

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

## Cocoa smoky off-flavour: A MS-based analytical decision maker for routine controls

### This is the author's manuscript

*Original Citation:*

*Availability:*

This version is available <http://hdl.handle.net/2318/1753103> since 2024-04-04T08:57:30Z

*Published version:*

DOI:10.1016/j.foodchem.2020.127691

*Terms of use:*

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

1 **Cocoa smoky off-flavour: a MS-based analytical decision maker for routine**  
2 **controls**

3 Camilla Scavarda<sup>1</sup>, Chiara Cordero<sup>1</sup>, Giulia Strocchi<sup>1</sup>, Cristian Bortolini<sup>2</sup>, Carlo Bicchi<sup>1</sup>, Erica  
4 Liberto\*<sup>1</sup>

5

6 <sup>1</sup> Dipartimento di Scienza e Tecnologia del Farmaco, Università degli Studi di Torino, Turin, Italy

7 <sup>2</sup> Soremartec Italia S.r.l., P.le P. Ferrero 1, 12051 Alba (CN), Italy

8

**10 Abstract**

11 Cocoa smoky off-flavour is generated from an inappropriate artificial drying applied on beans to  
12 speeding up the post-harvest process and it can affect the quality of the chocolate. The sensory  
13 tests are time-consuming, and at present, a fast analytical method to detect this defect in raw  
14 materials is not yet available. This study applies a HS-SPME-MS-enose in combination with  
15 chemometrics to obtain diagnostic mass-spectral patterns to detect smoked samples and/or as  
16 analytical decision maker. SIMCA models provide the best classification results, compared to PLS-  
17 DA, with sensitivities exceeding 90% and a high class specificity range of 89-100% depending on  
18 the matrix investigated (beans or liquors). Resulting diagnostic ions were related to phenolic  
19 derivatives. The discrimination ability of the method has been confirmed by a quantitative  
20 analysis through HS-SPME-GC-MS. HS-SPME-MS-enose turned out to be a fast, cost-effective and  
21 objective approach for high throughput analytical screening to discard defective cocoa samples.

22

23

24 **Keywords:** cocoa volatiles; smoky off-flavour; phenolic derivatives; HS-SPME-MS-enose;  
25 chemometrics; HS-SPME-GC-MS

## 26 1. INTRODUCTION

27 The quality of food is consumer-oriented and characterised by four closely interrelated  
28 dimensions: hedonic profile, healthy characteristics, economic convenience and the production  
29 process. Differences in quality assessment may have unfavourable consequences and lead to a  
30 product not fulfilling the expectations of consumers. Flavour is a key food quality attribute (e.g.  
31 safety and wholesomeness, physical characteristics, colour etc.) in cocoa-derived products. The  
32 flavour of chocolate heavily influences the pleasure derived from its consumption and evokes  
33 emotions, while also positively influencing the consumer's mood (Macht & Mueller, 2007;  
34 Spence, 2017; Wagner, Ahlstrom, Redden, Vickers, & Mann, 2014). The evaluation of flavour and  
35 its quality includes an appraisal of cocoa/chocolate intensity, particular flavour notes and the  
36 absence of flavour defects. Quality is dependent on how food products are handled at every  
37 touch point throughout the cocoa chain. Climate change, together with the global market  
38 pressure in response to growing demand (Eghbal, 2018) influence post-harvest processing and  
39 therefore affect the final flavour quality of the beans. The smoky off-flavour, in particular, can be  
40 generated from the inappropriate or poorly controlled artificial drying of the beans, which is  
41 performed to speed up moisture reduction (CABISCO/ECA/FCC, 2015; Perotti *et al.*, 2020; Serra  
42 Bonvehí & Ventura Coll, 1998). The smoky off-flavour also strongly affects the quality of finished  
43 chocolate (chocolate or confectionary) and cannot be eliminated during processing. This is  
44 predominantly a problem for cocoa beans from Cameroon, in West-Africa, where cocoa  
45 production originates from several small family farms, which must increasingly combat  
46 unfavourable climate change to make a profit (Eghbal, 2018; Statista, 2020; Wessel & Quist-  
47 Wessel, 2015). Some phenolic compounds, predominantly derived from lignin degradation by  
48 pyrolysis, have been related to the smoky note (Janairo & Amalin, 2018; Serra Bonvehí, 1998;  
49 Wang, Chambers, & Kan, 2018; Perotti *et al.*, 2020). Of the volatiles that make up the smoky note,

50 some components, such as guaiacols and methylphenols, have also been found to be cocoa key-  
51 aroma compounds, but their presence at high concentrations can negatively affect cocoa's  
52 sensory properties (Frauendorfer & Schieberle, 2006). In a previous article Perotti *et al.* (2020)  
53 chemically characterised the smoky off-flavour of cocoa beans and liquors using a metabolomic  
54 approach and a top-down strategy; HS-SPME was coupled with comprehensive GC equipped with  
55 a time of flight mass spectrometer (HS-SPME-GC×GC-TOF-MS) for use as a screening platform to  
56 identify informative odorants within a set of samples characterized as smoky and non-smoky by  
57 an industrial sensory panel (Perotti *et al.*, 2020). 1D-HS-SPME-GC-MS, which is used in a fully  
58 automated set-up coupled to both supervised and unsupervised chemometrics, was then used  
59 to monitor targeted discriminating compounds. Targeted smoky volatiles were then quantified  
60 using multiple headspace extraction (MHE) in cocoa beans, and a quantitative range of  
61 acceptability/rejection for the incoming cocoa beans was defined.

62 Despite the good overall performance of HS-SPME-GC-MS, the main drawback for quality control  
63 is still the long analysis time, which is unfavourable for high throughput data collection. Fast  
64 analytical methods that are based on direct injection and mass spectrometry (MS-based  
65 electronic nose or MS-enose) may be a solution for the rapid discrimination of smoky from non-  
66 smoky samples (Biasioli, 2016; Biasioli, Yeretjian, Märk, Dewulf, & Van Langenhove, 2011;  
67 Deuscher *et al.*, 2019; Liberto *et al.*, 2013; Liberto *et al.*, 2019). Unlike traditional electronic noses  
68 that are based on solid-state sensors, an MS-enose uses m/z ratios as chemical sensors and is  
69 more robust and reproducible, while also being unaffected by sample moisture (Loutfi,  
70 Coradeschi, Mani, Shankar, & Rayappan, 2015). The lack of specificity of an MS-enose, when  
71 compared to systems that include chromatographic separation, is compensated by its  
72 effectiveness, fast response, non-invasive operations and adequate sensitivity even without  
73 sample pre-treatment. On the other hand, the mass spectral fingerprints that are generated have

74 to be diagnostic of the sampled volatile fractions, and must provide characterising fragments  
75 with abundances suitable to discriminate and modelling the categories of the investigated  
76 samples (e.g. smoky vs non-smoky).

77 This study is therefore a step ahead of the top-down approach previously mentioned (Perotti *et*  
78 *al.*, 2020) to meet industry needs for speeding up analytical controls for primary materials  
79 acceptance, for industrial chocolate manufacturing, while supporting rejections with objective  
80 measurements and reliable data.

81 With this contribution, we would verify the following two hypotheses:

- 82 1. whether the contribute of spectral masses of the smoky volatiles, within a rich complex mass  
83 spectral fingerprint, is sufficiently diagnostic to be able to discriminate samples
- 84 2. whether this analytical approach, in combination with a discriminant mathematical model,  
85 can be exploited as an analytical decision maker (ADM) in a first screening control.

86 This study evaluates whether a HS-SPME-MS-enose is able to discriminate smoky from non-  
87 smoky cocoa in a significant set of samples. Our driving hypothesis is that, if diagnostic ions from  
88 smoky volatiles are produced at a suitable intensity, the proposed analytical approach may be a  
89 suitable ADM for cocoa batches received at the factory.

90

91

## 92 2. MATERIALS AND METHODS

### 93 2.1 Samples:

94 The sample set included 48 bean samples (n=23 smoky/rejected and n=25 non-smoky) and 176  
95 cocoa liquors (n= 42 smoky/rejected and n= 134 non-smoky) (*Theobroma cacao* L. main crop).  
96 Rejected and accepted samples here analysed have been selected based on the sensory tests. All  
97 samples were of commercial grade and compliant with the industrial quality control of  
98 Soremartec Italia srl (Alba, Italy). Cocoa-bean quality that was in agreement with FCC rules  
99 (Federation of Cocoa Commerce) and ISO 2451 was classified “standard beans”(FCC; ISO). Beans  
100 and liquors were directly sampled at the processing plants.

101

### 102 2.2 Head Space Solid Phase Micro Extraction sampling:

103 Volatiles were sampled using an automatic HS-SPME system installed on an MPS-2 multipurpose  
104 sampler controlled by Gerstel Maestro software (Gerstel, Mülheim a/d Ruhr, Germany), which  
105 was combined on-line with an Agilent 7890A GC coupled to a 5975B MS detector (Agilent, Little  
106 Falls, DE, USA).

107 Cocoa samples were ground in liquid nitrogen to give a homogeneous powder and then stored  
108 at -80°C until analysis. Cocoa powder (1.00 g) was weighed in headspace glass vials (20 mL),  
109 equilibrated for 5 min at 80°C and then sampled using HS-SPME for 10 min at 80 °C at a stirring  
110 speed of 350 rpm. **Sampling conditions and their optimization for smoky volatiles extraction were  
111 in agreement with the experimental results reported in a previous article (Perotti *et al.*, 2020).**

112 SPME fiber: Divinylbenzene/Carboxen/Polydimethylsiloxane (DVB/CAR/PDMS) df 50/30 µm - 2  
113 cm length from Merck (Bellefonte, PA, USA). Fibers were conditioned before use as  
114 recommended by the manufacturer. After sampling, the recovered analytes were thermally

115 desorbed, by heating the fibre for 3 min at 250 °C, into the GC injector body from where they  
116 were transferred on-line to the transfer capillary (Cordero *et al.*, 2019; Magnagna *et al.*, 2017,  
117 Magagna *et al.*,2018). All samples were analysed in duplicate.

118

### 119 2.3 MS-enose instrument set-up:

120 The GC oven and injector were maintained at 250 °C; injection mode, split; split ratio, 1/10;  
121 carrier gas, helium; flow rate, 0.4 mL/min; fibre desorption time and reconditioning, 3 min. The  
122 transfer column was uncoated deactivated fused silica tubing (dc = 0.10 mm, length = 6.70 m)  
123 from MEGA (Legnano, Italy).

124 MSD Conditions: ionisation, EI mode at 70 eV; temperatures: ion source: 230 °C, transfer line:  
125 280°C. Standard tuning was used and the scan range was set at m/z 35–350 with a scanning rate  
126 of 1,000 amu/s.

127

### 128 2.4 Data acquisition and elaboration

129 Data were acquired and processed using an Agilent MSD Chem Station ver. E.02.01.1177 (Agilent,  
130 Little Falls, DE, USA). Raw data were transformed using RapidDataInterpretation software by  
131 Gerstel (Gerstel, Mülheim a/d Ruhr, Germany). This is a post-run macro derived from the Gerstel  
132 Chemsensor add-on tool, which expands the scope of function of the Agilent ChemStation  
133 software. This step allows the 3-dimensional raw data supplied by mass spectrometry (retention  
134 time, m/z fragmentation and intensities) to be reduced to 2-dimensional data that can then be  
135 properly used by statistical software for further elaboration. Data obtained can be in the form of  
136 a \*.TIC file and/or a \*.DAT file (Mass Spectral Fingerprint). In both cases, sample intensities are  
137 summed; with the TIC method, as a function of the scans, and with the DAT method, as a function  
138 of the masses. The influence of retention time is removed from the data by summing the  
139 intensities of m/z value over the complete time period of the run. The intensities of a sample are



140 added as a function of the masses. The mass axis is very stable when compared to the retention  
141 time. The result is a diagram in which an intensity is assigned exactly to each mass (**Figure 1 SM**  
142 in supplementary material). When several samples are analysed, \*.DAT-file processing collects  
143 data of the whole sample set and provides a data matrix in which the rows report the samples  
144 and the columns report the intensity assigned to each mass fragment (Heiden *et al.*, 2002). Data  
145 matrices of 15,120 data points for beans (48 objects x 315 variables) and 110,880 data points for  
146 liquors (352 objects x 315 variables) were collected and elaborated further.

147 The \*.DAT file was used for data interpretation in this work. Raw data underwent pre-treatment  
148 that consisted of noise subtraction and internal normalisation of the signal from each sample  
149 versus the most intense ions ( $m/z$  43, basic peak taken as 1); they were subsequently pre-  
150 processed.

151 Chemometric analyses, Principal component analysis (PCA), SIMCA (Soft-Independent modelling  
152 of Class Analogy) and Partial Least Square Discriminant Analysis (PLS-DA) were carried out using  
153 Pirouette® (Comprehensive Chemometrics Modelling Software, version 4.5-2014) (Infometrix,  
154 Inc. Bothell, WA).

155

156

### 157 **3. RESULTS AND DISCUSSION**

158 The challenge in metabolomic studies is to simultaneously analyse a high number of metabolites  
159 that are of very different size, molecular weight, polarity and stability. Cocoa-bean volatiles are  
160 the final stage of the expression of changes in the metabolome that is the result of the production  
161 chain (Biasoli, 2016; Cevallos-Cevallos, Reyes-De-Corcuera, Etxeberria, Danyluk, & Rodrick, 2009;  
162 Ellis, 2019). They are influenced by external factors, e.g. climate and environmental conditions,

163 ripening, time of harvest, fermentation and post-harvest processing (and roasting for liquors).  
164 The physical and chemical interactions of all the compounds present in the volatile fraction  
165 therefore result in the final attributes of the product, which include sensory acceptability, quality,  
166 safety and shelf life (Acierno, Yener, Alewijn, Biasioli, & Van Ruth, 2016; Charles *et al.*, 2015;  
167 Liberto *et al.*, 2019). The identification of the volatiles and the interactions that are responsible  
168 for the characteristics of a food therefore facilitates the improved control and understanding of  
169 the food processes and systems that influence final-product characteristics (Gloess *et al.*, 2014;  
170 Liberto *et al.*, 2013; Lindinger *et al.*, 2008). However, chromatographic profiling and  
171 fingerprinting cannot be performed quickly enough to meet the requirements of a routine  
172 control system for the acceptability of incoming raw material at the production plant. This issue  
173 can be addressed by the direct injection of the sampled headspace into a mass spectrometer,  
174 which is here based on a quadrupole mass analyser (MS-enose).

175 As previously reported by Perotti *et al.* (Perotti *et al.*, 2020), in cocoa, smoky volatiles are, in  
176 general, present at trace levels, and diagnostic fragments are mandatory if their presence is to  
177 be confirmed; possible interference from isobaric ions that derive from components present in  
178 higher amounts may hamper the correct discrimination of smoky from non-smoky samples  
179 (hypothesis-driven approach). If their differentiation is still possible, other volatiles that are  
180 correlated to the smoky components must then be considered (hypothesis-generating  
181 approach). This last hypothesis is challenging and would require an in-depth investigation into  
182 the relationships between these components/ions, the smoky volatiles and the chemical  
183 fragmentation involved.

184 An explorative investigation by PCA was first carried out on bean and liquor samples to better  
185 understand the real potential, and possible limits, of the informative power of MS-enose  
186 fingerprinting in the description of samples, with or without the smoky defect.

187 *3.1 Explorative unsupervised data analysis on beans and liquors*

188 PCA results show fairly good separation between smoky (in red) and non-smoky (green) bean  
189 samples with an explained variance of 63.9% (**Figure 1a**). This is acceptable discrimination if we  
190 consider that the sample set was representative of different harvesting years (2017 and 2018),  
191 (evidenced along the first PC1) and of different harvest regions (Cameroon and Ecuador). In this  
192 case, the functional variables (harvest year and region) for the samples were kept to verify MS-  
193 enose fingerprinting's ability to discriminate the smoky defect even in presence of confounding  
194 external factors. The loadings plot (**Figure 1b**) shows several ions that are linked to the smoky  
195 characteristic, in particular on PC2, together with the harvest time on the first PC. On the other  
196 hand, the PCA on the whole data set of liquors (n=176) shows lower informative potential, in  
197 grouping, on the first two PCs, but demonstrates fairly good discrimination between smoky and  
198 non-smoky samples when further PCs are considered (**Figure 1c**). A possible explanation is that  
199 the roasting and the grinding processes heavily influence the presence of these components,  
200 either because of the effect of temperature and pressure or because of the physical changes in  
201 the matrix that are caused by the melting of the cocoa butter to give the cocoa pastes. However,  
202 the ions that describe the smoky clusters are the same regardless of whether beans or liquors  
203 are considered (**Figure 1b and 1d**).

204 The clusters shown in PCA suggest that supervised approaches should be adopted to define  
205 models for sample classification. In the next sections, two supervised approaches for the  
206 classification of bean and liquor samples, PLS-DA and SIMCA, will be considered.

207 *3.2 Supervised data analysis on beans and liquors*

208 PLS-DA is a discriminant classification that uses regression that is constructed between X, used  
209 as the predictor matrix the (m/z) variables, and the response Y, which indicates the category (ki);  
210 in this case,  $K=1$  for smoky and  $K=2$  for non-smoky beans. The discrimination rule is based on the

211 comparison of each row of the predicted matrix  $\hat{Y}$  with each pattern response vector. A sample,  
212  $i$ , is categorised into the class,  $k$ , whose pattern is the closest match. The matrix  $\hat{Y}$  is treated as  
213 the input data set for classification to evaluate the distance between a sample and a class pattern.  
214 However, this cannot be done directly because this matrix has a rank of  $K-1$ , and the  
215 corresponding covariance matrix is singular. Matrix  $\hat{Y}$  is decomposed using PCA, which reduces  
216 its dimension to  $K-1$ , to resolve this singularity. The scores matrix  $T$  represents a new data set to  
217 which a classification method can be applied.

218 SIMCA is considered soft class modelling because there are no hypotheses for the distribution of  
219 variables and their independence because each category model is developed independently and  
220 no information from the other categories is used. The mathematical model of each class is based  
221 on the principal components of the category, generally obtained as eigenvectors of the  
222 correlation coefficient matrix of the category. For a given class, the model dimensions are  
223 described either by a line (for one PC), by a plane (for two PCs) or by a hyper-plane (for more  
224 than two PCs). The range of scores onto such significant PCs defines the class space. New samples  
225 are projected in each PC space, which describes a specific class, and the F-test is used to evaluate  
226 the Euclidean distances of the objects from the model.

227 Class modelling differs from discriminant classification mainly because the focus is on a single  
228 category; the modelling approach characterises the class of interest against all the others. Class  
229 modelling may give a composite answer: 1) compliance with the model for one category only; 2)  
230 multiple compliance, with models of several categories; 3) non-classification: a new sample may  
231 be rejected by the models of all the categories under study.

232

### 233 3.2.1 Bean classification

234 To develop the classification model, the sample set was divided into a training set (n=41) and an  
235 external test set (n=24). PLS-DA results are reported in **Figure 2a and b**, which displays the scores  
236 and loadings plots of the training-set sample classification. The ion fragments with higher  
237 discriminant power for smoky beans are highlighted in yellow. These (m/z)s were found to be  
238 characteristic for several phenolic compounds and, in the current mode, they have a role in the  
239 description of the smoky-bean flavour. The classification ability for the calibration of the PLS-DA  
240 cross-validated model (CV=5) is 95.1%, as can be extrapolated from the confusion matrix in **Figure**  
241 **3**. Despite the good results on the training set, the model shows a correct prediction classification  
242 rate of only 83.3% when applied to the external test set, which was not used to train the model.  
243 However, the prediction displays good specificity for both classes and a sensitivity of 80% for the  
244 classification of the smoky flavour. This means that the model has quite a good ability to predict  
245 class for unknown samples with high specificity for the smoky flavour, but a lower specificity for  
246 non-smoky flavours, which can generate false negatives, i.e. indicating good samples when they  
247 are not. This must be taken into consideration as a risk for subsequent processing.

248 SIMCA modelling results are displayed in **Figure 4a**, which shows that the two groups are well  
249 recognizable. **Figure 4b** shows the class distance between smoky (CS1) and non-smoky (CS2)  
250 samples; Cooman's plot displays good separation between classes without overlapping, with the  
251 exception of sample "NSBCm2\_2" (i.e., non-smoky beans from Cameroon). The variables (i.e.,  
252 m/z ions) that have little or high importance for any class in the training set are shown in the  
253 modelling power in **Figure 4c**. Variables close to 1 have a high impact on the description of the  
254 training set. At the same time, it may be useful to know the best variables to classify the samples  
255 categories through the discriminant power. In particular, for each variable, it compares the

256 average residual variance of each class fit to all other classes, and the residual variance of all  
257 classes fit to themselves, thus providing an indication of how a variable discriminates between a  
258 “correct” and “incorrect” classification. variable value close to 0 indicates low discrimination  
259 ability, while a value that is much larger than 1 implies high discrimination power. **Figure 4d**  
260 shows the ion fragments with the highest discriminant power in the two sample categories. Of  
261 the high modelling and discriminant variables, some fragments ((m/z)) that are characteristic of  
262 several phenolic compounds have a role in the description of the smoky beans flavour. These  
263 compounds have also been identified as being responsible for smoky-hammy notes in other food  
264 matrices (Aprotosoaie, Vlad Luca, & Miron, 2016; CABISCO/ECA/FCC, 2015; Petričević, Marušić  
265 Radovčić, Lukić, Listeš, & Medić, 2018; Ridgway, Lalljie, & Smith, 2010; Serra Bonvehí & Ventura  
266 Coll, 1998). In particular, m/z 107-108-109, 124 and 138 are diagnostic for isomers of methyl  
267 phenols (*cresol isomers*) and *guaiacol*, while m/z 125-152-154 are diagnostic for *p-ethyl guaiacol*,  
268 *4-methyl-2,6-dimethoxyphenol* and *2,6-dimethoxyphenol*, and m/z 128 for *naphthalene* (Perotti  
269 *et al.*, 2020). The smoky-compound-related ions account for 40%, on average, of the whole gas  
270 chromatographic profile acquired using 1D-GC-MS.

271 Classification ability in calibration, for the SIMCA model, was 100%, as shown in **Figure 3a**. The  
272 developed model was applied to the external test set (not used to train the model) and showed  
273 global prediction classification with a sensitivity of 91.6% and a specificity of 100%, as shown in  
274 **Figure 3b** by the confusion matrix. In addition, the class specificity is excellent and the sensitivity  
275 of the model for both classes is above 90%. This means that class modelling is robust, without  
276 giving false-positive or false-negative sample classifications, and that only 8.40% of the samples  
277 should be investigated further by a sensory panel or by a confirmatory method to verify their  
278 acceptability (Perotti *et al.*, 2020).

279

280 *3.2.2 Liquor classification*

281 Liquor classification also need to be investigated because primary materials may either be beans  
282 or liquors at the cocoa production plant, depending on the country of origin and supplier.

283 The liquor sample set was unbalanced as it contained a higher number of the non-smoky  
284 category. Both classification approaches, however, showed similar prediction abilities, 92% and  
285 97% for PLS-DA and SIMCA respectively (**Figure 3b** and **Figure 5b** and **5e**). SIMCA training and  
286 test-set prediction is depicted in the hyper-plane of the first 3 PCs in **Figure 5a-b**, while PLS-DA  
287 prediction on the test set is displayed by the prediction plots of **Figure 5d** and **5e**. In the prediction  
288 plot of **Figure 5d**, reference lines (in light green) indicate the decision criterion space for class  
289 membership; only samples with Y values greater than 0.5 are called category members.  
290 Therefore, the red samples in the plot are all above the membership criterion for CS2, that is,  
291 class 2 (i.e., non-smoky liquors), which is on the X axis, while the green samples fall above the  
292 decision criterion for class 1 (e.g., smoky liquors), on the Y axis. **Figure 5e** clearly shows the  
293 predicted y values and the samples that are classified outside their group in the blue circles.

294 Discriminant variables for SIMCA, as shown in **Figure 5c**, were found to be m/z 152 and 154,  
295 which represent *phenethyl alcohol* and *p-ethyl guaiacol* (m/z=152) at 96% and *2,6-*  
296 *dimethoxyphenol* (m/z=154) at 81% on the 1D-GC-MS chromatographic profile. The same  
297 variables and m/z 107, 123, 137, 138 are diagnostic for isomers of *phenol*, *methyl phenols* (*cresol*  
298 *isomers*) and *p-ethylguaiacol* **Figure 5f**, whose abundance in the 1D-GC-MS chromatographic  
299 pattern of liquors for the above-cited volatiles is 62% (**Figure 2 SM**) (Perotti *et al.*, 2020). The  
300 mass spectra of several smoky components present in commercial mass spectral libraries (i.e.  
301 Wiley7N and NIST2014) are displayed in **Figure 3 SM** in the supplementary file.

302 Despite the unbalanced number of samples in the classes, the two supervised approaches show  
303 similar classification/modelling performance. However, SIMCA modelling has higher specificity  
304 for both classes and the highest sensitivity for CS1 (i.e. smoky samples). This means that a model  
305 for discrimination between smoky and non-smoky liquors is feasible and that it may be exploited  
306 as a second filter through the chocolate processing chain, for instance. Discrimination between  
307 good and bad beans can be considered the first step for conformity to the qualitative standard  
308 in cocoa-bean acceptance, also for economic reasons, while the second filter can be used on  
309 liquors to reinforce the reduction of the impact of the off-flavour on the final product, while  
310 maintaining its standard quality.

311 Samples that are excluded from the model or unclearly classified can be submitted to a  
312 confirmatory validated method (Cordero *et al.*, 2019; Perotti *et al.*, 2020).

313

### 314 *3.3 Qualitative and quantitative confirmation analytical method*

315 A confirmatory method that is based on a reference standard material is not feasible due to the  
316 lack of a cocoa smoky reference standard. Furthermore, several smoky volatiles are also  
317 endogenous components in beans and even more so in liquors. Spiking methods, performed via  
318 the addition of standard reference compounds, would falsify the quantitation due to their non-  
319 homogeneous distribution in the cocoa (beans and liquors), because of the heterogeneous  
320 nature of the solid matrix. Therefore, the sensory-driven screening method applied above has  
321 been cross-validated using a confirmatory qualitative-quantitative analysis; multiple  
322 headspace extraction (MHS-SPME) combined with 1D-GC-MS (Perotti *et al.*, 2020; Sgorbini *et al.*,  
323 2019). MHS-SPME-1D-GC-MS was validated in terms of its repeatability (intra-day repeatability)  
324 and intermediate precision (inter-day repeatability) for beans and liquors. The ISTD-normalised



325 analyte area responses of a set of 15 compounds were then processed using one-way ANOVA  
326 and Tukeys' comparative analysis on the data acquired from a quality-control smoky sample of  
327 both beans and liquors (QCs), on four different days in different weeks, analysed in triplicate.  
328 **Table 1** displays the intra-day (repeatability) and inter-day precision for the quantified smoky  
329 markers. Precision is expressed as RSD% on analyte normalized area responses. Results indicate  
330 good intra/inter-day precision for both beans and liquors. The Limit of Detection (LOD) was  
331 determined from the standard calibration curve, as  $LOD=3*Sa/b$ , where Sa is the standard  
332 deviation of the response and b is the slope of the calibration curve, while the Limit of  
333 Quantification LOQ was calculated as  $3.3*LOD$ . The LOD and LOQ of the smoky markers are also  
334 reported in **Table 1** together with the odour qualities, odour thresholds, target and qualifier ions  
335 that were used for their quantitation and qualification. The accuracy of the screening method  
336 was evaluated by cross-comparison with the quantitative results obtained using MHE-SPME-1D-  
337 GC-MS. **Figures 4 and 5 SM** display box plots of the quantitative results on non-matched or false  
338 positive samples (doubtful) that were obtained by class modelling. The quantitation of the  
339 selected markers allows an operative limit, of below 10 ng/g for beans and 100 ng/g for liquors,  
340 to be adopted for the acceptance of incoming cocoa samples, **Figure 6 SM** and **Table 1**.

341

#### 342 **4. CONCLUSION**

343 The HS-SPME-MS-enose, in combination with chemometric tools, delineate a successful work-  
344 flow for the detection of defective "smoky" samples (beans or liquors) and their discrimination  
345 from "non-smoky" ones, thus confirming our driving hypothesis. The validation of the models  
346 was performed at two levels: via internal validation (cross-validation), using a training set, and  
347 via an external validation test set. Diagnostic fragments of phenolic derivatives correlated to  
348 these models enable to assess the classification ability of the MS-enose method through a cross-

349 verified evaluation of the actual concentration of smoky compounds using MHE-SPME-1D-GC-  
350 MS. Despite 1) the compositional complexity of the bean and liquor volatilomes, 2) the low  
351 concentration of the smoky targeted markers compared to the major volatiles, and 3) the  
352 possible co-contribution of several different analytes to the fragment intensities, the sensitivity  
353 of the technique's multi-channel nature is sufficiently diagnostic, making it possible to  
354 mathematically model the variation in mass spectral fingerprints using multivariate regression  
355 procedures. In particular, the high specificity of the SIMCA models indicates that there is a low  
356 probability of false positive/negative classifications, although this occurred to a lesser extent for  
357 the smoky class in liquors. These features mean that the MS-enose can be exploited as an  
358 analytical decision maker for screening controls of both beans and liquors. Possibly unclearly or  
359 non-classified samples can be reasonably verified using a conventional analytical confirmation  
360 method via the quantitation of the smoky components. Moreover, the analytical system is  
361 versatile since it can be used for both a conventional setting for GC-MS and in MS-enose mode  
362 and is therefore suitable for a high-throughput, objective and cost-effective quality control.

### 363 **Aknowledgments**

364 The research was financially supported by Soremartec Italia s.r.l Alba (CN), Italy

365

366 **Table caption**

367 **Table 1.** Average concentrations of targeted smoky markers, respectively for smoky and non-  
368 smoky samples, together with their odour quality (The Good Scents Company Information  
369 System, 2018), odour threshold (ng/g orthonasal from oily matrix) as reported in the literature  
370 (Vichi, Romero, Tous, Tamames, & Buxaderas, 2008); \*OT in water (Biotechnology National  
371 Center, 2020; Buckpitt, Alan, Kephelopoulos, Stylianos, Koistinen, Kimmo, Kotzias, Dimitrios,  
372 Morawska, Lidia and Sagunski, 2010); Target ion (Ti) and qualifiers (Q1-Q2), adopted for their  
373 quantitation and identification, intra- inter-day precision, LOD and LOQ.

374 **Figure Captions**

375 **Figure 1.** PCA score plots of a) beans and b) liquors and the corresponding loading plots c) and  
376 d). Data were logarithmically (Log10) transformed and pre-processed by autoscaling. Red spots  
377 indicate smoky samples, green spots are non-smoky.

378 **Figure 2.** a) and b) PLS-DA score plot and loading plot of beans using an internal cross-validated  
379 method (CV=5), c) classification prediction of the external test set of samples. Data were  
380 transformed by autoscaling. Red spots indicate smoky samples, green spots are non-smoky.

381 **Figure 3.** Confusion matrices of bean and liquor classification in calibration (a) and in prediction  
382 on the test set (b), respectively for SIMCA and PLS-DA, together with sensitivity, specificity and  
383 correct classification rate values. CS1: Class 1, smoky beans, CS2: Class 2, non-smoky beans.

384 **Figure 4.** a) SIMCA beans classification scores plot. Data were transformed using a logarithmic  
385 scale and autoscaled. Explained variance on the first 3 PCs: 66.9%. b) Interclass distances  
386 between smoky and non-smoky samples, c) variables that impact on the modelling and d)  
387 discriminant variables in class classification. Red spots indicate smoky samples, green spots are  
388 non-smoky.

389 **Figure 5.** a-c) Liquor modelling using the SIMCA training set, test set and the discriminant  
390 variables of the model. d-f) PLS-DA prediction plot, predicted members in the two classes and  
391 discriminant variables for the classification. CS1: Class 1 smoky beans, CS2: Class 2 non-smoky  
392 beans.

393

#### 394 **Supplementary material**

395 **Figure 1 SM.** Flow-chart from raw-data acquisition to the raw-data matrix for suitable  
396 chemometrics. Total ion chromatogram (TIC) of bean-volatile fingerprints represents the 3-D  
397 data that is transformed, by the rapid data interpretation software, into 2-D data the Mass  
398 spectral fingerprint (\*.DAT) and the subsequent data matrix obtained from several mass spectral  
399 fingerprints.

400 **Figure 2 SM.** a) Liquor HS-SPME-GC-MS patterns of the TIC (Total ion current), and b) MIC (Mixed  
401 Ion Chromatogram: m/z 107, 123, 137, 138, 152, 154) with the recognised volatiles that  
402 contained diagnostic ions, set in the MIC.

403 **Figure 3 SM.** Mass-spectra fragments of smoky components from commercial mass spectral  
404 libraries (Wiley7N and NIST14)

405 **Figure 4.SM.** Box plots of the quantified smoky markers in false positive smoky beans

406 **Figure 5 SM.** Box plots of the quantified smoky markers in false positive smoky liquors

407 **Figure 6 SM.** Quantitative rejection ranges of the smoky markers (brown colour) for a) beans and  
408 b) liquors

409

410

411

412 References

413 Acierno, V., Yener, S., Alewijn, M., Biasioli, F., & Van Ruth, S. (2016). Factors contributing to the  
414 variation in the volatile composition of chocolate: Botanical and geographical origins of the  
415 cocoa beans, and brand-related formulation and processing. *Food Research International*,  
416 84, 86–95. <https://doi.org/10.1016/j.foodres.2016.03.022>

417 Aprotosoai, A. C., Vlad Luca, S., & Miron, A. (2016). Flavor Chemistry of Cocoa and Cocoa  
418 Products-An Overview. *Comprehensive Reviews in Food Science and Food Safety*, 15(1),  
419 73–91. <https://doi.org/10.1111/1541-4337.12180>

420 Biasioli, F. (2016). Volatilomics by direct injection mass spectrometry. In *in: Max Rubner*  
421 *Conference 2016: Food Metabolomics* (p. 12). DE. Retrieved 3 April 2020 from  
422 [https://www.openagrar.de/servlets/MCRFileNodeServlet/openagrar\\_derivate\\_00000261/](https://www.openagrar.de/servlets/MCRFileNodeServlet/openagrar_derivate_00000261/S1910.pdf)  
423 [S1910.pdf](https://www.openagrar.de/servlets/MCRFileNodeServlet/openagrar_derivate_00000261/S1910.pdf)

424 Biasioli, Franco, Yeretian, C., Märk, T. D., Dewulf, J., & Van Langenhove, H. (2011). Direct-  
425 injection mass spectrometry adds the time dimension to (B)VOC analysis. *TrAC - Trends in*  
426 *Analytical Chemistry*, 30(7), 1003-1017. <https://doi.org/10.1016/j.trac.2011.04.005>

427 Biotechnology National Center. (2020). pubchem. Retrieved from  
428 <https://pubchem.ncbi.nlm.nih.gov/> Accessed in February 2020.

429 Buckpitt, Alan, Kephelopoulos, Stylianos, Koistinen, Kimmo, Kotzias, Dimitrios, Morawska, Lidia  
430 and Sagunski, H. . (2010). Naphthalene - WHO Guidelines for Indoor Air Quality: Selected  
431 Pollutants - NCBI Bookshelf. Retrieved 3 April 2020, from  
432 <https://www.ncbi.nlm.nih.gov/books/NBK138704/>

433 CABISCO/ECA/FCC. (2015). *Cocoa Beans : Chocolate & Cocoa Industry Quality Requirements*.  
434 Retrieved from [http://www.cocoaquality.eu/data/Cocoa Beans Industry Quality](http://www.cocoaquality.eu/data/Cocoa%20Beans%20Industry%20Quality)

435 Requirements Apr 2016\_En.pdf Accessed January 2018.

436 Cevallos-Cevallos, J. M., Reyes-De-Corcuera, J. I., Etxeberria, E., Danyluk, M. D., & Rodrick, G. E.  
437 (2009). Metabolomic analysis in food science: a review. *Trends in Food Science and*  
438 *Technology*, 20, 557- 566. <https://doi.org/10.1016/j.tifs.2009.07.002>

439 Charles, M., Romano, A., Yener, S., Barnabà, M., Navarini, L., Märk, T. D., Navarini,  
440 L., Tilmann, D. M., Biasioli, F., Gasperi, F. (2015). Understanding flavour perception of  
441 espresso coffee by the combination of a dynamic sensory method and in-vivo nosespace  
442 analysis. *Food Research International*, 69, 9–20.  
443 <https://doi.org/10.1016/j.foodres.2014.11.036>

444 Cordero, C., Guglielmetti, A., Sgorbini, B., Bicchi, C., Allegrucci, E., Gobino, G., Baroux, L., Merle,  
445 P. (2019). Odorants quantitation in high-quality cocoa by multiple headspace solid phase  
446 micro-extraction: Adoption of FID-predicted response factors to extend method  
447 capabilities and information potential. *Analytica Chimica Acta.*, 1052, 190-201.  
448 <https://doi.org/10.1016/j.aca.2018.11.043>

449 Deucher, Z., Andriot, I., Sémon, E., Repoux, M., Preys, S., Roger, J.-M., Boulanger, R., Laboure,  
450 H. Le Quéré, J.-L. (2019). Volatile compounds profiling by using proton transfer reaction-  
451 time of flight-mass spectrometry (PTR-ToF-MS). The case study of dark chocolates  
452 organoleptic differences. *Journal of Mass Spectrometry*, 54(1), 92–119.  
453 <https://doi.org/10.1002/jms.4317>

454 Eghbal, M. (2018). Global Chocolate Industry: From Bean to Bar. Retrieved 8 January 2019, from  
455 <https://blog.euromonitor.com/global-chocolate-industry/>

456 Ellis, D. I. (2019). Omics approaches for food analysis and authentication. *Current Opinion in*  
457 *Food Science.*, 28, v-vi. <https://doi.org/10.1016/j.cofs.2019.11.011>

458 FCC. The Federation of Cocoa Commerce Services - Guidelines. Retrieved 2 February 2020,  
459 from <https://www.cocoaederation.com/services/guidelines>

460 Frauendorfer, F., & Schieberle, P. (2006). Identification of the key aroma compounds in cocoa  
461 powder based on molecular sensory correlations. *Journal of Agricultural and Food*  
462 *Chemistry*, 54(15), 5521–5529. <https://doi.org/10.1021/jf060728k>

463 Frauendorfer, F., & Schieberle, P. (2008). Changes in Key Aroma Compounds of Criollo Cocoa  
464 Beans During Roasting Changes in Key Aroma Compounds of Criollo Cocoa, 56(21) 10244-  
465 10251. <https://doi.org/10.1021/jf802098f>

466 Gloess, A. N., Vietri, A., Wieland, F., Smrke, S., Schönbacher, B., López, J. A. S., Petrozzi, S.,  
467 Bongers, S., Kozirowski, T., Yeretian, C. (2014). Evidence of different flavour formation  
468 dynamics by roasting coffee from different origins: On-line analysis with PTR-ToF-MS.  
469 *International Journal of Mass Spectrometry*, 365–366, 324–337.  
470 <https://doi.org/10.1016/j.ijms.2014.02.010>

471 Heiden, A. C., Gil, C., Kinton, V. R., Pfannkoch, E. A., Ramos, L. S., & Rohrback, B. (2002).  
472 *Classification of Food and Flavor Samples using a Chemical Sensor*. Gerstel application note  
473 n°7. <http://www.gerstel.es/pdf/p-cs-an-2002-07.pdf> Accessed January 2020.

474 ISO. ISO - ISO 2451:2017 - Cocoa beans — Specification and quality requirements. Retrieved 2  
475 February 2020, from <https://www.iso.org/standard/68202.html>

476 Janairo, J. I. B., & Amalin, D. M. (2018). Volatile chemical profile of cacao liquid smoke.  
477 *International Food Research Journal*, 25(1), 213–216.

478 Liberto, E, Ruosi, M. R., Cordero, C., Rubiolo, P., Bicchi, C., & Sgorbini, B. (2013). Non-separative  
479 headspace solid phase microextraction-mass spectrometry profile as a marker to monitor  
480 coffee roasting degree. *Journal of Agricultural and Food Chemistry*, 61(8), 1652-1660.  
481 <https://doi.org/10.1021/jf303067q>

482 Liberto, Erica, Bressanello, D., Strocchi, G., Cordero, C., Ruosi, M. R., Pellegrino, G., Bicchi, C.,  
483 Sgorbini, B. (2019). HS-SPME-MS-Enose Coupled with Chemometrics as an Analytical  
484 Decision Maker to Predict In-Cup Coffee Sensory Quality in Routine Controls: Possibilities

485 and Limits. *Molecules*, 24(24), 4515. <https://doi.org/10.3390/molecules24244515>

486 Lindinger, C., Labbe, D., Pollien, P., Rytz, A., Juillerat, M. A., Yeretian, C., & Blank, I. (2008).  
487 When Machine Tastes Coffee : Instrumental Espresso Coffee, 80(5), 1574–1581.

488 Macht, M., & Mueller, J. (2007). Immediate effects of chocolate on experimentally induced  
489 mood states. *Appetite*, 49, 667–674. <https://doi.org/10.1016/j.appet.2007.05.004>

490 Magagna, F., Guglielmetti, A., Liberto, E., Reichenbach, S. E., Allegrucci, E., Gobino, G., ...  
491 Cordero, C. (2017). Comprehensive Chemical Fingerprinting of High-Quality Cocoa at Early  
492 Stages of Processing: Effectiveness of Combined Untargeted and Targeted Approaches for  
493 Classification and Discrimination. *Journal of Agricultural and Food Chemistry*, 65(30),  
494 6329–6341. <https://doi.org/10.1021/acs.jafc.7b02167>

495 Magagna, F., Liberto, E., Reichenbach, S. E., Tao, Q., Carretta, A., Cobelli, L., Giardina, M., Bicchi,  
496 C., Cordero, C. (2018). Advanced fingerprinting of high-quality cocoa: Challenges in  
497 transferring methods from thermal to differential-flow modulated comprehensive two  
498 dimensional gas chromatography. *Journal of Chromatography. A*, 1535, 122–136.  
499 <https://doi.org/10.1016/j.chroma.2017.07.014>

500 Perotti, P., Cordero, C., Bortolini, C., Rubiolo, P., Bicchi, C., & Liberto, E. (2020). Cocoa smoky  
501 off-flavor: Chemical characterization and objective evaluation for quality control. *Food*  
502 *Chemistry*, 309, 126661. <https://doi.org/10.1016/j.foodchem.2019.125561>

503 Petričević, S., Marušić Radovčić, N., Lukić, K., Listeš, E., & Medić, H. (2018). Differentiation of  
504 dry-cured hams from different processing methods by means of volatile compounds,  
505 physico-chemical and sensory analysis. *Meat Science*, 137, 217–227.  
506 <https://doi.org/10.1016/J.MEATSCI.2017.12.001>

507 Ridgway, K., Lalljie, S. P. D., & Smith, R. M. (2010). Analysis of food taints and off-flavours: a  
508 review. *Food Additives & Contaminants: Part A*, 27(2), 146–168.  
509 <https://doi.org/10.1080/19440040903296840>



510 Serra Bonvehí, J., & Ventura Coll, F. (1998). Evaluation of Smoky Taste in Cocoa Powder. *Journal*  
511 *of Agricultural and Food Chemistry*, 46(2), 620-624. <https://doi.org/10.1021/jf9705429>

512 Sgorbini, B., Cagliero, C., Liberto, E., Rubiolo, P., Bicchi, C., & Cordero, C. (2019). Strategies for  
513 Accurate Quantitation of Volatiles from Foods and Plant-Origin Materials: A Challenging  
514 Task. *Journal of Agricultural and Food Chemistry*, 67(6), 1619-1630.  
515 <https://doi.org/10.1021/acs.jafc.8b06601>

516 Spence, C. (2017). Comfort food: A review. *International Journal of Gastronomy and Food*  
517 *Science*, 9, 105–109. <https://doi.org/10.1016/j.ijgfs.2017.07.001>

518 Statista. (2020). World cocoa production by country from 2012/2013 to 2018/2019. Retrieved 8  
519 January 2019, from [https://www.statista.com/statistics/263855/cocoa-bean-production-](https://www.statista.com/statistics/263855/cocoa-bean-production-worldwide-by-region/)  
520 [worldwide-by-region/](https://www.statista.com/statistics/263855/cocoa-bean-production-worldwide-by-region/)

521 The Good Scents Company Information System. (2018). The Good Scents Company (tgsc.).  
522 Retrieved 20 December 2019, from <http://www.thegoodscentscompany.com/>

523 Vichi, S., Romero, A., Tous, J., Tamames, E. L., & Buxaderas, S. (2008). Determination of volatile  
524 phenols in virgin olive oils and their sensory significance. *Journal of Chromatography A*,  
525 1211(1–2), 1–7. <https://doi.org/10.1016/j.chroma.2008.09.067>

526 Wagner, H. S., Ahlstrom, B., Redden, J. P., Vickers, Z., & Mann, T. (2014). The myth of comfort  
527 food. *Health Psychology*, 33(12), 1552–1557. <https://doi.org/10.1037/hea0000068>

528 Wang, H., Chambers, E., & Kan, J. (2018). Sensory Characteristics of Combinations of Phenolic  
529 Compounds Potentially Associated with Smoked Aroma in Foods. *Molecules (Basel,*  
530 *Switzerland)*, 23(8). <https://doi.org/10.3390/molecules23081867>

531 Wessel, M., & Quist-Wessel, P. M. F. (2015). Cocoa production in West Africa, a review and  
532 analysis of recent developments. *NJAS - Wageningen Journal of Life Sciences*, 74–75, 1–7.  
533 <https://doi.org/10.1016/j.njas.2015.09.001>

534

| Compounds                | Odor quality             | OT<br>(ng/g) | Ti and Q1-<br>Q2 (m/z)s | Beans               |           |                  |                  | Liquors             |             |                  |                  | LOQ<br>(ng/g) | LOD<br>(ng/g) |
|--------------------------|--------------------------|--------------|-------------------------|---------------------|-----------|------------------|------------------|---------------------|-------------|------------------|------------------|---------------|---------------|
|                          |                          |              |                         | Average conc (ng/g) |           | Precision        |                  | Average conc (ng/g) |             | Precision        |                  |               |               |
|                          |                          |              |                         | smoky               | non-smoky | Intraday<br>RSD% | Interday<br>RSD% | smoky               | non-smoky   | Intraday<br>RSD% | Interday<br>RSD% |               |               |
| Naphthalene              | Mothball-like            | 80*          | <b>128</b> ;64-102      | 32.5 ± 10.7         | 4.8 ± 3.3 | 3.0              | 6.7              | 41.8 ± 18.8         | 6.2 ± 1.3   | 0.3              | 3.6              | 3.0           | 0.9           |
| Guaiacol                 | medicinal, smoky, woody  | 10           | <b>109</b> ;81-124      | 68.6 ± 25.0         | 8.2 ± 3.8 | 4.6              | 8.0              | 364.8 ± 117.9       | 73.6 ± 15.0 | 1.7              | 3.4              | 3.1           | 0.9           |
| 2-Methoxy-4-methylphenol | phenolic, smoke-like     | 90*          | <b>138</b> ;95-123      | 63.8 ± 20.1         | -         | 1.8              | 6.0              | 370.7 ± 100.7       | 22.9 ± 12.8 | 1.7              | 3.4              | 5.8           | 1.7           |
| Phenol                   | phenolic, plastic rubber | 100          | <b>94</b> ;66-95        | 721.7 ± 482.2       | 5.7 ± 6.2 | 1.1              | 5.4              | 669.6 ± 149.3       | 59.2 ± 10.0 | 1.1              | 3.0              | 1.0           | 0.3           |
| <i>p</i> -Ethylguaiacol  | smoky bacon              | 50           | <b>137</b> ;122-152     | 82.9 ± 8.0          | -         | 3.5              | 7.6              | 346.6 ± 111.1       | 38.7 ± 20.2 | 1.8              | 7.0              | 32.5          | 9.7           |
| <i>p</i> -Cresol         | phenolic                 | 25           | <b>107</b> ;77-108      | 143 ± 47.9          | -         | 4.4              | 7.7              | 675.4 ± 200.4       | 12.9 ± 6.9  | 1.4              | 2.7              | 24.1          | 7.2           |

Figure 1

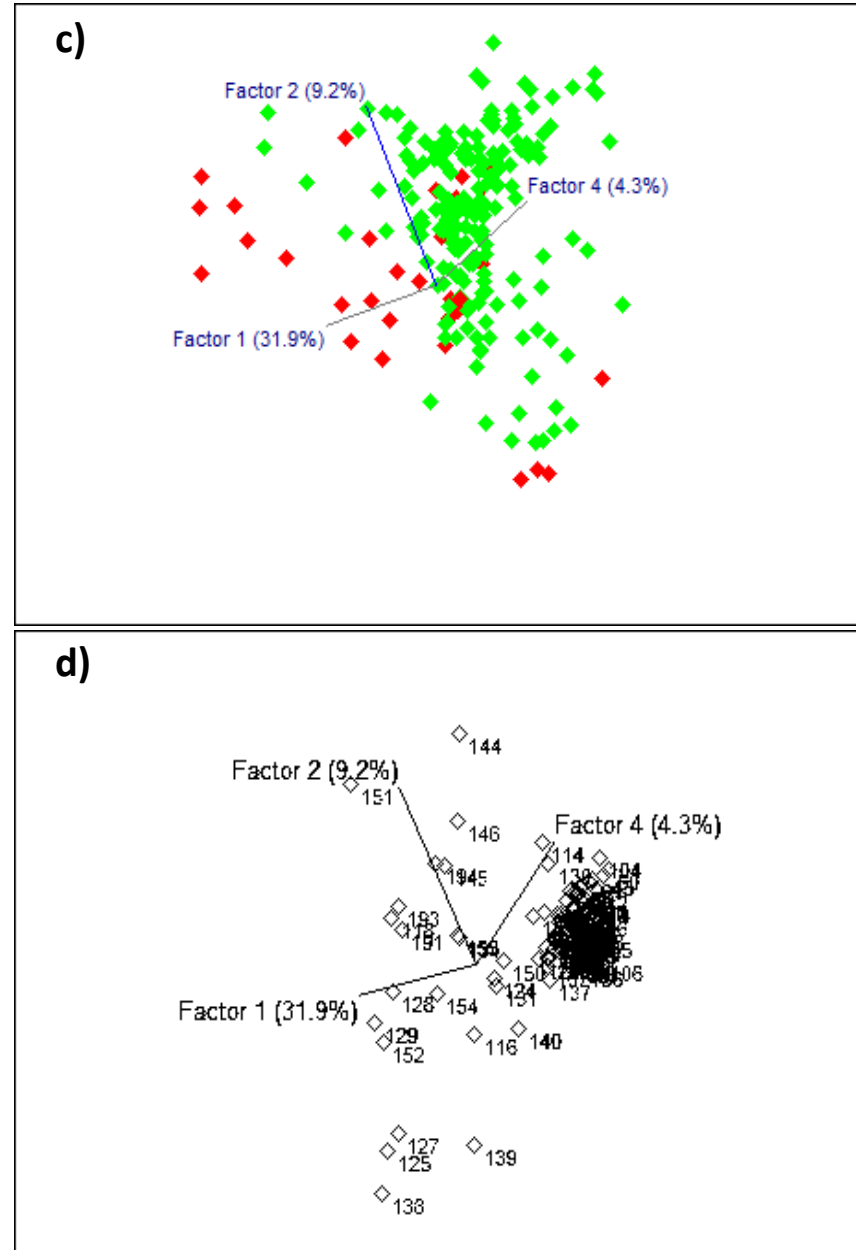
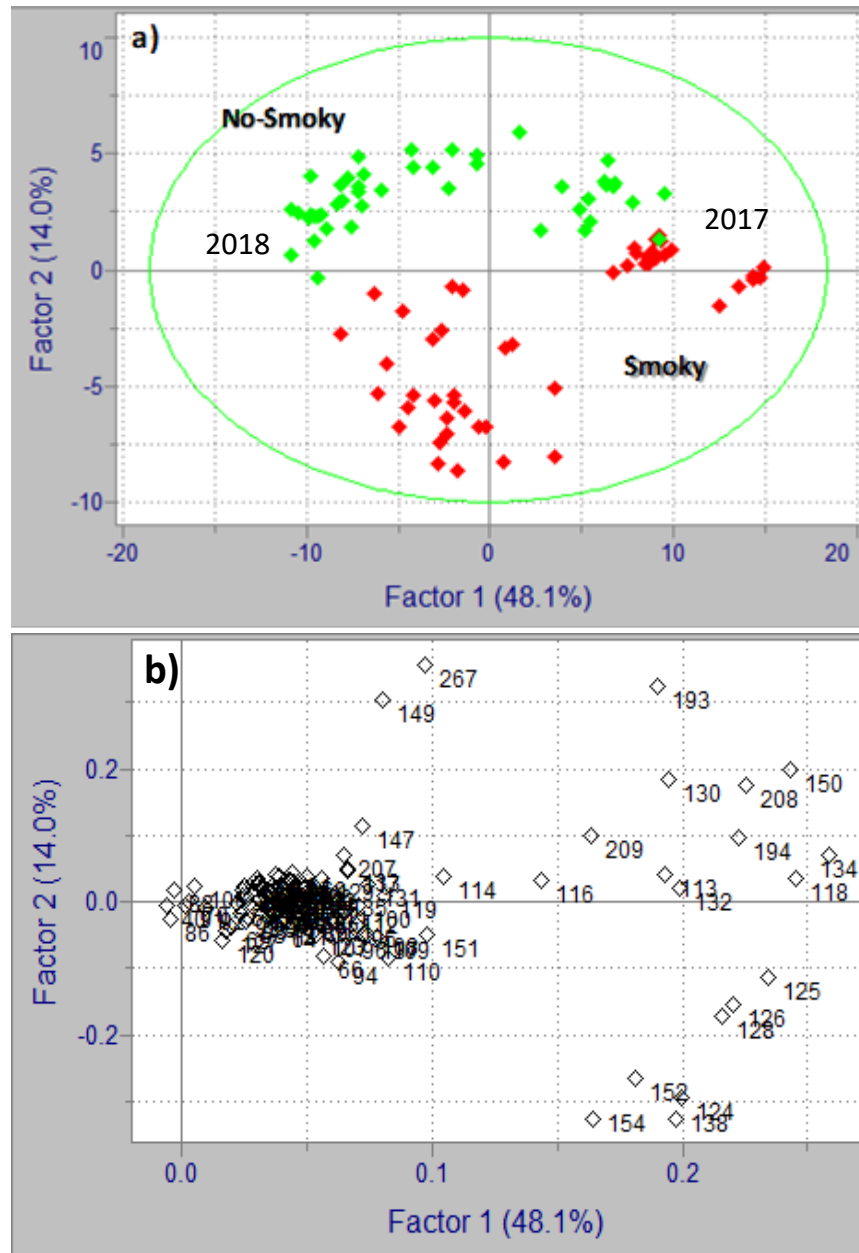


Figure 2

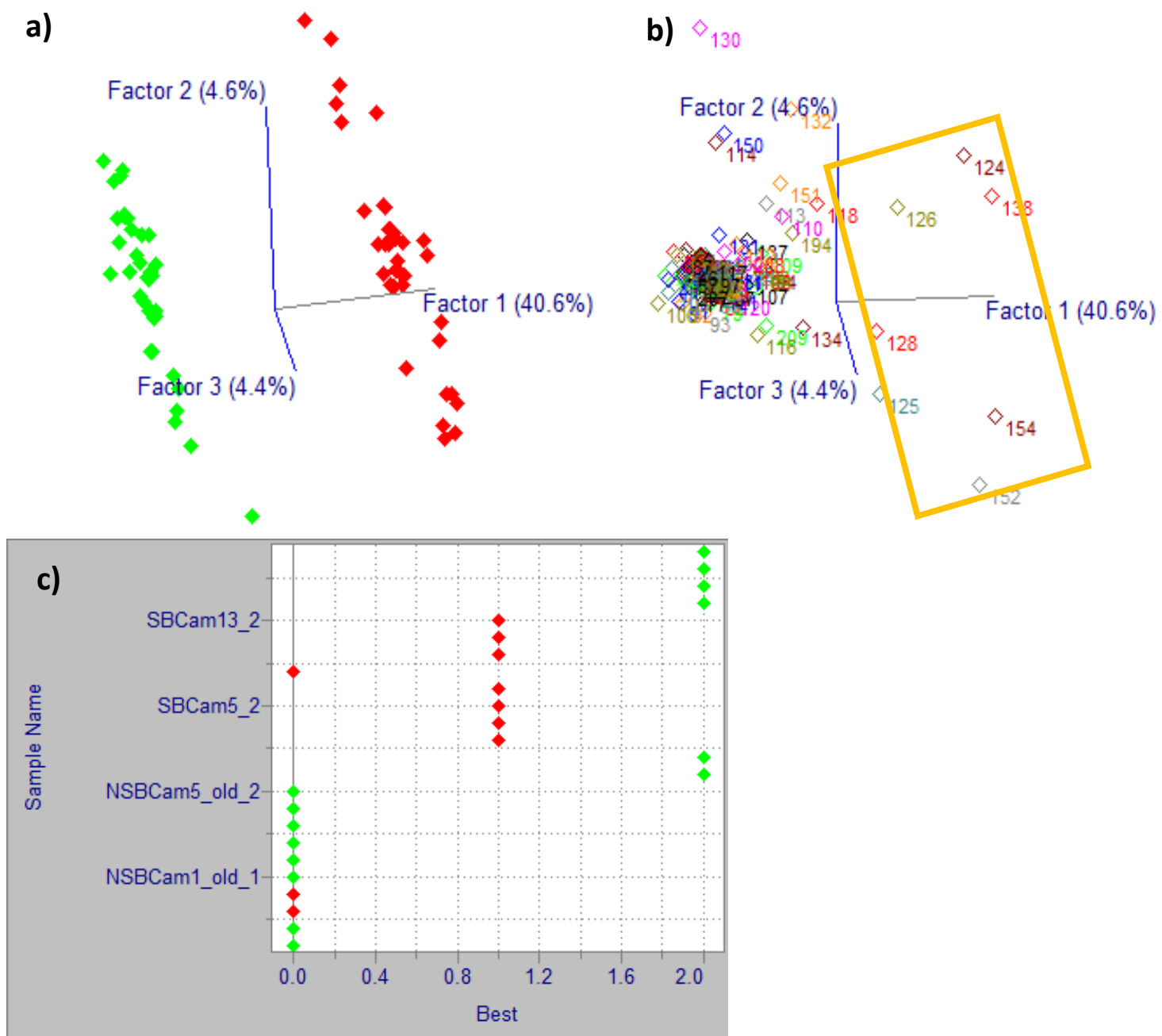


Figure 3

a)

**Training set misclassification: SIMCA**

|           | PredCS1@5          | PredCS2@5          | No match |
|-----------|--------------------|--------------------|----------|
| ActualCS1 | 21                 | 0                  | 0        |
| ActualCS2 | 0                  | 20                 | 0        |
|           | <i>Sensitivity</i> | <i>Specificity</i> | Total    |
| CS1       | 100                | 100                | 100%     |
| CS2       | 100                | 100                |          |

**Training set misclassification: SIMCA** 94.5

|           | PredCS1@5          | PredCS2@5          | No match |
|-----------|--------------------|--------------------|----------|
| ActualCS1 | 27                 | 1                  | 0        |
| ActualCS2 | 9                  | 146                | 4        |
|           | <i>Sensitivity</i> | <i>Specificity</i> | Total    |
| CS1       | 96.4               | 100                | 94.5%    |
| CS2       | 91.8               | 99                 |          |

**Training set misclassification: PLS-DA**

|           | PredCS1@5          | PredCS2@5          | No match |
|-----------|--------------------|--------------------|----------|
| ActualCS1 | 19                 | 0                  | 1        |
| ActualCS2 | 0                  | 20                 | 1        |
|           | <i>Sensitivity</i> | <i>Specificity</i> | Total    |
| CS1       | 95.0               | 100                | 95.1%    |
| CS2       | 95.2               | 100                |          |

**Training set misclassification: PLS-DA**

|           | Pred1@7            | Pred2@7            | No match |
|-----------|--------------------|--------------------|----------|
| ActualCS1 | 23                 | 4                  | 1        |
| ActualCS2 | 0                  | 151                | 8        |
|           | <i>Sensitivity</i> | <i>Specificity</i> | Total    |
| CS1       | 82.1               | 100                |          |
| CS2       | 94.9               | 97.0               | 97.7%    |



b)

**Test set misclassification: SIMCA**

|           | PredCS1@5          | PredCS2@5          | No match |
|-----------|--------------------|--------------------|----------|
| ActualCS1 | 9                  | 0                  | 1        |
| ActualCS2 | 0                  | 13                 | 1        |
| Unmodeled | 0                  | 0                  | 0        |
|           | <i>Sensitivity</i> | <i>Specificity</i> | Total    |
| CS1       | 90.0               | 100                | 91.6%    |
| CS2       | 92.8               | 100                |          |

**Test set misclassification: SIMCA**

|           | PredCS1@5          | PredCS2@5          | No match |
|-----------|--------------------|--------------------|----------|
| ActualCS1 | 8                  | 0                  | 0        |
| ActualCS2 | 1                  | 31                 | 0        |
| Unmodeled | 0                  | 0                  | 0        |
|           | <i>Sensitivity</i> | <i>Specificity</i> | Total    |
| CS1       | 100                | 89                 | 97.5     |
| CS2       | 96.8               | 100                |          |

**Test set misclassification: PLS-DA**

|           | PredCS1@5          | PredCS2@5          | No match |
|-----------|--------------------|--------------------|----------|
| ActualCS1 | 8                  | 2                  | 0        |
| ActualCS2 | 1                  | 12                 | 1        |
| Unmodeled | 0                  | 0                  | 0        |
|           | <i>Sensitivity</i> | <i>Specificity</i> | Total    |
| CS1       | 80.0               | 89                 | 83.3     |
| CS2       | 85.7               | 86                 |          |

**Test set misclassification: PLS-DA**

|           | PredCS1@6          | PredCS2@6          | No match |
|-----------|--------------------|--------------------|----------|
| ActualCS1 | 6                  | 2                  | 0        |
| ActualCS2 | 1                  | 31                 | 0        |
| Unmodeled | 0                  | 0                  | 0        |
|           | <i>Sensitivity</i> | <i>Specificity</i> | Total    |
| CS1       | 75.0               | 86.0               | 92.5     |
| CS2       | 96.8               | 94.0               |          |



Figure 4

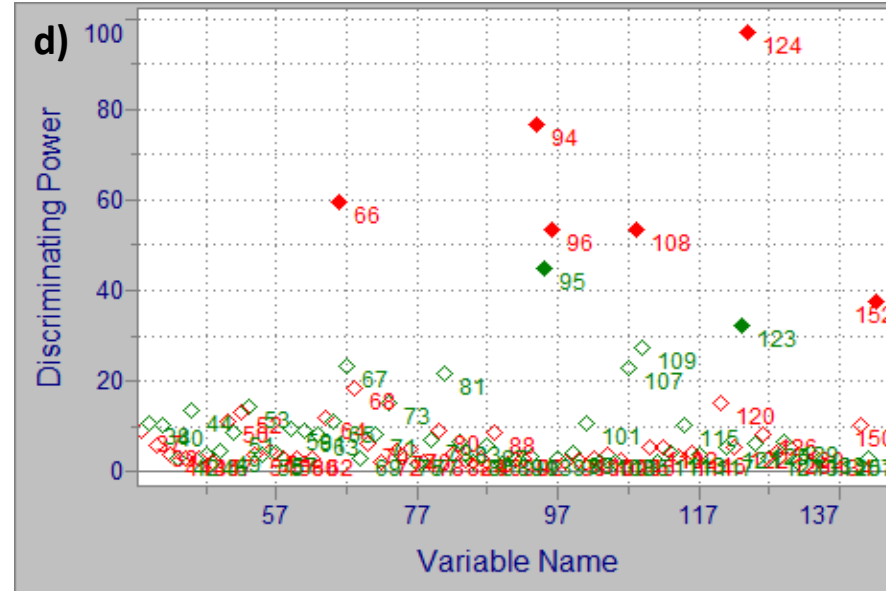
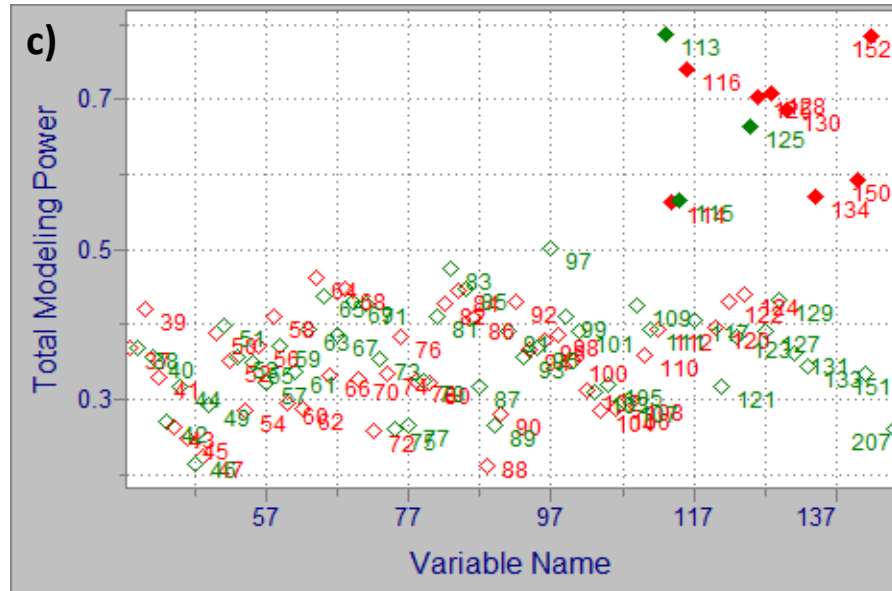
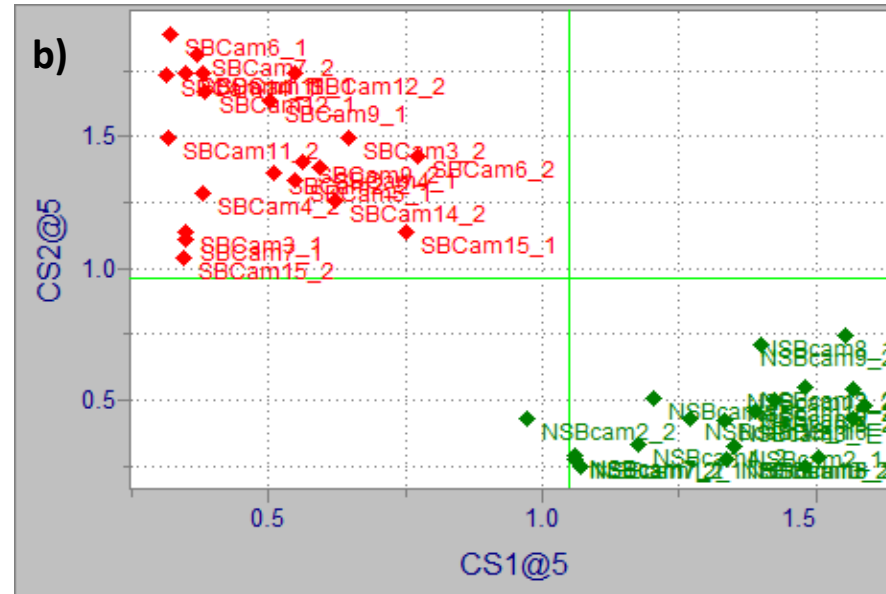
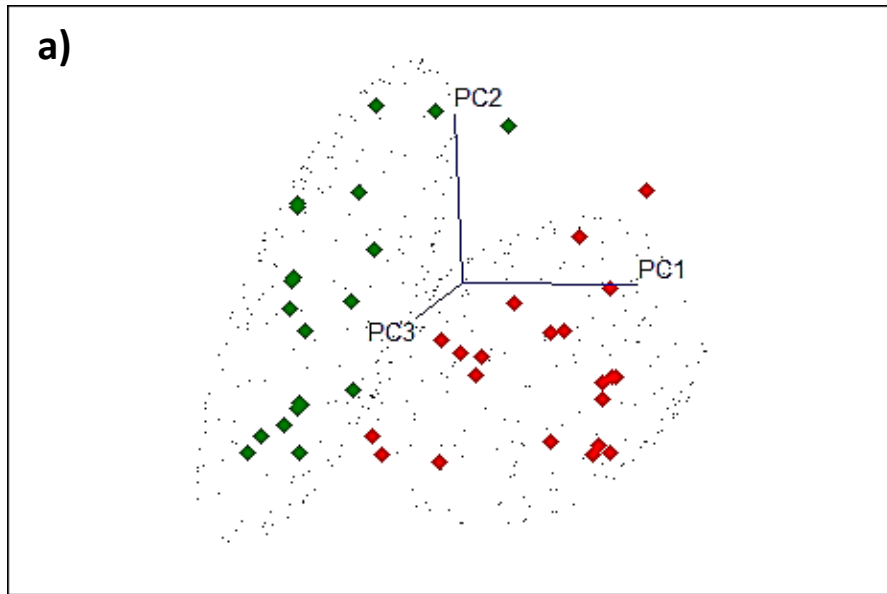
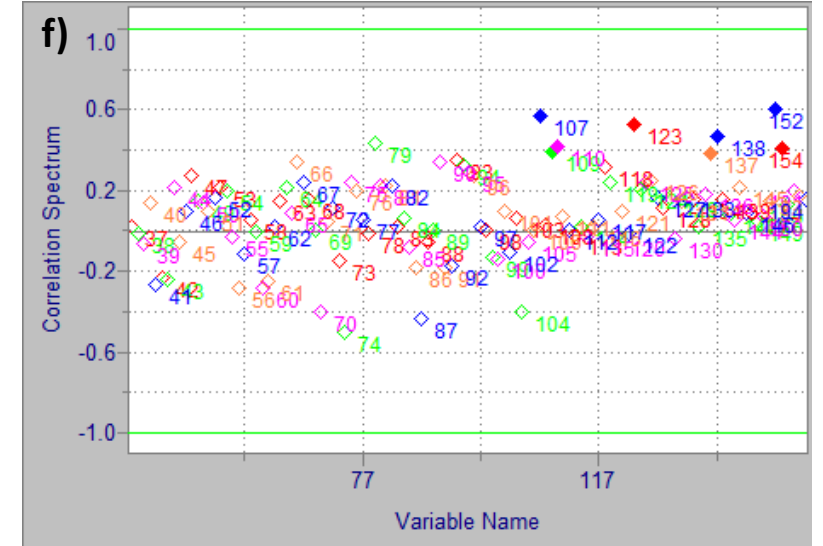
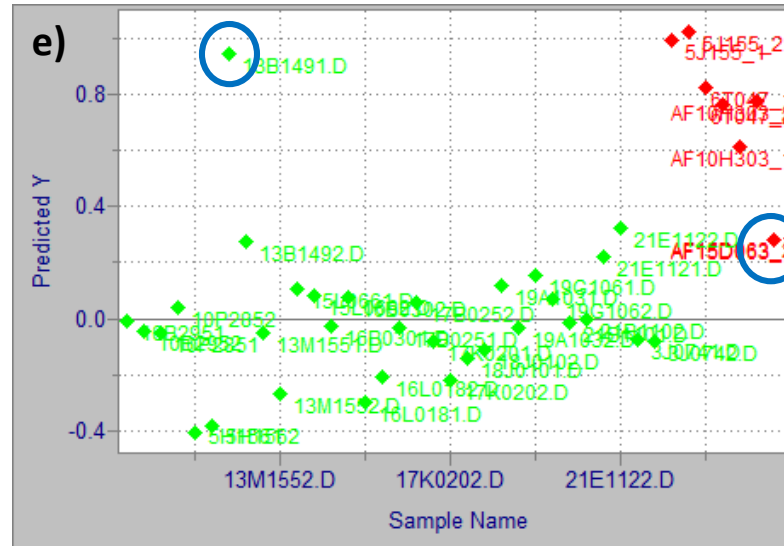
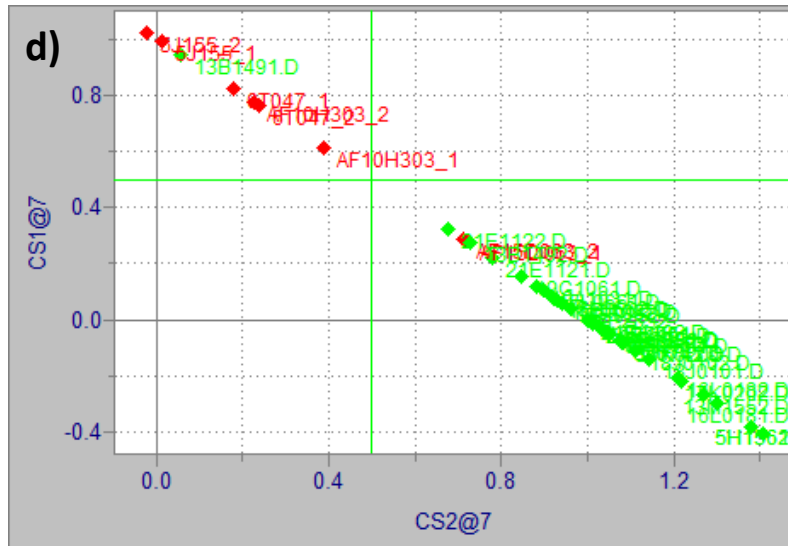
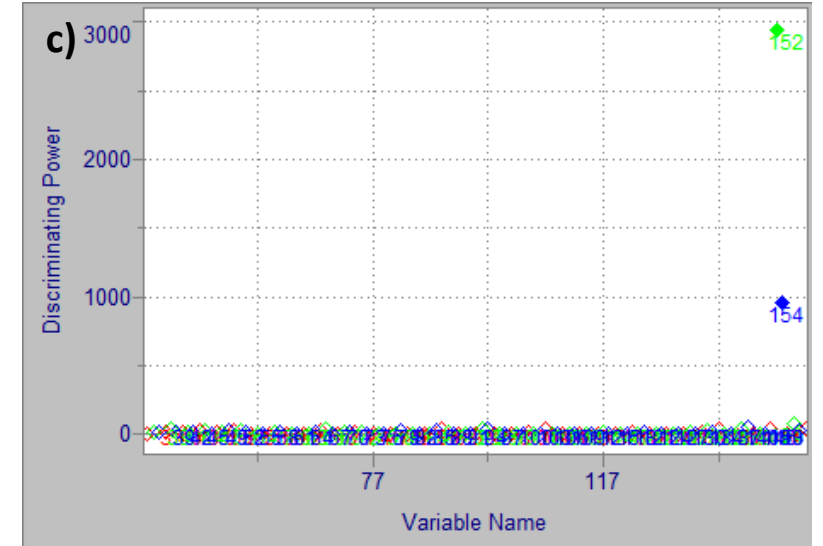
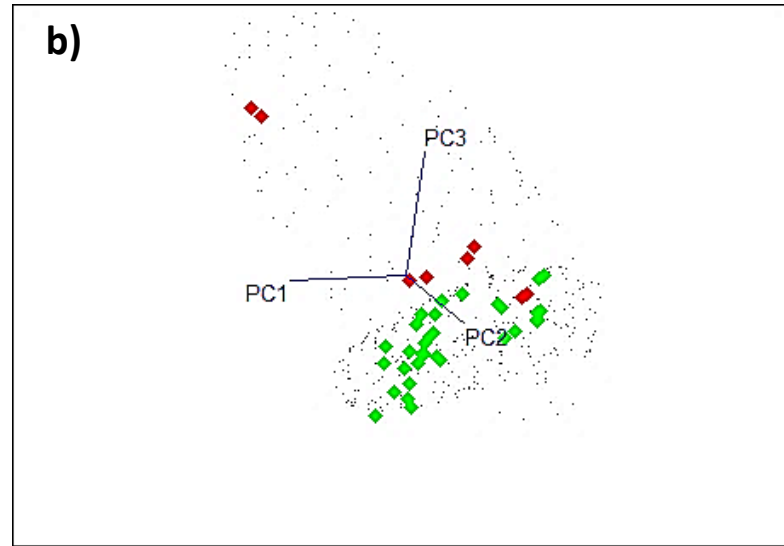
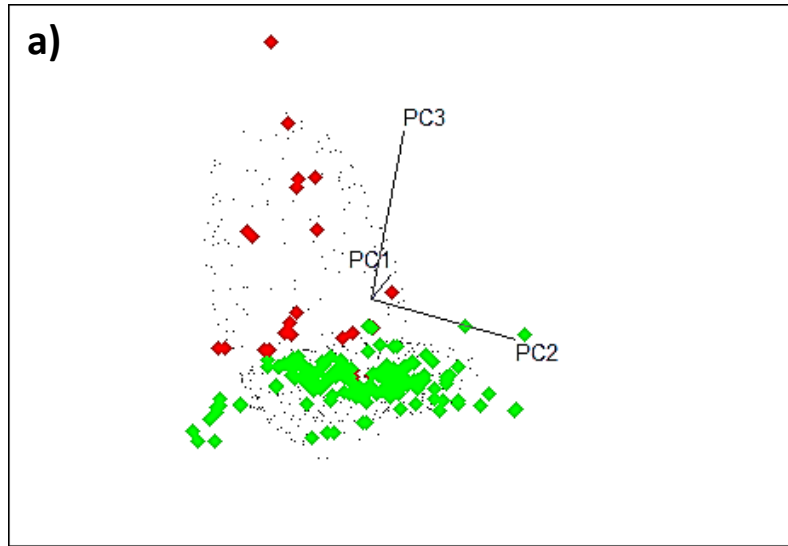


Figure 5



## **Cocoa smoky off-flavour: a MS-based analytical decision maker for routine controls**

Camilla Scavarda<sup>1</sup>, Chiara Cordero<sup>1</sup>, Giulia Strocchi<sup>1</sup>, Cristian Bortolini<sup>2</sup>, Carlo Bicchi<sup>1</sup>, Erica Liberto<sup>\*1</sup>

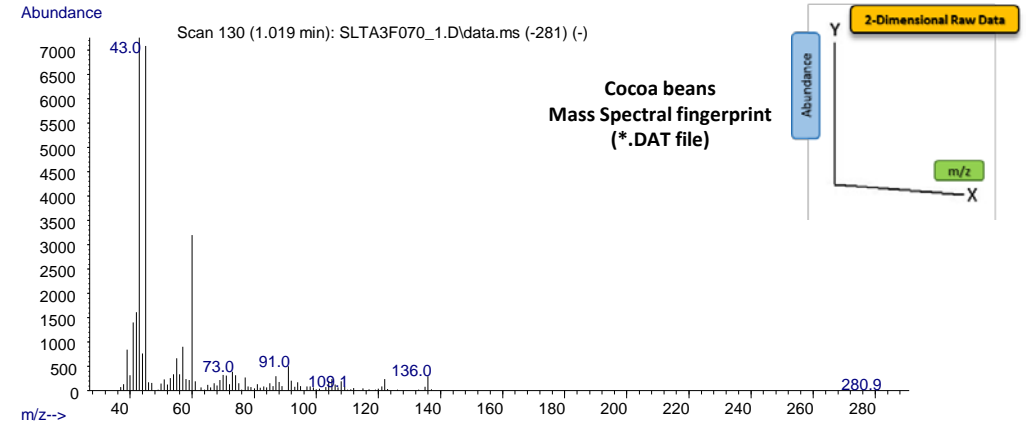
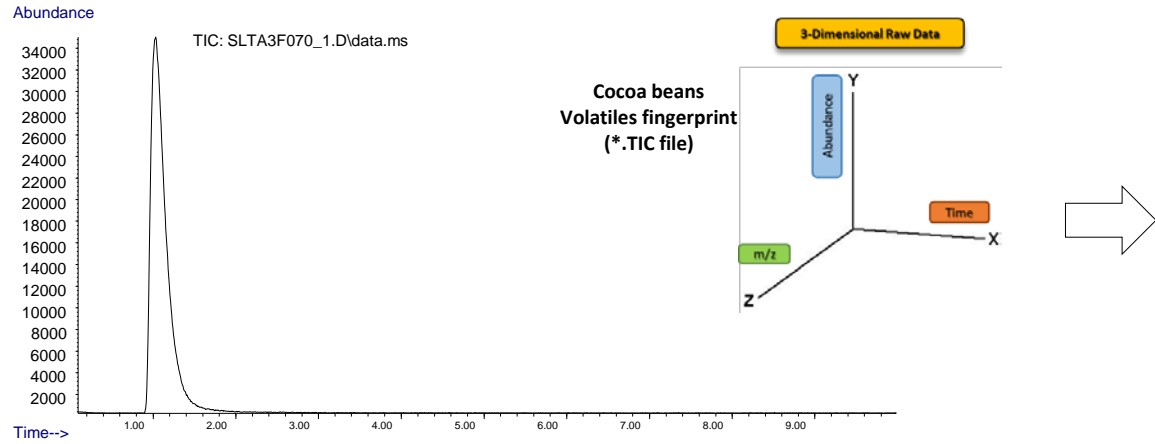
<sup>1</sup> Dipartimento di Scienza e Tecnologia del Farmaco, Università degli Studi di Torino, Turin, Italy

<sup>2</sup> Soremartec Italia S.r.l., P.le P. Ferrero 1, 12051 Alba (CN), Italy

\*Erica Liberto: [erica.liberto@unito.it](mailto:erica.liberto@unito.it); tel number: +390116707234



Figure 1 SM



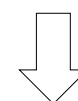
Pirouette - Liquos training.pir

File Edit Process Display Windows Help

onlyliquor

601.000000

|    | 11       | 12        | 13       | 14        | 15         | 16        | 17         | 18         | 19         | 20         | 21         | 22         | 23         | 24          | 25        | 26        | 27        | 28       |           |
|----|----------|-----------|----------|-----------|------------|-----------|------------|------------|------------|------------|------------|------------|------------|-------------|-----------|-----------|-----------|----------|-----------|
| 47 | 49       | 50        | 51       | 52        | 53         | 54        | 55         | 56         | 57         | 58         | 59         | 60         | 61         | 62          | 63        | 64        | 65        |          |           |
| 11 | NSL_5_G  | 3781.0000 | 595.0000 | 5448.0000 | 9466.0000  | 4120.0000 | 8650.0000  | 10388.0000 | 26222.0000 | 17094.0000 | 44245.0000 | 9886.0000  | 11918.0000 | 81603.0000  | 5546.0000 | 935.0000  | 1865.0000 | 525.0000 | 4718.0000 |
| 12 | NSL_5_G  | 3762.0000 | 588.0000 | 5203.0000 | 9003.0000  | 3800.0000 | 7763.0000  | 8339.0000  | 23891.0000 | 14608.0000 | 39338.0000 | 8668.0000  | 6391.0000  | 77301.0000  | 5542.0000 | 828.0000  | 1877.0000 | 385.0000 | 4246.0000 |
| 13 | NSL_10_P | 3510.0000 | 670.0000 | 5854.0000 | 10518.0000 | 5117.0000 | 10427.0000 | 16394.0000 | 24610.0000 | 17106.0000 | 44501.0000 | 9444.0000  | 8382.0000  | 63729.0000  | 4078.0000 | 592.0000  | 1365.0000 | 541.0000 | 3130.0000 |
| 14 | NSL_10_P | 3630.0000 | 879.0000 | 5889.0000 | 10389.0000 | 4911.0000 | 9905.0000  | 15533.0000 | 22924.0000 | 14685.0000 | 40509.0000 | 9159.0000  | 6419.0000  | 64161.0000  | 4012.0000 | 716.0000  | 1411.0000 | 478.0000 | 2974.0000 |
| 15 | NSL_5_Z  | 3824.0000 | 786.0000 | 5115.0000 | 9197.0000  | 4319.0000 | 9240.0000  | 12379.0000 | 22507.0000 | 13710.0000 | 37594.0000 | 8271.0000  | 8046.0000  | 63857.0000  | 3856.0000 | 608.0000  | 1167.0000 | 430.0000 | 3006.0000 |
| 16 | NSL_5_Z  | 3894.0000 | 755.0000 | 4952.0000 | 8828.0000  | 4146.0000 | 8847.0000  | 11739.0000 | 20794.0000 | 11740.0000 | 33137.0000 | 7995.0000  | 6445.0000  | 62258.0000  | 3577.0000 | 599.0000  | 1325.0000 | 490.0000 | 2766.0000 |
| 17 | NSL_5_B  | 3828.0000 | 767.0000 | 6308.0000 | 11066.0000 | 5072.0000 | 10244.0000 | 15920.0000 | 22939.0000 | 12670.0000 | 34982.0000 | 10191.0000 | 6761.0000  | 64365.0000  | 4291.0000 | 756.0000  | 1708.0000 | 536.0000 | 3425.0000 |
| 18 | NSL_5_B  | 3644.0000 | 728.0000 | 5970.0000 | 10827.0000 | 4967.0000 | 9982.0000  | 15422.0000 | 23224.0000 | 13605.0000 | 36526.0000 | 10159.0000 | 8103.0000  | 64058.0000  | 4324.0000 | 672.0000  | 1658.0000 | 525.0000 | 3406.0000 |
| 19 | SL_H_046 | 7979.0000 | 421.0000 | 4695.0000 | 8053.0000  | 4723.0000 | 10403.0000 | 17617.0000 | 20400.0000 | 10141.0000 | 32516.0000 | 5910.0000  | 7262.0000  | 157681.0000 | 6548.0000 | 1059.0000 | 1611.0000 | 456.0000 | 3784.0000 |
| 20 | SL_H_046 | 7769.0000 | 314.0000 | 4684.0000 | 8100.0000  | 4727.0000 | 10402.0000 | 17166.0000 | 19740.0000 | 9961.0000  | 31594.0000 | 5716.0000  | 6984.0000  | 153641.0000 | 6457.0000 | 996.0000  | 1577.0000 | 486.0000 | 3734.0000 |
| 21 | SL_E_154 | 5650.0000 | 453.0000 | 5746.0000 | 9917.0000  | 5064.0000 | 9975.0000  | 13845.0000 | 22861.0000 | 11957.0000 | 34573.0000 | 9419.0000  | 6671.0000  | 84903.0000  | 4617.0000 | 899.0000  | 1854.0000 | 575.0000 | 4146.0000 |
| 22 | SL_E_154 | 5634.0000 | 538.0000 | 5672.0000 | 9937.0000  | 5002.0000 | 10108.0000 | 14484.0000 | 22698.0000 | 11936.0000 | 34493.0000 | 9424.0000  | 7362.0000  | 85610.0000  | 4751.0000 | 863.0000  | 1879.0000 | 633.0000 | 4074.0000 |
| 23 | SI_N_154 | 6259.0000 | 569.0000 | 5083.0000 | 8741.0000  | 3940.0000 | 8710.0000  | 9143.0000  | 19138.0000 | 8066.0000  | 26543.0000 | 7099.0000  | 6272.0000  | 56706.0000  | 3578.0000 | 898.0000  | 1966.0000 | 552.0000 | 4574.0000 |



n samples acquired

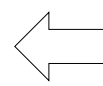
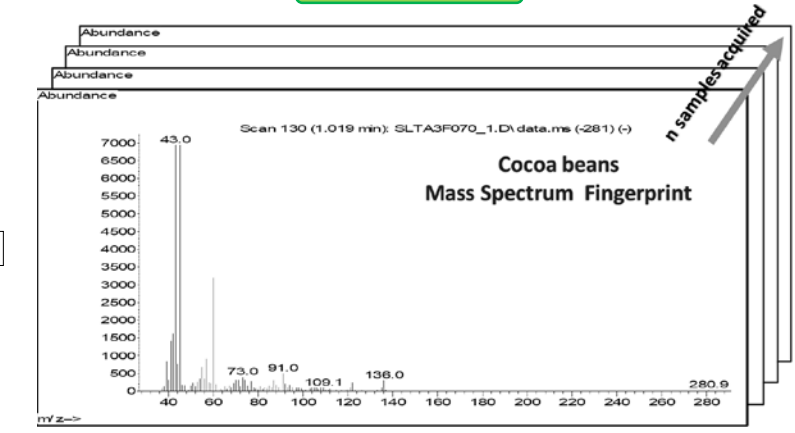


Figure 2 SM

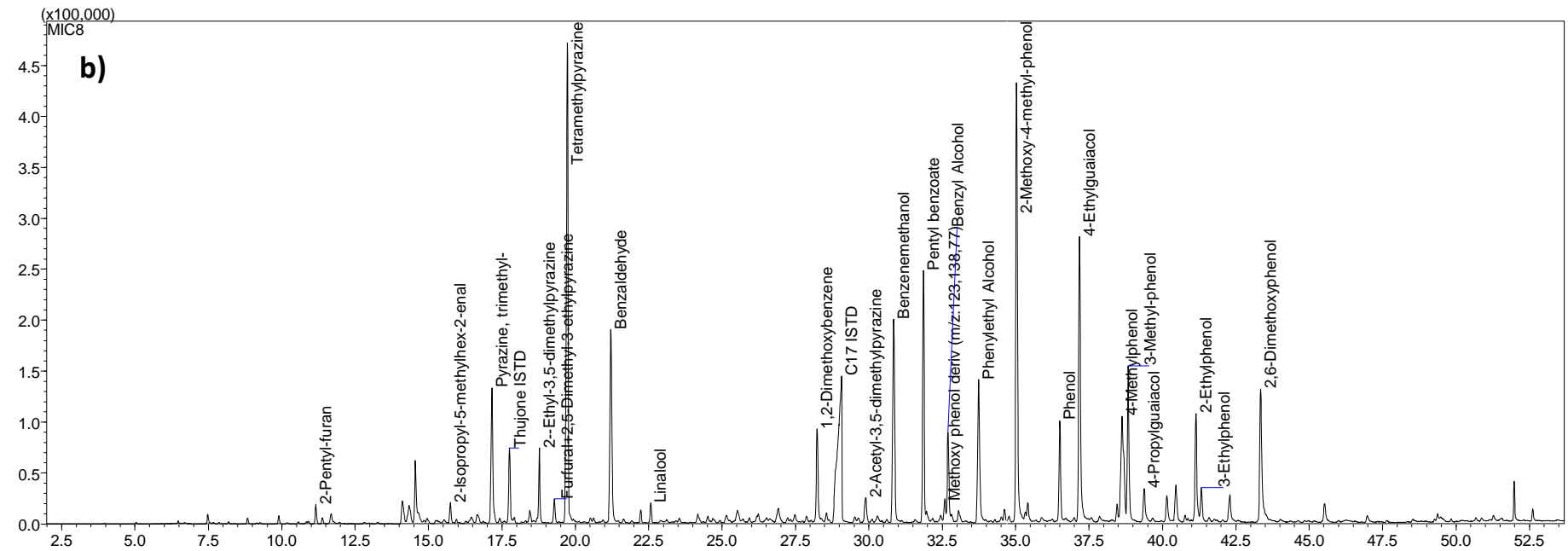
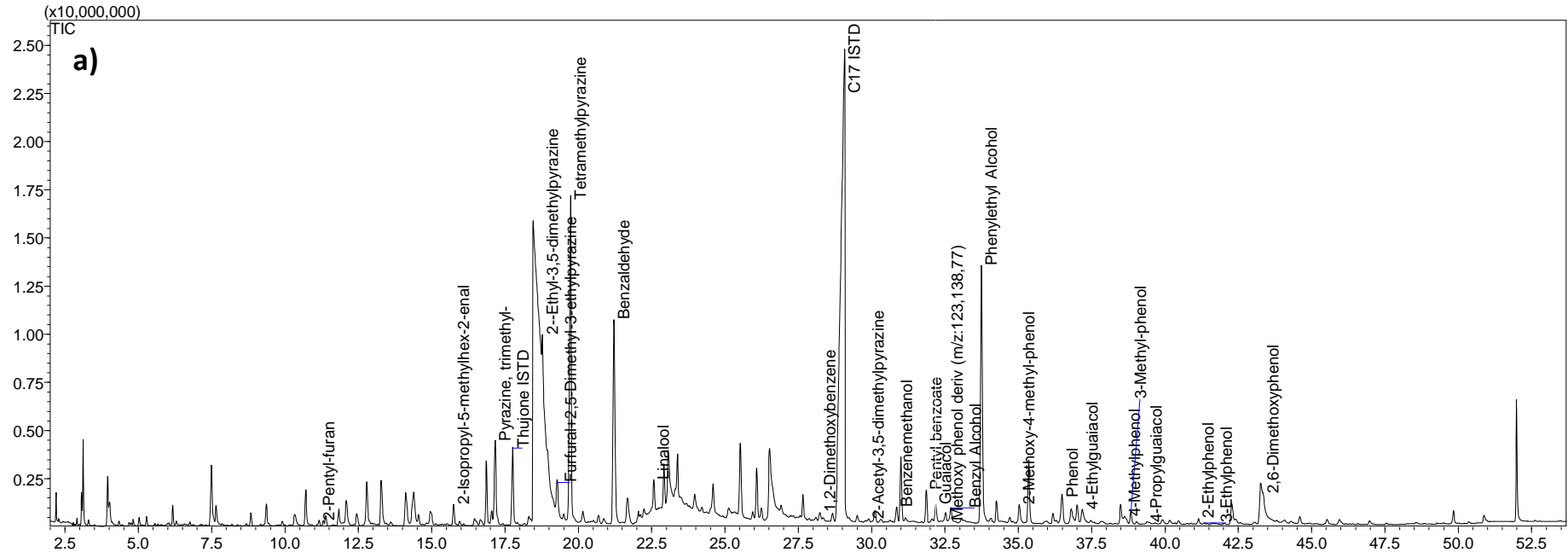
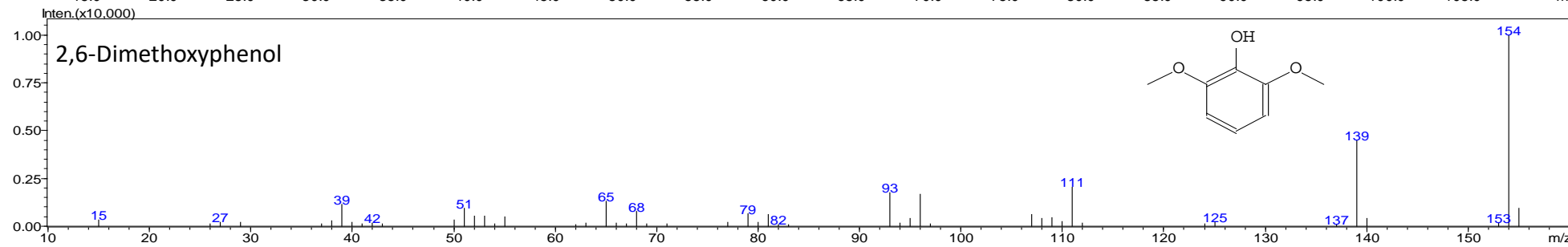
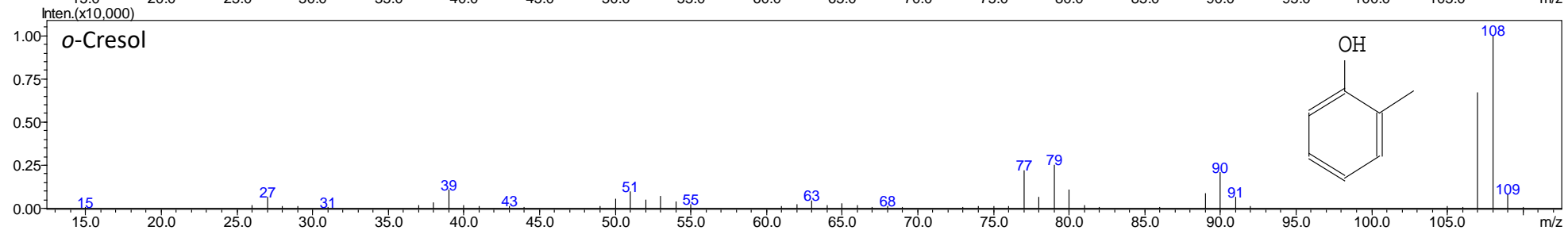
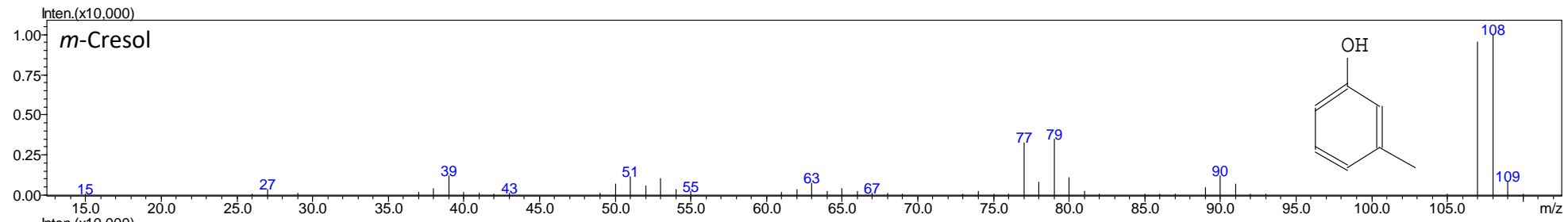
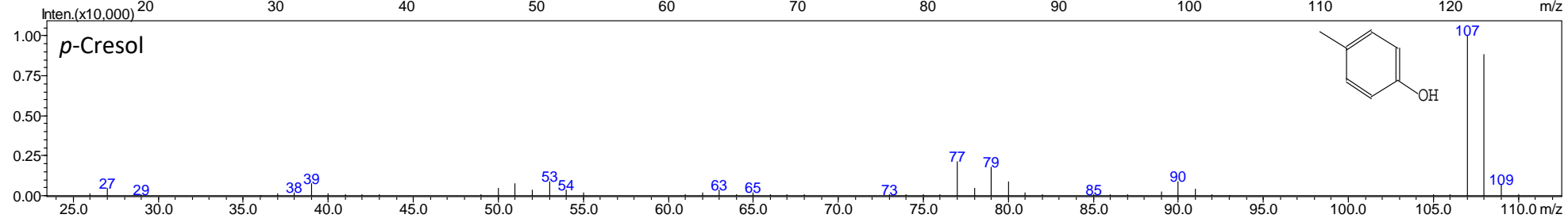
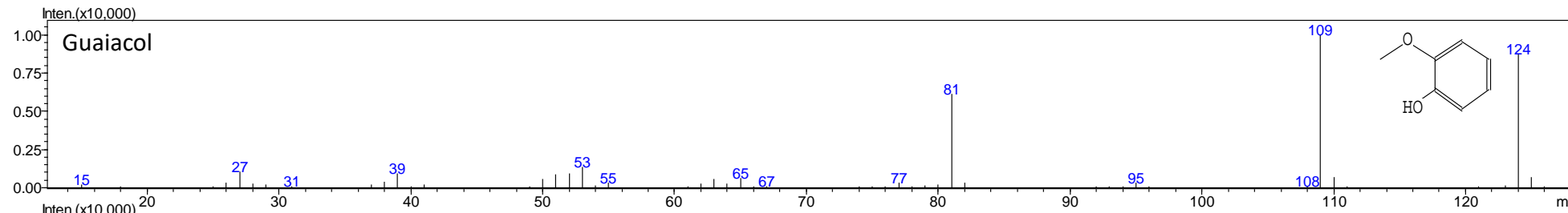


Figure 3SM



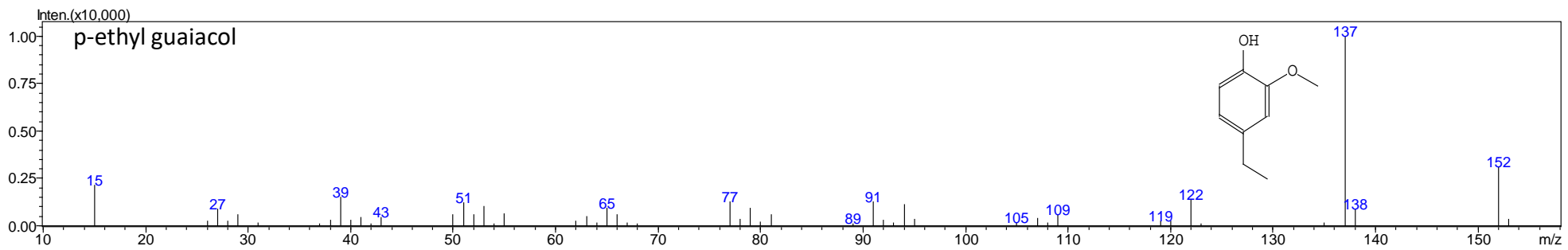
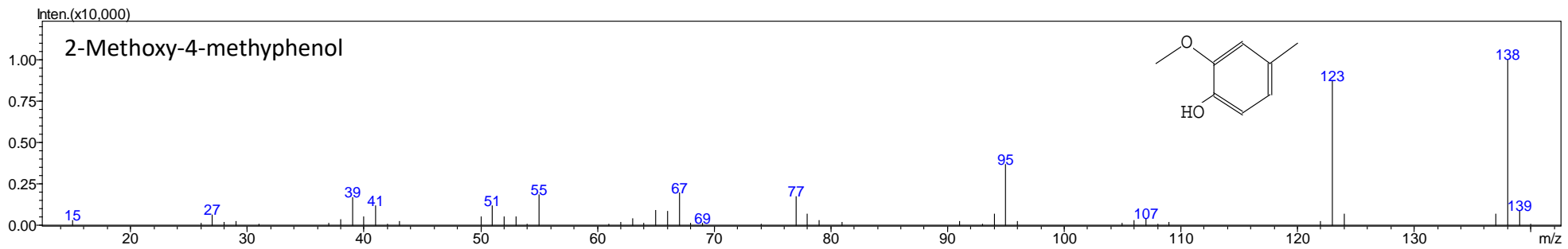
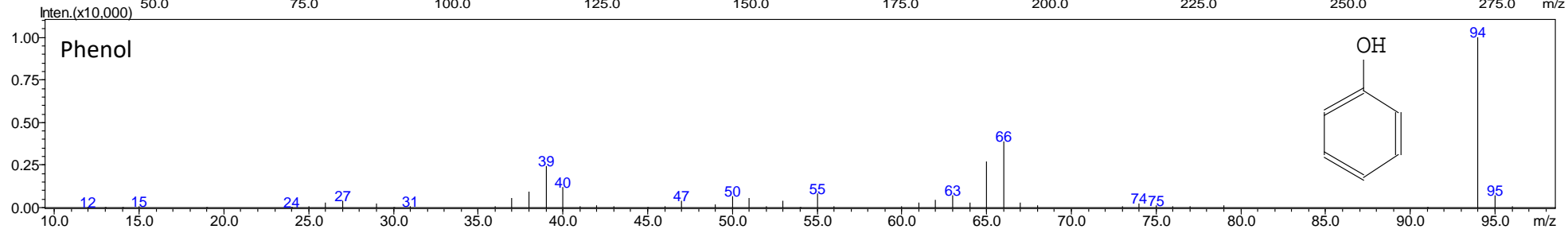
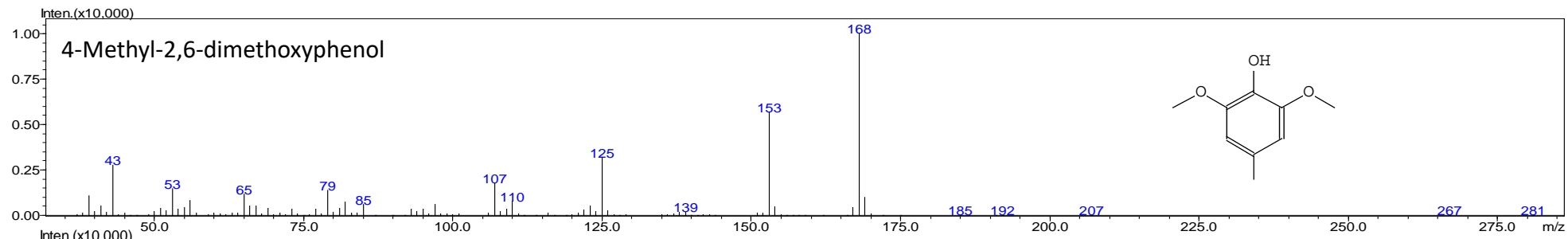
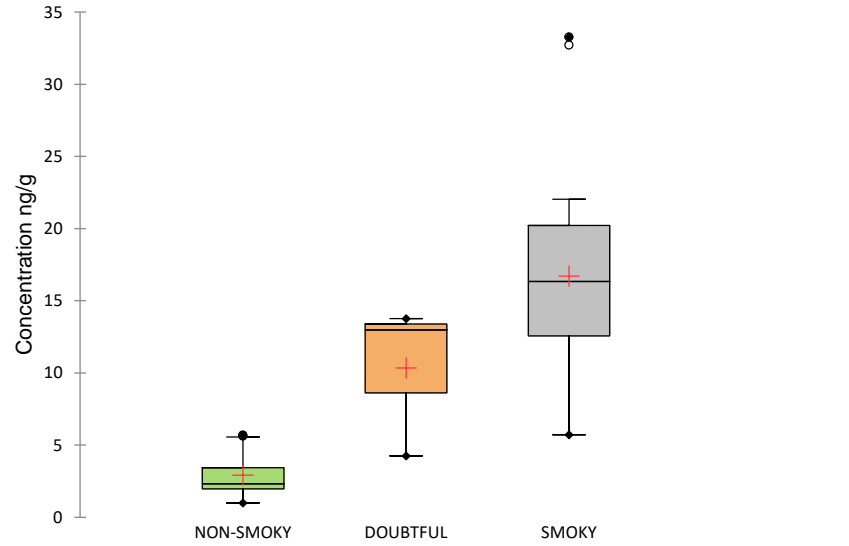
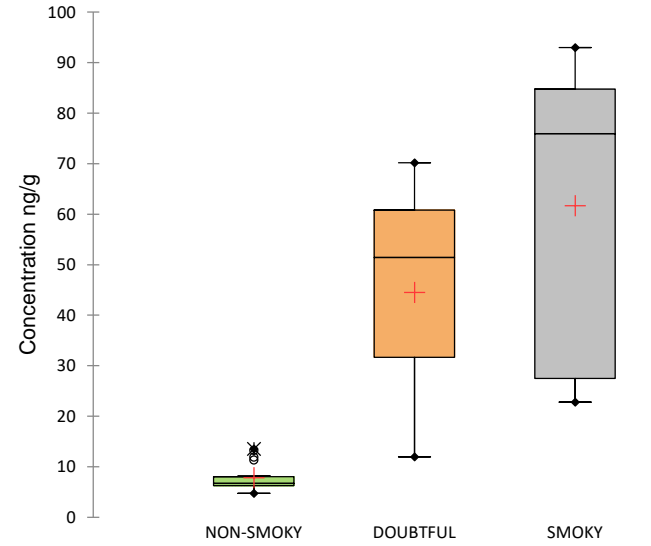


Figure 4 SM

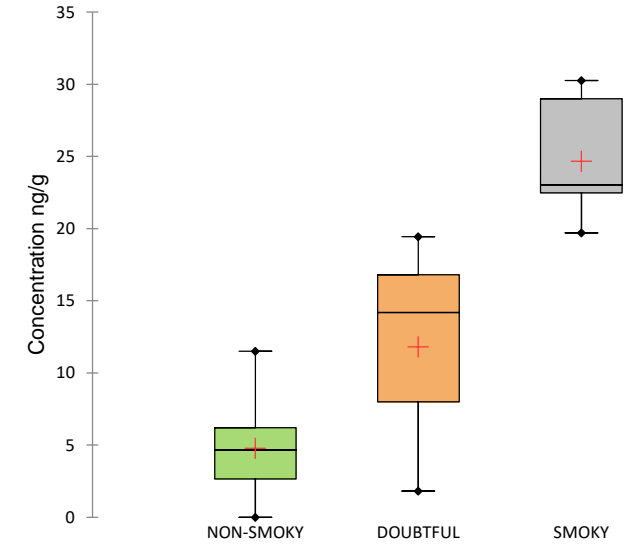
**Naphthalene**



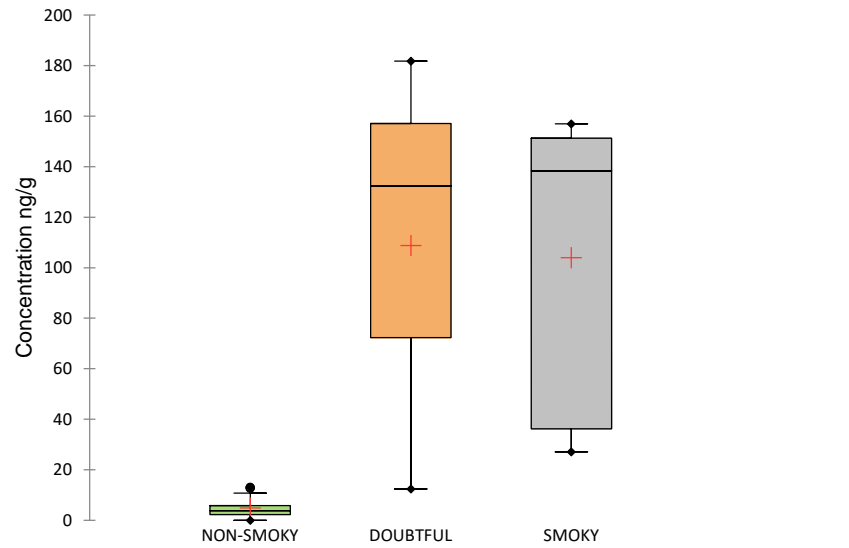
**Guaiacol**



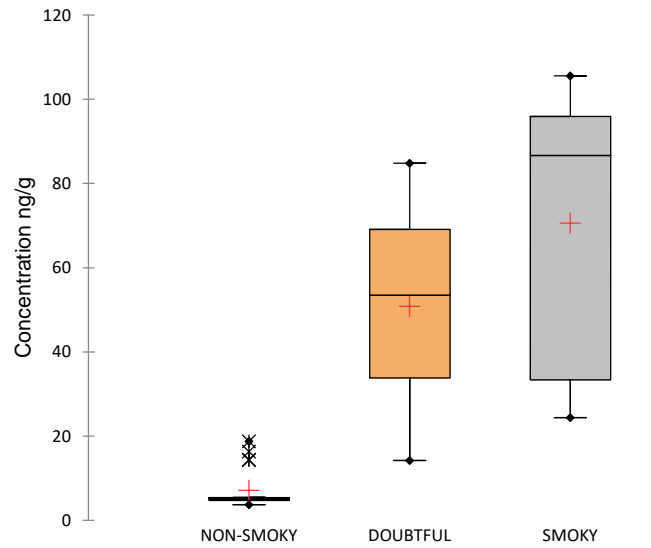
**2-Methoxy-4-methylphenol**



**Phenol**



**p-Ethylguaiacol**



**p-Cresol**

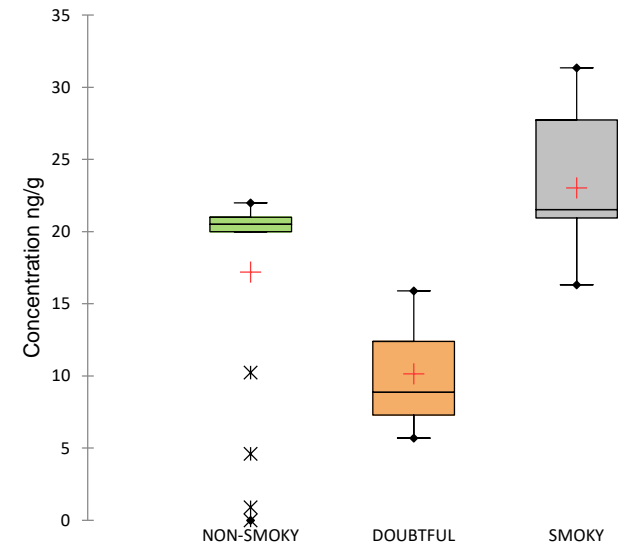


Figure 5 SM

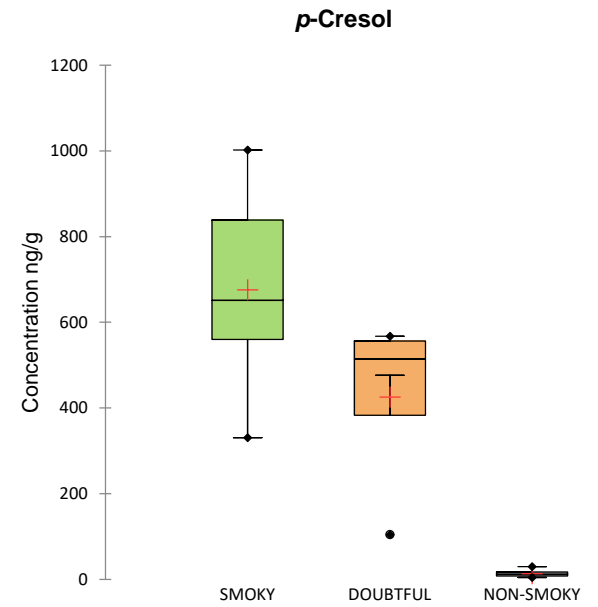
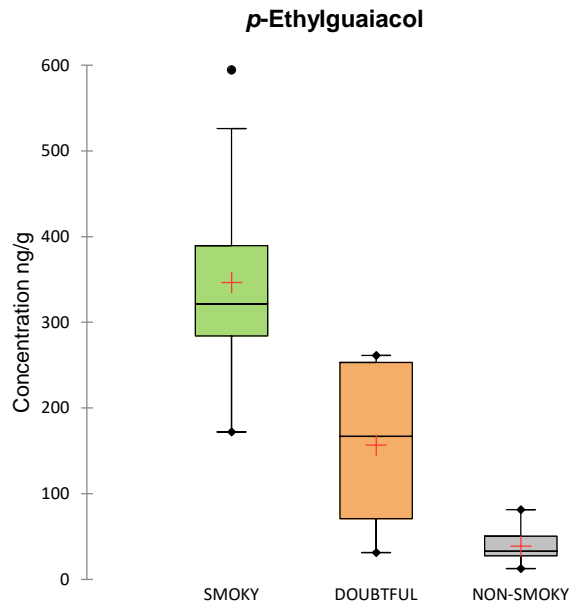
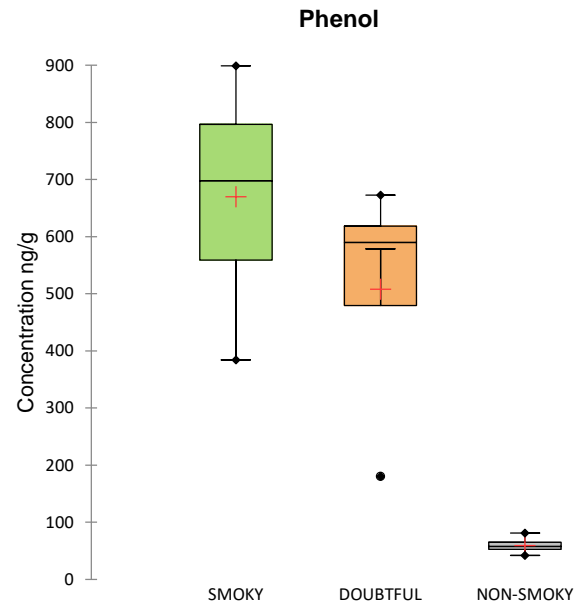
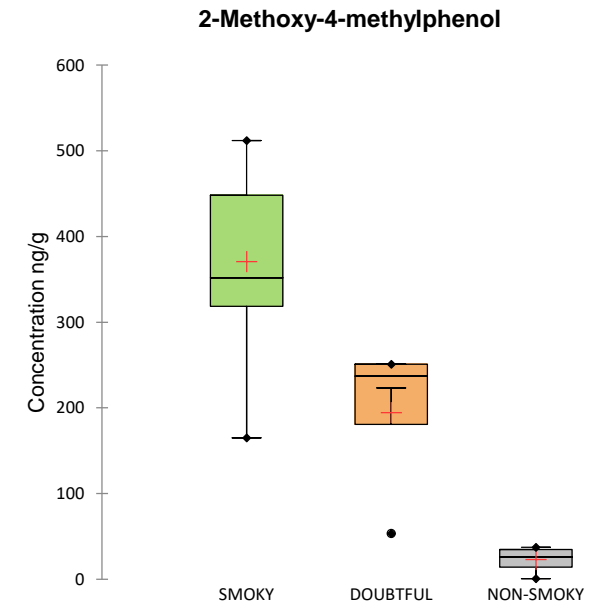
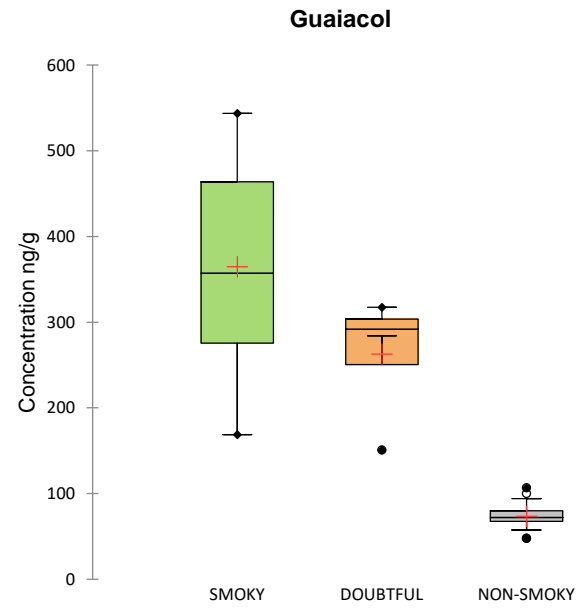
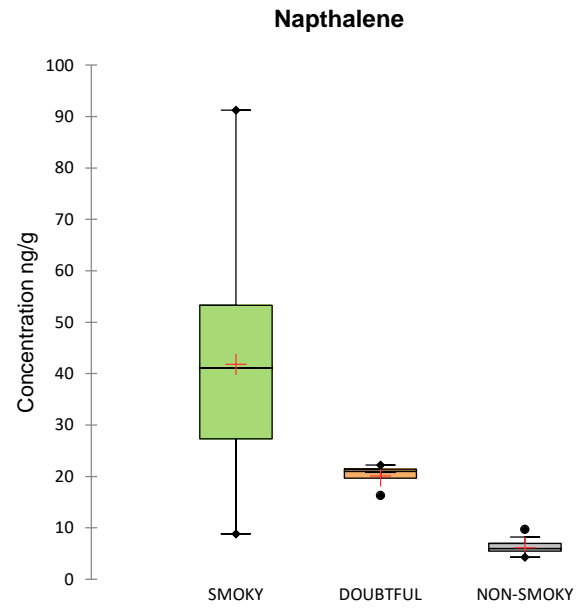


Figure 6 SM

