

# Standardized sampling applied for VOCs emitted by feces using a Micro-Chamber/Thermal Extractor ( $\mu$ -CTE)

**Ileana Andreea Ratiu**

"Raluca Ripan" Institute for Research in Chemistry, Babes-Bolyai University

**Radik Mametov**

Interdisciplinary Centre of Modern Technologies – BioSep, Nicolaus Copernicus University

**Tomasz Ligor**

Interdisciplinary Centre of Modern Technologies – BioSep, Nicolaus Copernicus University

**Bogusław Buszewski** (✉ [bbusz@chem.umk.pl](mailto:bbusz@chem.umk.pl))

Interdisciplinary Centre of Modern Technologies – BioSep, Nicolaus Copernicus University

---

## Research Article

**Keywords:** standardized sampling method, feces samples, Micro-Chamber/Thermal Extractor ( $\mu$ -CTE), Box–Behnken design

**Posted Date:** July 28th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-743655/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

VOCs (volatile organic compounds) are increasingly wished to be used in diagnosis of diseases. They present strategic advantages, beside classical methods used, such as simplicity and non-invasively of performant sample collection methods/ systems available our days. However, standardized sampling methods are required in order to achieve reproducible results. In the current study we developed a method to be used for feces sampling using a Micro-Chamber/Thermal Extractor ( $\mu$ -CTE). Design Expert software (with Box–Behnken design) was used to predict the solutions. Therefore by using the simulation experimental plan that was further experimentally verified, a sampling time of 19.6 min, at a sampling temperature of 30.6 °C by using a flow rate of 48.7 mL/min provided the higher response. The developed method was validated by using correlation tests and Network analysis, that both proved the validity of the developed model.

## 1. Introduction

VOCs emitted from feces are increasingly considered potential useful markers in diagnostic of different diseases; however standardization of sampling procedures is still in the incipient phase. Efforts for development of standardized sampling methods were made in different fields of metabolomics, and mostly for the case of breath samples<sup>1–5</sup>. Moreover, for breath analyses various sensors able to perform the analyses in real time, without any pre-concentration or storage were created in attempt to diagnose respiratory diseases<sup>6–8</sup> and a standardized breath sampler was developed<sup>1</sup>. Except breath samples, other matrices such as urine<sup>9–11</sup>, sweat<sup>12</sup>, tissues<sup>13</sup>, saliva<sup>14</sup> breast milk<sup>15</sup>, exudates<sup>16</sup> or bacteria<sup>17–19</sup> were investigated in research articles and discussed in reviews<sup>19–22</sup> for discover valuable biomarkers of certain diseases. In case of feces samples, literature in the field is related mostly to identification of VOCs, markers of colorectal-cancer<sup>23–26</sup>, but however, considerably less progress has been made compared with investigation of other biological matrices, especially breath and urine. Optimization of sampling parameters for human feces samples was realized in terms of investigating collection procedure, transport, and storage<sup>27–29</sup>. Extraction of VOCs from fecal samples was also optimized, in terms of time (10 vs 20 min), storage at room temperature of freezing, and amount of sample<sup>30</sup>. In terms of response, the number of peaks obtained after GC-MS analyses was considered<sup>30</sup>. Nevertheless, to the best of our knowledge no optimization of sampling parameters by using the Micro-Chamber/Thermal Extractor was realized before. Emitted components can be collected in a wide range of sorbent tubes and analyzed by standard analytical methods (usually using a GC-MS or a GC-FID equipped with a thermal-desorption (TD) unit). The  $\mu$ -CTE was designed for a widely range of applications including emissions screening of construction materials<sup>31</sup>, profiling the aroma and fragrance or foods and consumer products<sup>32</sup>. Nevertheless,  $\mu$ -CTE can be successfully used for other applications, including sampling of VOCs emitted by biological samples, as it was proved by the present study. In contrast to other studies, we used a complex optimization method, able to provide the best possible solutions, predicted by software as a result of experimental measurements and total interactions between three selected factors. The selected factors were as follow: extraction temperature, extraction time, and gas flow, while Design Expert software (with Box–Behnken design) was used to predict the solutions (responses).

The Box–Behnken design is an efficient and time saving optimization method which can provide a representation of polynomial response function that cannot be described by linear functions. It can be applied for optimization parameters in many fields, and it was previously used in investigation of extraction parameters in accelerated solvent extraction<sup>33</sup>, pressurized liquid extraction<sup>34</sup>, effect of process parameters on the microparticle production<sup>35</sup>, evaluation of shale gas recovery enhancement<sup>36</sup>, etc.

In contrast with the previous study that realized kindred experiments<sup>30</sup>, in addition to number of peaks we investigated as desired responses total peaks area and peaks intensities. Consequently, three models were created, and three final solutions were obtained. Our study included two steps: optimization of sampling parameters (the main aim of the study) and investigation of feces profiles coming from human volunteers of different ethnicities, genders, ages, with and without self-declared chronic diseases having different culinary habits. To the best of our knowledge the optimization of sampling parameters for  $\mu$ -CTE was not investigated before for neither in case of feces samples, nor other biological samples, and this is why the present work represents an original approach.

## 2. Materials And Methods

### 2.1. Materials and chemicals involved

Analytical standards: 4-bromofluorobenzene, alkanes (C8 to C26), methanol and hexane with purity  $\geq 99.99\%$ , disposable aluminum dish, (20 mL) to be placed in the sampling chamber were all purchased from Markes International (UK). Sterile containers for sampling were bought from pharmacy.

### 2.2. Samples collection and preparation

A total number of 20 volunteers with different ethnicity (European, Asian, African, Latin American), all of them living in Poland at the moment of samples collection were involved in the experiment. A written informed consent was handed to each participant, who then completed a short questionnaire. The samples were collected by self-collection in sterile container provided by the researchers. The donor allotted a code to his/her own sample, which was written on the sample container and in the questionnaire. The samples were provided in the sterile container that was packed in aluminum foil. All samples were collected together, while the questionnaires were placed by the donors in a blank folder. That helped in linking a sample with a given questionnaire, and in keeping anonymous the identity of donor. However, relevant information about volunteers, extracted from the questionnaires is provided in Table S1. Because some questions were not mandatory, Table S1 includes less complex information in case of some volunteers. Immediately after collecting, the samples were stored in the home fridge (between 0 and 10 hours) and transported in the lab. Here the samples were stored at  $-80^{\circ}\text{C}$ , until the total number of 20 was collected (3 days period). Further the samples were defrosted; from each donor 2 samples of 0.25 g were packed and prepared to be used for analysis of VOCs, while 1 gram was kept for building the pool of samples. In case of 2 volunteers the sample amount was not enough to be added in the pool, while in case of one more the added amount was less than 1 gram. Consequently, the pool contained the mixture of 18 samples only. The pool was very well homogenized and divided in individual samples of 0.25 g each. The resulted samples were again stored at  $-80^{\circ}\text{C}$  and kept until were used for sampling of VOCs. For optimization of sampling parameters only pool samples were used, while later when VOCs screening of each sample was made. Pool samples were also analyzed in this step and used as quality controls.

### 2.3. Sampling procedure using Micro-Chamber/Thermal Extractor ( $\mu$ -CTE)

The Micro-Chamber/Thermal Extractor ( $\mu$ -CTE) produced by Markes International is a compact portable unit 31.5 cm high, 12 cm wide and 50 cm deep, with a total mass 10.2 kg. It is equipped with up to six small cylindrical chambers (28mm deep and 45mm in diameter) where the sample can be placed. An adjustable flow rate (clean air, nitrogen, helium) between 10 and 500 mL/min and temperatures from ambient up to  $120^{\circ}\text{C}$  can be used for sampling. The  $\mu$ -CTE (Fig. 1, part A) was designed for sampling of semi-volatile or volatile organic compounds released from various materials, from up to six samples simultaneously.

Electrical properties are as follow: maximum power: 360 W, line voltage 100–240 V, frequency 50–60 Hz, while the input inrush current is < 25 A (cold start). It can be operated in ambient temperature between 15 to 30°C, and humidity range 5 to 95% <sup>32</sup> Emitted components can be collected in a wide range of sorbent tubes and analyzed by standard analytical methods (usually using a GC–MS/FID system equipped with a thermal-desorption (TD) unit).

In the present study, before sampling, each sample taken out of the freezer, placed in the sampling chamber. Temperature and gas flow were according with the designed runs from Design expert. Subsequently, the adsorbent tube (manufactured by Markes International, sorbent type: Tenax TA/Carbograph 5TD (Bio Monitoring), Stainless Steel, 1/4" i.d., length 9 cm) was placed on the top of a sampling chamber (Fig. 1, part B). The gas flow made possible the passing of emitted VOCs through the adsorbent tube. The Micro-Chamber/Thermal Extractor (Markes International) was connected to cylinder with 99.99 % nitrogen (Air Liquide, Poland). This method achieved dynamic head-space sampling of the VOCs associated with related metabolites. After sampling, the analyses started immediately without another future storage.

## 2.4. GC-MS system used

The analysis was carried out using Agilent 7820A GC coupled with mass spectrometer Agilent 5977B MSD (Agilent Technologies, Waldbronn, Germany). GC was equipped with DB-5 MS capillary column (30 m × 0.25 mm × 1 µm). Helium was used as the carrier gas in constant flow mode at 1 ml/min. The column temperature was programmed as follows, the initial temperature of 40°C was held for 5 min, and then the temperature was increased by 10°C/min to 310°C and maintained for 4 min at this value. The mass spectrometer was operating at EI (70 eV) mode. Ion source temperature was set to 230°C and transfer line were set to 250°C, acquisition frequency 2.9 scans/sec and mass range 50–550 amu. The dual stage thermal desorber Markes TD-100 (Markes International, Bridgend, UK) was coupled to chromatographic column by heated transfer line equipped with deactivated capillary (2 m × 0.25 mm) and micro union. The TD conditions were as follows sorbent tube desorption temperature 280°C for 5 min, at desorption flow 50 cm<sup>3</sup> min<sup>-1</sup>. Cold trap desorption at 300°C for 3 min at 2 cm<sup>3</sup> min<sup>-1</sup> flow. Trap heating rate was set to 100°C min<sup>-1</sup>, and transfer line temperature 200°C.

## 2.5. Quality control, system cleaning and stability checking

Pool samples were used as quality controls. After each four regular samples, one pool was analyzed. Every morning before starting the analyses the instrument was subjected to a cleaning method that included graduate heating of the column up to 300°C and purging with gas (total time of the method was 30 minutes). If the background level was not considered satisfactory the cleaning method was repeated. Before using, sorbent tubes were also conditioned in the TD unit, at 320°C for 2 hours with helium flow rate 95 mL/min. Retention index containing a mixture of alkanes (C8 to C26, prepared in hexane) were also run daily before starting the analyses. By using the known retention times of standards, system stability was verified. Internal standard (IS), BFB (4-bromofluorobenzene) in methanol was used and added to samples at fixed concentration of 100 ng mL<sup>-1</sup>. After optimization of parameters, when VOCs of samples were investigated, the ratio between IS and each detected VOC was used to compare the obtained profiles. In this way, the possible errors resulting from non-reproducibility between sorbent tubes, were eliminated. Eventually, background of clean blank tubes was analyzed. The resulted components originated from blanks (coming from sorbent tubes, column bleeding and septa), pools and 20 involved volunteers, accounting for the number of 205 VOCs are shown in Table S2, presented as supplementary materials.

## 2.6. Statistical analyses

Response surface methodology (RSM, Design-Expert v.11, Stat-Ease, Minneapolis, Minnesota, USA), with Box-Behnken design, an efficient and time saving optimization method was used for optimization of sampling

parameters. Pearson correlation analyses, and hierarchical clusters analysis were created using IBM SPSS statistical software, package version 21. Network analysis was performed using R studio with console (version 3.6.3, Boston, MA, USA). Where applicable, microsoft Power Point 2010 was used to label and combine the figures.

## 2.7. Ethical Considerations:

This study was conducted according to the guidelines of the Declaration of Helsinki and approved by Bio-Ethical Commission of Collegium Medicum in Bydgoszcz of the Nicolaus Copernicus University in Toruń, according with the agreement number KB 49/2018 – 16.01.2018. Written informed consent was obtained from all participants.

## 3. Results And Discussion

### 3.1. Sampling optimization

#### 3.1.1. Involved factors and fitting equations

In our study, Box-Behnken design was involved in attempt to optimize the sampling procedure for feces samples using a  $\mu$ -CTE produced by Markes International. Tested parameters were as follow: extraction temperature, sampling time and gas flow. All of them were selected as coded independent variables (factors) at three levels: minimum (-1), central (0) and maximum (1). The factors were tested by carrying out 17 random runs (each one in duplicate), including five central points. The used values were: temperature: 24°C, 32°C, 40°C; sampling time: 10 min, 20 min and 30 min; gas flow: 20 mL/min, 35 mL/min and 50 mL/min. The effect of factors was observed for: number of total detected peaks, total peaks area and total intensity of the peaks.

Four polynomial models: linear, two-factor interaction (2FI), quadratic and cubic model, were statistically evaluated. The linear and 2FI models were not significant, while the cubic model was aliased. This means that there were not enough unique design points to independently estimate all the coefficients for those models. Thus, the quadratic model was selected to build the response surface in the subsequent optimization process.

To assess the impact of each factor on the observed responses, the second-order polynomial model (Eq. 1) described the relation between independent variables and the obtained responses.

$$Y = \beta_0 + \sum \beta_i X_i + \sum \beta_{ii} X_i^2 + \sum \beta_{ij} X_i X_j \quad (1)$$

where, Y is the obtained response,  $X_i$  and  $X_j$  are the independent factors,  $\beta_0$ ,  $\beta_i$ ,  $\beta_{ii}$ , and  $\beta_{ij}$  are the regression coefficients for intercept, linear, quadratic, and interaction coefficients terms.

The second order polynomial equations (Eq. 2 to 4) were used to express the obtained responses as a function of coded and independent variables, where Y = response;  $X_1$ : Extraction temperature (°C);  $X_2$ : extraction time (minutes);  $X_3$ : flow rate (mL/min);

$$Y1 \text{ [number of the peaks]} = 122.60 - 7.87 X_1 + 7.81 X_2 + 4.69 X_3 + 7.37 X_1 X_2 + 6.88 X_1 X_3 - 7.00 X_2 X_3 - 27.05 X_1^2 - 5.68 X_2^2 + 7.58 X_3^2 \quad (2)$$

$$Y2 \text{ [peak areas]} = 4.61E + 07 - 1.98E + 06 X_1 + 7.31E + 06 X_2 + 6.13E + 06 X_3 - 3.18E + 06 X_1 X_2 - 3.22E + 06 X_1 X_3 - 6.54E + 06 X_2 X_3 - 1.67E + 07 X_1^2 - 5.82E + 06 X_2^2 - 7.36E + 06 X_3^2 \quad (3)$$

$$Y_2 [\text{peak intensity}] = 1.04E + 07 - 4.76E + 05 X_1 + 1.98E + 06 X_2 + 1.45E + 06 X_3 - 5.33E + 05 X_1 X_2 - 4.91E + 05 X_1 X_3 - 7.14E + 05 X_2 X_3 - 3.75E + 06 X_1^2 - 3.28E + 05 X_2^2 - 9.42E + 05 X_3^2 \quad (4)$$

### 3.1.2. Analysis of variance (ANOVA) for quadric polynomial model

For evaluation of quadratic model criteria such as: assessment of “adjusted R-squared” and “predicted R-squared” were applied. The “Predicted R-Squared” was in reasonable agreement with the “adjusted R-Squared”, while the recorded difference was less than 0.2 in all three cases (see Table S3). However, ANOVA test was used for evaluation of goodness-of-fit of the proposed design and demonstrated that all three models were significant based on  $R^2$  (0.9385 to 0.9665), lack of fit ( $p > 0.05$ ) and p-values (0.0018 to 0.0002). The  $R^2$  values indicate that between 93% up to 96% of the changes observed in the responses were obtained under the combination of factors and could be explained by the model equations. Furthermore, low values of coefficient of variance (between 4.7 % and 15.69 %) were obtained, which indicates an adequate precision of the experimental values. The p-values lower than 0.0500 indicates a corresponding model, while smaller p-value are associated with more significant result. Our obtained p-values show that there is only between 0.02% up to 0.18% chance that large F-values could occur due to the noise in experiment. This implies that all designed quadratic models are highly significant. Lack of Fit (and the obtained F values) are other important index in evaluation of model reliability because the regression equation and coefficient of determination are evaluated to test the model fitting. Our obtained F values (between 0.34 and 1.05) indicates that lack of fit is not significant compared to pure error. Detailed ANOVA test results for the quadratic polynomial model are presented in Table S3.

The actual versus predicted responses are plotted in Fig. 2, part I. The actual values are obtained data, while the predicted values are produced by the model. By observing Fig. 2 part I, we can understand that all three models (A, B and C) are able to predict well the actual response. Residuals are defined as deviation between predicted and actual values. A good model is expecting that the obtained responses will follow a normal distribution along the straight line (see Fig. 2 part II). If an evident S-shape curve is formed, this denotes that the residuals are not following a normal distribution; consequently an inappropriate model was obtained. According with Fig. 2 part II, the Studentized residuals follow a normal distribution. The results presented in Table S3 and Fig. 2 indicates that there is not any evidence pointing out possible errors regarding the adequacy of the model.

### 3.1.3. The influence of investigated factors on the obtained responses

#### 3.1.3.1. Single factor influence

The influence of each investigated factor on the obtained responses is depicted in Fig. 3. Part A is presenting the influence of temperature variation when sampling time and gas flow were kept constant at central points (sampling time 20 min and gas flow 35 mL/min). As highlighted, for all three models the best responses were obtained around the value of 32°C. Presumably, this means that lowest temperature (24°C) is not enough to fully volatilize the components, while the highest (i.e. 40°C) can produce a series of problems that lead to the loss of volatiles, their conversion in more different types of volatiles and/or some other specific issues. For example, it was reported that microbial fermentation processes are increasing at higher temperatures<sup>37,38</sup>, while different patterns of amino-acids decomposition by bacteria in biological samples were also referred<sup>16,19</sup>. However, since VOCs have specific optimal temperatures of volatilization, optimal temperature choice depends on VOC of interest, which may differ according with desired applications and on system used for analysis. In our case, we could observe in the chromatograms of samples analyzed at 40°C unseparated (overlapped) wide peaks. Nevertheless, in the mentioned chromatograms both

peaks intensity and peaks area were lower compared with the chromatograms recorded at lower temperatures. Regarding the targeted VOCs, we were more interested in highly volatile short chain alcohols acids and other components containing nitrogen and sulfur.

In Fig. 3 part B the influence of sampling time is presented when temperature and gas flow were constant (temperature 32°C and gas flow 35 mL/min). It can be observed that in all three cases sampling time has a direct influence on the obtained responses, longer sampling time giving higher responses. In case of number of the peaks and peaks area the trend line arrived at the plateau, while in case of peaks intensity is still increasing. This presumably can mean that a longer sampling time could have given a higher response.

When gas flow was investigated at constant sampling time (20 min) and temperature (32°C), the trend line was rising with increasing of flow in case of peaks number (Fig. 3 part C). In case of peaks area, the plateau was reached between 38 and 44 mL/min, followed by a slight decrease of the trend line when the flow was increased more than 45 mL/min. However, continuous increasing tendency was observed in case of peaks intensity, with plateau tentative at the maximum flow. Consequently, we can affirm in this step that longer sampling time as well as higher gas flow and temperature around 30°C will facilitate the obtaining of the best possible results

### 3.1.3.2. Two factors influence

Preliminary solutions were obtained after single factor influence checking. Nevertheless, the main aim of our study was to develop a method that can provide the best solution to be used by a combination of the mentioned factors, once in the sampling process the variables can be influenced between themselves. Consequently, two factors influence, when a third one (named actual factor) is kept constant at the central point is highlighted in Fig. 4 and will be further discussed. In case of peaks number (Fig. 4 part A), when the actual factor was gas flow (kept constant at 35 mL/min) it was observed that a middle temperature (around 30°C) and middle extraction time (approximately 20 min) provided the higher response (A1). By maintaining the extraction time constant, middle temperature and high gas flow (> 45 mL/min) provided the best solutions (A2). Eventually, at constant temperature (32°C) it was observed that highest gas flow (tending to the maximum value), could lead to the detection of the maximum possible number of peaks, while the extraction time did not present significant influence (A3). These findings highlight that in case of detected number of peaks, the most relevant factor is flow rate. Moreover, by observing the last “umbrella” of Fig. 4, part A (A3) it is suggested that the usage of higher gas flow could have led to a higher number of detected peaks.

For peaks area (Fig. 4, part B) a very good fitted model was obtained, once the figures shape is like well contoured umbrellas, and the color scale in going from the lowest to the highest values (blue to red). It was highlighted that when the actual factor was gas flow, extraction time between 25 and 30 minutes at about 32°C is giving a good response (B1). If the actual factor was considered extraction time, a temperature around 30°C and a gas flow of 40 mL/min provided the highest response (B2). In the third case (B3), when the actual factor was temperature, the highest extraction time (30 min), when using a gas flow between 35 and 40 mL/min provided the highest peaks area. By observing Fig. 4, part B3 it is highlighted that longer extraction time could have led to the recording of higher peak areas. In case of this model, gas flow and extraction time were more influential factors than temperature.

Finally, the influence of factors on peaks intensity is presented in Fig. 4, part C. Part C1 (gas flow = constant at 35 mL/min) is showing that at highest extraction time (30 min) and middle temperature (around 30°C) is leading to the highest intensities. Moreover, it may be observed that in Part C1 that longer extraction time can provide peaks of higher intensity. Nevertheless, high gas flow (50 mL/min) and temperature near to 30°C provided good solution, although a gas flow > 50 mL/min would probably have produced more intense peaks, as shown in Part C2. It can be

observed in Part C3 (where the actual factor was temperature) that a gas flow between 35 and 45 mL/min and a sampling time  $\geq 30$  minutes led to obtaining of maximum intensity, although a higher extraction time could have resulted in more intense peaks. These findings suggest that in case of peaks intensity the most relevant factor was extraction time.

### 3.1.4. Obtained solutions

The final obtained solutions were predicted by the software as a result of experimental measurements and total interactions between the three selected factors. The predicted solutions, that were verified than experimentally for each model, as presented in Table 1.

Table 1  
Obtained solutions as a result of optimization of sampling parameters

Approach	Temp (°C)	Extraction time (min)	Gas flow (mL/min)	Predicted response	Standard error	Desirability	Measured response		
							R1 (No of peaks)	R2 (Area of peaks)	R3 (Intensity of peaks)
Number of peaks	<b>30.661</b>	<b>19.612</b>	<b>48.687</b>	132.678	2.879	1.000	<b>137</b>	<b>4.60E+07</b>	<b>1.31E+07</b>
Area of peaks	30.948	25.528	38.004	4.89E+07	2.25E+06	0.936	128	5.51E+07	1.67E+07
Intensity of peaks	30.572	29.242	46.887	1.23E+07	5.96E+05	1.000	91	4.41E+07	1.17E+07
R1 – the obtained response when the solution acquired for peaks number was used									

R2 – the obtained response when the solution acquired for peaks area was used

R1 – the obtained response when the solution acquired for peaks intensity was used

Because the software affords for selection of solutions (in term of goal and importance), thus allowing the prioritization of options according to needs, certain criteria were chosen in attempt to obtain the best solutions, that will fit our model. Thereby, maximum importance was accorded to each factor (temperature, sampling time and gas flow), while the selected goal was “in range”. To the obtained response we assigned also maximum importance and the selected goal was “maximize”, while in case of error, the selected goal was “in range” and the selected importance was medium. According to Table 1, the most suitable parameters to be used are the parameters obtained according with peak numbers (highlighted in bold).

### 3.2. Testing of developed model

In order to test the validity of the developed model, Correlation Analysis and Network analysis were run. In case of correlation, parametric tests were used once they are “a statistical tool expecting a linear correlation between the investigated variables coming from the same source, and/or they have the same measure unit”<sup>39</sup>. Two different approaches were tackled: correlation between all 205 detected components in the whole matrix (including detected components appearing in pool samples, in samples from each volunteer and in blanks) – represented in Fig. 5; and correlation between all 20 included volunteers and pool samples – as presented in Fig. 6. Network analysis highlighted in Fig. 7 is presenting discrimination between the investigated groups, act that confirms the validity of the model.

### 3.2.1. Correlation based on detected components

A Pearson moment product correlation was used to highlight the correlation that occurs between the detected components. Because the matrix was complex (including 205 VOCs) a snap shot representation only could be presented in Fig. 5 in the form of a heat map. The correlation values significant at 0.01 level were between  $r(205) = -0.446$  to  $0.999$ , which pointed out in a low up to very strong positive or negative correlation. Moreover, the correlation values significant at a 0.05 level accounted for  $r(205) = -0.560$  to  $0.426$ , denotative of a moderate correlation. The strong and very strong correlation that was achieved when  $p$  was significant at 0.01 level ( $p = 0.01$ ), was a positive correlation. All the components detected in blanks presented a positive correlation with each other ( $r(205) = 0.145$ ,  $p = 0.05$  up to  $0.982$ ,  $p = 0.01$ ). With exceptions the components detected in pooled samples (accounting the number of 146) positively correlated the detected components. An example was for example, ethyl acetate and 1,3,5-trifluorobenzene, that did not present any correlation with the pool samples, but they were however highlighted as components originated from blank tubes.

Some other examples were VOCs detected in case of limited number of volunteers only, in relatively small amounts, reason why they was not detectable in pool samples. Less often than positive correlation, negative correlation ( $r(205) = -0.56$ , up to  $-0.15$ ,  $p = 0.05$ ) was also observed. Components that exhibited the most often negative correlation with other present in the matrix were: 2-propanamine 2-methyl-; fumaronitrile; trifluoromethyltrimethylsilane; 2,5-hexanediol; cyclotetrasiloxane, octamethyl-; cyclopentasiloxane, decamethyl-; biphenyl; dibutyl phthalate; cyclohexane, isothiocyanato-; heptasiloxane, 1,1,3,3,5,5,7,7,9,9,11,11,13,13-benzene, 1,1'-(4,4-dimethyl-1-butene-1,4-diyl)bis-tetradecamethyl- and five more that could not be identified, and remained unknown. They were generally associated with components coming from tubes, septa and columns bleeding. These findings suggest that the background may affect the obtained results, however generally low values of negative correlation were observed.

However, the most of the VOCs common between pooled samples and volunteers' samples presented moderate or low positively correlation. According with these findings we can assume the validity of the model, while we find positive correlation between samples and pools when investigating the detected output (the VOCs). This denotes also that the model worked correct and the pools were representative.

### 3.2.2. Correlation based on investigated samples

As a second approach, a Pearson moment correlation was conducted also to check the correlation and the level of significance between pool samples and samples coming from each volunteer. The whole VOCs profile of each sample was used in this boarding. The correlation matrix was designed based on the hierarchical clusters' analyses obtained using the method "average linkage between groups" within the interval "squared Euclidian distance" and presented in Fig. 6. The dendrogram of clusters' analysis shows the formation of seven main clusters with different levels of significance. The samples coming from 11 volunteers fused together in one cluster with similar distance level in the left part of dendrogram, finding that suggest that they had kindred VOCs composition. The samples #20, #3, #4, #14 and #5 clustered next to them, highlighting less similarities compared with the first 11 mentioned. Nevertheless, positioned at the opposite site, in a rather arbitrarily manner were the clusters corresponding to volunteers #17, #9, #15 and #1, all of them with different levels of significance. These denote that the VOCs composition of the mention samples presented the most different features when compared with others. All five pool samples with the same level of similarity were segregated, between the other clusters, drawing actually the border between those with kindred VOCs profiles and those that presented most different features. Observing pools positioning in Fig. 6, we can assume that they were representative for the batch of analyzed samples, and consequently we presume also the validity of the developed model, when the analyzed samples were considered as variables.

Regarding the heat map it can be observed that the highest correlation occurred, as expected between the pool samples ( $r(23) = 0.762$  up to  $0.970$ ,  $p = 0.01$ ). The pools were also well correlated with the samples. However they were less correlated with the samples coming from the volunteers #1, #3 and #6. These samples were associated with the three cases where the amount of sample was insufficient to add to the pool. Except of this, positively correlation from low to very high was accounted between samples originated from volunteers. As a general conclusion, all these findings indicate that there is relevant correlation between the investigated samples, both when the investigated variables were the detected VOCs and when each sample was investigated based on the entire profile, even if these characteristics are far different and in appearance not linked-up with each other.

### **3.2.3. Discrimination between the groups based on Network Analysis**

The dataset representing distributions of 205 detected VOCs was used to build the network analysis model (Fig. 7). Consequently, based on the obtained profiles and using the incidence of the peaks, a network analysis was created in order to obtain an exploration of the data. R studio with console (version 3.6.3, Boston, MA, USA) was used for network analyses. The model separated the three investigated groups, by leading into the formation of three cluster groups (group of pools, the group of volunteers and the blanks). In addition, VOCs detected just in volunteers' samples have been dispersed around the mentioned group (purple circles), common VOCs between pools and volunteers were located mostly between the two groups, and the VOCs detected in blanks were scattered into the left-down part (black diamonds). Regarding the number of VOCs used in network analysis, 2 were specific for blanks, 59 for volunteers, while 146 VOCs were common between the pools and volunteers. We presume that most of VOCs presented in Fig. 7 are endogenous generated by the organism, as a normal process of metabolism. However, certainly part of VOCs are exogenous absorbed by the organism from the environment and eliminated through feces. However, variability in VOCs coming from different subjects was observed, and this phenomenon is related to several factors, among which diet, living style, personal habits, etc.

## **Conclusions**

A standardized sampling method for VOCs emitted from feces by using as sampling system a Micro-Chamber/Thermal Extractor ( $\mu$ -CTE) able to provide the samples collection in controlled conditions was developed in the present study. Design Expert software (with Box–Behnken design) was used as simulation software for predicting the solutions that were eventually verified experimentally. The developed model proved to be correct according with statistical analysis (ANOVA tests, Actual vs Predicted Values and Normal plot of Residuals). Consequently, the sampling time of 19.6 min, at a sampling temperature of  $30.6\text{ }^{\circ}\text{C}$  by using a flow rate of  $48.7\text{ mL/min}$  provided the higher response. The developed method was validated by using correlation tests and Network analysis, that both proved the validity of the developed model.

## **Abbreviations**

$\mu$ -CTE - Micro-Chamber/Thermal Extractor

## **Declarations**

**Funding:** This work was supported by Airborne Biomarkers for Colorectal Cancer, project within program ERA-NET: TRANSCAN-2, Joint Transnational Call for Proposals 2016 on “Minimally and non-invasive methods for early detection and/or progression of cancer”, Transcan-067, KE, no. ERA-NET TRANSCAN/02/2018

**Conflict of interest:** the authors declare no conflict of interest

**Authors contribution:** Conceptualization: IAR; Data curation: IAR; Formal analysis: RM, TL, BB, IAR; Funding acquisition BB; Investigation RM, IAR, TL; Methodology: RM, IAR; Project administration: BB, TL; Resources: BB; Software: IAR; Supervision: IAR, TL, BB; Validation: RM, IAR; Writing - original draft: IAR; Writing - review & editing: RM, TL, BB.

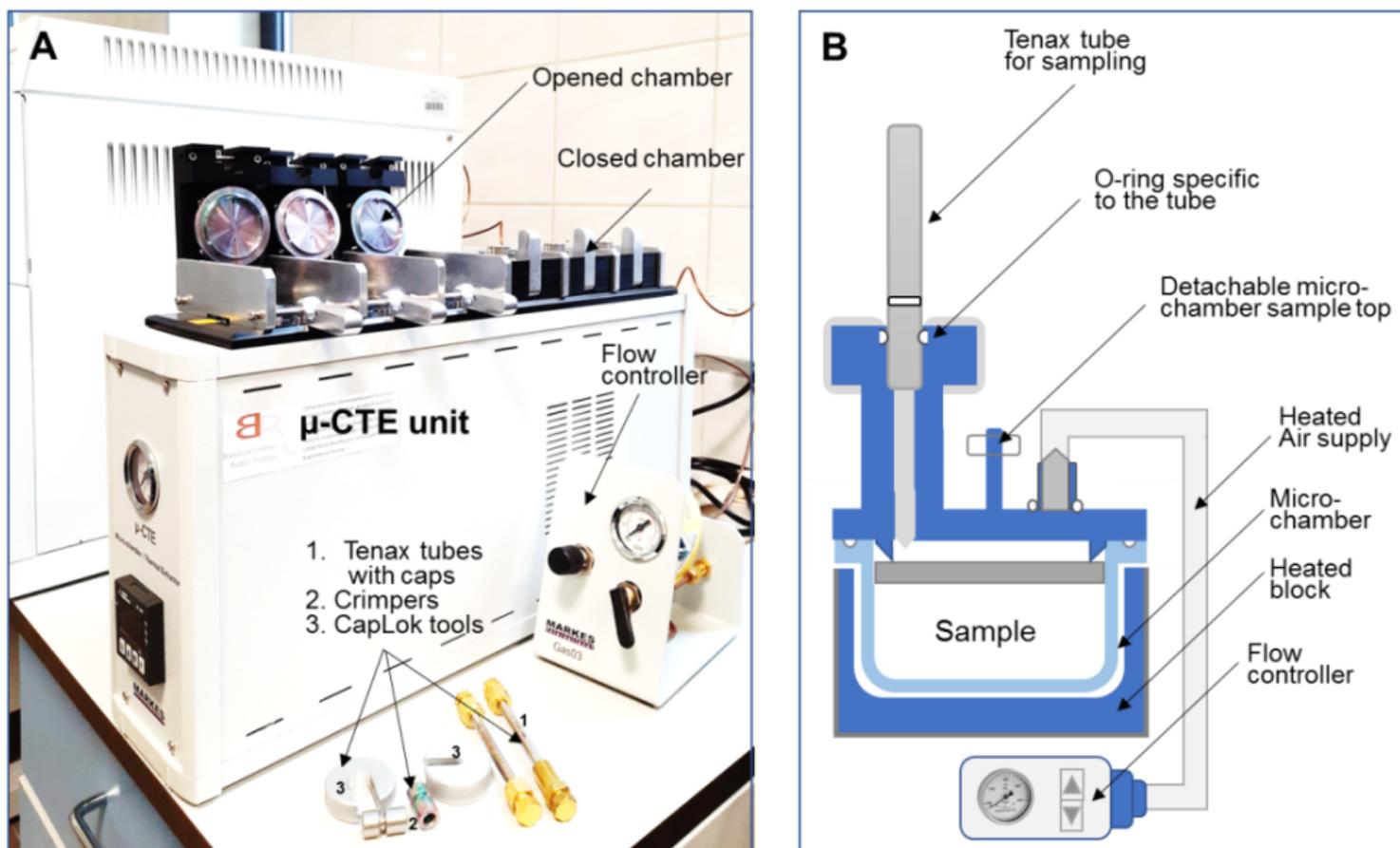
## References

1. Harshman, S. W. *et al.* Characterization of standardized breath sampling for off-line field use. *J. Breath Res.* **14**, (2020).
2. Fens, N. *et al.* External validation of exhaled breath profiling using an electronic nose in the discrimination of asthma with fixed airways obstruction and chronic obstructive pulmonary disease. *Clin. Exp. Allergy*, **41**, 1371–1378 (2011).
3. Phillips, M. *et al.* Blinded validation of breath biomarkers of lung cancer, a potential ancillary to chest CT screening. *PLoS One*, **10**, 1–17 (2015).
4. Ratiu, I. A., Ligor, T., Bocos-Bintintan, V., Mayhew, C. A. & Buszewski, B. Volatile Organic Compounds in Exhaled Breath as Fingerprints of Lung Cancer, Asthma and COPD. *J. Clin. Med*, **10**, 32 (2020).
5. Monedeiro, F., Milanowski, M. M., Ratiu, I. A., Bro, B. & Ligor, T. Needle Trap Device - GC - MS for Characterization of Lung Diseases Based on Breath VOC Profiles(2021).
6. Timms, C., Thomas, P. S. & Yates, D. H. Detection of gastro-oesophageal reflux disease (GORD) in patients with obstructive lung disease using exhaled breath profiling. *J. Breath Res.* **6**, (2012).
7. Hattesoehl, A. D. M. *et al.* Discrimination between COPD patients with and without alpha 1-antitrypsin deficiency using an electronic nose., **16**, 1258–1264 (2011).
8. Bannier, M. A. G. E., Van De Kant, K. D. G., Jöbsis, Q. & Dompeling, E. Feasibility and diagnostic accuracy of an electronic nose in children with asthma and cystic fibrosis. *J. Breath Res.* **13**, (2019).
9. McFarlane, M. *et al.* Urinary volatile organic compounds and faecal microbiome profiles in colorectal cancer. *Color. Dis*, **21**, 1259–1269 (2019).
10. Maurer, D. L. *et al.* Screening of Microbial Volatile Organic Compounds for Detection of Disease in Cattle: Development of Lab-scale Method. *Sci. Rep*, **9**, 1–14 (2019).
11. Tomasz Ligor, J., Zawadzka, G., Straczynski, Rosa, M. & González Paredes, A. W. P. & Ileana Andreea Ratiu, M. M. Searching for Potential Markers of Glomerulopathy in Urine by. *Molecules* **26**, (2021).
12. Monedeiro, F., dos Reis, R. B., Peria, F. M., Sares, C. T. G. & De Martinis, B. S. Investigation of sweat VOC profiles in assessment of cancer biomarkers using HS-GC-MS. *J. Breath Res*, **14**, 026009 (2020).
13. Ratiu, I. A. *et al.* GC-MS application in determination of volatile profiles emitted by infected and uninfected human tissue. *J. Breath Res.* **13**, (2019).
14. Monedeiro, F. *et al.* VOC profiles of saliva in assessment of halitosis and submandibular abscesses using HS-SPME-GC/MS technique. *Molecules* **24**, (2019).
15. Pajewska-Szmyt, M., Sinkiewicz-Darol, E. & Gadzała-Kopciuch, R. The impact of environmental pollution on the quality of mother's milk. *Environ. Sci. Pollut. Res*, **26**, 7405–7427 (2019).
16. Ratiu, I. A. *et al.* Features of infected versus uninfected chemical profiles released from human exudates. *Stud. Univ. Babeş-Bolyai Chem*, **64**, 207–216 (2019).

17. Andreea, R. I., Victor, B. B., Matthew, T. & Hugo, M. V. T. P. Discrimination of chemical profiles of some bacterial species by analysing culture headspace air samples using TD-GC/MS. *Curr. Anal. Chem*, **10**, 488–497 (2014).
18. Bocos-Bintintan, V., Thomas, C. L. P. & Ratiu, I. A. Sensors' array of aspiration ion mobility spectrometer as a tool for bacteria discrimination., **206**, 120233 (2020).
19. Ratiu, I. A. *et al.* An Optimistic Vision of Future: Diagnosis of Bacterial Infections by Sensing Their Associated Volatile Organic Compounds. *Crit. Rev. Anal. Chem*, **0**, 1–12 (2019).
20. Dospinescu, V. M., Tiele, A. & Covington, J. A. Sniffing out urinary tract infection-diagnosis based on volatile organic compounds and smell profile. *Biosensors*, **10**, 1–28 (2020).
21. Mametov, R., Ratiu, I. A., Monedeiro, F., Ligor, T. & Buszewski, B. Evolution and Evaluation of GC Columns. *Crit. Rev. Anal. Chem*, **0**, 1–24 (2019).
22. Ratiu, I. A., Ligor, T., Bocos-Bintintan, V. & Buszewski, B. Mass spectrometric techniques for the analysis of volatile organic compounds emitted from bacteria. *Bioanalysis*, **9**, 1069–1092 (2017).
23. Monedeiro, F., Monedeiro-Milanowski, M., Ligor, T. & Buszewski, B. A Review of GC-Based Analysis of Non-Invasive Biomarkers of Colorectal Cancer and Related Pathways. *J. Clin. Med*, **9**, 3191 (2020).
24. Bond, A. *et al.* Volatile organic compounds emitted from faeces as a biomarker for colorectal cancer. *Aliment. Pharmacol. Ther*, **49**, 1005–1012 (2019).
25. Janfaza, S., Khorsand, B., Nikkhah, M. & Zahiri, J. Digging deeper into volatile organic compounds associated with cancer. *Biol. Methods Protoc*, **4**, 1–11 (2019).
26. Hanna, G. B., Boshier, P. R., Markar, S. R. & Romano, A. Accuracy and Methodologic Challenges of Volatile Organic Compound-Based Exhaled Breath Tests for Cancer Diagnosis: A Systematic Review and Meta-analysis. *JAMA Oncol*, **5**, 1–11 (2019).
27. Biswas, S. *et al.* A Practive Faeces Collection Protocol for Multidisciplinary Research in Wildlife Science. *Curr. Sci*, **116**, 1878 (2019).
28. Hogue, S. R., Gomez, M. F., da Silva, W. V. & Pierce, C. M. A Customized At-Home Stool Collection Protocol for Use in Microbiome Studies Conducted in Cancer Patient Populations. *Microb. Ecol*, **78**, 1030–1034 (2019).
29. Wu, W. K. *et al.* Optimization of fecal sample processing for microbiome study – The journey from bathroom to bench. *J. Formos. Med. Assoc*, **118**, 545–555 (2019).
30. Ahmed, I. An Efficient Method Development for Extraction of Faecal Volatile Organic Compounds for GC-MS Analysis. *Biomed. J. Sci. Tech. Res*, **22**, 16596–16603 (2019).
31. Schripp, T. *et al.* A microscale device for measuring emissions from materials for indoor use. *Anal. Bioanal. Chem*, **387**, 1907–1919 (2007).
32. Micro-Chamber/Thermal Extractor. *Markes International* vol. QUI-1022 1–16  
file:///C:/Users/UMK/Downloads/qui\_1022\_u\_cte\_operators\_manual\_and\_addendum.pdf (2011).
33. Al-Suod, H. *et al.* Supercritical fluid extraction in isolation of cyclitols and sugars from chamomile flowers. *J. Sep. Sci*, **42**, 3243–3252 (2019).
34. Al-Suod, H., Ratiu, I. A., Górecki, R. & Buszewski, B. Pressurized liquid extraction of cyclitols and sugars: optimization of extraction parameters and selective separation. *J. Sep. Sci*, **42**, 1265–1272 (2019).
35. Hsu, Y. T. & Su, C. S. Application of box–behnken design to investigate the effect of process parameters on the microparticle production of ethenzamide through the rapid expansion of the supercritical solutions process. *Pharmaceutics* **12**, (2020).

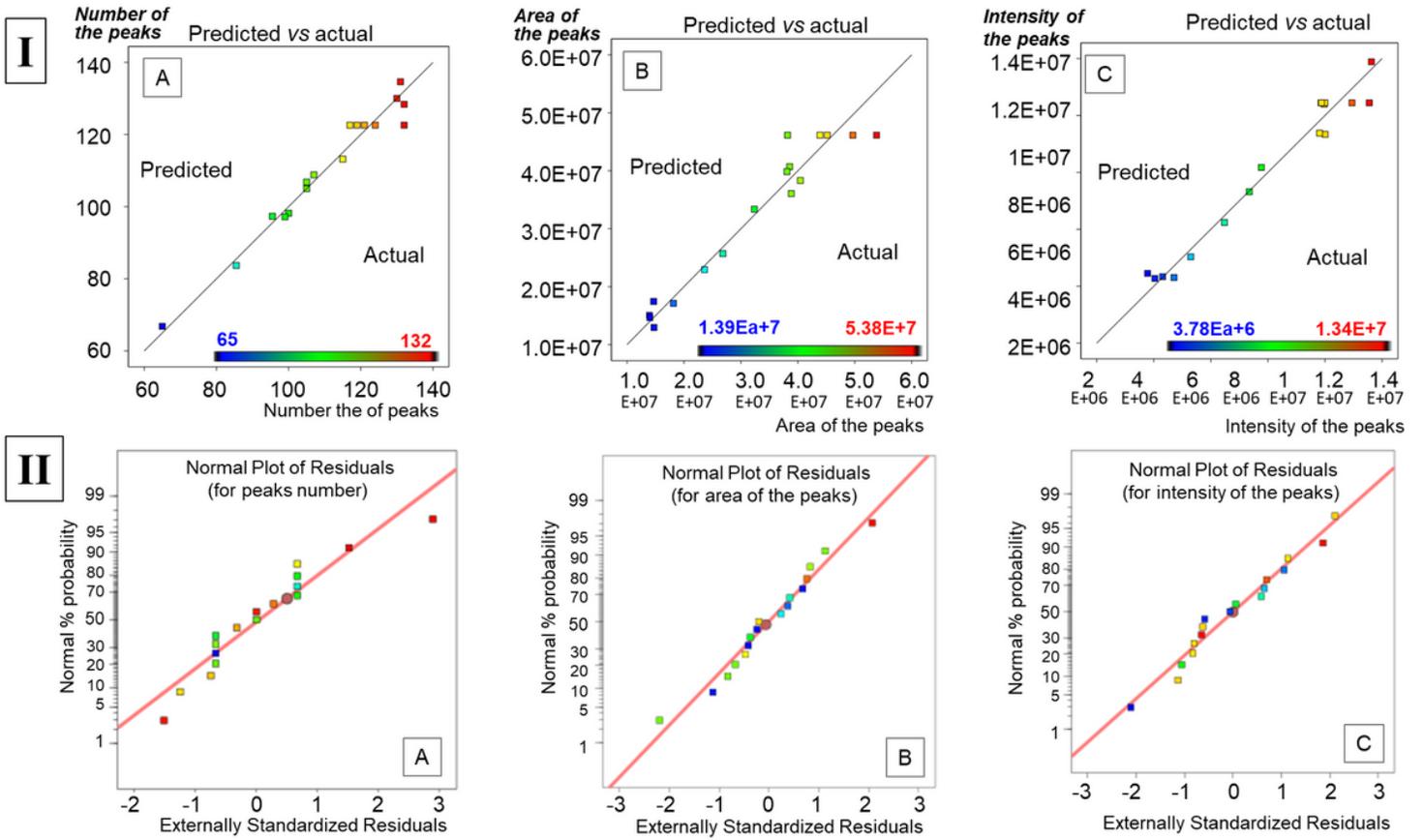
- 36. Liu, J., Wang, J., Leung, C. & Gao, F. A multi-parameter optimization model for the evaluation of shale gas recovery enhancement. *Energies* 11(2018).
- 37. Gratton, J. *et al.* Optimized Sample Handling Strategy for Metabolic Profiling of Human Feces. *Anal. Chem*, **88**, 4661–4668 (2016).
- 38. Tedjo, D. I. *et al.* The effect of sampling and storage on the fecal microbiota composition in healthy and diseased subjects. *PLoS One* **10**, (2015).
- 39. Ratiu, I. A., Al-Suod, H., Bukowska, M., Ligor, M. & Buszewski, B. Correlation study of honey regarding their physicochemical properties and sugars and cyclitols content. *Molecules* **25**, (2020).

## Figures



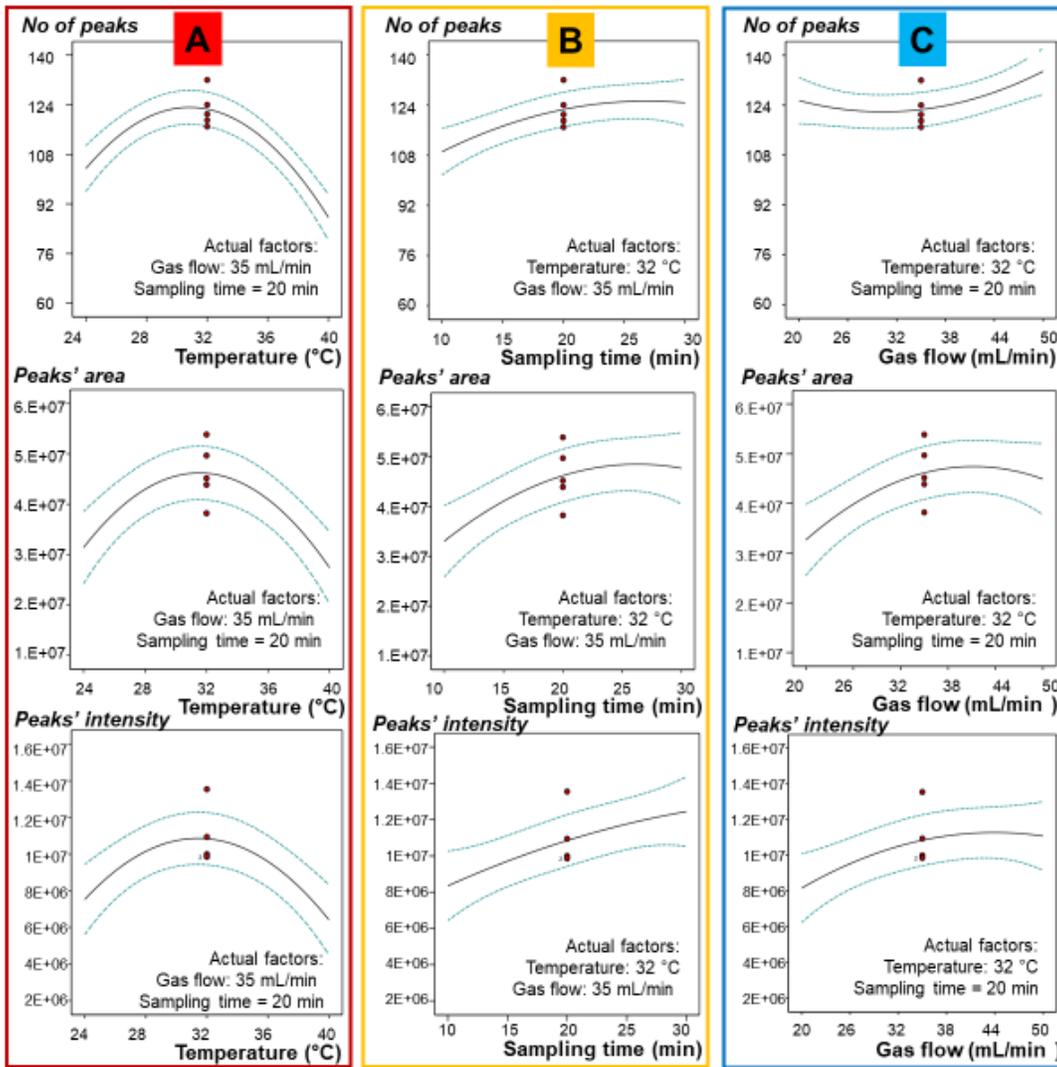
**Figure 1**

Micro-Chamber/Thermal Extractor presentation, where: A - Photo presenting μ-CTE unit and accessories & B - Schematic showing sampling procedure.



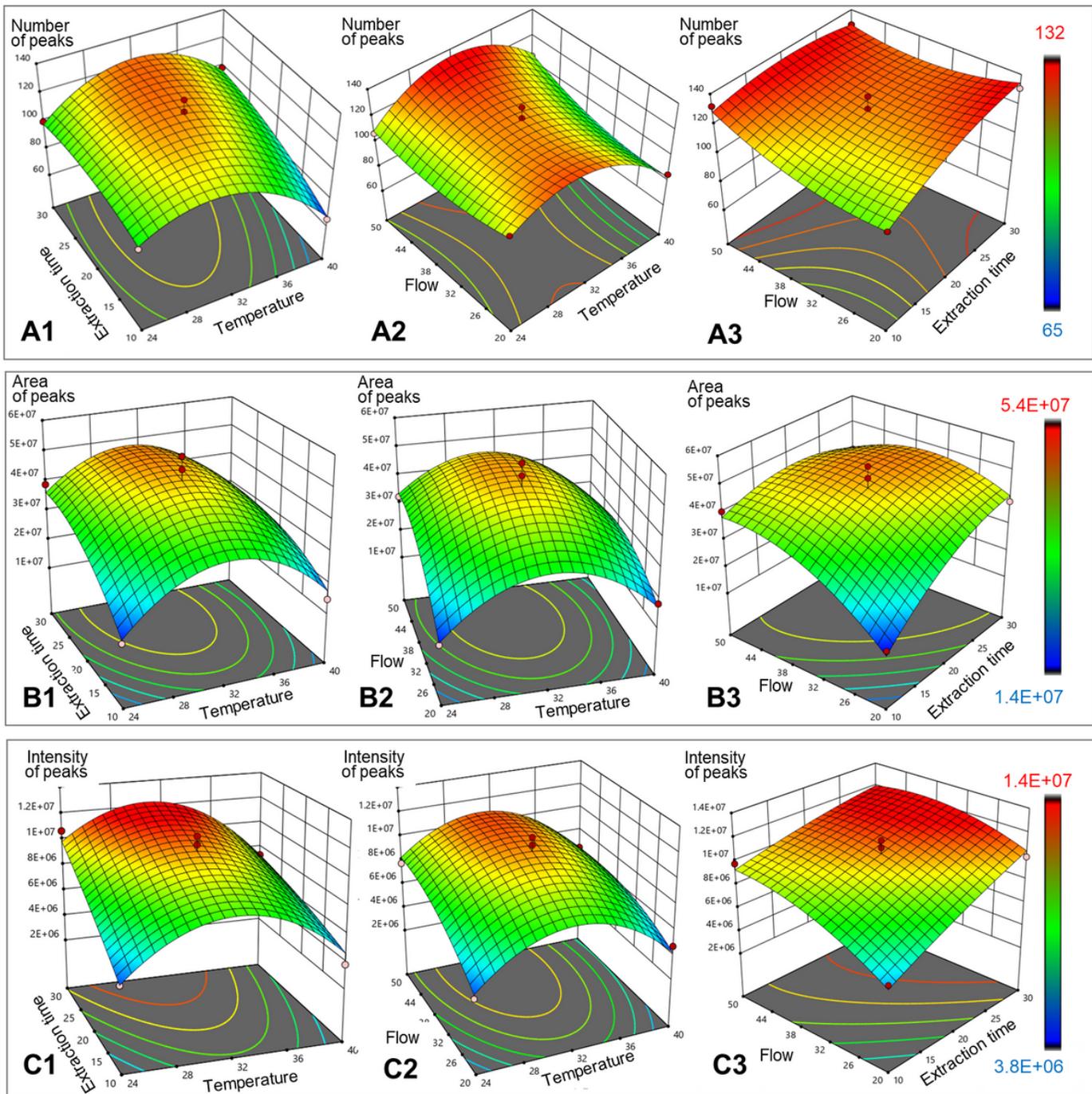
**Figure 2**

Evaluation of the adequacy of the three investigated models (created for: number of the peaks (A), area of the peaks (B) and intensity of the peaks (C)), where Part I: Actual vs predicted values & Part II: Normal plot of residuals at different production period



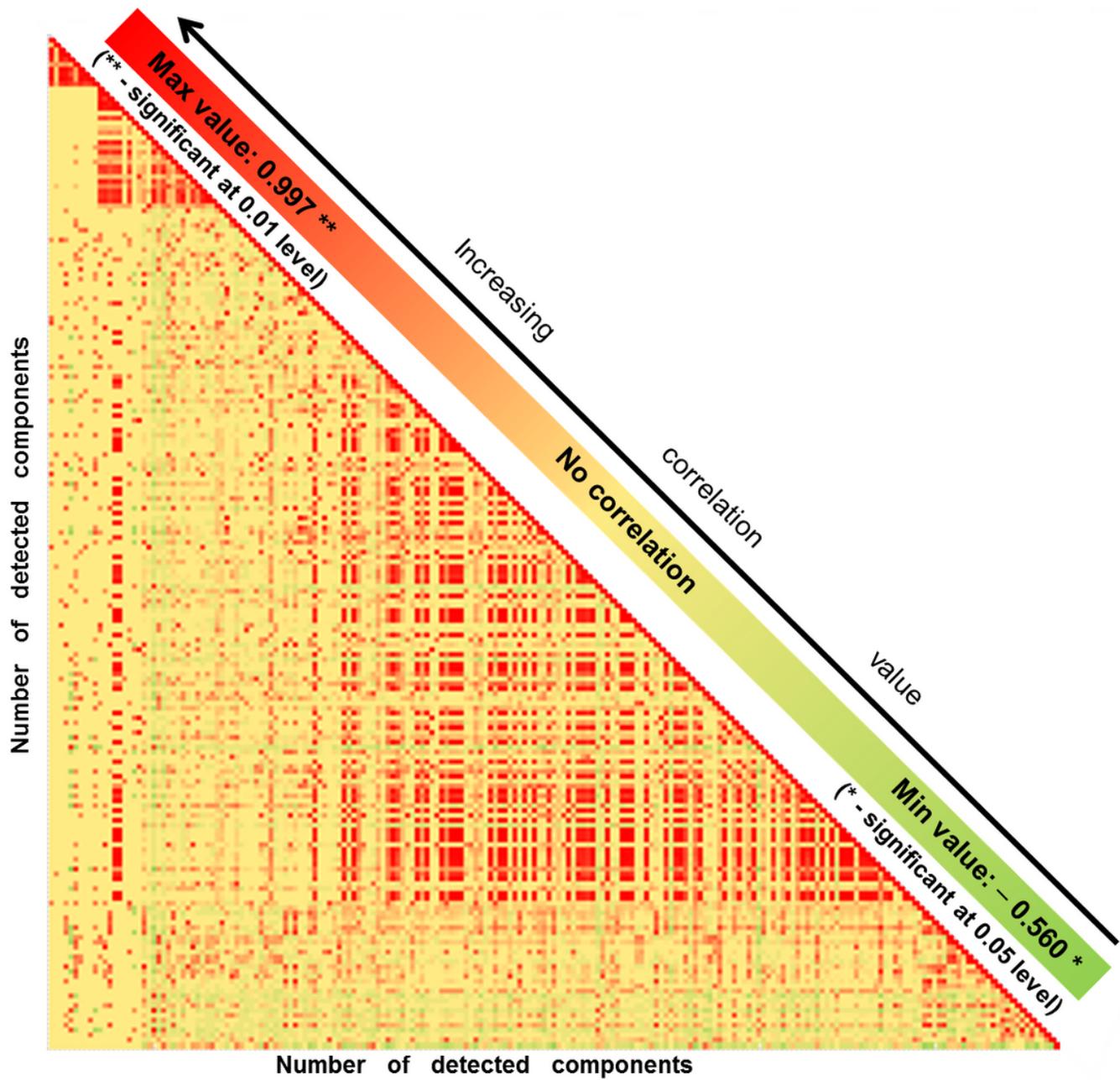
**Figure 3**

The influence of investigated factors on the obtained responses, where: A - influence of temperature, B influence of sampling time and C – influence of gas flow)



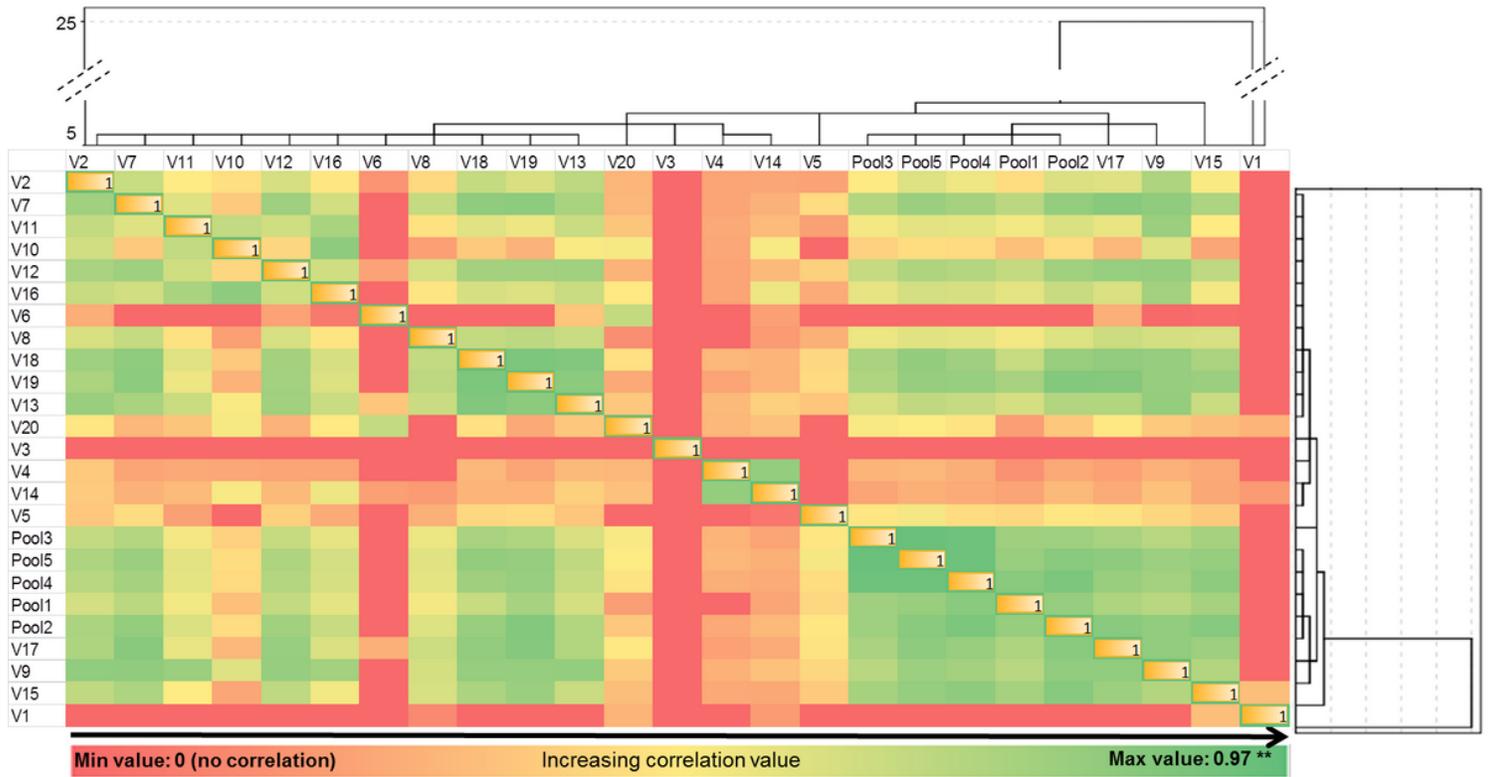
**Figure 4**

3D surface response for interaction effect of two-parameters according with: A – number of the peaks (A1 – extraction time vs temperature, A2 – flow vs temperature, A3 – flow vs extraction time), B – area of the peaks (B1 – extraction time vs temperature, B2 – flow vs temperature, B3 – flow vs extraction time), C – intensity of the peaks (C1 – extraction time vs temperature, C2 – flow vs temperature, C3 – flow vs extraction time)



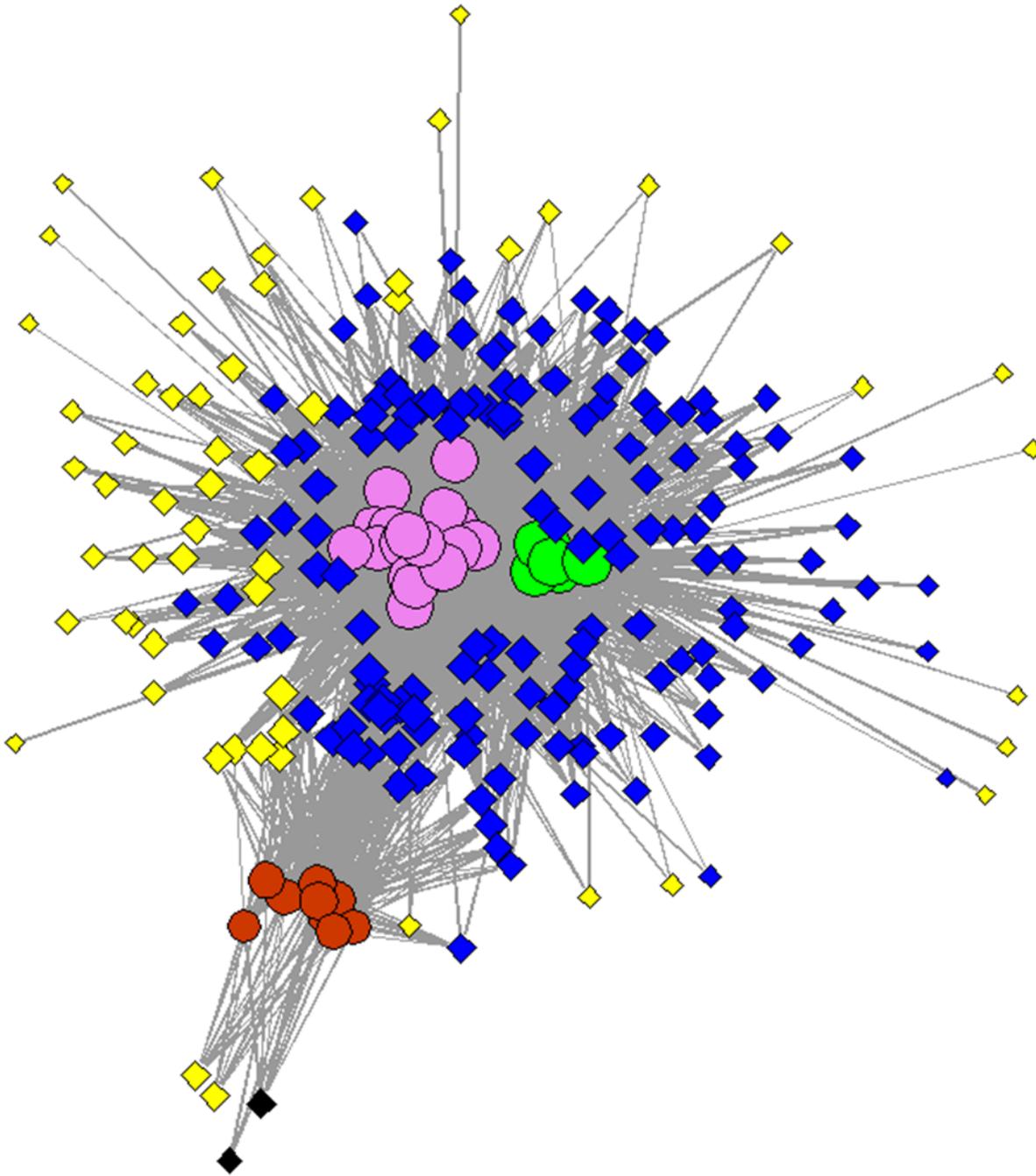
**Figure 5**

Correlation matrix presenting the significance of the 205 detected VOCs, highlighted in the form of a heat map whose shades represent the correlation values (green – negative correlation, yellow – no correlation, red – positive correlation).



**Figure 6**

Heat map representing the correlation between the investigated variables according with hierarchical clusters analyses



**Figure 7**

Network analyses representing separation between the three investigated groups based on the recorder VOCs profiles coming from pool samples (green circles), included volunteers (purple circles) and blank tubes (red circles). The blue diamonds are representing the components common between pools and volunteers, the yellow diamonds are highlighting the VOCs appearing in volunteers only, while the black ones were detected in blanks only.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS1.docx](#)

- [TableS2Matrixofdetectedmolecularfeatures.xlsx](#)
- [TableS3.docx](#)