Machine Learning-Based System for Matrix Effect Correction in Gas Chromatography

Jorge Soria Romeo, [1]Paúl Durán Sánchez, [2]Javier Lacasta Miguel

[1]Catalysis and Reactor Engineering Group (CREG), [2]Advanced Information Systems Laboratory (IAAA)
Instituto de Investigación en Ingeniería de Aragón (I3A)
Universidad de Zaragoza, Mariano Esquillor s/n, 50018, Zaragoza, Spain.
Tel. +34-601014177, e-mail: *872016@unizar.es*

Abstract

Gas chromatography of multicomponent gaseous mixtures suffers from the *matrix effect*, causing significant measurement errors in methanation processes. To address this challenge, this study evaluates various predictive models designed to determine component concentrations from chromatographic intensities across three different chromatographs (A, B and C).

Introduction

Accurate measurement of gas stream concentrations is essential for industrial applications, particularly CO_2 methanation into synthetic natural gas using green hydrogen.

Under ideal conditions, the chromatographic intensity of each gas follows a direct proportionality with its concentration, allowing univariate calibration of each component. However, this assumption fails due to the matrix effect: remaining gases in the mixture alter the measured signal of the target gas, requiring multivariate calibration.

In catalytic methanation, these interferences lead to elemental mass-balance errors exceeding 30%, well above the acceptable 5% experimental limit. Moreover, the matrix effect is equipment-specific: the three chromatographs require distinct calibration datasets and independent models.

The objective of this research is to evaluate predictive models for accurate gas concentration estimation, while deploying a graphical user interface to execute the selected models and provide additional utility features.

Methodology

It consists of data filtering, rigorous model training and evaluation, and user application development.

Outlier and Out-of-Distribution Detection

Data cleaning involved two distinct phases. First, Dixon's Q Test was used to detect intra-group anomalies within five-replicate experimental groups. Second, Isolation Forest algorithm was applied to systematically identify anomalous samples. To visually inspect this cleaning process, the identified group anomalies can be projected onto a two-dimensional PCA score plot (see Figure 1).

To handle future anomalous data, an Out-of-Distribution detection pipeline was proposed, combining the Mahalanobis Distance with an Isolation Forest Score.

Model Training and Validation

To prevent evaluation bias and data leakage, model selection and hyperparameter tuning were performed using Nested Cross-Validation with GroupKFold. Model performance was evaluated using Root Mean Squared Error and Mean Absolute Error (MAE).

Web Application and Space-Filling Sampling

A web application has been developed to store trained models and their associated calibration dataset. The tool can be used for: (i) predicting concentrations from new chromatographic intensities, (ii) computing metrics and producing plots, and (iii) executing an OOD analysis pipeline. A key functionality is the space-filling module, which proposes new experimental samples to improve calibration set coverage using the Kennard & Stone algorithm.

Results

Calibration data analysis refuted the assumptions of an ideal system. Cross-correlation matrices between univariate model residuals and gas concentrations (see Figure 2) demonstrated systematic interferences, proving the presence of the matrix effect.

Several predictive models were trained, including multivariable regression, tree-based models, kernel methods and neural networks. The introduction of domain knowledge (forcing the predicted concentration to zero if the measured intensity is zero) improved physical consistency of all models. Ultimately, the robust Huber Regression and Support Vector Regression (SVR) demonstrated best generalization capabilities, obtaining the lowest MAE.

Table 1. Validation MAE (x10^{-3}) of best model per family

	Chromatograph		
Model	A	B	C
Huber Robust Regression	1.00	2.19	1.75
XGBoost	7.86	6.06	4.31
SVR (RBF kernel)	0.92	2.23	1.42
Wide & Deep (Neural netw.)	1.75	2.92	1.92

Once the best models were selected based on MAE, elemental balance errors were computed on independent test experiments as an additional final validation. Figure 3 shows the evolution of oxygen, hydrogen and carbon balance errors over time, where mean errors remain below the 5% target, in an example methanation experiment using a FeNi catalyst.

Conclusions

Machine learning successfully models non-linear matrix effects in chromatography. Contrary to conventional assumptions, simpler models like SVR and robust regression outperform complex neural networks. Their lower structural complexity prevents overfitting under moderate sample sizes, thereby ensuring the physical integrity of elemental balances during experimentation.

Acknowledgements

This research has been partially funded by the Grants and Scholarships Program of the Aragon Institute for Engineering Research (I3A). It emerged as a multidisciplinary collaboration project between the Catalysis and Reactor Engineering Group and the Advanced Information Systems Laboratory.

REFERENCES

[1]. MERCADER, V. D. et al. 2026. Intensifying synthetic natural gas production by functionalization of a NiFe/γ-Al_2O_3 catalyst with alkaline and alkaline-earth materials. *Fuel*. Vol. 406, art. no. 13669. DOI 10.1016/j.fuel.2026.13669.

[2]. INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY. *Matrix effect* [online]. Version 5.0.0. 2025 [accessed: 19 May 2026]. Available: https://doi.org/10.1351/goldbook.M03759

[3]. GOMES, C., CLAEYS-BRUNO, M. and SERGENT, M. 2018. Space-filling designs for mixtures. *Chemometrics and Intelligent Laboratory Systems*, vol. 174, pp. 111-127. DOI 10.1016/j.chemolab.2018.01.013.

[4]. DEAN, R. B. and DIXON, W. J. 1951. Simplified Statistics for Small Numbers of Observations. *Analytical Chemistry*, vol. 23, no. 4, pp. 636-638. DOI 10.1021/ac60052a025.

FIGURES

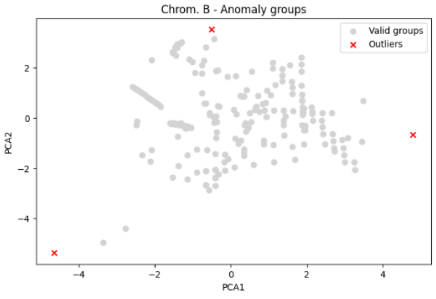

Figure 1. Outliers' detection using Isolation Forest in Chrom. B

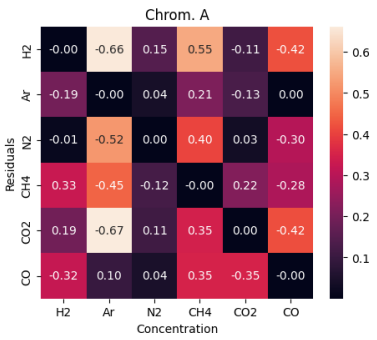

Figure 2. Cross-correlation matrix between univariate model residuals and gas concentrations in Chrom. A

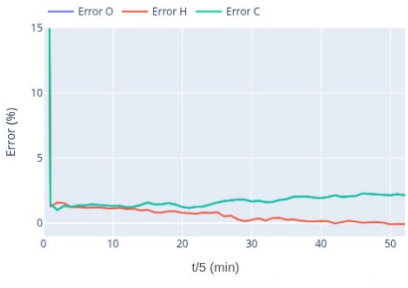

Figure 3. Example of elemental balance error evolution over time for a methanation of CO_2 experiment using FeNi catalyst