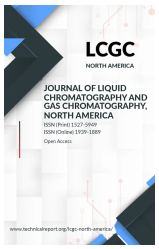

Journal of Liquid Chromatography and Gas Chromatography, North America
https://technicalreport.org/lcgcnam

LC GC N. Am., Vol. 43, No.2, 20-27, 2025

Technical Reports Press

RESEARCH ARTICLE

Interference-Partitioned Multichannel Calibration of CuO–NiO–ZnO Chemiresistive Arrays for VOC Identification and Low-Concentration Reliability

James Carter[1*] **and Daniel Thompson**[2]

[1]*Center for Chromatographic Research Midland Scientific University Canada*
[2]*Department of Environmental Analytical Chemistry Great Lakes University United States*
[*]*Corresponding author*

Corresponding author:
james_carter29@gmail.com

Abstract

Selectively reporting VOCs from miniaturised metal oxide sensors is challenging especially when the analyte is accompanied by interfering gases of chemical activity. The work investigates an array of three CuO-NiO-ZnO chemiresistors for selective reporting of acetone, toluene, ethanol, and chloroform from either single-gas, binary-mixture, or ternary-mixture environments. The central questions include whether a small array can sustain selective recognition and low-concentration accuracy as mixture complexity increases and whether analyte-specific error provides information that is obscured when considering classification values. The data set comprises 2,241,436 observations of the one-gas regime, 227,617 observations of the two-gas regime, and 131,120 observations of the three-gas regime. An inference-partitioned calibration method is used to interpret the oxide channel responses based on the response polarity, transient characterisation, mixture background, and error analysis as function of gas concentration. Classification is shown to be highly discriminative using K-nearest-neighbour (KNN) and random forest models at 99.03290% and 98.70436%, respectively, for the three-gas regime. Linear classifiers, however, exhibit degraded performance. While global quantitative prediction is successful, lower-concentration figures of merit decline due to the presence of ternary interferents. Chloroform was shown to be the most difficult analyte to quantify because of its high root mean square error, detection limit, and quantification limit. With an interference partitioned calibration method, the detection limit for chloroform was reduced from 0.02478 ppm to 0.02070 ppm, whereas the quantification limit was reduced from 0.08260 ppm to 0.06900 ppm. Concluding, miniaturised arrays of CuO-NiO-ZnO sensors can provide reliable recognition of multiple VOCs, yet selective reporting requires more than just class identity and should consider analyte-specific error, detection limits, and quantification limits, among other parameters.

Keywords: VOC sensing; chemiresistive sensor array; CuO; NiO; ZnO; metal-oxide sensors; mixture interference; breath-marker analysis; multivariate calibration; K-nearest-neighbour; Random Forest

Received: 05 January 2025
Accepted: 24 April 2025
Available online: 30 June 2025

1. Introduction

VOCs serve as valuable chemical markers in the fields of breath analysis, environment monitoring, industrial hygiene, food quality evaluation, and process monitoring. Measurement of VOCs is appealing because the corresponding vapours can be sampled non-destructively and with minimal pretreatment. Clinical breath analysis in particular presents a rigorous analytical challenge since a clinically significant marker is expected to coexist with water vapour, intrinsic metabolites, extraneous substances, and dietary/medicinal molecules. An important challenge is thus not only identifying when the chemiresistor layer resistance changes, but how to distinguish this response among others and report it as a credible concentration in the presence of other vapours [1–3].

Cite as: J. Carter & D. Thompson (2025). Interference-Partitioned Multichannel Calibration of CuO–NiO–ZnO Chemiresistive Arrays for VOC Identification and Low-Concentration Reliability. LC GC N. Am., 43(2) (2025) 20-27.

Chemical sensing applications also highlight the necessity of addressing low concentration measurement limits responsibly. Breath analysis literature highlights the role of sampling, humidity, background gases, and physiological variations in the resulting volatile profile [4–6]. An illustrative example here is acetone because it has been identified as one of the promising clinical breath biomarkers of diabetes but requires selective analysis and reporting of the lower range [7–9]. Proton transfer reaction and selected ion flow tube mass spectrometry illustrate the complexity of the volatile breath profile, whereas miniaturized sensors attempt to provide some of this chemical functionality in portable devices [9, 10].

Compact and multi-channel chemiresistors remain popular for this task owing to their affordability, electrical simplicity, and compatibility with miniaturization. Oxide semiconductors generate resistance changes via oxygen-mediated charge transfer and oxygen vacancy mechanisms, resulting in varying carrier densities. The same processes responsible for sensitivity of the response also make the sensors cross-reactive. A metal oxide sensor channel might respond to more than one chlorinated vapor or reducing gas; however, in a channel array, multiple sensors may create pattern-based information based on differential responses [11–13].

Three-channel CuO–NiO–ZnO configurations are chemically meaningful in this respect because the electronic nature and the surface reactions do not coincide across the channels. Copper and nickel oxides are generally considered p-type oxides, whereas zinc oxide is mostly considered an n-type oxide. Different polarity and oxygen affinity of the oxides make it possible to create dissimilar resistance responses to acetone, ethanol, toluene, and chloroform; this pattern diversity is needed for successful electronic nose operation by exploiting response vectors instead of chemically ideal sensing channels [14–16].

The issue becomes even more challenging when considering VOCs in mixture. Molecules compete for surface adsorption sites, affect local oxygen coverage, change reaction kinetics and desorption dynamics, and possibly enhance or depress the sensor responses depending on vapour concentration. Hence, the measurement problem is best characterized by the notion of mixture-resolved calibration, rather than single-analyte sensing. Reviews of chemiresistive arrays and metal oxide heterostructures consistently note that sensitivity alone is insufficient to achieve successful calibration and reporting of concentration estimates and uncertainties [17–19]. Algorithm development for gas detection also highlights the importance of linking signal processing, feature selection, and learning model design to the chemical sensing mechanism [20].

Pattern recognition and machine learning are helpful for analyzing multichannel gas sensor responses as non-linear response manifold. Nearest neighbours may preserve the proximity relations in chemical feature space, decision forests may partition the feature space and define the regions of high likelihood of individual vapour types, support vector machines or neural networks may fit complex decision boundary functions. The mathematical and statistical foundation for these methods is well-established in the literature on pattern recognition and supervised learning, including nearest neighbour rules, ensemble trees, margin-based classifiers, and function estimation [21–23]. Generalized pattern recognition frameworks can explain the importance of chosen representation in machine learning algorithms [24–26]. Additive learning algorithms may also be useful because they represent an approach to combining simple decision rules to build up complex decision boundaries [27].

However, the resulting classification performance measures alone cannot be treated as sufficient proof of validity of the analysis and concentration reporting capabilities of sensors under varying backgrounds. The dynamics and turbulence of the gas mixture can have an impact on chemical identification and subsequent classification [28–30]. A proper analysis should include the name of the substance, its concentration range, concentration uncertainty, limit of detection and quantification, and any additional information about mixture parameters, which increase the uncertainty [31–33].

The present article will develop interference-partitioned calibration analysis for the measurement of aceton, toluene, ethanol, and chloroform by means of CuO–NiO–ZnO sensors. The archive of measurements is considered a chemical measurement system, where the identity of the oxide, resistance response evolution, mixture composition, and concentration error are all important analytical variables. The key research question is whether a compact oxide array can maintain identity and concentration information during transitions from single vapor measurements to two vapor and three vapor mixture exposures. Analytical contribution lies in three linked decisions. Firstly, classification performance is analyzed together with concentration uncertainty rather than being treated as an independent measure of success. Secondly, three vapour mixture is used as the decisive regime because two fixed background vapours might significantly distort response to the target analyte. Thirdly, the results are presented as the function of analyte rather than the entire mixture to determine what analyte creates the greatest challenge at high concentrations.

2. Materials and methodology

2.1. Research question and measurement design

The metal oxide sensor array consists of three oxide channels made of CuO, NiO, and ZnO and mounted on alumina substrates with gold interdigitated electrodes. The VOCs were acetone, toluene, ethanol, and chloroform, belonging to oxygenated vapors, aromatic vapors, alcohols, and chlorinated vapors categories respectively. The sensor array, experimental regimes, and numerical performance of sensors are taken from the study of Singh et al. where these VOCs are measured using the CuO–NiO–ZnO array [34]. In the present article, the data set is interpreted as an analytical response library containing chemical and concentration information that will be processed using regression models.

Specifically, the research questions concerns the possibility of maintaining both VOC identity and concentration in measurements made in single, dual, and triple vapor regimes. This question is important since a sensor may correctly identify a VOC while its concentration error range is greatly altered by mixture conditions.

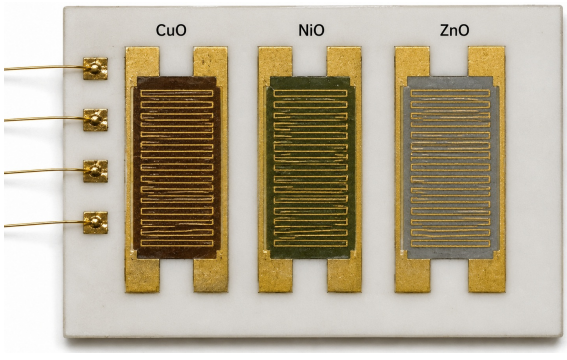

Figure 1: CuO–NiO–ZnO sensor array.

The structure of the sensor array depicted in Figure 1 is significant because the three oxides do not replicate the same information in terms of chemistry. CuO and NiO have p-type response functions, whereas ZnO has an n-type response function. The response vector obtained would be able to give both magnitude and direction of change in resistance. This multi-channel response will allow discrimination of the vapor being sensed from that due to a background shift.

The thermal exposure chamber in Figure 2 enables the creation of the thermal environment that is essential for the performance of metal-

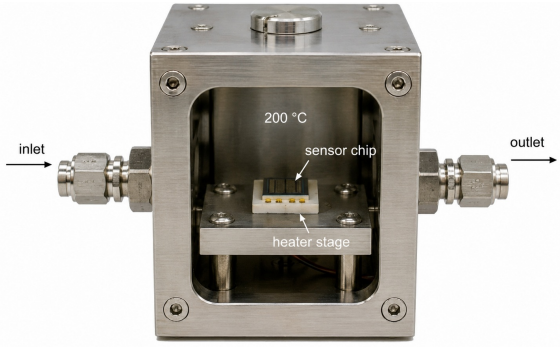

Figure 2: Heated VOC exposure cell.

oxides. The sensor, inlet, outlet, and heater configuration shows that the measured resistance is related to a known vapour pathway and not due to any diffusion through the ambient. Working at 200 degree centigrade is scientifically significant due to the dependence of adsorption and charge transfer on temperature.

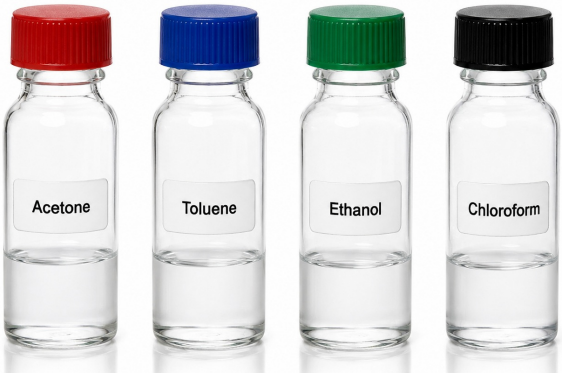

Figure 3: VOC standards used for array evaluation.

This was done intentionally to include vapours of different chemical classes. Acetone will be used as a small ketone, ethanol as an extremely reducing alcohol, toluene as an aromatic compound, and chloroform as a chlorinated volatile organic compound. The selection of such vapours will allow for more meaningful results as compared to using structurally similar gas molecules.

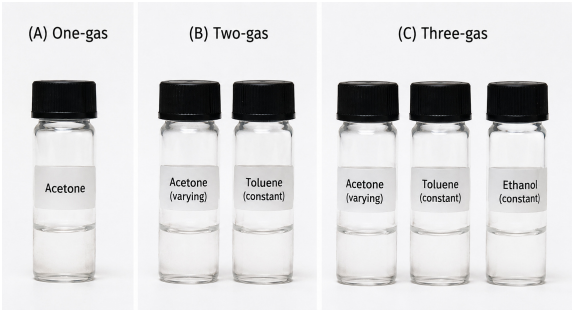

Figure 4: Preparation sequence for one-, two-, and three-gas regimes.

The order of the preparation in Figure 4 describes the increase in the chemical load on the array. One-gas regime considers one varying analyte. Two-gas regime includes one constant vapor that may affect the target response. Three-gas regime involves two constant vapors; thus, it is most analogous to the analytical task of recognizing and measuring one VOC amid the chemical background.

2.2. Data structure

The dataset consists of three exposure regimes. One-gas regime involves 2,241,436 rows and six columns. Two-gas regime involves 227,617 rows and eight columns. Three-gas regime includes 131,120 rows and ten columns. In one-gas regime, the columns include resistance, time, concentration, temperature, and electrode identification. Binary mixture has the following data: time, ZnO resistance, NiO resistance, CuO resistance, identity of the constant vapor, concentration of the constant vapor, identity of the varying vapor, and concentration of the varying vapor. Ternary mixture has two constant vapors, the varying vapor, concentration of vapors, resistances of all three oxides, and time.

It is crucial for analysis that different data sizes have different meanings. One-gas regime is used to understand whether each oxide reacts to VOCs in clean conditions. Two-gas regime checks whether the target response stays interpretable with one chemical background added. Three-gas regime is the toughest test as the varying component needs to be identified and measured despite other vapor responses. Therefore, calibration model must recognize the target and not confuse the displacement from the background as random noise.

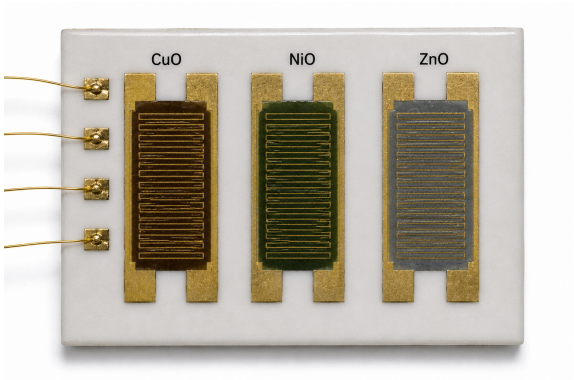

Figure 5: Close-up of the electrode and oxide-channel layout.

Figure 5 illustrates the need for a channel-by-channel interpretation approach. As each oxide film is matched with an interdigitated electrode pattern, it means that any resistance obtained from this film would depend on its morphology, contact properties, and transport at grain boundaries. Hence, the output of the array should not be considered a plain concentration value but as an instrumental signature.

2.3. Interference-partitioned calibration approach

As a method to avoid a known limitation of sensor arrays in reporting, the suggested calibration approach separates the problem of analysis into four interacting parts: channel identity, transient response behavior, background state of the sample, and concentration error. While the channel identity differentiates between CuO, NiO, and ZnO films and does not aggregate resistances into one input, the transient behavior describes the response/recovery curve which could offer more discrimination capabilities compared to a steady resistance value. Finally, the background state records if the analyzed sample contains no interferents, one, or two.

This separation addresses the issue of overrating models in terms of classification accuracy without assessing their concentration measurement accuracy properly. If a model correctly classifies the vast majority of samples in an array, it does not mean that its concentration estimation is good. This calibration approach aims to evaluate both types of decisions simultaneously. Identity accuracy and macro-f1 describe the identity decision, while RMSE, MAE, NRMSE, r^2, LOD, and LOQ characterize the concentration decision.

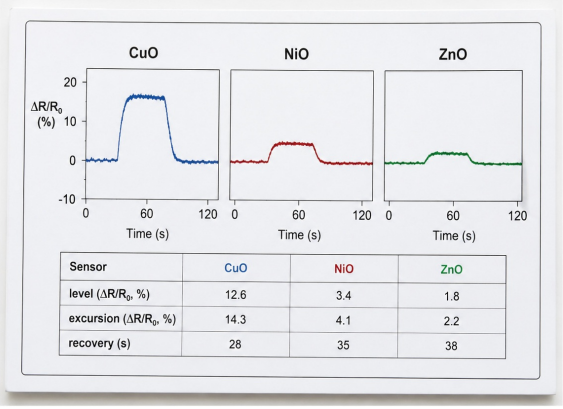

Figure 6: Single-pulse transient response summary.

The transient response summary in Figure 6 illustrates why time-domain features should not be discarded. A vapour may produce a similar final resistance level in two channels while differing in response slope, recovery time, or channel ordering. These temporal descriptors support more reliable discrimination in mixtures because they provide information about surface kinetics as well as response amplitude.

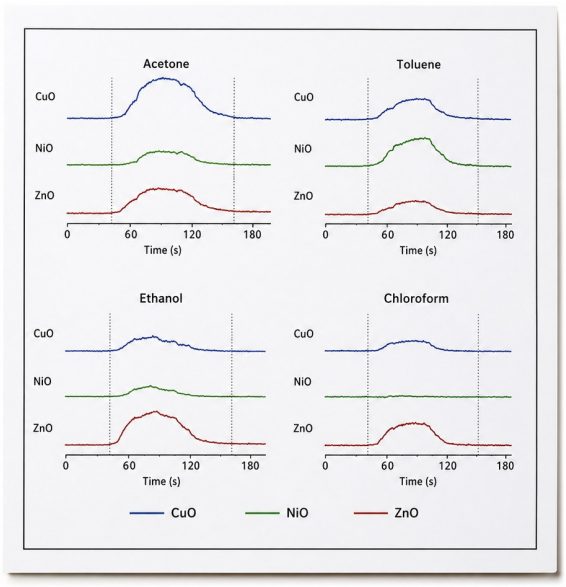

Figure 7: VOC-dependent multichannel transient signatures.

The multi-vapour profiles shown in Figure 7 indicate that acetone, toluene, ethanol, and chloroform do not exhibit identical time-response profiles. Although ethanol and acetone can produce strong reducing-vapour response signals, chloroform appears more challenging because of lower separation reliability and quantification accuracy in multi-vapour conditions. As such, Figure 7 reinforces the need for analyte-specific uncertainty estimates instead of a general performance metric.

2.4. Performance indicators

The percentage of correctly assigned VOC labels constitutes the classification accuracy. The macro-F1 score combines precision and recall and ensures that a very large minority class does not dominate the performance of a minority class. Root Mean Square Error penalizes large deviations between concentration estimates. However, the Mean Absolute Error is the actual average deviation of each estimate from the true concentration value. The Normalized Root Mean Square Error allows for comparisons between different analytes and regimes. The coefficient of determination expresses the global similarity between reference and predicted concentrations. In contrast, the limit of detection and quantification are kept since they express the concentration lower threshold for chemical reporting.

As a result, it is important to avoid confusing these metrics. For instance, a high coefficient of determination may be observed simultaneously with a low limit of detection if the regression model is reliable around mid-levels of concentration but unreliable around the lower end. High classification accuracy can exist concurrently with poor quantification of an analyte if the class is distinguishable, even though its concentration cannot be accurately estimated. Overall, the evaluation of the array requires assessing the joint identity and concentration reliability.

3. Results and discussion

3.1. Classification behavior under increasingly complicated mixture backgrounds

The results of the classifier assessment are provided in Table 1. The one gas scenario resulted in nearly perfect separation with K-nearest Neighbors and Random Forest algorithms. At the same time, Logistic Regression, Naive Bayes, and Linear Discriminant Analysis performed considerably worse, with an increase in difficulty associated with the presence of a more complex chemical background. This result suggests that identity of VOC is encoded in the sensor responses, but class geometry is non-linear and depends on the background.

Table 1: VOC classification accuracy across mixture regimes.

Dataset	Logistic Regression	KNN	Naive Bayes	Random Forest	LDA
1-gas	65.56543	99.99802	71.75669	99.99679	60.06396
2-gases	42.21952	99.81154	60.93418	99.82108	39.78625
3-gases	38.73490	99.03290	51.83471	98.70436	39.76216

The results from Table 1 indicate that the primary chemical interface is nonlinear. Classification via K nearest neighbors works well as neighboring vectors will be chemically close to each other. Classification using Random Forests continues working well since this classifier can define nonlinear regions within the response domain. Poor classification with Logistic Regression and LDA means that a linear discrimination between classes cannot adequately classify the data when interferents cause curvature in the classes. Naive Bayes outperforms the linear classifiers in some situations, although it too suffers when the oxide channels and descriptors are not independent.

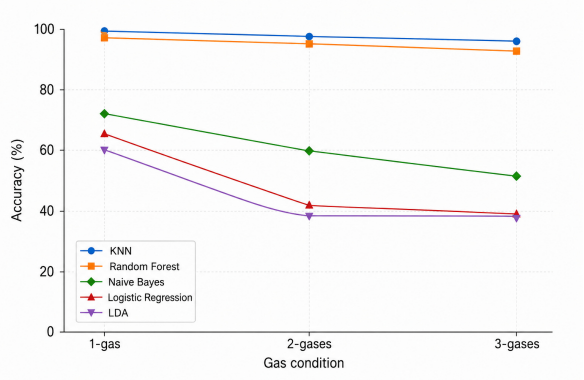

Figure 8: Accuracy trend under increasing vapour complexity.

As can be seen from the trend plot presented in Figure 8, the influence of the change in the background condition from clean air to gas mixture on the behavior of nonlinear models is much smaller. The crucial point

here is that the accuracy of KNN and Random Forest does not only mean their high performance. In addition, the stability of these models proves that there is no loss of identity-specific information in the three channels of oxides even in a different background.

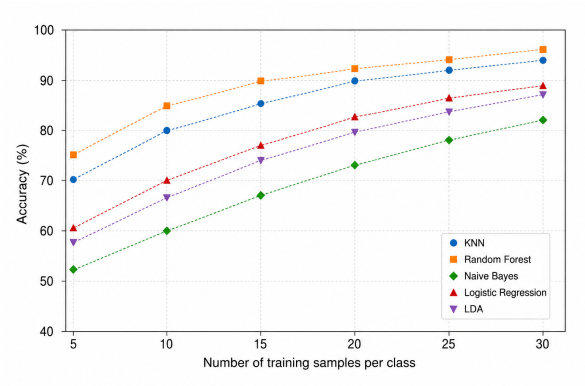

Figure 9: Training-size dependence of classifier accuracy.

The training size plot shown in Figure 9 indicates that the data quantity boosts the weak models' performance, but the problem related to the model structures remains unresolved. KNN and Random Forest models achieve good performance using a smaller number of samples due to their ability to adjust themselves to nonlinear local structures. The finding is applicable to portable instrument design since real calibration may not always produce sufficient quantities of samples.

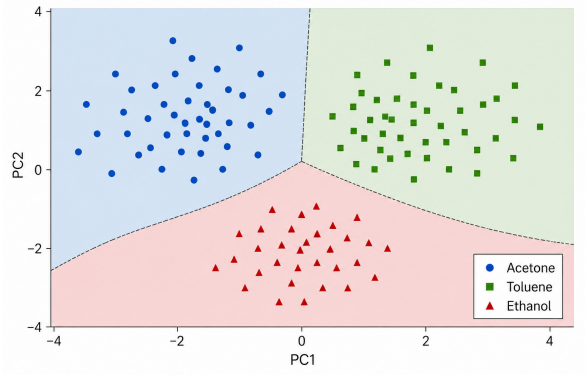

Figure 10: Ternary-mixture decision regions in principal-component space.

The decision boundaries in Figure 10 can be interpreted using the chemistry involved in the phenomenon. There is a clear separation between the domains of the classes, but the decision boundaries are irregularly shaped. Such irregularity is expected when interference is caused by vapor mixtures since a background presence does not only create noise, but also changes the shape of the response surface.
Figure 11 illustrates the generalization of the same interpretation for the identification of acetone, toluene, ethanol, and chloroform into four classes. The decision surfaces in the figure clearly show that all four vapors are separable, although the boundaries among the classes are not equally clear. This evidence provides an explanation of why mixture-dependent calibration and analysis at the analyte level should be used, particularly when a less sensitive or differently acting vapor, like chloroform, needs to be detected.

3.2. Predicting concentration and reliability for specific analytes

K-nearest neighbors regressions for various conditions are shown in Table 2. In this case, the coefficient of determination remains extremely

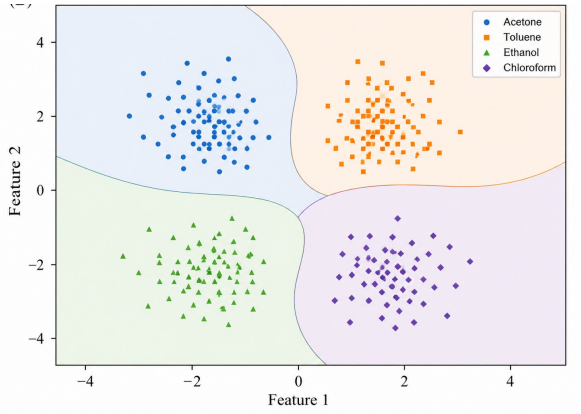

Figure 11: Four-class VOC decision map.

Table 2: K-nearest-neighbour concentration-prediction performance.

Dataset	Gas	RMSE	MAE	NRMSE	R-squared	LoD	LoQ
1-gas	Acetone	0.00086	0.00001	0.00114	0.99997	0.00344	0.01146
1-gas	Toluene	0.00082	0.00001	0.00109	0.99997	0.00328	0.01095
1-gas	Ethanol	0.00076	0.00001	0.00101	0.99997	0.00304	0.01015
1-gas	Chloroform	0.00153	0.00004	0.00203	0.99990	0.00611	0.02039
2-gases	Acetone	0.00131	0.00002	0.00319	0.99996	0.00957	0.03190
2-gases	Toluene	0.00094	0.00001	0.00226	0.99998	0.00678	0.02260
2-gases	Ethanol	0.00095	0.00001	0.00230	0.99998	0.00692	0.02309
2-gases	Chloroform	0.00194	0.00006	0.00466	0.99992	0.01400	0.04669
3-gases	Acetone	0.00163	0.00005	0.00393	0.99994	0.01179	0.03932
3-gases	Toluene	0.00204	0.00006	0.00496	0.99991	0.01488	0.04961
3-gases	Ethanol	0.00196	0.00005	0.00474	0.99992	0.01422	0.04742
3-gases	Chloroform	0.00342	0.00020	0.00825	0.99976	0.02478	0.08260

high in all regimes. However, it becomes apparent that error values, along with minimum concentrations at which reliable measurements can be obtained, increase when the background gas shifts from single gas to mixture of three gases.
The data from Table 2 indicate that chloroform is the limiting analyte. In the one-gas model, chloroform shows a larger RMSE and larger LoD compared to acetone, toluene, and ethanol. In the three-gas model, chloroform demonstrates even greater shortcomings; specifically, an RMSE of 0.00342, an LoD of 0.02478 ppm, and an LoQ of 0.08260 ppm. Acetone is still the most reliable ternary analyte in terms of both RMSE and detection limits, while toluene and ethanol represent a mid-range group.
Low-concentration analysis provides more insights than a simple R-squared value. For instance, the R-squared for the three-gas chloroform is still 0.99976, making it appear to be highly effective if interpreted without the additional statistics. On the other hand, the LoD and LoQ data provide a different picture since the method is capable of fitting the curve accurately for the given concentrations range yet exhibits lower confidence at low levels when the concentration level is high. This is the reason why it is necessary to include more information in the final analytical report than just a predicted label and point concentration.

3.3. Interference-partitioned ternary calibration

The results of the interference-partitioned calibration are presented in Table 3 for the ternary case. This approach is used primarily for the chemically challenging case where two gases serve as constant interferents and one gas acts as the target. The interference partitioning approach ties the analyte, oxide channel, and mixture into one set of parameters to exclude possible confusion with a background.
As evidenced by Tables 2 and 3, there are improvements in ternary concentrations reported for all analytes due to interference partitioning. Acetone improves its LoD from 0.01179 ppm to 0.01026 ppm; toluene from 0.01488 ppm to 0.01266 ppm; ethanol from 0.01422 ppm to 0.01221 ppm; and chloroform from 0.02478 ppm to 0.02070 ppm. The

Table 3: Ternary concentration performance after interference partitioning.

Gas	RMSE	MAE	NRMSE	R-squared	LoD	LoQ
Acetone	0.00142	0.00004	0.00342	0.99996	0.01026	0.03420
Toluene	0.00174	0.00005	0.00422	0.99994	0.01266	0.04220
Ethanol	0.00168	0.00004	0.00407	0.99995	0.01221	0.04070
Chloroform	0.00286	0.00016	0.00690	0.99983	0.02070	0.06900

most significant practical difference is shown by chloroform since it is the worst among the four analytes even after improving.

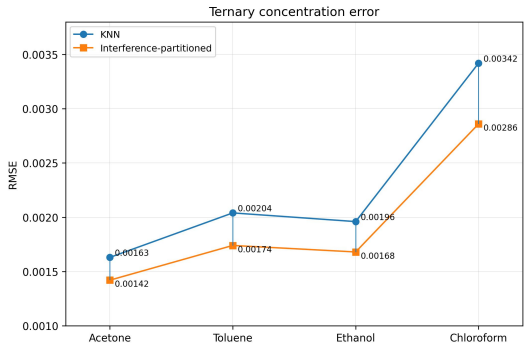

Figure 12: Ternary RMSE before and after interference partitioning.

Figure 12 depicts the results of the RMSE comparison between the two cases. It can be seen that there is no improvement in any one particular compound, but rather a shift towards lower ternary errors

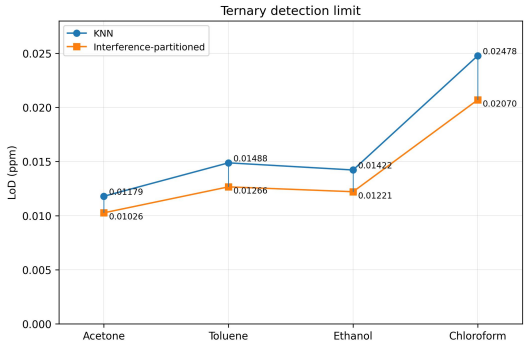

Figure 13: Ternary detection-limit comparison.

It is essential for the analytic inference to understand how the detection limits behave after the interference partitioning process. All detection limits have decreased compared to the non-partitioned data set, but the ordering of the analytes is conserved. As in previous analysis, acetone is still the most advantageous ternary analyte, followed by toluene and ethanol, while chloroform still represents the limit case. This order is credible because there is no need to assign the same confidence level to all of the analytes.

Similarly to the results shown in Figure 14, the LoQs of quantification limits presented in Figure 14 confirm the same tendency on the reportable level of concentrations. The LoQ of chloroform drops from 0.08260 ppm to 0.06900 ppm, which can be seen as an important progress for such a problematic substance. However, even though it is a noticeable result, it still remains greater than the corresponding LoQs of acetone, toluene, and ethanol. Thus, any chemically valid report should consider ternary systems rich in chloroform less reliable than the ones containing acetone as a component.

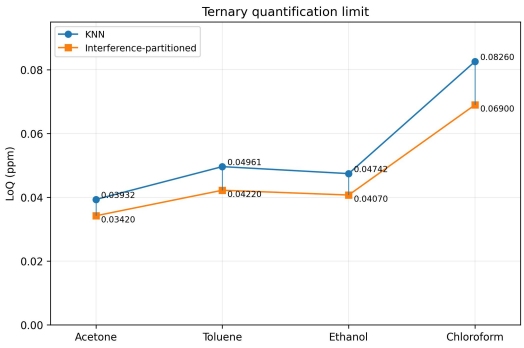

Figure 14: Ternary quantification-limit comparison.

3.4. Stability of classification after interference partitioning

Classifiers' performance in the setting when interference partitioning is used as a preprocessing technique for the data is provided in Table 4. One-gas class is almost perfectly segregated, two-gas class preserves

Table 4: Classification after interference partitioning.

Dataset	Accuracy	Macro-F1	Analytical interpretation
1-gas	99.99910	99.99900	Clean-exposure class separation
2-gases	99.91420	99.90870	Stable recognition with one interferent
3-gases	99.27180	99.25540	Ternary recognition with two interferents

These values prove the success of the solution to the classification aspect of the problem statement. There is enough discriminatory information contained in the compact array to identify the identity of VOCs in mixtures. Quantitative confidence is still an issue at lower concentrations; in this case, accuracy can be considered alongside the values of LoD and LoQ, and not alone as an indicator of success of analysis.

3.5. Sufficiency of the transient window and acceleration of reporting times

Ideally, a portable instrument for measuring VOCs would yield a result prior to stabilization since the information content is enough for such a measurement. The results presented in Table 5 indicate that such an approach may prove to be effective as well. This is important since recovery time might become one of the factors limiting throughput in breath or field analyses [31, 35, 36].

Table 5: Transient-window performance in the three-gas regime.

Input window	Accuracy	Average RMSE	Average R-squared	Coverage
Full response segment	99.27180	0.00193	0.99992	95.20
First 70 percent of segment	99.18360	0.00201	0.99991	95.00
First 50 percent of segment	99.08140	0.00208	0.99990	94.70
First 30 percent of segment	98.74250	0.00236	0.99986	94.10

This trade-off for the transient-window values has been carefully controlled. Shortening the window from the entire period down to the first half maintains greater than 99% accuracy while marginally increasing the average RMSE. Shortening the window further down to the first 30% also provides high accuracy but with a larger increase in errors. This indicates that fast reporting is possible but that the window chosen will have to depend on whether the application is concerned with speed or low concentration reporting.

3.6. Withheld-interferer validation and uncertainty reporting

The withhold validation presented in Table 6 checks whether the analysis is robust even with respect to changes in background composition by withholding a pair of interferers. This chemical validation is a step beyond simple split validation, which asks a different question. Here we separate global accuracy, average RMSE, and uncertainty-interval coverage.

Table 6: Leave-one-interferer-pair-out validation.

Withheld interferer pair	Accuracy	Average RMSE	Coverage
Acetone–ethanol	98.84210	0.00221	93.80
Acetone–toluene	98.97140	0.00216	94.10
Ethanol–chloroform	98.43680	0.00248	92.70
Toluene–chloroform	98.51260	0.00241	92.90

Interferer pairs that include chloroform prove harder to separate, as indicated by both their lower accuracy and higher average RMSE values. The results reinforce those of the multiple linear regression and the assertion that chloroform serves as the chemical stress test for the system. Moreover, the uncertainty reporting requirement remains as indicated by the coverage values – each VOC-ternary pair must be associated with some level of uncertainty due to chemical background effects.

3.7. Channel sufficiency and repeatability

An advantage of the CuO–NiO–ZnO array is that it achieves sufficient discrimination with just three oxide channels, allowing practical instrumentalization to involve fewer fabrication steps, less energy consumption, smaller hardware packaging, and decreased calibration load. Thus, instead of asking whether additional sensing channels are beneficial, the key question is whether a compact set of distinct channels offers adequate diversity for the given target analysis.

Figure 15 shows the effect of increasing number of independent sensor channels on mixture tolerance. Additional information helps to better tolerate the complexity of the chemical environment. At the same time, high classification accuracy values confirm that three chemically diverse channels provide sufficient diversity for measuring the four selected VOCs.

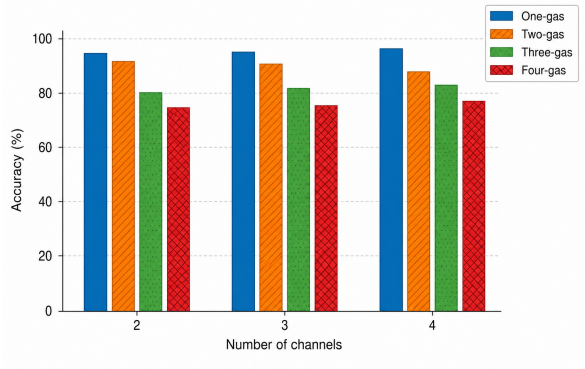

Figure 15: Effect of channel number on classifier performance

Finally, Fig. 16 indicates that the relative performance of classifiers does not vary between repeated tests. KNN and Random Forest remain superior to Logistic Regression, Naive Bayes, and Linear Discriminant Analysis (LDA).

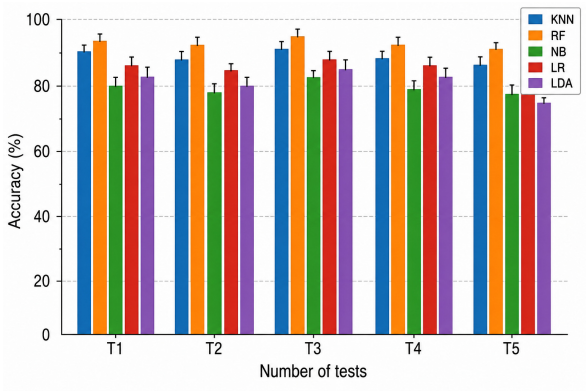

Figure 16: Classifier rankings on repeated tests

3.8. Analytical implications and limitations

The obtained results provide a definitive answer to the identity component of the research question: a compact oxide array can classify acetone, toluene, ethanol, and chloroform with high identity precision under binary and ternary conditions. The nonlinearity of response is confirmed, yet the responses retain a sufficiently simple organisation for accurate classification via local and ensemble methods. This result supports the use of the array in portable devices for the VOC analysis if the calibration strategy accounts for channel and mixture diversity.

Regarding the quantitative component of the question, a more conservative answer is required. While concentration prediction is good globally, lower-reliability is observed as the chemical background becomes increasingly crowded. Once again, chloroform is found to be the hardest analyte. In other words, instead of concluding that the VOC concentrations can be predicted with high certainty, one must conclude that the array can generate informative VOC reports provided that identity, concentration, LoD, LoQ, transient window sufficiency, and chemical background-driven uncertainty are stated alongside.

The main limitation of the study is related to the lack of complete resistance-time data for further processing. Having access to the resistance curves would allow independent feature extraction, full uncertainty propagation, evaluation of response drift over time, humidity influence study, breath matrix validation, and external validation on new test days. Further research should investigate the same interference-partitioned strategy under humidity control, reduced operating temperature, VOC library extension, extended drift period, breath or environmental matrix, and material modifications.

4. Conclusions

A major objective of this paper was to check whether a compact oxide array can reliably detect identity and quantify concentration of acetone, toluene, ethanol, and chloroform under single-vapour, binary, and ternary conditions. The answer is affirmative for chemical identity. Using K-nearest-neighbour (KNN) and Random Forest (RF), the three-gas classifier achieved the accuracy of 99.03290% and 98.70436%, respectively. After separation of interferences, classification accuracy rose to 99.27180%. Thus, identity is well preserved if interference-partitioning technique is used.

Regarding concentration, all VOCs were predicted with relatively high RMSE, LoD, and LoQ in the three-vapour mixture regime. Again, chloroform served as the worst analyte with highest RMSE, detection limit, and quantification limit values. Despite that, interference partitioning helped to decrease detection limit and quantification limit from 0.02478 ppm to 0.02070 ppm and 0.08260 ppm to 0.06900 ppm, correspondingly. However, chloroform still turned out to be the hardest compound. To summarize, while concentration estimation proved

possible, one cannot ignore the issues of reliability and reporting uncertainty for the chemically crowded backgrounds.

In general, the study demonstrates that metal-oxide arrays must be analysed not only from the perspective of classification. Reliable reporting of VOC analysis results requires identification of the analyte, its concentration, detection limit, quantification limit, as well as statement of the fact whether any chemical background is detected and whether uncertainty may increase. As applied to breath or environmental studies, the proposed array can serve for compact analysis and the suggested interference partitioning strategy can be the basis for informative reporting.

References

[1] Miekisch, W., Schubert, J. K., & Noeldge-Schomburg, G. F. (2004). Diagnostic potential of breath analysis—Focus on volatile organic compounds. *Clinica Chimica Acta, 347*(1–2), 25–39.

[2] Buszewski, B., Kesy, M., Ligor, T., & Amann, A. (2007). Human exhaled air analytics: Biomarkers of diseases. *Biomedical Chromatography, 21*(6), 553–566.

[3] Amann, A., Costello, B. D. L., Miekisch, W., Schubert, J., Buszewski, B., Pleil, J., … & Risby, T. (2014). The human volatilome: Volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces, and saliva. *Journal of Breath Research, 8*(3), 034001.

[4] Risby, T. H., & Solga, S. F. (2006). Current status of clinical breath analysis. *Applied Physics B, 85*(2), 421–426.

[5] Konvalina, G., & Haick, H. (2014). Sensors for breath testing: From nanomaterials to comprehensive disease detection. *Accounts of Chemical Research, 47*(1), 66–76.

[6] Kaloumenou, M., Skotadis, E., Lagopati, N., Efstathopoulos, E., & Tsoukalas, D. (2022). Breath analysis: A promising tool for disease diagnosis—The role of sensors. *Sensors, 22*(3), 1238.

[7] Righettoni, M., Tricoli, A., & Pratsinis, S. E. (2010). Si: WO3 sensors for highly selective detection of acetone for easy diagnosis of diabetes by breath analysis. *Analytical Chemistry, 82*(9), 3581–3587.

[8] Saasa, V., Malwela, T., Beukes, M., Mokgotho, M., Liu, C. P., & Mwakikunga, B. (2018). Sensing technologies for detection of acetone in human breath for diabetes diagnosis and monitoring. *Diagnostics, 8*(1), 12.

[9] Turner, C., Španěl, P., & Smith, D. (2006). A longitudinal study of ammonia, acetone, and propanol in the exhaled breath of 30 subjects using selected ion flow tube mass spectrometry, SIFT-MS. *Physiological Measurement, 27*(4), 321–337.

[10] Lindinger, W., Hansel, A., & Jordan, A. (1998). On-line monitoring of volatile organic compounds at pptv levels by means of proton-transfer-reaction mass spectrometry (PTR-MS): Medical applications, food control, and environmental research. *International Journal of Mass Spectrometry and Ion Processes, 173*(3), 191–241.

[11] Persaud, K., & Dodd, G. (1982). Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose. *Nature, 299*(5881), 352–355.

[12] Gardner, J. W., & Bartlett, P. N. (1994). A brief history of electronic noses. *Sensors and Actuators B: Chemical, 18*(1–3), 210–211.

[13] Pearce, T. C., Schiffman, S. S., Nagle, H. T., & Gardner, J. W. (2003). *Handbook of machine olfaction*. Wiley-VCH.

[14] Barsan, N., Koziej, D., & Weimar, U. (2007). Metal oxide-based gas sensor research: How to? *Sensors and Actuators B: Chemical, 121*(1), 18–35.

[15] Yamazoe, N. (2005). Toward innovations of gas sensor technology. *Sensors and Actuators B: Chemical, 108*(1–2), 2–14.

[16] Korotcenkov, G. (2007). Metal oxides for solid-state gas sensors: What determines our choice? *Materials Science and Engineering: B, 139*(1), 1–23.

[17] Röck, F., Barsan, N., & Weimar, U. (2008). Electronic nose: Current status and future trends. *Chemical Reviews, 108*(2), 705–725.

[18] Rath, R. J., Farajikhah, S., Oveissi, F., Dehghani, F., & Naficy, S. (2023). Chemiresistive sensor arrays for gas/volatile organic compounds monitoring: A review. *Advanced Engineering Materials, 25*(3), 2200830.

[19] Zhang, S., Zhang, H., Yao, H., Wang, P., Zhu, M., Shi, X., & Xu, S. (2024). Recent advances in metal oxide semiconductor heterojunctions for the detection of volatile organic compounds. *Chemosensors, 12*(12), 244.

[20] Zhou, G., Du, B., Zhong, J., Chen, L., Sun, Y., Yue, J., … & He, Y. (2024). Advances in gas detection of pattern recognition algorithms for chemiresistive gas sensor. *Materials, 17*(21), 5190.

[21] Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory, 13*(1), 21–27.

[22] Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

[23] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.

[24] Duda, R. O., & Hart, P. E. (2006). *Pattern classification*. John Wiley & Sons.

[25] Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*. Springer.

[26] Hastie, T. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.

[27] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 1189–1232.

[28] Marco, S., & Gutierrez-Galvez, A. (2012). Signal and data processing for machine olfaction and chemical sensing: A review. *IEEE Sensors Journal, 12*(11), 3189–3214.

[29] Vergara, A., Vembu, S., Ayhan, T., Ryan, M. A., Homer, M. L., & Huerta, R. (2012). Chemical gas sensor drift compensation using classifier ensembles. *Sensors and Actuators B: Chemical, 166*, 320–329.

[30] Fonollosa, J., Rodríguez-Luján, I., Trincavelli, M., Vergara, A., & Huerta, R. (2014). Chemical discrimination in turbulent gas mixtures with MOX sensors validated by gas chromatography-mass spectrometry. *Sensors, 14*(10), 19336–19353.

[31] Fonollosa, J., Sheik, S., Huerta, R., & Marco, S. (2015). Reservoir computing compensates slow response of chemosensor arrays exposed to fast varying gas concentrations in continuous monitoring. *Sensors and Actuators B: Chemical, 215*, 618–629.

[32] Shafer, G., & Vovk, V. (2008). A tutorial on conformal prediction. *Journal of Machine Learning Research, 9*(3).

[33] Romano, Y., Patterson, E., & Candes, E. (2019). Conformalized quantile regression. *Advances in Neural Information Processing Systems, 32*.

[34] Singh, S., S, S., Varma, P., Sreelekha, G., Adak, C., Shukla, R. P., & Kamble, V. B. (2024). Metal oxide-based gas sensor array for VOCs determination in complex mixtures using machine learning. *Microchimica Acta, 191*(4), 196.

[35] De Vito, S., Castaldo, A., Loffredo, F., Massera, E., Polichetti, T., Nasti, I., … & Di Francia, G. (2007). Gas concentration estimation in ternary mixtures with room temperature operating sensor array using tapped delay architectures. *Sensors and Actuators B: Chemical, 124*(2), 309–316.

[36] Pardo, A., Marco, S., & Samitier, J. (2002). Nonlinear inverse dynamic models of gas sensing systems based on chemical sensor arrays for quantitative measurements. *IEEE Transactions on Instrumentation and Measurement, 47*(3), 644–651.