterns generated by interactions between the gas mixture and the sensing surfaces. Such a configuration allowed the system to capture complex changes in indoor air composition associated with fungal contamination and microbial volatile organic compounds.

Methods

Description of measuring procedure

Prior to each measurement, the sensor system was purged with clean reference air in order to stabilize the baseline signal and remove residual compounds from previous analyses. The flushing process was carried out using synthetic air supplied from a gas cylinder, ensuring minimal background interference.

During the measurement procedure, air samples were collected using a sampling bag and introduced into the sensor chamber. The sensor responses were recorded over a defined time interval, allowing for the acquisition of stable and repeatable signal patterns. Each measurement cycle included a baseline stabilization phase, an exposure phase, and a recovery phase, during which the sensors returned to initial conditions.

The recorded sensor signals were preprocessed prior to model development. The preprocessing procedure included normalization, outlier removal, and synchronization with the reference GC/MS measurements. These steps ensured consistency between sensor responses and chromatographic data and enabled the application of

statistical regression and machine learning models for mVOC concentration estimation. The overall research workflow is presented in Figure 1.

Data processing

The data obtained from the sensor arrays were further analyzed using both conventional statistical methods and advanced predictive modeling techniques. In particular, linear regression models were applied as a baseline approach to describe the relationship between sensor responses and the concentrations of selected mVOCs.

Linear regression

Regression modeling aimed to estimate particular substance concentration (dependent variable) based on the sensor signals (independent variables). The fundamental form of linear regression is expressed as follows:

$$\gamma = \beta_0 + \sum_{i=1}^{n} \beta_i X_i + \varepsilon \qquad (1)$$

where: γ – estimated (ethylbenzene, cumene and limonene) concentration, X_i – response of the *i- this sensor*, β_i -regression coefficients, ε- random component.

Artificial intelligence methods

In order to improve prediction accuracy, machine learning techniques were applied, including support vector machines (SVM) and multilayer perceptron (MLP) neural networks [17–19]. The SVM method is based on the construction of an optimal decision function that minimizes prediction error while maintaining good generalization ability. In regression tasks, SVM enables the identification of nonlinear relationships between sensor responses and the concentrations of analyzed compounds through the application of kernel functions.

The MLP model represents a feed-forward artificial neural network composed of interconnected neurons arranged in input, hidden, and output layers. During the training process, the network adjusts connection weights iteratively, allowing the model to capture complex nonlinear dependencies occurring within multidimensional sensor datasets. Owing to its adaptive structure, the MLP approach is particularly suitable for the analysis of gas sensor array signals characterized by high variability and mutual interactions between variables [20,21].

The recorded sensor signals were further processed and compared with the reference GC/MS data. Preprocessing steps included normalization, removal of outliers, and temporal alignment of sensor responses with chromatographic measurements. These procedures ensured consistency between datasets and enabled the application of regression and machine learning models for the estimation of mVOC concentrations. Reference measurements of ethylbenzene, cumene, and limonene concentrations were performed using solid-phase microextraction coupled with gas chromatography and mass spectrometry (SPME-GC/MS) [22].

The application of both approaches enabled a comparative evaluation of traditional statistical and machine learning methods for the estimation of selected microbial volatile organic compounds, namely ethylbenzene, cumene, and limonene.

Prior to model development, the recorded sensor responses were subjected to preprocessing procedures including normalization and extraction of representative signal features. For each sensor, averaged relative resistance values obtained during the stabilized phase of the measurement signal were used as input variables for

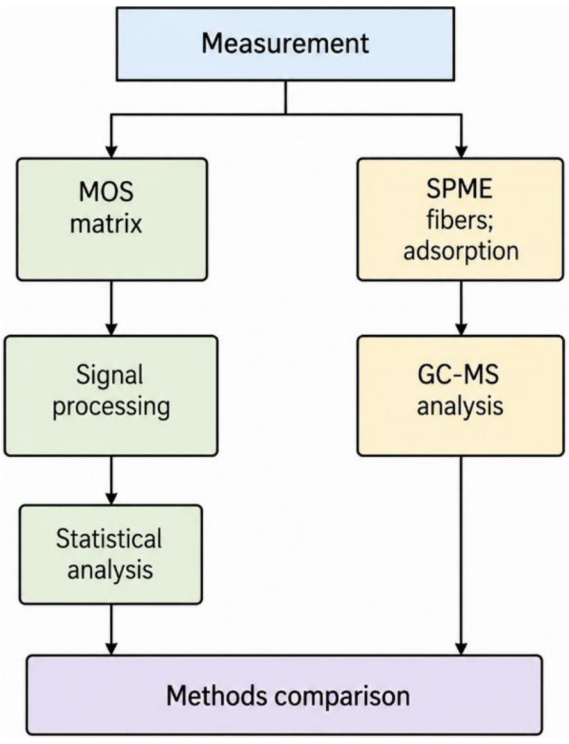

Figure 1. Schematic overview of the experimental procedure used in the MOS and EC series

statistical and machine learning analyses. This approach reduced the influence of short-term fluctuations and improved the repeatability of the obtained datasets [23]. To ensure reliable model evaluation despite the limited number of samples, a 5-fold cross-validation procedure was applied. The dataset was randomly divided into five subsets, and during each iteration, four subsets were used for model training while the remaining subset was used for validation. The final performance metrics were calculated as the average values obtained across all validation folds. This approach reduced the risk of overfitting and improved the robustness of model assessment.

The predictive performance of the developed models was assessed using commonly applied statistical indicators, including the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE).

RESULTS

Data from gas chromatography

The data presented in Table 1 summarize the concentrations of ethylbenzene, cumene, and limonene obtained from SPME-GC/MS analyses performed using SPME fibers exposed in the investigated indoor environments. The results indicate that limonene exhibited the highest concentration levels and variability among the analyzed compounds, whereas cumene occurred predominantly at trace concentrations. Ethylbenzene was characterized by moderate and relatively stable concentration values across the examined samples. The results are arranged in descending order according to limonene concentration, which showed the widest concentration range among the analyzed compounds. This arrangement enables a clearer comparison of the variability of selected microbial volatile organic compounds across the investigated indoor environments.

Data from gas sensor array

The sensor array signals consisted of time-dependent resistance changes recorded for individual sensors during exposure to indoor air samples. Each investigated environment was represented by a multidimensional dataset composed of resistance responses obtained from the sensors included in the matrix. Variations in signal intensity and temporal response profiles reflected differences in the composition of volatile compounds present in the analyzed indoor environments. Exemplary signal achieved during in situ tests by sensor array is presented in Figure 2.

The obtained sensor responses demonstrated clear variability between the analyzed indoor environments. Differences were observed both in the magnitude of relative resistance changes

Table 1. Concentrations of selected microbial volatile organic compounds identified in individual indoor environments using SPME-GC/MS analysis

Rooms	Ethylbenzene [µg/dm³]	Cumene [µg/dm³]	Limonene [µg/dm³]
R1	0.0091	0.0008	0.1452
R2	0.0014	0.0001	0.0484
R3	0.0154	0.0016	0.0283
R4	0.0070	0.0	0.0005
R5	0.0044	0.0004	0.0004
R6	0.0137	0.0000	0.0202
R7	0.0259	0.0003	0.0007
R8	0.0019	0.0002	0.0139
R9	0.0074	0.0016	0.0000
R10	0.0026	0.0003	0.0100
R11	0.0049	0.0003	0.0077
R12	0.0079	0.0003	0.0025
R13	0.0023	0.0004	0.0019
R14	0.0032	0.0000	0.0015
R15	0.0001	0.0000	0.0008

and in the temporal characteristics of the sensor signals. Selected sensors showed pronounced responses during exposure to samples characterized by elevated concentrations of microbial volatile organic compounds, whereas lower signal intensities were recorded for reference environments with limited microbiological contamination [24].

The multidimensional character of the recorded signals enabled the identification of distinct response patterns associated with different air quality conditions. Variations in signal profiles between individual rooms confirmed the sensitivity of the applied sensor matrix to changes in the composition of volatile compounds present in indoor air. The processed sensor datasets were subsequently used for the development of regression and machine learning models aimed at estimating ethylbenzene, cumene, and limonene concentrations.

DISCUSSION

Analysis of reference data from gas chromatography

The comparative analysis of the obtained results is presented in Table 2, which

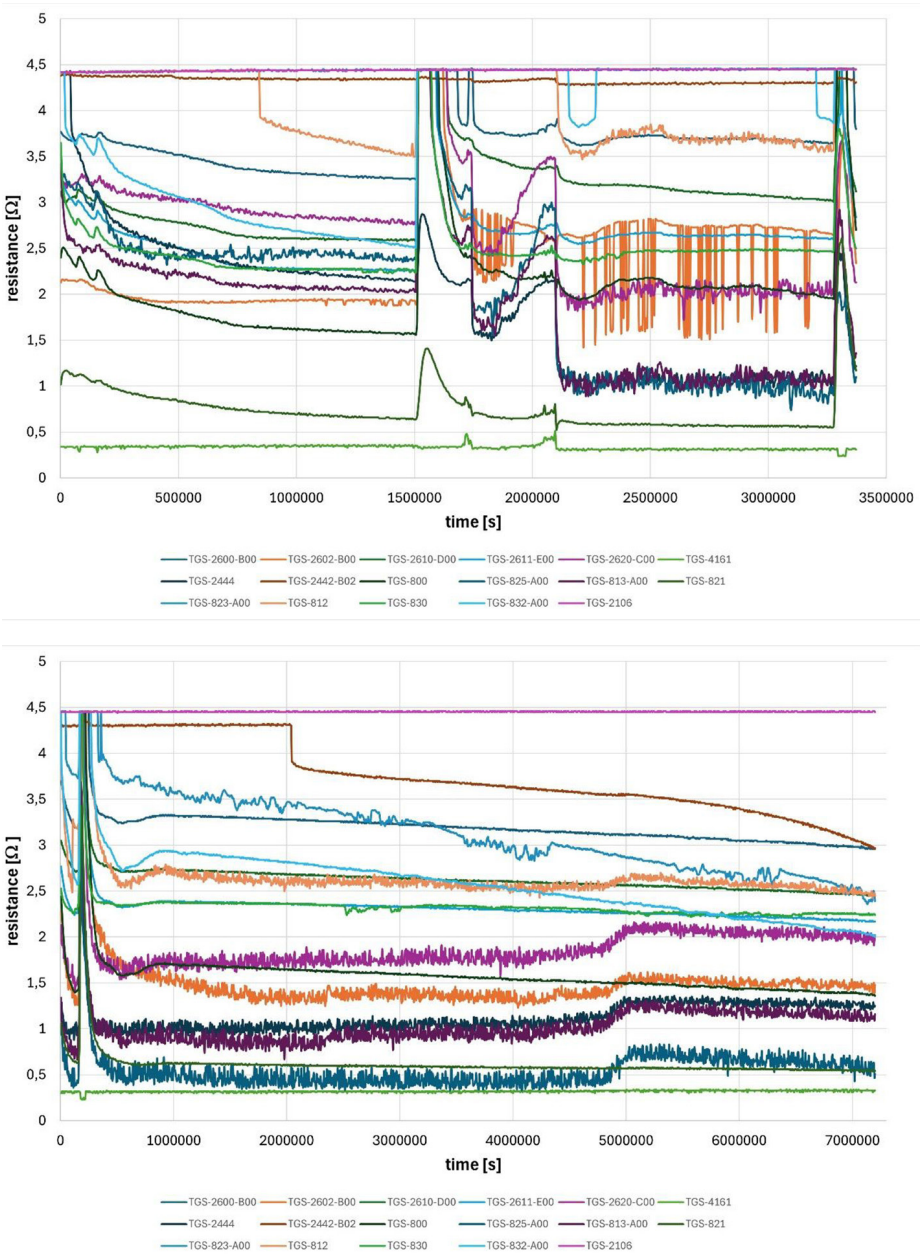

Figure 2. Exemplary raw signal achieved by the applied gas sensor array within in situ tests

summarizes the statistical characteristics of the analyzed compounds together with the performance of the developed estimation models. The table includes the minimum (Min), maximum (Max), and mean (Mean) concentrations determined using the GC/MS method, as well as additional statistical parameters such as median values, lower and upper quartiles (Q1 and Q3), and standard deviation (SD). The presented data indicate that limonene was characterized by the highest concentration levels and the greatest variability among the analyzed compounds, whereas cumene occurred predominantly at trace concentrations. Despite these differences, satisfactory estimation results were achieved for all investigated substances using sensor-based predictive models.

The descriptive statistics of the reference SPME-GC/MS measurements indicate clear differences in the concentration ranges of the analyzed compounds. Ethylbenzene occurred at moderate levels, with concentrations ranging from 0.0001 to 0.0259 µg/dm³. Its mean value was 0.0071 µg/dm³, while the median was 0.0049 µg/dm³, indicating moderate variability between the investigated rooms.

Cumene was detected at the lowest concentration levels among the analyzed compounds. In several rooms, its concentration was equal to 0.0000 µg/dm³, while the maximum value reached 0.0016 µg/dm³. This confirms that cumene occurred mainly at trace levels, which may affect the stability of its quantitative estimation using sensor-based models.

Limonene showed the widest concentration range, from 0.0008 to 0.1452 µg/dm³, and the highest mean concentration among the analyzed compounds. The elevated maximum value and relatively high standard deviation indicate substantial differences between individual indoor environments. These differences suggest that limonene emissions varied considerably depending on the conditions present in particular rooms. Limonene, characterized by higher concentrations and larger variability, generated stronger sensor responses, facilitating model development. Cumene occurred mainly at trace levels, resulting in weaker signal-to-noise ratios. Ethylbenzene showed intermediate behavior.

Overall, the reference GC/MS data show that the analyzed compounds differed markedly in concentration level and variability. These differences were important for further model development, as compounds occurring at trace levels may be more difficult to estimate accurately than compounds characterized by stronger and more variable concentration signals [18,19].

Analyses data from gas sensor arrays

Table 3 presents the averaged values of relative resistance obtained for individual sensors of the MOS matrix during measurements performed in the investigated indoor environments [25]. In addition to mean sensor responses, the table also includes standard deviation values describing the variability of signals recorded for particular sensors.

Linear regression models

The statistical parameters obtained for the linear regression models developed for ethylbenzene, cumene, and limonene are presented in Table 4. The results indicate that the regression models achieved different levels of prediction accuracy depending on the analyzed compound.

The highest coefficient of determination (R^2=0.850) was obtained for cumene, indicating the best agreement between predicted and reference values among the analyzed substances. Slightly lower, but still satisfactory, prediction performance was observed for ethylbenzene (R^2=0.793). In contrast, the limonene model was characterized by the lowest coefficient of determination (R^2= 0.588), suggesting greater variability and reduced prediction stability for this compound.

Similar relationships were observed for the error indicators RMSE and MAE. The lowest prediction errors were obtained for cumene, whereas

Table 2. Statistics of the GC-MS estimation of particular volatile organic compounds concentrations

Compound	Min [µg/dm³]	Max [µg/dm³]	Mean [µg/dm³]	Median [µg/dm³]	Q1 [µg/dm³]	Q3 [µg/dm³]	SD [µg/dm³]
Ethylbenzene	0.0001	0.0259	0.0078	0.0049	0.0023	0.0091	0.0065
Cumene	0.0000	0.0016	0.0004	0.0003	0.0001	0.0005	0.0004
Limonene	0.0008	0.1452	0.0239	0.0139	0.0025	0.0207	0.0362

Table 3. Mean values of relative resistance and standard deviation obtained for individual sensors of the MOS sensor array

Number of sensor	Relative resistance [-]	Standard deviation [-]
s1	0.545	0.083
s2	0.227	0.123
s3	0.481	0.086
s4	0.445	0.073
s5	0.435	0.174
s6	0.081	0.011
s7	0.214	0.122
s8	0.221	0.118
s9	0.381	0.065
s10	0.206	0.106
s11	0.223	0.122
s12	0.262	0.150
s13	0.513	0.151
s14	0.390	0.308
s15	0.523	0.083
s16	0.587	0.175
s17	0.919	0.228

the highest values were observed for limonene. The obtained results indicate that the effectiveness of linear regression models depended on the concentration variability and signal characteristics of the analyzed compounds.

The performance of the developed linear regression models was evaluated using the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). Among the analyzed compounds, cumene achieved the highest prediction accuracy (R^2 = 0.850), followed by ethylbenzene (R^2=0.793). In contrast, limonene showed the lowest model performance (R^2=0.588) and the highest prediction errors. These results suggest that compounds characterized by higher concentration variability were more difficult to estimate using linear regression methods. information enabling quantitative estimation of selected microbial volatile organic compounds.

AI models

The prediction performance obtained using the support vector machine (SVM) approach is summarized in Table 5. The table presents the statistical parameters describing the accuracy of concentration estimation for the analyzed compounds.svm

The application of artificial intelligence methods resulted in lower prediction errors compared to the classical linear regression models. In particular, the SVM approach achieved very high coefficients of determination together with low RMSE and MAE values for all analyzed compounds.

The MLP (Table 6) models also provided satisfactory estimation accuracy, especially for ethylbenzene and limonene. Overall, the obtained results indicate that machine learning techniques enabled more effective interpretation of multidimensional sensor signals and improved the estimation of microbial volatile organic compound concentrations in comparison with linear regression methods.

Comparison of model performance

The comparison of statistical parameters obtained for the developed models indicates that machine learning approaches generally provided

Table 5. Statistical parameters obtained for SVM models

Compound	RMSE	R^2	MAE
Ethylbenzene	0.001	0.995	0.001
Cumene	0.000	0.994	0.000
Limonene	0.003	0.996	0.003

Table 6. Statistical parameters obtained for MLP models

Compound	RMSE	R^2	MAE
Ethylbenzene	0.002	0.932	0.001
Cumene	0.001	0.083	0.001
Limonene	0.003	0.995	0.002

Table 4. Statistical parameters obtained for linear regression models developed for selected mVOCs

Compound	R^2	RMSE [µg/dm^3]	MAE [µg/dm^3]
Ethylbenzene	0.793	0.003	0.002
Cumene	0.850	0.000	0.000
Limonene	0.588	0.025	0.018

Table 7. Comparison of statistical parameters obtained for different estimation models

Compound	Model	R^2	RMSE [μg/dm³]	MAE [μg/dm³]
Ethylbenzene	Linear regression	0.793	0.003	0.002
Ethylbenzene	SVM	0.995	0.001	0.001
Ethylbenzene	MLP	0.932	0.002	0.001
Cumene	Linear regression	0.850	0.000	0.000
Cumene	SVM	0.994	0.000	0.000
Cumene	MLP	0.083	0.001	0.001
Limonene	Linear regression	0.588	0.025	0.018
Limonene	SVM	0.996	0.003	0.003
Limonene	MLP	0.995	0.003	0.002

higher prediction accuracy than classical linear regression models. Among the analyzed methods, the SVM models achieved the best overall performance, characterized by the highest coefficients of determination and the lowest prediction errors for all investigated compounds.

The MLP models also demonstrated high estimation accuracy for ethylbenzene and limonene. However, a significantly lower R^2 value was observed for cumene, suggesting reduced model stability for compounds occurring at very low concentration levels.

In contrast, the linear regression models were characterized by lower prediction accuracy, particularly for limonene, for which the highest RMSE and MAE values were obtained. The results confirm that nonlinear machine learning methods are more effective for the interpretation of multidimensional gas sensor array data and estimation of microbial volatile organic compound concentrations.

The comparison of the developed models demonstrated clear differences in prediction performance depending on the applied analytical approach and the investigated compound. Among all analyzed methods, the SVM models achieved the highest prediction accuracy, with coefficients of determination close to 1.0 and consistently low RMSE and MAE values for ethylbenzene, cumene, and limonene.

The MLP models also provided satisfactory estimation results, particularly for ethylbenzene and limonene. However, significantly lower prediction accuracy was observed for cumene, indicating reduced effectiveness of the neural network model for compounds occurring at trace concentration levels.

The low R^2 value obtained for the MLP model in the case of cumene may be attributed to the extremely narrow concentration range and low variance of the target variable. Neural networks generally require larger datasets and stronger signal variability to learn robust nonlinear relationships. Consequently, the MLP model was less effective for cumene than both SVM and linear regression.

In comparison, the linear regression models were characterized by lower coefficients of determination and higher prediction errors, especially in the case of limonene. The obtained results indicate that nonlinear machine learning approaches are more suitable for the interpretation of multidimensional gas sensor array signals and for the estimation of microbial volatile organic compounds in indoor air environments (Table 7).

CONCLUSIONS

The conducted study demonstrated the potential applicability of MOS gas sensor arrays combined with statistical and machine learning methods for the indirect estimation of selected microbial volatile organic compounds occurring in indoor environments. The proposed approach enabled the analysis of multidimensional sensor responses associated with varying levels of microbiological contamination observed in the investigated rooms.

Reference concentrations of ethylbenzene, cumene, and limonene determined using the SPME-GC/MS method revealed noticeable differences in both concentration levels and variability of the analyzed compounds. Among the investigated substances, limonene exhibited the highest concentration values and the widest variability range, whereas cumene was detected mainly at trace levels. These differences significantly

influenced the effectiveness of the developed estimation models and the quality of prediction results obtained from sensor signals.

The performed analyses confirmed that the applied MOS sensor matrix was sensitive to changes in the composition of indoor air and capable of generating characteristic multidimensional response patterns corresponding to different environmental conditions. The obtained sensor signals allowed the development of predictive models enabling quantitative estimation of selected mVOCs.

Comparison of the investigated modeling approaches demonstrated that machine learning methods provided substantially better prediction accuracy than classical linear regression models. Among all analyzed techniques, the support vector machine approach achieved the highest coefficients of determination together with the lowest RMSE and MAE values for all investigated compounds. The Multi-Layer Perceptron models also provided satisfactory results for ethylbenzene and limonene, although reduced prediction accuracy was observed for cumene, likely due to its very low concentration levels and limited variability within the analyzed dataset.

The obtained results indicate that nonlinear machine learning approaches are particularly suitable for the interpretation of multidimensional gas sensor array signals and estimation of microbial volatile organic compounds in indoor environments. The proposed methodology may therefore provide a basis for the development of rapid, low-cost, and non-invasive systems supporting indoor air quality assessment and preliminary detection of microbiological contamination in buildings.

REFERENCES

1. Wilson AD, Baietto M. Applications and advances in electronic-nose technologies. Sensors. 29 czerwca 2009;9(7):5099–148. https://doi.org/10.3390/s90705099
2. Röck F, Barsan N, Weimar U. Electronic Nose: Current Status and Future Trends. Chem Rev. 1 lutego 2008;108(2):705–25. https://doi.org/10.1021/cr068121q
3. Persaud K, Dodd G. Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose. Nature. września 1982;299(5881):352–5. https://doi.org/10.1038/299352a0
4. Li Z, Askim JR, Suslick KS. The optoelectronic nose: Colorimetric and fluorometric sensor arrays. Chem Rev. 9 stycznia 2019;119(1):231–92. https://doi.org/10.1021/acs.chemrev.8b00226
5. Korpi A, Järnberg J, Pasanen AL. Microbial volatile organic compounds. Critical Reviews in Toxicology. 1 lutego 2009;39(2):139–93. https://doi.org/10.1080/10408440802291497 PubMed PMID: 19204852
6. Kiani S, Minaei S, Ghasemi-Varnamkhasti M. Fusion of artificial senses as a robust approach to food quality assessment. Journal of Food Engineering. 1 lutego 2016;171:230–9. https://doi.org/10.1016/j.jfoodeng.2015.10.007
7. Capelli L, Sironi S, Rosso RD. Electronic noses for environmental monitoring applications. Sensors. 24 października 2014;14(11):19979–20007. https://doi.org/10.3390/s141119979
8. Haugen JE, Kvaal K. Electronic nose and artificial neural network. Meat Science. 1 stycznia 1998;Meat Consumption and Culture 44th International Congress of Meat Science and Technology49:S273–86. https://doi.org/10.1016/S0309-1740(98)90054-7
9. Gardner JW, Bartlett PN. A brief history of electronic noses. Sensors and Actuators B: Chemical. 1 marca 1994;18(1):210–1. https://doi.org/10.1016/0925-4005(94)87085-3
10. Miekisch W, Schubert JK, Noeldge-Schomburg GFE. Diagnostic potential of breath analysis—focus on volatile organic compounds. Clinica Chimica Acta. 1 września 2004;347(1):25–39. https://doi.org/10.1016/j.cccn.2004.04.023
11. Romain AC, Nicolas J. Long term stability of metal oxide-based gas sensors for e-nose environmental applications: An overview. Sensors and Actuators B: Chemical. 29 kwietnia 2010;Selected Papers from the 13th International Symposium on Olfaction and Electronic Nose146(2):502–6. https://doi.org/10.1016/j.snb.2009.12.027
12. Breiman L. Random Forests. Machine Learning. 1 października 2001;45(1):5–32. https://doi.org/10.1023/A:1010933404324
13. Zhang C, Qian L, Zeng W. MOS based gas sensor in detection of volatile organic compounds: A review. Sensors and Actuators A: Physical. 16 października 2025;393:116818. https://doi.org/10.1016/j.sna.2025.116818
14. Pang Z, Chen Q, Han W, Zheng L. Value-centric design of the internet-of-things solution for food supply chain: Value creation, sensor portfolio and information fusion. Inf Syst Front. 1 kwietnia 2015;17(2):289–319. https://doi.org/10.1007/s10796-012-9374-9
15. Singh S, S S, Varma P, Sreelekha G, Adak C, Shukla RP, et al. Metal oxide-based gas sensor

array for VOCs determination in complex mixtures using machine learning. Microchim Acta. 13 marca 2024;191(4):196. https://doi.org/10.1007/s00604-024-06258-8

16. Wiley.com [Internet]. [cytowane 13 maja 2026]. Statistical Learning Theory | Wiley. Dostępne na: https://www.wiley.com/en-us/Statistical+Learning+Theory-p-9780471030034
17. Guz Ł, Łagód G, Jaromin-Gleń K, Suchorab Z, Sobczuk H, Bieganowski A. Application of gas sensor arrays in assessment of wastewater purification effects. Sensors. 23 grudnia 2014;15(1):1–21. https://doi.org/10.3390/s150100001
18. Gutierrez-Osuna R. Pattern analysis for machine olfaction: a review. IEEE Sensors Journal. czerwca 2002;2(3):189–202. https://doi.org/10.1109/JSEN.2002.800688
19. Marco S, Gutierrez-Galvez A. Signal and data processing for machine olfaction and chemical sensing: A review. IEEE Sensors Journal. listopada 2012;12(11):3189–214. https://doi.org/10.1109/JSEN.2012.2192920
20. Li G, Park S, Kim H, Hilal M, Cai Z. Artificial intelligence-enabled MOS gas sensors: Towards Selective and intelligent detection in complex environments. Sensors and Actuators A: Physical. 1 kwietnia 2026;400:117502. https://doi.org/10.1016/j.sna.2026.117502
21. Nguyen NV, Nguyen VC, Le HM, Nguyen VH. Machine learning-enhanced multi-gas discrimination with a miniaturized MOS sensor array. J Comput Electron. 15 czerwca 2025;24(4):117. https://doi.org/10.1007/s10825-025-02354-x
22. Gutarowska B, Piotrowska M. Methods of mycological analysis in buildings. Building and Environment. 1 kwietnia 2007;42(4):1843–50. https://doi.org/10.1016/j.buildenv.2006.02.015
23. James D, Scott SM, Ali Z, O'Hare WT. Chemical sensors for electronic nose systems. Microchim Acta. 1 lutego 2005;149(1):1–17. https://doi.org/10.1007/s00604-004-0291-6
24. Han J, Li H, Cheng J, Ma X, Fu Y. Advances in metal oxide semiconductor gas sensor arrays based on machine learning algorithms. J Mater Chem C. 27 lutego 2025;13(9):4285–303. https://doi.org/10.1039/D4TC05220J
25. Veríssimo MIS. A critical review of the analytical performance of the most recent MOS-based gas sensors for indoor air quality monitoring of WHO priority pollutants. TrAC Trends in Analytical Chemistry. 1 września 2024;178:117813. https://doi.org/10.1016/j.trac.2024.117813