ACS CHEMICAL HEALTH & SAFETY

Open Access

pubs.acs.org/acschas

Article

Solvent Emissions Prediction from Laboratory Fume Hoods: A Preliminary Study

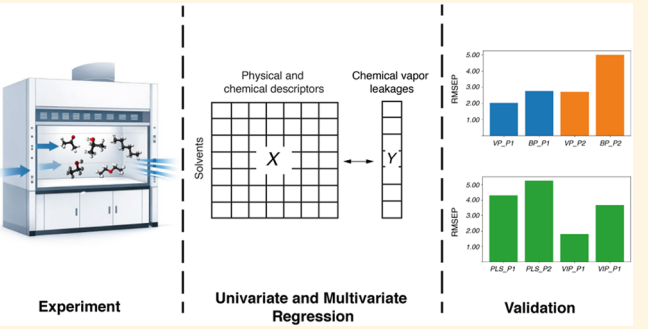
Raffaele Emanuele Russo, Giacomo Seccacini, Paolo Cognigni, Martina Fattobene, Silvia Zamponi, Paolo Conti, Marco Ortelli, and Mario Berrettoni*

Cite This: *ACS Chem. Health Saf.* 2026, 33, 138–146

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: The correct use and performance assessment of chemical fume hoods are essential to minimize operator exposure to hazardous volatile compounds. In this study, a chemometric strategy was developed to predict solvent emissions based on molecular and physicochemical descriptors. A small but representative set of solvents (ethanol, diethyl ether, isopropanol, *n*-hexane, and acetone) was selected according to safety profiles and structural diversity. Controlled evaporation tests were carried out using a dynamic containment setup inspired by the EN 14175 outer plane methodology. Emissions were experimentally quantified under two standardized conditions. Initially, univariate regression models were developed using Ordinary Least Squares, with vapor pressure and boiling point as single predictors. Partial Least Squares regression was therefore applied to check ten chemical−physical molecular descriptors; variable selection reduced the number to the previous two. The reduced model improves prediction accuracy under both experimental conditions, reducing the error in prediction from 4.30 to 1.48 for the first setting and from 5.25 to 3.67 for the second setting. These results demonstrate that solvent emissions from laboratory fume hoods can be reliably predicted using chemometric models based on molecular descriptors. The proposed approach offers a valuable tool for emission assessment, informed solvent selection, and chemical risk management in laboratory environments.



KEYWORDS: *emission assessment, chemometric strategy, laboratory fume hood, molecular descriptors, regression models*

1. INTRODUCTION

The exposure to volatile chemicals and nanoparticles poses a critical health risk to operators in both academic and industrial environments. Acute or chronic inhalation of solvents and aerosols can cause short and long-term toxic effects, damaging organs such as lungs, liver, kidneys or leading to neurological disorders and even cancer.[1,2] To mitigate such exposures, the laboratory chemical fume hood represents one of the most widespread and relevant collective protective tools. It acts as a dynamic barrier system, containing and evacuating hazardous substances through controlled aspiration flow. Nowadays, different types of fume hoods have been developed to accommodate different usage needs and laboratory environments. These include Constant Air Volume (CAV) hoods,[3,4] variable Air Volume (VAV) hoods,[5] air Curtain (AC) hoods,[5−7] and range hoods (RH),[6,8,9] each with specific design features and airflow control strategies. However, fume hoods prevent operators from breathing toxic chemicals only if they are used in the correct way. A good protection from dangerous vapors only occurs if the hood is placed in a part of the laboratory where the overall air flows (interference from the outside environment, eddy currents, wind speed and conditions of air supply and return) do not disturb the aspiration of the system of the fume hood and if the operators are aware of the proper everyday use of the hood; other factors affecting the correct operation of the hood are its geometry and the presence of many containers of chemicals and large equipment which can disturb the airflow inside the hood.[10,11] A proper maintenance and yearly testing of the exhaust capability of the fume hood is a good way to ensure safety for every person inside a laboratory. The correct operation of the fume hood is assessed through standard measurements established and described by national and international organizations, such as the ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers) 110, the DIN (Deutsches Institut für Normung) 12924 and the EN (European Norm) 14175[12]:

Received: September 26, 2025
Revised: December 22, 2025
Accepted: January 7, 2026
Published: January 15, 2026

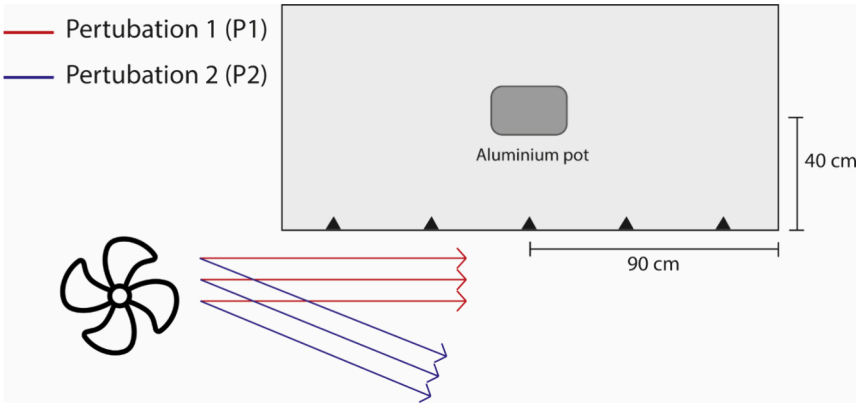


Figure 1. Schematic representation of the experimental setup used to induce airflow perturbations during solvent emission measurements. The aluminum pot containing the solvent was centrally placed inside the fume hood. The black filled triangles indicate the sampling nozzles used to collect vapor at multiple points along the front plane of the hood.

- face velocity test: it consists in measuring the velocity of air flowing perpendicular to the hood front, with different sash heights, in order to assess the air flow uniformity across the entire front of the hood;[13]
- smoke test, allowing the visualization of the air flows inside the hood;
- containment test: it is a quantitative measure of the gas escaping from the hood, providing useful information about its extraction system. The measure is done by releasing a tracer gas inside the hood, commonly sulfur hexafluoride, SF_6, since it is odorless, colorless, nontoxic, nonflammable, noncorrosive and nearly absent in the environment.[13,14]

Outer plane test refers to a containment test performed outside the plane of the fume hood's sash, simulating user interaction. Through a well-defined measurement system, this dynamic test determines the hood's ability to contain chemical vapors, enabling the estimation of volatile substances that escape from the hood front and inhaled by the operator in front of it.

In recent years, several key factors have been identified as influencing fume hood containment.[15] These include geometric and operational parameters such as sash height, face velocity, the presence of internal vortices, source location, and the positioning of experimental equipment, all of which significantly affect containment.[3,12,16−19] Operator-related factors, including body movement and heat emission, have also been shown to disturb airflow and promote leakage of hazardous substances.[18−20] Furthermore, room-level ventilation and external air drafts can significantly affect hood performance and contribute to loss of containment.[9,21]

While the reported previous studies focus on how external factors and user behavior affect containment, the present work aims to shift the perspective toward the intrinsic properties of the chemicals themselves. Thus, integrating controlled emission data is possible to propose a predictive approach based on molecular and physicochemical descriptors. Moreover, the study was intentionally designed to simulate perturbed conditions, such as the presence of cross drafts, which are well-known to compromise fume hood containment. These conditions represent realistic worst-case scenarios that may frequently occur in laboratory settings.

2. MATERIALS AND METHODS

2.1. Experimental Procedure

To simulate challenging operating conditions, controlled airflow disturbances were introduced using an external fan. The experimental setup was designed to reproduce worst-case scenarios for operator exposure, maintaining a fixed face velocity to reflect standardized testing conditions. The investigation was done by evaluating the outflow of five different solvents (ethanol, acetone, isopropanol, hexane and diethyl ether) with respect to ten molecular and physical-chemical descriptors to identify potential correlations that could be used to develop a predictive model. All the experiments were conducted on a VAV fume hood, with internal floor dimensions of 180 cm wide × 80 cm deep. All reagents used in this study were purchased from Carlo Erba and used as received. The solvents were placed in the internal floor of the fume hood, in an aluminum pot with a surface of 103.5 cm^2 and 1 cm high, for a total solvent volume of about 100 mL. The aluminum pot was placed on the hood plane at 40 cm from the front edge, and the vapor spill was sampled outside the fume hood by placing the nozzles at a distance of 5 cm from the front. (Figure 1).

The chemical vapors leakages were measured in three different conditions with the opened sash:

1. the first condition was without any external perturbation modifying the normal air flow inside the hood (P0);
2. for the second condition, a strong perturbation was created in order to disturb the hood's internal air flow and force the solvent vapors to leak out (P1);
3. for the third condition, a softer perturbation was created, but still strong enough to cause some vapors to leak out from the hood (P2).

The perturbations were induced by placing a fan next to the fume hood; for the strong perturbation, the fan air flow was directed in parallel with respect to the sash, while, for the milder one, the fan air flow was directed externally with respect to the sash plane. The scheme describing the different perturbations is reported in Figure 1.

The P0 test in the absence of perturbation represents the ideal condition, in which no solvent leakage were recorded (<LOD). The perturbations, on the other hand, were introduced during the measurements with the aim of forcing the chemical vapors to escape from the hood, in order to simulate the worst conditions. In this scenario, an operator positioned in front of the hood could be more exposed to hazardous substances, thus the disturbances were intended to simulate the maximum possible vapor leakage, establishing a reference for the worst-case scenario. In this way, these data are critical for future development of predictive models aimed at estimating solvent emissions under the most unfavorable operating conditions. Before conducting the emission experiments, the fume hood was standardized and characterized according to the EN 14175-3:2019 standard. The airflow rate was maintained at 609 m^3/h in both test conditions. Under

https://doi.org/10.1021/acs.chas.5c00179
ACS Chem. Health Saf. 2026, 33, 138−146

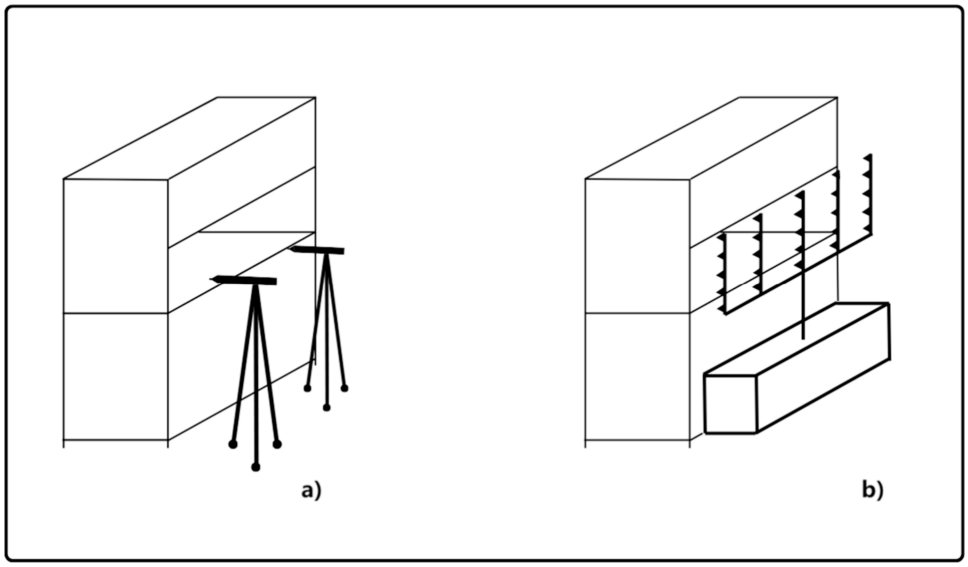


Figure 2. Schematic representation of the two sampling systems used: (a) the simplest one, the leak is determined precisely using a Tiger placed on a stand; (b) a grid with 25 nozzles and a uniform aspiration system (EN 14175).

the strong perturbation scenario (P1), the average face velocity was 0.24 m/s, while the room air velocity measured was 0.85 m/s. In the moderate perturbation condition (P2), the face velocity remained constant at 0.24 m/s, but the room air velocity was 0.13 m/s.

The measurements were initially taken using spot sampling, placing two different stands with Tiger^{LT} portable gas detectors (ION Science) directly in front of the hood at the height of the workbench and equidistant from each other (Figure 2a). However, the results obtained are inconsistent due to the fluid dynamics of the aspiration front. In other words, the use of the Tiger gas detector provided only localized, point-based measurements that did not adequately reflect the real behavior of solvent vapors or the overall containment performance of the fume hood. For this reason, a more comprehensive and standardized approach was adopted, in accordance with EN 14175. For this reason, solvent leakages were measured by placing in front of the hood's sash a multipoint sampling grid, made by 25 sampling nozzles drawing in the vapors generated, covering the entire frontal plane of the fume hood (Figure 2b).

The sampling system was connected to a Miran 1B2 Portable Ambient Air Analyzer Infrared Spectrophotometer for the quantitative evaluation of vapor leakages from the fume hood frontal plane, set with an inlet flow of 4.5 L/min. To avoid a purely theoretical assessment of solvent evaporation, which is strongly influenced by fluid-dynamic parameters, the vapors suctioned at the hood outlet were conveyed into a collection system with a known flow rate and connected to the Tiger instrument to ensure constant flow for quantitative measurements. The experimental data set was constructed by measuring the emissions of the selected solvents under controlled and standardized conditions inside a chemical fume hood like described before. Ethanol, diethyl ether, *n*-hexane, and isopropyl alcohol were used to construct the training set, while acetone was employed as an external validation compound to test the predictive performance of the model.

Ten chemical-physical parameters are used to model in the system, described below:

1. Vapor pressure (VP), expressed in mmHg, quantifies the tendency of a compound to evaporate at a given temperature and is directly related to its volatility.
2. The topological polar surface area of the molecule (TPSA), measured in Å^2, accounts for the surface area associated with polar atoms, typically oxygen and nitrogen, and provides information about the compound's capacity to engage in hydrogen bonding and its interaction with the surrounding environment.
3. The number of heavy atoms, count of all non-hydrogen atoms, reflects the molecular size and overall complexity of the structure,
4. The number of bonds provides an estimate of the internal bonds.
5. The number of hydrogen bond donors indicates the number of hydrogen atoms attached to electronegative atoms capable of forming hydrogen bonds, a feature that can affect the intermolecular interactions and the extent to which a substance tends to remain in the liquid phase.
6. The Parachor (Paracoro), as defined by Sugden, is a physicochemical descriptor that combines molar volume with a function of the compound's surface tension.
7. The boiling point temperature (BP), expressed in degrees Celsius, is the temperature at which a compound changes from a liquid to a gaseous state at standard pressure, and is inversely proportional to its propensity to evaporate under ambient conditions.
8. Molecular weight, expressed in g/mol, is a fundamental structural property that can influence vaporization dynamics.
9. Vapor density, relative to air (air = 1), provides insight into the behavior of the vapor phase once released, especially in terms of dispersion and potential accumulation in the laboratory environment.
10. Liquid density, expressed in g/mL, describes the mass-to-volume ratio of the substance in its liquid state and may influence the rate of evaporation depending on the experimental setup.

The above parameters were the predictors of Partial Least Squares (PLS) regression models used to model the emission measurements (P1, P2). The selected descriptors were chosen based on established scientific principles related to solvent volatility.[22−25] Each variable, such as vapor pressure, boiling point, molecular weight, and hydrogen bonding capacity, has documented relevance in influencing evaporation rates, intermolecular interactions, and the overall tendency of a compound to transition into the gas phase under ambient conditions.[26−29]

2.2. Chemometric Approach

To investigate the relationship between the molecular properties of solvents and their emission behavior from laboratory fume hoods, both univariate and multivariate regression approaches were explored. A set of molecular and physicochemical descriptors was used as predictors to assess whether statistically significant relationships could be established with the measured emissions under controlled experimental conditions. Although the data set was limited in size, the study aimed to evaluate the

https://doi.org/10.1021/acs.chas.5c00179
ACS Chem. Health Saf. 2026, 33, 138−146

Table 1. Experimental Data Matrix[a]

solvent	vapor pressure (mmHg)	topological polar surface area (Å^2)	heavy atom count	bond count	hydrogen bond donor count	paracoro	boiling point (°C)	molecular weight (g/mol)	vapor density (Air = 1)	density (g/cm^3)	P0 (ppm)	P1 (ppm)	P2 (ppm)
ethanol	59.3	20.2	3	7	1	126.26	78.2	46.07	1.59	0.79	<LOD (2.00 ppm)	8.66	8.22
ethanol	59.3	20.2	3	7	1	126.26	78.2	46.07	1.59	0.79	<LOD (2.00 ppm)	12.49	6.35
ethanol	59.3	20.2	3	7	1	126.26	78.2	46.07	1.59	0.79	<LOD (2.00 ppm)	11.75	7.29
diethyl ether	538	9.2	5	14	0	212.16	34.6	74.12	2.55	0.71	<LOD (2.00 ppm)	33.64	25.81
diethyl ether	538	9.2	5	14	0	212.16	34.6	74.12	2.55	0.71	<LOD (2.00 ppm)	35.63	27.77
diethyl ether	538	9.2	5	14	0	212.16	34.6	74.12	2.55	0.71	<LOD (2.00 ppm)	32.83	27.91
n-hexane	153	0.0	6	19	0	268.36	68.7	86.12	2.97	0.66	<LOD (1.00 ppm)	16.14	11.59
n-hexane	153	0.0	6	19	0	268.36	68.7	86.12	2.97	0.66	<LOD (1.00 ppm)	15.28	12.35
n-hexane	153	0.0	6	19	0	268.36	68.7	86.12	2.97	0.66	<LOD (1.00 ppm)	16.70	10.63
isopropyl alcohol	45.4	20.2	4	11	1	162.72	82.3	60.10	2.07	0.79	<LOD (1.00 ppm)	6.71	3.65
isopropyl alcohol	45.4	20.2	4	11	1	162.72	82.3	60.10	2.07	0.79	<LOD (1.00 ppm)	6.34	3.85
isopropyl alcohol	45.4	20.2	4	11	1	162.72	82.3	60.10	2.07	0.79	<LOD (1.00 ppm)	7.30	3.43
acetone	231	17.1	4	9	0	162.27	56.0	58.10	2.01	0.79	<LOD (4.00 ppm)	21.29	15.36
acetone	231	17.1	4	9	0	162.27	56.0	58.10	2.01	0.79	<LOD (4.00 ppm)	18.09	10.30
acetone	231	17.1	4	9	0	162.27	56.0	58.10	2.01	0.79	<LOD (4.00 ppm)	20.80	–

[a]15 samples corresponding to 5 solvents, each described by 10 molecular and physicochemical descriptors.

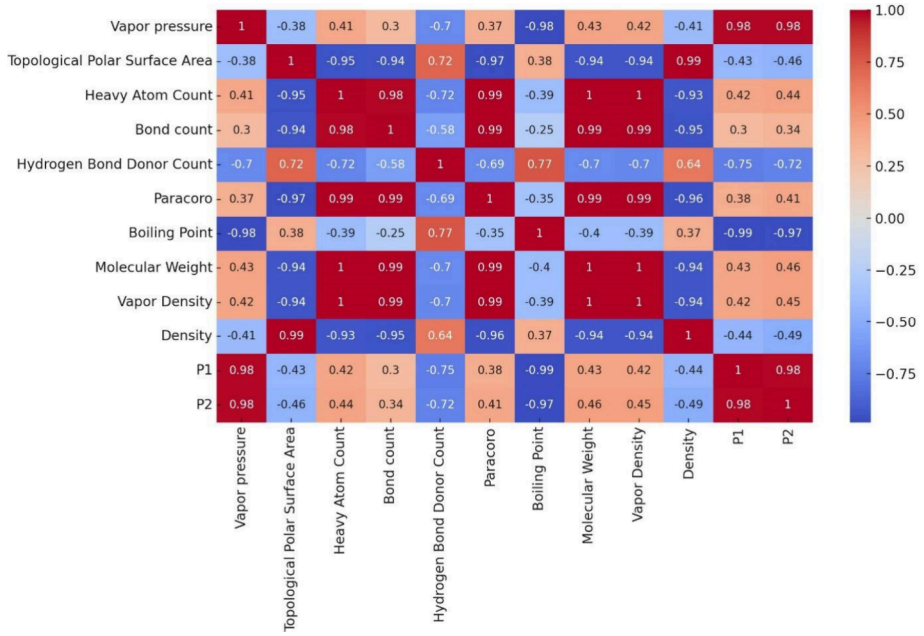


Figure 3. Correlation matrix of the descriptors and responses.

predictive power of individual descriptors, selected based on their correlation with the response variables, as well as to assess the added value of combining multiple variables within a chemometric modeling framework.

Univariate models were developed using the classical Ordinary Least Squares (OLS) method, which allowed us to quantify the predictive contribution of the most strongly correlated descriptors. Multivariate models, by contrast, were built using Partial Least Squares (PLS) regression,[30−32] Prior to model construction, all predictor variables were autoscaled to ensure comparability, normalize their contributions, and mitigate differences in scale and measurement units. PLS is especially advantageous when the number of predictors exceeds the number of observations and when multicollinearity is present, as it projects both the predictor (X) and response (Y) spaces into a latent

https://doi.org/10.1021/acs.chas.5c00179
ACS Chem. Health Saf. 2026, 33, 138−146

Table 2. Summary of Univariate Regression Models (OLS)[a]

model	regression equation	R^2	factor significance (desired p-value < 0.05)	lack of Fit (desired p-value > 0.05)	RMSEP
P1	P1 = 6.73 + 0.05 × Vapor Pressure	96.91%	<0.0001	0.014	2.03
	P1 = 53.38 − 0.55 × Boiling Point	98.40%	<0.0001	0.197	2.77
P2	P2 = 3.58 + 0.04 × Vapor Pressure	97.40%	<0.0001	0.003	2.72
	P2 = 43.84 − 0.48 × Boiling Point	98.66%	<0.0001	0.090	5.01

[a]Factor significance refers to the significance of the regression coefficient (slope of the predictor variable x) as assessed by analysis of variance (ANOVA).

Table 3. External Prediction Results for Acetone (95% CI)

response	predictor	experimental values	predicted value	95% confidence interval
P1	vapor pressure	21.29, 18.09, 20.80	18.60	[17.29, 19.91]
	boiling point	21.29, 18.09, 20.80	22.45	[21.40, 23.50]
P2	vapor pressure	15.36, 10.30	13.83	[12.79, 14.86]
	boiling point	15.36, 10.30	17.15	[16.38, 17.91]

structure that maximizes their covariance. The PLS algorithm is based on the following decomposition:

$$X = TP^T + E, \quad Y = UC^T + G \quad U = TD \tag{1}$$

It performs the orthogonal decomposition of the two matrices X (predictors) and Y (responses) maximizing the correlation between the scores from the two blocks through an inner relation. T and U are the score matrices for the predictors and the responses, respectively; they represent the projections of the original data onto the latent variable spaces. P and C are the corresponding loading matrices, that is, the weights indicating the contribution of each original variable to the latent components. E and G are the residual matrices, capturing the portion of variance in X and Y not explained by the model.

VIP (Variable Importance in Projection) coefficients are used to select important variables. VIP coefficients quantify the importance of each predictor in explaining the variances of both the independent and response variables; thus, the variables with the largest VIP value (close to or >1) are retained because the most important ones are discarded.[30]

To avoid the "sample replicate trap" and to reduce the risk of model overfitting, the optimal number of latent variables was determined using a continuous block cross-validation strategy. In this approach, submodels were built by excluding contiguous blocks of m samples at a time, where m corresponds to the number of experimental replicates per compound. Thus, the number of latent variables was selected by minimizing the Root Mean Square Error of Cross-Validation (RMSECV).[33] Moreover, the absence of overfitting has been evaluated by doing three permutation tests (the Wilcoxon signed-rank test, the Sign test, and the randomization t test) using the residuals in cross-validation, because they are more sensitive to detect overfitting. The test involved 1000 permutations, during which the values of the response variables (Y) were randomly shuffled through the experiments while keeping the predictor matrix (X) unchanged.[34,35]

3. RESULTS AND DISCUSSION

The full data set (X and Y data) used for model development is presented in Table 1. The P0 condition, corresponding to standard fume hood operation without external perturbations, was excluded from the analysis. For some solvents, no measurable emissions were detected, while for others, a constant low emission value was observed. In both cases, the limited variability and the low concentration levels (often below 1 ppm) reflect the expected behavior of an effectively operating fume hood and would not contribute meaningfully to regression modeling.

A Pearson correlation matrix was computed to explore the linear relationships between the ten physicochemical molecular descriptors both with each other and with the two response variables, P1 and P2 (Figure 3). This preliminary analysis aimed to identify which individual descriptors were most strongly associated with solvent emissions and to evaluate the presence of multicollinearity. The analysis revealed that vapor pressure was strongly and positively correlated with both response variables (r = 0.98 for both P1 and P2), while boiling point exhibited a strong negative correlation (r = −0.99 for P1 and −0.98 for P2). These results are consistent with the physicochemical interpretation of these variables: solvents with higher vapor pressure or lower boiling points tend to evaporate more readily and therefore generate higher emission levels. These two descriptors emerged as the most promising candidates for univariate regression modeling.

Moreover, the descriptors themselves were found to be strongly intercorrelated, which supports the use of multivariate techniques such as PLS. Overall, this finding justifies the development of both univariate and multivariate models, enabling a direct comparison of their predictive performance.

3.1. Univariate Regression

Following the correlation analysis, which highlighted vapor pressure and boiling point as the most significantly correlated descriptors with the emission responses, univariate regression models were developed using these variables individually as predictors.

Each model demonstrated a very high degree of explained variance (R^2 > 96%) and highly significant regression coefficient (p < 0.0001), confirming the presence of a strong linear relationship between the selected descriptors and solvent emission behavior.

To ensure the reliability of the OLS models, standard diagnostic checks were performed. Residuals from all four models were evaluated and found to follow a normal distribution, as confirmed by the Shapiro−Wilk and Anderson−Darling tests (p > 0.05 in all cases). Visual inspection of the residual plots further supported this conclusion, showing random scatter around zero without evident patterns or trends, thus validating the assumptions of normality and independence. The assumption of homoscedasticity was also assessed through formal statistical tests for equality of variances, specifically Levene's and Bartlett's tests. In all four models, the resulting p-values were >0.05, indicating no significant differences in variance across the groups defined by the predictor values.

However, the lack-of-fit analysis revealed notable differences in model adequacy: while the models based on boiling point showed no significant lack-of-fit (p > 0.05), those based on vapor pressure exhibited statistically significant lack-of-fit (p < 0.05).

https://doi.org/10.1021/acs.chas.5c00179
ACS Chem. Health Saf. 2026, 33, 138−146

Table 4. PLS Model Performance: Comparison between Full and VIP-Selected Models for Responses P1 and P2

model ID	LVs	*X*-block variance (%)	*Y*-block variance (%)	RMSEC	RMSECV	RMSEP
P1 – full model	2	98.39	98.61	1.23	7.55	4.30
P2 – full model	2	98.39	99.05	0.88	6.50	5.25
P1 – VIP-selected	1	99.87	97.78	1.56	4.52	1.48
P2 – VIP-selected	1	99.87	98.25	1.19	3.63	3.67

Table 5. Predicted vs. Experimental Values for the External Validation Compound (Acetone)[a]

model	replicate	measured value	predicted (full)	residual (full)	predicted (VIP)	residual (VIP)
P1	1	21.29	22.73	1.45	20.53	−0.77
	2	18.09	22.73	4.64	20.53	2.44
	3	20.80	22.73	1.93	20.53	−0.27
P2	1	15.36	17.44	2.09	15.49	0.13
	2	10.30	17.44	7.13	15.49	5.19

[a]Residuals and 95% confidence intervals of the measured values are included for both full and VIP-selected models.

This suggests that the relationship between boiling point and emission is more stably linear within the studied range, whereas vapor pressure may introduce deviations from linearity under the same conditions. The key performance indicators of the univariate models are summarized in Table 2

To further assess the predictive performance of the univariate models, acetone was used as an external validation compound like described in section 2.1. The predicted values obtained from both vapor pressure and boiling point models were compared with the available experimental replicates of P1 ($n = 3$) and P2 ($n = 2$), as shown in Table 3.

For both responses, the models provided predictions that lie within the experimental range and fall inside the 95% confidence intervals derived from the replicates.

In summary, while univariate models based on vapor pressure demonstrated strong correlations and good predictive capability, the presence of significant lack-of-fit ($p < 0.05$ in both cases) suggests that the relationship is not fully captured by a linear model. This indicates potential nonlinearity or the influence of additional unaccounted factors. This outcome may also be influenced by the limited number of replicates ($n = 3$), which affects the degrees of freedom and consequently the sensitivity of the lack-of-fit test. A larger number of experimental replicates would likely provide more accurate estimates of variability and could reduce the likelihood of detecting statistically significant lack-of-fit due to random variation alone. This limitation has been acknowledged, and future work will aim to include expanded data sets to improve the robustness of model evaluation. On the other hand, univariate models based on boiling point showed statistically well-behaved diagnostics and produced predicted values that fell within the 95% confidence intervals of the experimental measurements.

To further explore the potential for improved predictive performance and to capture additional complexity in the emission behavior of solvents, a multivariate modeling strategy was adopted. While univariate models, particularly those based on boiling point, demonstrated reliable predictions and good statistical performance, the strong correlation between vapor pressure and boiling point (see Figure 3) raised the question of whether their combined use in a multivariate framework could enhance model robustness and generalization. PLS regression was thus employed to evaluate the synergic contribution of these and other descriptors, while effectively addressing collinearity and minimizing the risk of overfitting.

3.2. Multivariate Regression

Given the limitations observed in the univariate approach it became evident that a more comprehensive modeling strategy was required. Solvent emission behavior can be likely influenced by a combination of molecular and physicochemical properties that cannot be adequately captured by a single descriptor. To this end, a multivariate regression approach was adopted to integrate multiple descriptors simultaneously and better account for the system's complexity.

The goal of the multivariate analysis is

> to determine whether incorporating additional descriptors could improve the predictive performance compared to the univariate models, and
>
> to explore the contribution of each variable to the overall model through latent structures.

The PLS models were constructed using the full set of molecular descriptors, and variable selection was later applied using VIP scores. Vapor pressure and boiling point were retained in the simplified (VIP-selected) models as they contributed most significantly to the prediction of solvent emissions. Their selection is chemically meaningful, as both vapor pressure and boiling point are directly related to the volatility and phase transition behavior of organic compounds, which are key factors in emission phenomena under fume hood conditions.

Table 4 summarizes the key performance indicators for all models, including the number of latent variables (LVs), the cumulative variance explained in the *X* and *Y* blocks, and the root-mean-square errors for calibration (RMSEC), cross-validation (RMSECV), and external prediction (RMSEP).

Overall, all models explained a high proportion of variance in both the *X*-block and *Y*-block.

The full model for response P1 achieved good calibration statistics, with an RMSEC of 1.23 and a R^2_cal of 0.99 and R^2_CV of 0.85. A similar pattern was observed for the full P2 model, where RMSEC was as low as 0.88 with an R^2_cal and R^2_CV equal to 0.99 and 0.85 respectively. However, their cross-validation and prediction errors (RMSECV and RMSEP) were comparatively higher.

In contrast, the VIP-selected models, which retained only vapor pressure and boiling point as predictors, achieved better performance in cross validation and external prediction. For instance, the P1VIP model reduced RMSEP from 4.30 to 1.48, RMSECV from 7.55 to 4.52, while maintaining a high level of explained variance. Moreover, the model reaches a significantly

https://doi.org/10.1021/acs.chas.5c00179
ACS Chem. Health Saf. 2026, 33, 138−146

higher R^2_cal and R^2_CV value of 0.98 and 0.95, respectively. Similarly, the P2VIP model showed a higher R^2_cal of 0.98 and R^2_CV of 0.96 and an RMSEP of 3.67 versus 5.25 in the full model. These results demonstrate a clear benefit in terms of predictive generalization when dimensionality is reduced by excluding less informative variables. To evaluate the predictive performance of the developed PLS models, the experimental and predicted values for the validation set were compared, as reported in Table 5. For each compound, the predicted values from both the full model and the VIP-selected model were examined, and the corresponding residuals (defined as the difference between predicted and measured values) were calculated. Additionally, the 95% confidence intervals (CIs) for the measured values were calculated based on replicate standard deviations and appropriate *t*-distribution values.

For P1, the mean of the measured values was 20.06 with a standard deviation of 1.72, yielding a 95% CI of [15.78 − 24.33]. The prediction obtained from the VIP model (20.53) fell well within this interval and showed smaller residuals across all replicates (−0.76, 2.44, −0.27) that is a better performance than the full model. Similarly, for P2, the average measured value was 12.83 with a standard deviation of 3.58 and wider confidence interval due to higher variability between the two replicates. The prediction from the VIP-selected model (15.49) was closer to the observed data (residues 0.13 and 5.19) compared to the full model (residues 2.08 and 7.13).

Eqs 2 and 3 describes the PLS regression models obtained for the two experimental conditions (P1 and P2), Each model is reported in its original units:

$$P1 = 30.07 + 0.03 \times VP - 0.28 \times BP \quad (2)$$

$$P2 = 23.72 + 0.02 \times VP - 0.24 \times BP \quad (3)$$

The models display different intercepts while retaining highly similar linear coefficients. This difference in intercepts could not reflect any change in the underlying physicochemical mechanism but instead arises from the distinct mean emission levels associated with the two airflow perturbations. Specifically, the strong perturbation applied in P1 leads to an upward shift in the baseline emission level, whereas the milder perturbation in P2 produces lower average emissions. Thus, the intercept term probably accounts for the aerodynamic influence imposed by the perturbation intensity. These results demonstrate that even a simple two-variable PLS model can effectively describe emission behavior under perturbed containment conditions, offering a robust and interpretable alternative to more complex modeling approaches. Table 6 shows the results of the permutation test, previously described, applied to every model. In the full models, the self-prediction residuals showed significance across all three statistical tests, with *p*-values well below 0.05, that is the model well predict the data. However, for the cross-validated residuals, more sensitive to overfitting, the full models did not reach statistical significance, with *p*-values slightly exceeding the 0.05 threshold, indicating a possible overfitting of the model. In contrast, the VIP-selected models passed all permutation tests for both self-prediction and cross-validation residuals indicating that in this case, both models well predict.

Therefore, once the PLS models were developed, they were compared with the univariate OLS models to evaluate their relative performance. RMSEP is typically used to compare models built on the same data set (Tables 2 and 4), with the model yielding the lowest RMSEP generally considered the most accurate. However, in chemometric modeling a statistically rigorous comparison based on the evaluation of prediction bias and confidence intervals on SEP (Standard Error of Prediction) ratios is generally recommended to ensure reliable conclusions.[36] However, in this study it would be unreasonable to apply such statistical procedures as the number of test samples is insufficient to support a meaningful statistical comparison.

Table 6. Permutation Test Results: *p*-Values from Three Statistical Tests (Wilcoxon, Sign Test, and Randomized T-test) for Both Self-Prediction and Cross-Validation Residuals[d]

case	model ID	LVs	test 1[a]	test 2[b]	test 3[c]
self-prediction	P1 − full	2	0.002	0.010	0.007
cross-validated	P1 − full	2	0.075	0.129	0.121
self-prediction	P2 − full	2	0.002	0.007	0.006
cross-validated	P2 − full	2	0.070	0.111	0.123
self-prediction	P1 − VIP	1	0.003	0.016	0.007
cross-validated	P1 − VIP	1	0.005	0.022	0.010
self-prediction	P2 − VIP	1	0.002	0.010	0.007
cross-validated	P2 − VIP	1	0.004	0.015	0.009

[a]Test 1: Pairwise Wilcoxon signed rank test. [b]Test 2: Pairwise sign test. [c]Test 3: Randomization *t* test. [d]$p < 0.05$ indicates statistical significance.

Nevertheless, some important considerations can be drawn from the available data. Among the three regression models tested, the univariate OLS model based on vapor pressure can be treated with caution as it showed a significant lack-of-fit. This suggests that a single descriptor is not sufficient to capture the complexity of solvent evaporation behavior, and therefore a multivariate approach such as PLS should be preferred. The OLS model based on boiling point, on the other hand, exhibited better statistical diagnostics, with no significant lack-of-fit and predictions consistent with the experimental confidence intervals. However, the PLS models show lower RMSEP values, suggesting a better overall fit and higher predictive accuracy. Both univariate and PLS methods seem to give better models when measurements are performed under strong perturbation (P1) while in case of mild perturbation (P2) we are able to model the leakage with a lower precision.

4. CONCLUSIONS

This study explored the use of molecular descriptors to predict solvent emissions from laboratory fume hoods, applying both univariate (OLS) and multivariate (PLS) regression techniques. While vapor pressure and boiling point individually showed strong linear correlations with emission responses, the resulting univariate models revealed distinct limitations. These issues highlighted the need to explore multivariate modeling strategies capable of capturing more complex interactions among molecular descriptors. PLS models integrating multiple descriptors, especially after VIP-based variable selection, provided more accurate and robust predictions. Notably, although the scaled PLS models for the two perturbation conditions (P1 and P2) shared the same structure, the intercepts of the unscaled models differed significantly. This shift could reflect the influence of the containment conditions on baseline emission levels, and confirms that the same physicochemical properties can manifest differently depending on airflow dynamics. Nonetheless the data set used in this study was limited to five solvents by current university safety regulations, future efforts should focus on expanding the solvent data set to improve model robustness. The ability to predict fume hood leaks based on chemical-physical descriptors depends on the

https://doi.org/10.1021/acs.chas.5c00179
ACS Chem. Health Saf. 2026, 33, 138−146

measurement conditions applied but in all the cases investigated two descriptors, boiling point and vapor pressure, are sufficient to obtain predictive models. Overall, this study provides a solid starting point for the development of data-driven tools to predict solvent hood leakage in laboratory environments. The proposed approach has shown promising results, demonstrating that chemometric modeling based on molecular descriptors can effectively support exposure assessment. With further implementation, refinement and validation, this methodology could become a valuable asset for chemical risk management, informed solvent selection, and the promotion of safer and more sustainable laboratory practices.

AUTHOR INFORMATION

Corresponding Author

Mario Berrettoni − *School of Science and Technology, Chemistry Division, University of Camerino, Camerino, MC 62032, Italy;* orcid.org/0000-0002-2273-035X; Email: mario.berrettoni@unicam.it

Authors

Raffaele Emanuele Russo − *School of Science and Technology, Chemistry Division, University of Camerino, Camerino, MC 62032, Italy;* orcid.org/0000-0001-6362-6201

Giacomo Seccacini − *School of Science and Technology, Chemistry Division, University of Camerino, Camerino, MC 62032, Italy*

Paolo Cognigni − *School of Science and Technology, Chemistry Division, University of Camerino, Camerino, MC 62032, Italy*

Martina Fattobene − *School of Science and Technology, Chemistry Division, University of Camerino, Camerino, MC 62032, Italy*

Silvia Zamponi − *School of Science and Technology, Chemistry Division, University of Camerino, Camerino, MC 62032, Italy*

Paolo Conti − *School of Science and Technology, Chemistry Division, University of Camerino, Camerino, MC 62032, Italy;* orcid.org/0000-0002-5469-5173

Marco Ortelli − *Stargate Innovation Srl, Limbiate, MB 20812, Italy*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.chas.5c00179

Author Contributions

CRediT: **Raffaele Emanuele Russo** data curation, formal analysis, software, validation, visualization, writing - original draft; **Giacomo Seccacini** data curation, investigation, methodology, writing - original draft; **Paolo Cognigni** data curation, investigation, methodology, validation, visualization, writing - original draft; **Martina Fattobene** supervision, validation, writing - review & editing; **Silvia Zamponi** project administration, supervision, writing - review & editing; **Paolo Conti** data curation, formal analysis, software, writing - review & editing; **Marco Ortelli** conceptualization, resources, supervision, writing - review & editing; **Mario Berrettoni** conceptualization, funding acquisition, supervision, validation, writing - review & editing.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We sincerely thank Marco Ortelli from Stargate Innovation Srl for providing us with the instruments for sampling and analysis, and for partially supporting the funding of this study.

REFERENCES

(1) Vlaanderen, J.; Straif, K.; Pukkala, E.; Kauppinen, T.; Kyyrönen, P.; Martinsen, J. I.; Kjaerheim, K.; Tryggvadottir, L.; Hansen, J.; Sparén, P.; Weiderpass, E. Occupational Exposure to Trichloroethylene and Perchloroethylene and the Risk of Lymphoma, Liver, and Kidney Cancer in Four Nordic Countries. *Occup Environ. Med.* **2013**, *70* (6), 393−401.

(2) Davardoost, F.; Kahforoushan, D. Health Risk Assessment of VOC Emissions in Laboratory Rooms via a Modeling Approach. *Environmental Science and Pollution Research* **2018**, *25* (18), 17890−17900.

(3) Tseng, L.-C.; Huang, R. F.; Chen, C.-C. Significance of Face Velocity Fluctuation in Relation to Laboratory Fume Hood Performance. *Ind. Health* **2010**, *48*, 43−51.

(4) Tseng, L. C.; Huang, R. F.; Chen, C. C.; Chang, C. P. Correlation between Airflow Patterns and Performance of a Laboratory Fume Hood. *J. Occup Environ. Hyg* **2006**, *3* (12), 694−706.

(5) Tsai, S. J.; Huang, R. F.; Ellenbecker, M. J. Airborne Nanoparticle Exposures While Using Constant-Flow, Constant-Velocity, and Air-Curtain-Isolated Fume Hoods. *Ann. Occup. Hyg.* **2010**, *54* (1), 78−87.

(6) Huang, R. F.; Nian, Y.-C.; Chen, J.-K. Static Condition Differences in Conventional and Inclined Air-Curtain Range Hood Flow and Spillage Characteristics. *Environ. Eng. Sci.* **2010**, *27* (6), 513−522.

(7) Li, X.; Jiang, Y.; Zhu, J.; Wang, L.; Zhang, M.; Xu, X.; Fang, Z.; Zhuo, Y.; Zhao, X.; Li, Z.; Cao, Y. Air Curtain Dust-Collecting Technology: Investigation of Industrial Application in Tobacco Factory of the Air Curtain Dust-Collecting System. *Process Safety and Environmental Protection* **2021**, *149*, 676−683.

(8) Logachev, K.; Ziganshin, A.; Kryukova, O.; Averkova, O.; Kryukov, I.; Gol'tsov, A. Improving Dust Capture Efficiency with Local Exhaust Hoods in Manicure Shops. *Build Environ* **2020**, *181*, No. 107124.

(9) Chen, J.-K.; Fung Huang, R.; Hsin, P.-Y.; Hsu, C. M.; Chen, C.-W. Flow and Containment Characteristics of an Air-Curtain Fume Hood Operated at High Temperatures. *Ind. Health* **2012**, *50*, 103−114.

(10) Palluzi, R. "why Don't We Just Put It in the Hood?": Issues with Degrading Hood Effectiveness Due to Equipment Placement. *ACS Chemical Health and Safety* **2020**, *27* (1), 52−59.

(11) Yu-Jie, W.; Li-Hsin, C.; Tzu-Ping, T. The Study on Influencing Factors of Airflow in Laboratory Fume Hood. In *E3S Web of Conferences*; EDP Sciences, 2023; Vol. *396*.

(12) Chen, K.; Wang, W.; Zhang, W. Investigation of Influential Factors on Laboratory Fume Hood Containment Performance. *Sci. Technol. Built Environ* **2020**, *26* (3), 387−399.

(13) Tseng, L. C.; Huang, R. F.; Chen, C. C.; Chang, C. P. Aerodynamics and Performance Verifications of Test Methods for Laboratory Fume Cupboards. *Ann. Occup. Hyg.* **2007**, *51* (2), 173−187.

(14) Lin, G. Y.; Lee, Y. M.; Chiu, B. W.; Hsu, H. Y.; Chen, C. W.; Tsai, C. J. A Comprehensive Study Combing Experiment and CFD Simulation on the Fume Hood Performance for Nanoparticles and Isopropanol Control. *Process Safety and Environmental Protection* **2024**, *186*, 361−375.

(15) Ahn, K.; Woskie, S.; DiBerardinis, L.; Ellenbecker, M. A Review of Published Quantitative Experimental Studies on Factors Affecting Laboratory Fume Hood Performance. *J. Occup Environ. Hyg* **2008**, *5* (11), 735−753.

(16) Reiman, J.; Joao, R. V.; Volin, C. E.; Fernandez, J. S.; Party, E.; Gershey, E. L. Fume Hood Performance: Slot Settings and Containment Testing. *Appl. Occup. Environ. Hyg.* **1998**, *13* (11), 766−769.

(17) Zhao, K.; Liang, S.; Cheng, F.; Cheng, Y.; Meng, C. Design Analysis and Optimization of a Novel Air-Curtain Fume Hood Based on Attached Air Jet. *J. Build. Eng.* **2025**, *99*, No. 111605.

(18) Shen, C.; Dunn, K. H.; Woskie, S. R.; Bennett, J. S.; Ellenbecker, M. J.; Dandy, D. S.; Tsai, C. S.-J. The Effect of the Body Wake and Operator Motion on the Containment of Nanometer-Scale Airborne Substances Using a Conventional Fume Hood and Specially Designed Enclosing Hood: A Comparison Using Computational Fluid Dynamics. *J. Nanopart. Res.* **2022**, *24* (4), 79.
(19) Liu, D.; Meng, C.; Chen, J.; Li, L. The Effect of the Thermal Plume Generated by Body Heat Dissipation on the Containment of Fume Hood. *Indoor and Built Environment* **2021**, *30* (10), 1846−1856.
(20) Muta, R.; Ito, K. Impact of Heat Generation and Use of Experimental Instruments in a Fume Hood on Pollutant Capture Efficiency. *Japan Architectural Review* **2022**, *5* (4), 702−713.
(21) Dunn, K. H.; Tsai, C. S. J.; Woskie, S. R.; Bennett, J. S.; Garcia, A.; Ellenbecker, M. J. Evaluation of Leakage from Fume Hoods Using Tracer Gas, Tracer Nanoparticles and Nanopowder Handling Test Methodologies. *J. Occup Environ. Hyg.* **2014**, *11* (10), D164−D173.
(22) Feng, X.; Xiong, J.; Wang, Q.; Jia, Q.; Yan, F. Boiling Point Enhanced Vapor Pressure Prediction by Antoine-Based Quantitative Structure−Property Relationship. *Chem. Eng. Sci.* **2025**, *306*, No. 121269.
(23) Hyun Nam, J.; Lee, S.; Jo, S.; Kim, J.; Lee, J.; Koo, J.; Lee, B.; Jeong, K.; Yu, D. Improving Vapor Pressure Prediction Through Integration of Multiple Molecular Representations: A Super Learner Approach. *J. Chemom.* **2025**, *39* (2), 1.
(24) Yang, X.; Ren, S.; Wang, Y.; Yang, G.; Li, Y.; Li, C.; Wang, L.; Yao, L.; Wang, L. Volatility Parametrization of Low-Volatile Components of Ambient Organic Aerosols Based on Molecular Formulas. *Environ. Sci. Technol.* **2023**, *57* (31), 11595−11604.
(25) Chantada-Vázquez, M. D. P.; García Vence, M.; Serna, A.; Núñez, C.; Bravo, S. B. SWATH-MS Protocols in Human Dis-Eases. In *Shotgun Proteomics*; Humana: New York, NY, 2021; pp 105-141
(26) Naef, R.; Acree, W. E. Calculation of the Vapour Pressure of Organic Molecules by Means of a Group-Additivity Method and Their Resultant Gibbs Free Energy and Entropy of Vaporization at 298.15 k. *Molecules* **2021**, *26* (4), 1045.
(27) Dearden, J. C. Quantitative Structure-Property Relationships for Prediction of Boiling Point, Vapor Pressure, and Melting Point. *Environ. Toxicol. Chem.* **2003**, *22* (8), 1696−1709.
(28) Balaban, A. T.; Basak, S. C.; Mills, D. *Issue in Honor of Prof Development of Quantitative Structure-Activity Relationship Models for Vapor Pressure Estimation Using Computed Molecular Descriptors*, 2001.
(29) Duchowicz, P. R.; Fioressi, S. E.; Bacelo, D. E.; Quispe, A. Q.; Yapu, E. L.; Castañeta, H. QSPR Predicting the Vapor Pressure of Pesticides into High/Low Volatility Classes. *Environmental Science and Pollution Research* **2024**, *31* (1), 1395−1402.
(30) Wold, S.; Sjostrom, M.; Eriksson, L. PLS-Regression: A Basic Tool of Chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109−130.
(31) Wold, S.; Martens, H.; Wold, H. The Multivariate Calibration Problem in Chemistry Solved by the PLS Method. In *Matrix Pencils. Lecture Notes in Mathematics*; Kågström, B.; Ruhe, A., Eds.; 1983; p 973.
(32) Gentle, J. E.; Härdle, W. K.; Mori, Y. *Springer Handbooks of Computational Statistics Series Editors*. http://www.springer.com/series/7286.
(33) Pérez-Guaita, D.; Kuligowski, J.; Garrigues, S.; Quintás, G.; Wood, B. R. Assessment of the Statistical Significance of Classifications in Infrared Spectroscopy Based Diagnostic Models. *Analyst* **2015**, *140* (7), 2422−2427.
(34) Thomas, E. V. Non-Parametric Statistical Methods for Multivariate Calibration Model Selection and Comparison. *J. Chemom* **2003**, *17* (12), 653−659.
(35) Van Der Voet, H. *Chemometrics and Intelligent Laboratory Systems Comparing the Predictive Accuracy of Models Using a Simple Randomization Test*, 1994; Vol. *25*.
(36) Duponchel, L.; Fabre, C.; Bousquet, B.; Motto-Ros, V. Statistical Comparison of Predictive Models for Quantitative Analysis and Classification in the Framework of LIBS Spectroscopy: A Tutorial. *Spectrochim. Acta, Part B* **2023**, *208*, No. 106776.

https://doi.org/10.1021/acs.chas.5c00179
ACS Chem. Health Saf. 2026, 33, 138−146