# Gas chromatography-mass spectrometry based untargeted volatolomics for smoked seafood classification

Leticia Lacalle-Bergeron [a], Tania Portolés [a,*], Carlos Sales [a], M. Carmen Corell [b], Fernando Domínguez [c], Joaquim Beltrán [a], Juan Vicente Sancho [a], Félix Hernández [a]

[a] Enviromental and Public Health Analytical Chemistry, Research Institute for Pesticides and Water (IUPA), University Jaume I, Av. Sos Baynat S/N, 12071 Castellón de la Plana, Spain
[b] Sea Delight Europe, S.L, C/ Sao Paulo, 14 Planta 2ª Oficina n°3-P.I. El Sebadal, 35008 Las Palmas de Gran Canaria, Spain
[c] Food Safety Unit, Centre of Public Health of Castellón, Av. del Mar 12, 12003 Castellón de la Plana, Spain

ARTICLE INFO

ABSTRACT

With the increase of the demand of low flavouring smoked seafood products, there is a need of methodologies able to distinguish between different seafood treatments, as not all of them are allowed in all markers.

Following this objective, in the present work an untargeted volatolomics approach was applied to identify volatile markers that demonstrate that Cold smoked products can be distinguished from Tasteless smoke neither Carbon monoxide treated seafood, which are prohibited in the European Union.

The use of dynamic headspace for the volatile extraction followed by thermal desorption in combination with Gas Chromatography (GC) coupled to single quadrupole Mass Spectrometry (MS) has been employed for the determination of volatile composition of smoked fish. Data processing consisted on the use of PARADISe software, applied for GC/MS data treatment, followed by the multivariate analysis with PLS_Toolbox (MATLAB), and finally the creation and validation of statistical classification model.

All 107 variables obtained allowed the construction of a model reaching the correct classification of 97% of the blind samples, while a simplified model with only 11 variables correctly classified up to 93% of the blind samples. These 11 compounds were elucidated to develop subsequent target volatolomics approaches, if needed. Ordered according to the importance in the classification model, the elucidated compounds were: 3-methyl-cyclopentanone, ethylbenzene, 2-methyl-2-cyclopenten-1-one, 2-methyl-benzofuran, furfuryl alcohol, 2-acetyl-furan, acetophenone, guaiacol, 1-hydroxy-2-butanone, 4-vinylguaiacol and acetoin.

The results demonstrated the great potential of untargeted volatolomics for smoked seafood treatments classification.

## 1. Introduction

The smoking treatment of food products has been applied for food preservation since ancient times. This technique allows the preservation of fish by drying and by adding naturally produced microbistatic constituents from the wood smoke. Nowadays, the smoking techniques have been evolving and the aim of smoking, in addition to preservation, is to develop particular flavour, colour and texture characteristics derived from the burned wood that is used in the process (Varlet, Serot, & Prost, 2009). To this aim, different parameters, such as the type of wood and time of exposure to smoke, among others, have to be optimized by the industry to obtain a certain type of flavour, flavour intensity and product

quality (Jónsdóttir, Ólafsdóttir, Chanie, & Haugen, 2008).

Alternative processes have emerged, such as the treatment with carbon monoxide (CO) or filtered wood smoke (tasteless smoke (TS), clear smoke…). These techniques are based on the use of CO as colour-stabilizer, maintaining and enhancing the red colour associated with a fresh aspect of the fish flesh, particularly in tuna, and delaying the browning that usually appears with product aging (Barstad, Alvik, & Løvaas, 2006; Bartolucci et al., 2010; Gokoglu, 2020). Nevertheless, filtered wood smoke and CO-treatments of fish are not permitted in the European Union; moreover, CO is excluded from the list of allowed food additives (Regulation (EC) No 1333/2008). This measure was taken because consumers could be confused about the freshness of the product

---

(Directive 91/493/EEC). In the case of histidine-rich fishes, the fraudulent use of these treatments may increase the risk of histamine intoxication (Bartolucci et al., 2010; Dalgaard, Emborg, Kjølby, Sørensen, & Ballin, 2008).

The increase in the fish demanded for preparation of *sushi* in EU has favoured new cold smoking treatments and the application of different degrees of smoking to get different flavours levels. However, because of the faint organoleptic properties of some of them, they can be mistaken by CO or TS smoked products and therefore be rejected at European customs.

The chemicals responsible for the sensory attributes of smoked fish products are mainly volatile organic compounds (VOCs), such as phenols, furan-like compounds, aldehydes or ketones, among others (Varlet et al., 2009). Accordingly, the analysis of the smoked chemical profile is commonly made by gas chromatography (GC), which in combination with mass spectrometry (GC/MS) allows a sensitive determination with great identification capability able to detect and identify the volatile compounds that characterize the flavour of studied samples.

The extraction technique is a matter of concern, as this step is essential to obtain reliable data and full characterization of the volatile profile. Automatable direct headspace injection (HS) based techniques are commonly used as it implies low sample manipulation, simple, cheap and fast option. Although it suffers from low sensitivity for some compounds present at low concentrations in the vapour phase and the analysis parameters should be carefully optimized to get reproducible results (Soria, García-Sarrió, Ruiz-Matute, & Sanz, 2017). Headspace-solid phase microextraction (HS-SPME) has been already used for VOCs analysis of smoked food products (Marušić Radovčić, Vidaček, Janči, & Medić, 2016; Saldaña et al., 2019; Vidal, Goicoechea, Manzanos, & Guillén, 2017) with a good pre-concentration factor and solventless. In contrast, the adsorption capability is highly dependent on sample matrix and the coating of the fibre (Płotka-Wasylka, Szczepańska, Owczarek, & Namieśnik, 2017). On the contrary, dynamic HS with sorbent trapping (DHS-P&T) captures the VOCs present in the sample on a solid sorbent with the aid of an inert gas flow for continuous extraction. DHS-P&T allows to increase the volatiles recovery, preconcentrating most of them and therefore enhancing the sensitivity, with good efficiency and low sample manipulation (Soria et al., 2017; Thomsen et al., 2016). Analytes are transferred to the GC system via thermal desorption and then cryo-focused into the GC injector, which leads to an additional increase of the sensitivity due to the complete transfer of the extracted analytes. This technique has been successfully applied to analyse VOCs in food matrices including smoked food (Dirinck, Schreyen, & Schamp, 1977; Fredes et al., 2016; Huang et al., 2019; Sales, Portolés, Johnsen, Danielsen, & Beltran, 2019; Soria, Martínez-Castro, & Sanz, 2008; Thomsen et al., 2016).

The determination of the volatile profile has been widely applied for food characterization (Beaulieu & Lea, 2006; Ben Brahim et al., 2018; Jónsdóttir et al., 2008; Sérot, Baron, Knockaert, & Vallet, 2004). Commonly, a targeted approach is used (i.e. focusing the (quantitative) analysis on a limited list of target compounds), which can provide biased information because the compounds of interest have to be pre-selected and it provides incomplete information on the samples composition. Oppositely to target approaches, the non-targeted metabolomics have great potential in the volatile fingerprinting determination. The metabolomic fingerprinting is defined as *"the unbiased, global screening approach to classify samples based on metabolite patterns or "fingerprints" that change in response to disease, environmental or genetic perturbations with the ultimate goal to identify discriminating metabolites"* (Dettmer, Aronov, & Hammock, 2007). When the object of study is the highly and semi volatile fraction of molecules, it is called volatolomics. This approach was firstly implemented in the health area (Bouhlel et al., 2017; Broza, Mochalski, Ruzsanyi, Amann, & Haick, 2015), but it is gaining relevance in food related areas, such as food quality control or food authenticity (Abou-el-karam, Ratel, Kondjoyan, Truan, & Engel, 2017; Sales et al., 2019). In untargeted volatolomics, the chromatographic analysis must be robust, with adequate peak resolution, and good retention time and peak shape reproducibility. To obtain overall information from the analytes present in samples, and to identify those that might be useful markers for metabolomics, the MS acquisition must be performed in full scan mode (Garcia & Barbas, 2011).

Data processing is of special relevance in non-targeted metabolomics/volatolomics, as reflected in the amount of papers and software reported on metabolomics studies. Some of these software are MzMine (Sales et al., 2017), XCMS (Gil-Solsona et al., 2016), MetAlign (Tomita, Nakamura, & Okada, 2018) or ADMIS (Dudzik et al., 2017) among others. Using these software tools the main objective is to detect the thousands of signals through the chromatogram and obtaining chromatographic peaks at different $m/z$. The peak picking and deconvolution allow to detect relevant $m/z$ values associated with a specific retention time, i.e. with a specific component, with a minimum area or intensity. When all samples have been submitted to peak picking, a retention time alignment is performed to match the peaks across the samples (Dudzik, Barbas-Bernardos, García, & Barbas, 2018).

PARAllel FACtor Analysis2 (PARAFAC2) (Johnsen, Amigo, Skov, & Bro, 2014) based Deconvolution and Identification System (PARADISe) (Johnsen, Skou, Khakimov, & Bro, 2017) have appeared recently as new and innovative application for GC-EI-MS data processing. Differently to other tools, such as XCMS or MzMine, PARADISe performs automatic tentative peak identification based on deconvoluted EI mass spectra in combination with the NIST library. Therefore, it reduces the data matrix as well as the time consumed in statistical analysis and elucidation steps.

In the present study, a volatolomics approach based on GC/MS analysis has been applied to develop a classification model to differentiate between fish product samples (tuna and swordfish) that were submitted to different smoking processes, with modifications in the type of treatment and its intensity. An untargeted volatolomics approach has been applied in this work, contrarily to the works reported until now on VOCs in smoked fish that were based on targeted strategies.

## 2. Materials and methods

### 2.1. Chemicals and reagents

Internal standard toluene-D8 ($\geq$99%) was purchased from Sigma Aldrich (Madrid, Spain). Hexane (trace Analysis quality (AT) GC) was provided by Scharlau (Barcelona, Spain). External standards of volatile compounds, supplied by Supelco (Sigma-Aldrich and Fluka; Barcelona, Spain) as pure compound (92–99.5%) were applied for signal deviation correction purposes and identity confirmation: Toluene, hexanal, butyl acetate, trans-2-hexen-1-al, benzaldehyde, 3-carene, β-cyclocitral, eugenol, β-ionone, 2-isobutylthiazole, (Z)-2-heptenal, α-terpineol, ethylbenzene, (+/−)-3-methylcyclopentanone, acetoin (3-hydroxy-2-butanone) , 1-hydroxy-2-butanone, 2-furyl methyl ketone (2-acetylfuran), 2-methylbenzofuran, furfuryl alcohol (2-furanmethanol), 2-methyl-2-cyclopenten-1-one, acetophenone, 2-methoxy-phenol (guaicol) and 2-methoxy-4-vinylpehnol (4-vinyl guaiacol). Tenax® TA TDU glass thermal desorption tubes 60/80mesh, O.D. 6 mm × 4 mm (i.d.) × L 60 mm, used as traps were purchased from Gerstel (Mülheim an der Ruhr, Germany)

### 2.2. Smoked fish samples

The smoked fish samples were provided by Sea Delight Europe, SL. This company has patented a new method of cold smoked where the temperature of the process is maintained at 4 °C with the objective of avoiding the generation of histamine in blue fish, which appears at temperatures above 4.4 °C. With this method, three types of Cold Smoked seafood products are produced: (i) light smoke grade: the time of flavoured wood smoke exposition is short and the smoked flavour obtained is very light; (ii) medium smoke grade: the exposition to the flavoured wood smoke is larger and the flavour obtained is moderate;

(iii) full cure (classic) smoked grade: the product has strong flavour and aroma due to the 48 h curation. Sea Delight produces CO and Tasteless smoked seafood products that are commercialised in USA and Canada, respectively.

A total of 300 samples were used: 20 samples of tuna and 20 samples of swordfish for the Light Cold Smoke (LCS) treatment, and 26 tuna and 26 swordfish samples for each of the other smoking treatments: Tasteless smoke (TS), Carbon monoxide smoke (CO), Full Cure Cold Smoke (FCS), Medium Cold Smoke (MCS) and raw samples (no smoking treatment) (NAT). Samples were stored in freezer at $-25\ °C$ until the extraction.

### 2.3. Purge-and-trap extraction

Smoked fish samples were defrosted at room temperature (24 °C) and triturated before extraction. Then, 5 g of sample were weighed into a 150 mL conical flask. The volatile's extraction procedure was based on our previous works (Beltran et al., 2006; Sales et al., 2019). The flask was immediately closed with a glass tap with two connection tubes: one for the dry $N_2$ gas entrance and the other for the exit connected to the sorbent Tenax® TA TDU trap tubes (Fig. S1). The sorbent trap tubes were previously spiked with 10 μL of 50 μg mL$^{-1}$ internal standard to correct for potential extraction deviations. Sample extraction was carried out at 40 °C for 60 min (immersed in a water bath) with a dry nitrogen (99.7%) flow of 100 mL min$^{-1}$ to perform the purge process. Finally, the sorbent trap tubes were thermally desorbed with the aid of a TDU into the GC/MS.

The samples were randomly analysed in order to avoid bias in the methodology, performing 18 extractions per day of the samples defrosted immediately before extraction. Quality Control (QCs) samples commonly used in metabolomics (i.e. a pool of samples to monitor the performance of the metabolomic workflow) could not be performed due to the difficulty to achieve an average representative and homogeneous mix of samples, and to the absence of sample extracts because extraction was made directly in phase gas. Alternatively, replicate thermal desorption traps were spiked with 10 μL of a mix of volatile compounds at 50 μg mL$^{-1}$, which were processed at the beginning and at the end of the sequence batch, and every 6 samples, for correction of the instrument deviation.

### 2.4. GC-EI-MS analysis

An Agilent 6890 Plus Series gas chromatograph coupled to a quadrupole mass spectrometer, Agilent 5973 N Mass Selective Detector, with an electron ionization (EI) source and MPS2 autosampler from Gerstel (Linthicum, MD, USA) was used for VOCs analysis. The GC separation was carried out on a 30 m × 0.25 mm DB-WAXETR (0.25 μm film thickness) capillary column (J&W Scientific, Folsom, CA, USA), with helium at a constant flow of 1 mL min$^{-1}$ as carrier gas. The column temperature program started at 40 °C for 3 min; then increased to 160 °C at 5 °C min$^{-1}$ and held for 2 min; then increased to 260 °C at 40 °C min$^{-1}$ and held for 1.50 min (total chromatographic run 32 min).

The injection system comprised two devices; a thermal desorption unit (TDU) and CIS 4 PTV injector. The sorbent traps, used for sample extraction, were thermally desorbed in the TDU in splitless mode using a desorption program that started at 50 °C (1 min equilibrium time), then increased to 260 °C at 12 °C s$^{-1}$ and held for 8 min; the transfer line temperature was 260 °C. The CIS4 PTV was equipped with a Tenax® TA packed liner and temperature program started at 40 °C during 1 min and then the temperature increased at 12 °C s$^{-1}$ to 260 °C and held for 8 min. In Fig. S2 a total ion chromatogram obtained from a LCS tuna sample after the described method is shown. As it can be observed, proper chromatographic peak shape is obtained even for the early eluting peaks, although no crio-focusing was used by means of a gap column.

### 2.5. Data treatment

The GC/MS data, acquired in full SCAN mode, were converted to ". cdf" data format thanks to the Chemstation® (Agilent) *export to .AIA* function and pre-processed using the PARADISe software. After loading the exported data to PARADISe software, around 150 time intervals or regions of interest (ROIs) were defined manually along the chromatogram, taking into account the peak shape in total ion chromatogram, when visible, and avoiding empty spaces between intervals. For each ROI, software calculates a model with a maximum of 8 components and 50,000 maximum iterations in order to resolve the underlying and, possibly, overlapping compounds. Once the model for each interval was created it was optimized with the selection of as many compounds as possible, providing a model fitting and model consistency over 95% as well as background removal and avoiding model overfitting. The final report was created with the list of compounds and their peak area for each sample in *.xls* format. Since a mixture of external standards were injected every 6 samples, the peak areas were normalized with the area of the closest compound in retention time of the nearest external standard mixture injection, to correct the differences due to instrumental drift, to finally be scaled applying pareto-scaling (van den Berg, Hoefsloot, Westerhuis, Smilde, & van der Werf, 2006).

The statistical multivariate analysis was performed with MATLAB environment (version R2013a, The Mathworks, Natick, MA) along with PLS_Toolbox (version 7.5.2, Eigenvector Research, Wenatchee, WA).

## 3. Results and discussion

### 3.1. Volatile extraction procedure performance

Both, direct HS injection and DHS-P&T, were tested for extraction of VOCs. Both systems had been previously used in our laboratory for extraction of volatile compounds in food or food products (Beltran et al., 2006; Fredes et al., 2016; Sales et al., 2019). Although HS-SPME has demonstrated to be a valuable technique for volatile extraction, automatization equipment connected to the GC/MS was not available in our laboratory. Therefore, direct HS injection and DHS-P&T were assayed analysing three replicates of each type of smoked tuna (LCS, MCS, FCS, TS and CO) and raw tuna (NAT) under the same conditions, performing the subsequent analysis by GC/MS in full scan mode.

The results were clearly better (both in terms of sensitivity and number of detected compounds) with the DHS-P&T. The performance obtained with HS static procedure was poorer as regards the number of peaks and the sensitivity reached, which was far below the DHS-P&T which sensitivity was favoured by the pre-concentration factor of the dynamic process, especially for those compounds with low vapour pressure. This fact is in accordance with previous studies (Beltran et al., 2006; Fredes et al., 2016), where a more exhaustive comparison of the available sample treatment for VOCs extraction were performed, as in Sales et al., 2019 (Sales et al., 2019) where the extraction performance of DHS-P&T was tested by obtaining up to 1000 times more sensitivity for most of volatile components compared with a static technique headspace-stir bar sorptive extraction (SHS-SBSE). All together demonstrates the advantages of DHS-P&T over the static headspace techniques.

### 3.2. Processing of GC/MS data with PARADISe software

Data processing started with GC/MS data conversion to ".cdf" (netCDF) format using Chemsation® by Agilent Technologies. Due to its potential in gas chromatography applications and based on our previous experience in our laboratory, PARADISe (Johnsen et al., 2014, 2017; Khakimov et al., 2016) was used for peak picking and subsequent alignment of retention times to match the peaks across samples.

To this purpose, the chromatogram was first divided into 150 ROIs and each ROI was individually modelled using the PARAFAC2 algorithm

(Harshman, 1972) for peak deconvolution based on the mass spectra and the intensity of the signals. The model validation was conducted following Khakimov et al. (Khakimov et al., 2016) recommendations. For each ROI, the model fitting was tested to a maximum of eight components, selecting the optimal number of components based on a good model fit and core consistency (both over 95%), noise removal and low residuals, but avoiding model overfitting while obtaining well resolved peaks. Among the components picked, only the ones with a robust match with the NIST08 mass spectral library were retained in the final peak table, removing those coming from the baseline or column bleed. The models optimized by PARADISe using the 300 samples ended in a total of 107 components tentatively identified, recorded with their area in an .*xls* data table. Its capability for spectra deconvolution (even with unit mass resolution), for distinction between the signals from baseline or column bleeding, and for co-eluting components detection, allow PARADISe to get a consistent data matrix that simplifies the following steps, including the statistical analysis. For further information about PARADISe processing see (Sales et al., 2019).

### 3.3. Discriminant analysis by multivariate statistics

After internal standard normalization and pareto-scaling of the peak data, the next step was the multivariate statistical analysis. After the aforementioned data transformation, the dataset was divided into two groups: 80% of samples (considering each smoking treatment) for model training and the remaining 20% for model validation.

As an unbiased data exploratory analysis, a principal components analysis (PCA) was carried out considering the training data set. Fig. 1 shows the score plot of the first two components of the same PCA, where PC1 and PC2 explain the 43.95% and 9.94% of the variance respectively, labelled by a) the two species of fish analysed and b) the different smoking treatments. From Fig. 1a, there was not significant inherent separation between the samples of the two fish species studied (tuna or swordfish). Thus, all the samples from the same treatment can be grouped in order to study the effect of the different treatments, regardless of the fish species. Regarding Fig. 1b, intrinsic separation between the smoking treatments can be observed. The full cure (FCS) and medium (MCS) cold smoke treatment groups (CS) could be distinguished easily (green squares and light blue triangles, respectively) from the other samples (the non-cold smoke group, No-CS) along the first component (PC1). However, Light Cold Smoke (LCS) (dark blue triangles) could not be differentiated from the rest of the low to no treatment samples: Tasteless (TS), CO and Untreated fish (NAT).

After PCA analysis, the partial least squares discriminant analysis (PLS-DA) was applied, which considers additional information about the

groups to be classified (including the smoking treatment in the input information) (Fig. 2). The PLS-DA score plot of the latent variables 1 and 2 (LV1 vs LV2) (Fig. 2a) showed a clear separation between FCS and MCS from the remaining groups (and between them) along the LV1 axis. However, LCS group was still close to the non-cold smoke groups (TS, CO and NAT) although its differentiation improved regarding PCA since a gradual separation along LV2 could start to be observed. This distinction between the LCS and the No-CS groups was better noticed in the PLS-DA 3D (Fig. 2b and c).

In order to get information on relevant compounds related to the cold smoking process, PLS-DA between two groups was applied: the target class was "CS" (FCS, MCS and LCS) and the non-target class was "No-CS" (TS, CO and NAT). In order to verify the accuracy of the model, the classification of the samples for the validation set was performed directly by the software. The PLS-DA was built using 4 latent variables (LV) and it explained 57.52% variance. The classification plot obtained from this PLS-DA model is shown in Fig. 3, where samples were labelled by CS vs No-CS (Fig. 3a) and by the different smoking processes (Fig. 3b). It can be observed that all No-CS samples were at the same level, and they could be attributed a similar (and low) organoleptic load. Regarding the CS group, there is a slight variation depending on the time of exposure to smoke and therefore on the intensity level of the smoke flavour and odour, being the FCS samples at the top, MCS at lower level and LCS just above the threshold. This implies that LCS has low intensity organoleptic properties associated to cold smoke, but analysing the volatile composition it can be differentiated from other treatments that are not allowed in the European Union (TS and CO).

Confusion matrix summarizes the PLS-DA results of the training and validation set of samples (Fig. 4). The model was able to correctly classify 100% of the Full Cure cold smoke, the 97.6% of the Medium cold smoke and 87.5% of the Light cold smoke of the training samples. This model was then evaluated with the remaining 20% of the samples, where 100% of the Full Cure and Medium and 87.5% of the Light cold smoked samples selected for evaluation were correctly classified as "Cold Smoked". It is worth noting that this model was highly efficient because all different CS treatments could be differentiated from other NO-CS processes that are forbidden in EU.

### 3.4. Volatile profile-based classification model building and evaluation for different seafood treatments

The classification model was made with all 107 variables present in the samples. To generate a simpler model easier to apply in future targeted routine approach, most significant variables were selected according to the VIP score (Variable Importance in the Projection). This
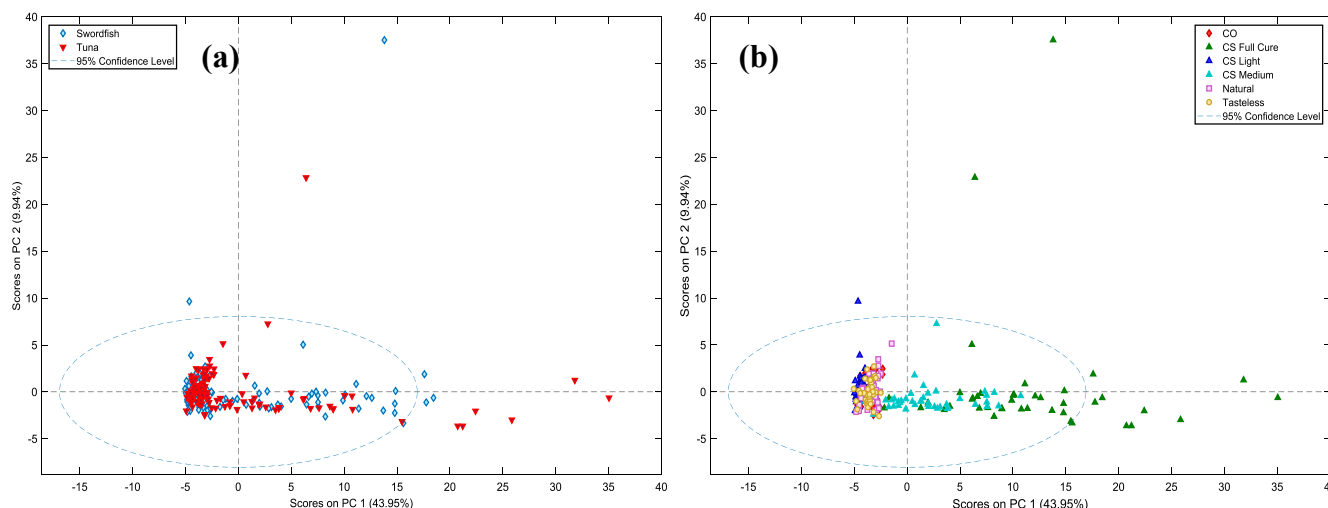


**Fig. 1.** PCA score plots of the acquired data for the method training (a) coloured by fish species and (b) by smoking treatments.
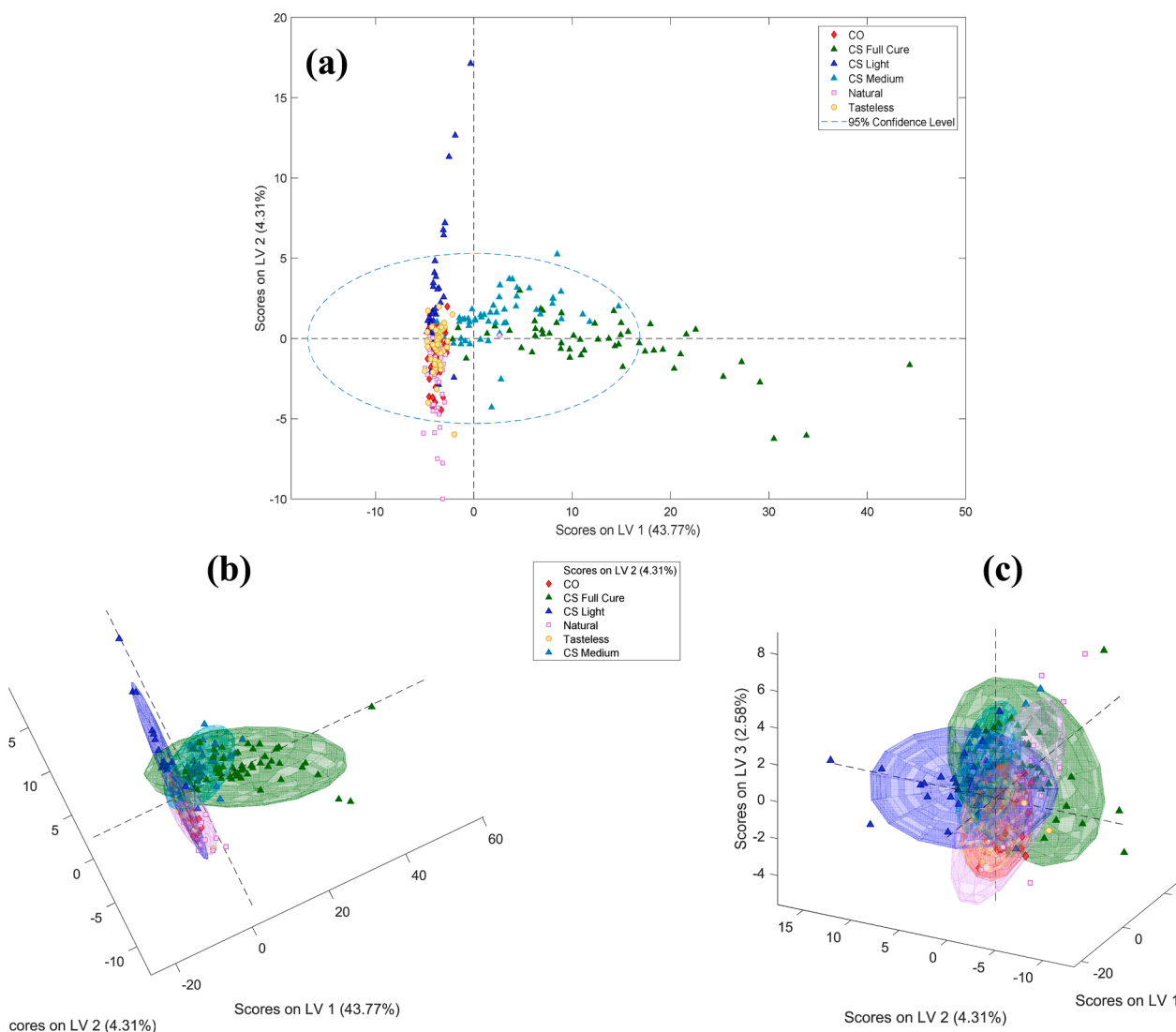
**Fig. 2.** PLS-DA score plots of the acquired data for the method training (a) 2D plot plane LV1vs LV2 and (b) and (c) 3D score plots for LV1, LV2 and LV3.

score summarizes the contribution a component makes to the PLS-DA model. In order to select as few variables as possible, the 29 features with VIP ≥ 1 were firstly selected to build the PLS-DA model. On the basis of the satisfactory results obtained, the reduction of variables, applying the threshold VIP ≥ 1 for the new scores, continued progressively, until the model failed. Finally, the number of variables could be reduced down to 11 with the model still remaining consistent. The Table 1 lists the main components selected as tentative markers for subsequent confirmation.

Considering only the 11 tentative markers selected a new model was built and the classification plot is shown on Fig. 5 and the confusion matrix on Fig. 6. With the 11 variables model, the group of samples are slightly (positively or negatively) affected by the variable reduction, being the Light cold smoke the most affected (81.3% were correctly classified in the training samples in comparison to the 87.5% of the previous model). Nevertheless, this is still an acceptable result and it allowed a satisfactory classification of the samples randomly selected for evaluation. Thus, 100% FCS and MCS, and 87.5% of the LCS samples, were satisfactorily classified as "Cold Smoke" samples with the developed model, reaching the same results as for the 107 classification model.

### 3.5. Cold smoke-related compounds identification in the simplified classification model

Identification of the 11 compounds used in the simplified model is crucial for the development of targeted methods in future works, and for validation of the classification model by analysing a large amount of samples and blind samples prior to its routine analysis application.

An interesting contribution of PARADISe software is that automatically performs a comparison between the deconvoluted spectra and the NIST EI mass spectra library (in this case NIST08) giving the best fitted candidate. In order to increase the confidence in the identification, Linear Retention Indices (LRIs) were calculated for each compound using a C7-C20 alkane mixture. The tentative identification for a compound was assigned when the NIST match for this compound was over 800; and the LRI match with the NIST library was below ± 20. Finally, the identity of the 11 markers was confirmed by the injection of reference standards under sample identical conditions, and subsequent comparison of spectra and retention times, according to the criteria of our laboratory and the *Chemical Analysis Working group (CAWG) Metabolomics Standards Initiative (MSI)* (Sumner et al., 2007). Results are shown in Table 1 with their molecular formula, the detected molecular ion, the NIST match, retention time and LRI. In the case of the 4-vinyl-guaiacol, the retention time fell out of the C20 alkane range, thus the LRI could not be calculated. However, its identity was confirmed with the
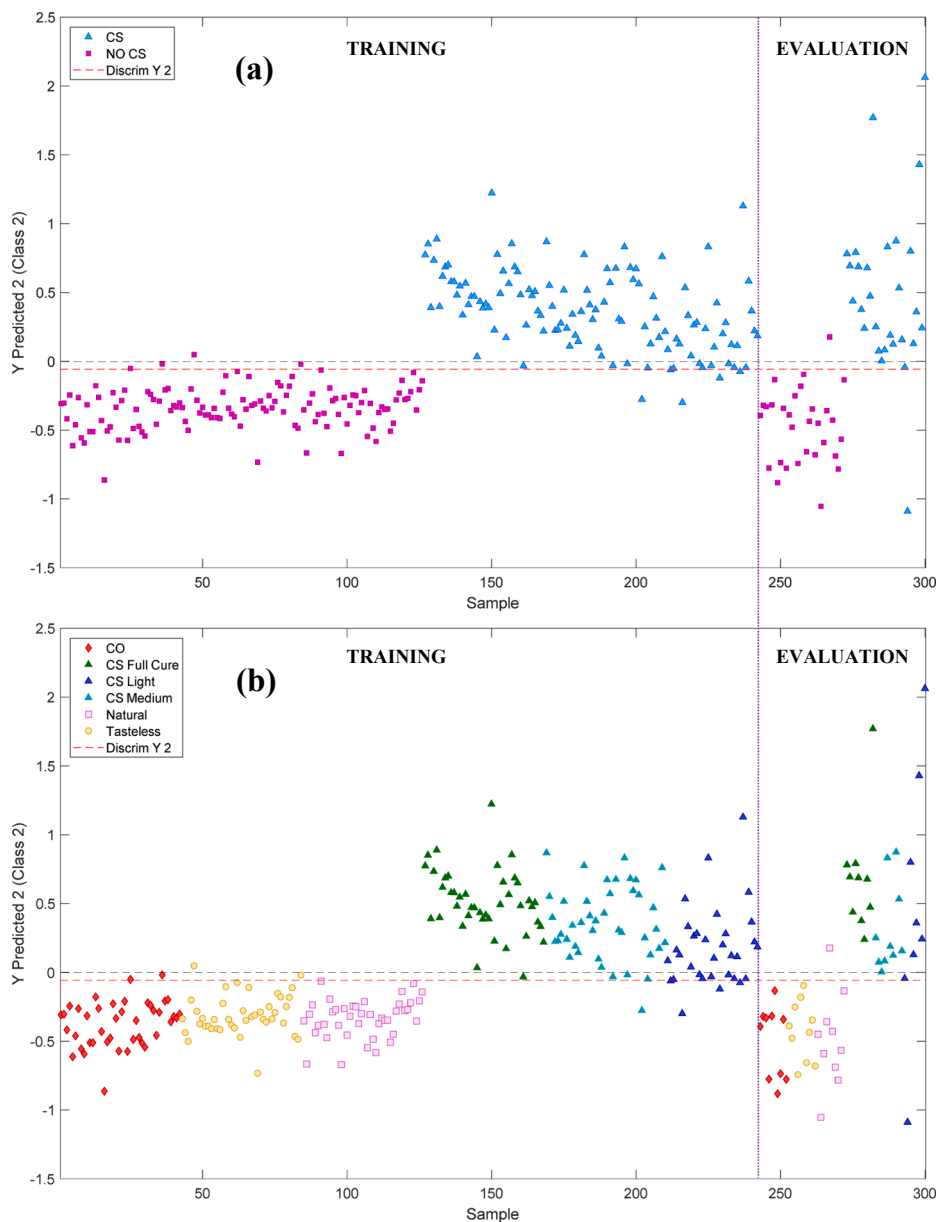
**Fig. 3.** Model sample classification with the training and the evaluation set of samples coloured (a) by Cold smoked or non-cold smoked and (b) by the different smoking treatments.

reference standard injection. Our results were compared with the available literature in the field of organoleptic cold smoke composition, which reports that ketones, phenolic derivatives and furan derivative compounds are associated with smoke flavour and odour (Gómez-Estaca, Gómez-Guillén, Montero, Sopelana, & Guillén, 2011; M.D. Guillén, Errecalde, Salmerón, & Casas, 2006; Jónsdóttir et al., 2008). The compound with higher importance in our 11 variables classification model was 3-methyl-cyclopentanone, previously identified as wood smoke component together with other ketones such as 2-methyl-2-cyclo-penten-1-one, 3-hydroxy-2-butanone, 1-hydroxy-2-butanone and ace-tophenone (Vidal et al., 2017). Guaicol and guaicol derivatives, such as 4-vinyl guiacol has been previously reported as derived from wood py-rolysis and with a high importance in the smoke flavour grade. Associated with this property are also furan derivatives compounds, such as 2-acetylfuran, 2-methyl-benzofuran and 2-furanmethanol (Jónsdóttir et al., 2008). Finally, hydrocarbons, like ethylbenzene, part of the wood smoke constituents (María D. Guillén & Errecalde, 2002) were also found as marker although no references about its flavour contribution

was found in the literature.

## 4. Conclusions

The use of untargeted volatolomics based on DHS-P&T for volatile extraction and subsequent analysis by GC/MS has allowed to obtain relevant information regarding the volatile composition of smoked fish samples, responsible for the classification of smoking technique applied. The use of PARADISe has allowed robust peak detection, cleaner spectra and, in combination with NIST libraries, an efficient tentative identifi-cation. Using this methodology a classification model has been devel-oped able to distinguish samples with "Cold Smoked" treatment (Full Cure, Medium and Light smoked) from those without "Cold Smoked" treatment (Tasteless, CO and untreated), and it has allowed to build a consistent statistical model for correct classification. The model built with all the 107 detected compounds allowed the correct classification of 96.3% of the blind samples, while using the simplified model, based on only 11 identified compounds, 95% of the blind samples were still

| | | OBJECTIVE | |
|---|---|---|---|
| | | **NO Cold Smoked** | **Cold Smoked** |
| NO CS | CO | 100 % | 0 |
| | Tasteless | 100 % | 0 |
| | Natural | 100 % | 0 |
| CS | CS Full Cure | 0 | 100 % |
| | CS Medium | 0 | 100 % |
| | CS Light | 0 | 100 % |

| | TRAINING SET | | | | | EVALUATION SET | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **NO Cold Smoked** | | **Cold Smoked** | | | **NO Cold Smoked** | | **Cold Smoked** | |
| CO | 40/42 | 95.2 % | 2/42 | 4.8 % | CO | 10/10 | 100 % | 0/10 | 0 % |
| Tasteless | 40/42 | 95.2 % | 2/42 | 4.8 % | Tasteless | 10/10 | 100 % | 0/10 | 0 % |
| Natural | 41/42 | 97.6 % | 1/42 | 2.4 % | Natural | 9/10 | 90 % | 1/10 | 10 % |
| CS Full Cure | 0/42 | 0 % | 42/42 | 100 % | CS Full Cure | 0/10 | 0 % | 10/10 | 100 % |
| CS Medium | 1/42 | 2.4 % | 40/42 | 97.6 % | CS Medium | 0/10 | 0 % | 10/10 | 100 % |
| CS Light | 4/32 | 12.5 % | 28/32 | 87.5 % | CS Light | 1/8 | 0 % | 7/8 | 87.5 % |

**Fig. 4.** Confusion matrix showing the comparison between the objective (up) of sample classification and the results (down) Training and evaluation set of samples after processing the samples through the entire developed procedure. The classification model was developed based on all the variables obtained (model score −0.05).

**Table 1**
GC/MS measurements for the identified markers for the Cold smoke reduced classification model.

| VIP value | MARKER | Molecular formula | M+· (*m/z*) | NIST Match | $t_R$ (min) | Linear Retention Index (LRI) Experimental | LRI Reported in NIST library (*) |
|---|---|---|---|---|---|---|---|
| 2.50 | 3-methyl-cyclopentanone | C6H10O | 98 | 802 | 10.86 | 1210 | 1218 |
| 2.33 | Ethylbenzene | C8H10 | 106 | 911 | 8.51 | 1125 | 1122 |
| 1.94 | 2-methyl-2-cyclopenten-1-one | C6H8O | 96 | 975 | 15.13 | 1366 | 1367 |
| 1.94 | 2- methyl-benzofuran | C9H8O | 132 | 820 | 20.91 | 1593 | 1589 |
| 1.92 | Furfuryl alcohol | C5H6O2 | 98 | 945 | 22.46 | 1659 | 1659 |
| 1.85 | 2-acetylfuran | C6H6O2 | 110 | 897 | 18.73 | 1504 | 1501 |
| 1.85 | Acetophenone | C8H8O | 120 | 844 | 22.24 | 1649 | 1649 |
| 1.67 | Guaiacol | C7H8O2 | 124 | 936 | 26.91 | 1863 | 1862 |
| 1.56 | 1-hhydroxy-2-butanone | C4H8O2 | 88 | 874 | 15.42 | 1376 | 1375 |
| 1.48 | 4-vinylguaicol | C9H10O2 | 150 | 833 | 30.06 | (**) | (**) |
| 1.01 | Acetoin | C4H8O2 | 88 | 827 | 12.93 | 1285 | 1285 |

(*) The Linear Retention Index (LRI) were obtained for each compound from NIST Library (https://webbook.nist.gov/) according to the most similar column and chromatographic conditions.
(**)The retention time fell out of the alkane range, the Linear Retention Index (LRI) could not be calculated.

correctly classified. The confirmation of the identity of these 11 compounds with their reference standards will allow in the near future to develop a targeted method to be implemented in routine analysis.

The model was developed to classify fish flesh treated with cold smoking at 4 °C and other non-cold smoked treatments. In this work, relative chromatographic areas, referred to internal standard, were used for classification purposes. For future routine applications, reference standards should be used in order to calculate the concentration of each marker in the samples, following a simple target quantitative approach.

The possibility to apply the developed methodology to other fish flesh samples treated with cold smoking is promising but should be further studied and validated. Thus, the markers proposed for the cold smoked seafood analysed in this work may not be the most suitable for other cold smoking processes performed by other companies. This is due to the variability between the cold smoke treatments applied (type of wood, exposure time, temperature, salinity…).

**CRediT authorship contribution statement**

**Leticia Lacalle-Bergeron:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Funding acquisition. **Tania Portolés:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Carlos Sales:** Methodology, Software, Formal analysis, Investigation, Visualization. **M. Carmen Corell:** Conceptualization, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project administration, Funding acquisition. **Fernando Domínguez:** Conceptualization, Resources, Writing - review & editing, Visualization, Project administration. **Joaquim Beltrán:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft. **Juan Vicente Sancho:** Conceptualization, Methodology, Formal analysis, Resources, Data curation, Writing -
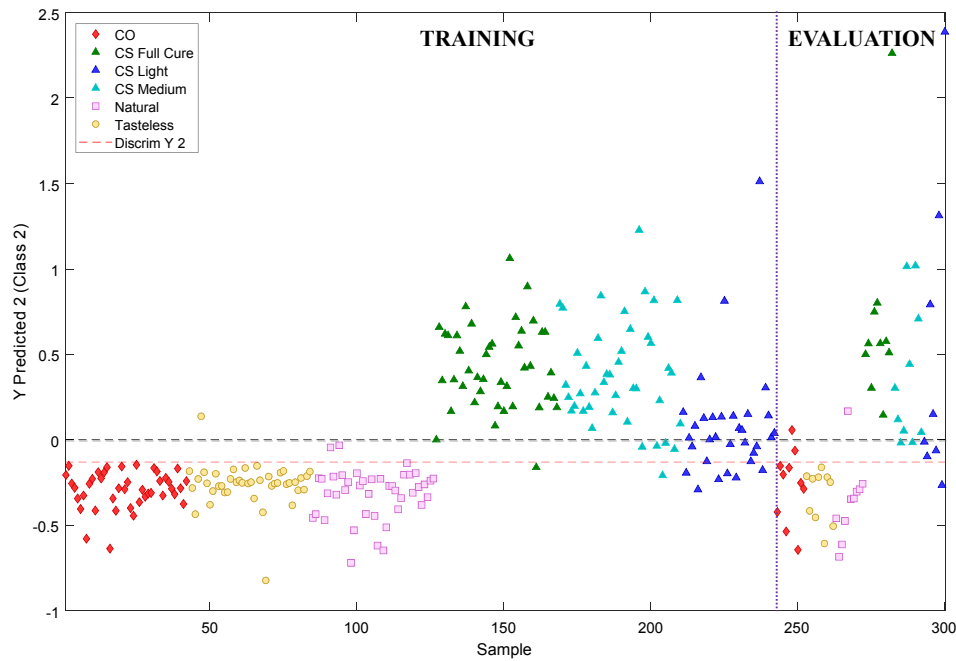
**Fig. 5.** Model sample classification with 11 variables with the training and the evaluation set of samples coloured by the different smoking treatments.

| | | OBJECTIVE | |
|---|---|---|---|
| | | **NO Cold Smoked** | **Cold Smoked** |
| NO CS | CO | 100 % | 0 |
| | Tasteless | 100 % | 0 |
| | Natural | 100 % | 0 |
| CS | CS Full Cure | 0 | 100 % |
| | CS Medium | 0 | 100 % |
| | CS Light | 0 | 100 % |

| TRAINING SET | | | | | EVALUATION SET | | | |
|---|---|---|---|---|---|---|---|---|
| | **NO Cold Smoked** | | **Cold Smoked** | | | **NO Cold Smoked** | | **Cold Smoked** |
| CO | 42/42 | 100 % | 0/42 | 0 % | CO | 8/10 | 80 % | 2/10 | 20 % |
| Tasteless | 41/42 | 97.6 % | 1/42 | 2.4 % | Tasteless | 10/10 | 100 % | 0/10 | 0 % |
| Natural | 40/42 | 95.2 % | 2/42 | 4.8 % | Natural | 9/10 | 90 % | 1/10 | 10 % |
| CS Full Cure | 1/42 | 2.4 % | 41/42 | 97.6 % | CS Full Cure | 0/10 | 0 % | 10/10 | 100 % |
| CS Medium | 1/42 | 2.4 % | 40/42 | 97.6 % | CS Medium | 0/10 | 0 % | 10/10 | 100 % |
| CS Light | 6/32 | 18.7 % | 26/32 | 81.3 % | CS Light | 1/8 | 12.5 % | 7/8 | 87.5 % |

**Fig. 6.** Confusion matrix showing the comparison between the objective (up) of sample classification and the results (down) Training and evaluation set of samples after processing the samples through the entire developed procedure. The classification model was developed based on the 11 most relevant variables obtained (model score −0.1).

original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Félix Hernández:** Validation, Resources, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary material

Experimental design used for DHS-P&T extraction for smoked fish samples. Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodres.2020.109698.

## References

Abou-el-karam, S., Ratel, J., Kondjoyan, N., Truan, C., & Engel, E. (2017). Marker discovery in volatolomics based on systematic alignment of GC-MS signals: Application to food authentication. *Analytica Chimica Acta, 991*, 58–67. https://doi.org/10.1016/j.aca.2017.08.019.

Barstad, H., Alvik, A. C., & Løvaas, E. (2006). Seafood research from fish to dish. In J. B. Luten, C. Jacobsen, K. Bekaert, A. Sæbø, & J. Oehlenschläger (Eds.), *Seafood research from fish to dish: Quality, safety and processing of wild and farmed seafood.* Wageningen Academic Publishers. https://doi.org/10.3920/978-90-8686-581-9.

Bartolucci, G., Droghetti, E., Focardi, C., Bambagiotti-Alberti, M., Nocentini, M., & Smulevich, G. (2010). High throughput headspace GC-MS quantitative method to measure the amount of carbon monoxide in treated tuna fish. *Journal of Mass Spectrometry, 45*(9), 1041–1045. https://doi.org/10.1002/jms.1798.

Beaulieu, J. C., & Lea, J. M. (2006). Characterization and semiquantitative analysis of volatiles in seedless watermelon varieties using solid-phase microextraction. *Journal of Agricultural and Food Chemistry, 54*(20), 7789–7793. https://doi.org/10.1021/jf060663l.

Beltran, J., Serrano, E., López, F. J., Peruga, A., Valcarcel, M., & Rosello, S. (2006). Comparison of two quantitative GC–MS methods for analysis of tomato aroma based on purge-and-trap and on solid-phase microextraction. *Analytical and Bioanalytical Chemistry, 385*(7), 1255–1264. https://doi.org/10.1007/s00216-006-0410-9.

Ben Brahim, S., Amanpour, A., Chtourou, F., Kelebek, H., Selli, S., & Bouaziz, M. (2018). Gas chromatography-mass spectrometry–olfactometry to control the aroma fingerprint of extra virgin olive oil from three tunisian cultivars at three harvest times. *Journal of Agricultural and Food Chemistry, 66*(11), 2851–2861. https://doi.org/10.1021/acs.jafc.7b06090.

Bouhlel, J., Ratel, J., Abouelkaram, S., Mercier, F., Travel, A., Baéza, E., … Engel, E. (2017). Solid-phase microextraction set-up for the analysis of liver volatolome to detect livestock exposure to micropollutants. *Journal of Chromatography A, 1497*, 9–18. https://doi.org/10.1016/j.chroma.2017.03.008.

Broza, Y. Y., Mochalski, P., Ruzsanyi, V., Amann, A., & Haick, H. (2015). Hybrid volatolomics and disease detection. *Angewandte Chemie International Edition, 54*(38), 11036–11048. https://doi.org/10.1002/anie.201500153.

Dalgaard, P., Emborg, J., Kjølby, A., Sørensen, N. D., & Ballin, N. Z. (2008). Histamine and biogenic amines: formation and importance in seafood. In Improving seafood products for the consumer (pp. 292–324). Elsevier. https://doi.org/10.1533/9781845694586.3.292.

Dettmer, K., Aronov, P. A., & Hammock, B. D. (2007). Mass spectrometry-based metabolomics. *Mass Spectrometry Reviews, 26*(1), 51–78. https://doi.org/10.1002/mas.20108.

Dirinck, P., Schreyen, L., & Schamp, N. (1977). Aroma quality evaluation of tomatoes, apples, and strawberries. *Journal of Agricultural and Food Chemistry, 25*(4), 759–763. https://doi.org/10.1021/jf60212a039.

Dudzik, D., Barbas-Bernardos, C., García, A., & Barbas, C. (2018). Quality assurance procedures for mass spectrometry untargeted metabolomics. a review. *Journal of Pharmaceutical and Biomedical Analysis, 147*, 149–173. https://doi.org/10.1016/j.jpba.2017.07.044.

Dudzik, D., Zorawski, M., Skotnicki, M., Zarzycki, W., García, A., Angulo, S., … Ramos, M. P. (2017). GC–MS based gestational diabetes mellitus longitudinal study: Identification of 2-and 3-hydroxybutyrate as potential prognostic biomarkers. *Journal of Pharmaceutical and Biomedical Analysis, 144*, 90–98. https://doi.org/10.1016/j.jpba.2017.02.056.

Fredes, A., Sales, C., Barreda, M., Valcárcel, M., Roselló, S., & Beltrán, J. (2016). Quantification of prominent volatile compounds responsible for muskmelon and watermelon aroma by purge and trap extraction followed by gas chromatography–mass spectrometry determination. *Food Chemistry, 190*, 689–700. https://doi.org/10.1016/j.foodchem.2015.06.011.

Garcia, A., & Barbas, C. (2011). Gas chromatography-mass spectrometry(GC-MS)-based metabolomics. In T. O. Metz (Ed.), Metabolic profiling, methods and protocols (Vol. 708, pp. 191–204). https://doi.org/10.1007/978-1-61737-985-7.

Gil-Solsona, R., Raro, M., Sales, C., Lacalle, L., Díaz, P., Ibáñez, M., … Hernández, F. J. (2016). Metabolomic approach for Extra virgin olive oil origin discrimination making use of ultra-high performance liquid chromatography – Quadrupole time-of-flight mass spectrometry. *Food Control, 70*, 350–359. https://doi.org/10.1016/j.foodcont.2016.06.008.

Gokoglu, N. (2020). Innovations in seafood packaging technologies: A review. *Food Reviews International, 36*(4), 340–366. https://doi.org/10.1080/87559129.2019.1649689.

Gómez-Estaca, J., Gómez-Guillén, M. C., Montero, P., Sopelana, P., & Guillén, M. D. (2011). Oxidative stability, volatile components and polycyclic aromatic hydrocarbons of cold-smoked sardine (*Sardina pilchardus*) and dolphinfish (*Coryphaena hippurus*). *LWT - Food Science and Technology, 44*(6), 1517–1524. https://doi.org/10.1016/j.lwt.2011.02.006.

Guillén, M. D., Errecalde, M. C., Salmerón, J., & Casas, C. (2006). Headspace volatile components of smoked swordfish (*Xiphias gladius*) and cod (*Gadus morhua*) detected by means of solid phase microextraction and gas chromatography–mass spectrometry. *Food Chemistry, 94*(1), 151–156. https://doi.org/10.1016/j.foodchem.2005.01.014.

Guillén, María D., & Errecalde, M. C. (2002). Volatile components of raw and smoked black bream (*Brama raii*) and rainbow trout (*Oncorhynchus mykiss*) studied by means of solid phase microextraction and gas chromatography/mass spectrometry. *Journal of the Science of Food and Agriculture, 82*(9), 945–952. https://doi.org/10.1002/jsfa.1128.

Harshman, R. A. (1972). PARAFAC2: Mathematical and technical notes. *UCLA Working Papers in Phonetics, 22*(10), 30–44.

Huang, X. H., Qi, L. B., Fu, B. S., Chen, Z. H., Zhang, Y. Y., Du, M., … Qin, L. (2019). Flavor formation in different production steps during the processing of cold-smoked Spanish mackerel. *Food Chemistry, 286*, 241–249. https://doi.org/10.1016/j.foodchem.2019.01.211.

Johnsen, L. G., Amigo, J. M., Skov, T., & Bro, R. (2014). Automated resolution of overlapping peaks in chromatographic data. *Journal of Chemometrics, 28*(2), 71–82. https://doi.org/10.1002/cem.2575.

Johnsen, L. G., Skou, P. B., Khakimov, B., & Bro, R. (2017). Gas chromatography – mass spectrometry data processing made easy. *Journal of Chromatography A, 1503*, 57–64. https://doi.org/10.1016/j.chroma.2017.04.052.

Jónsdóttir, R., Ólafsdóttir, G., Chanie, E., & Haugen, J. E. (2008). Volatile compounds suitable for rapid detection as quality indicators of cold smoked salmon (*Salmo salar*). *Food Chemistry, 109*(1), 184–195. https://doi.org/10.1016/j.foodchem.2007.12.006.

Khakimov, B., Mongi, R. J., Sørensen, K. M., Ndabikunze, B. K., Chove, B. E., & Engelsen, S. B. (2016). A comprehensive and comparative GC–MS metabolomics study of non-volatiles in Tanzanian grown mango, pineapple, jackfruit, baobab and tamarind fruits. *Food Chemistry, 213*, 691–699. https://doi.org/10.1016/j.foodchem.2016.07.005.

Marušić Radovčić, N., Vidaček, S., Janči, T., & Medić, H. (2016). Characterization of volatile compounds, physico-chemical and sensory characteristics of smoked dry-cured ham. *Journal of Food Science and Technology, 53*(11), 4093–4105. https://doi.org/10.1007/s13197-016-2418-2.

Płotka-Wasylka, J., Szczepańska, N., Owczarek, K., & Namieśnik, J. (2017). Miniaturized solid phase extraction. *Comprehensive Analytical Chemistry, 76*, 279–318. https://doi.org/10.1016/bs.coac.2017.03.001.

Saldaña, E., Saldarriaga, L., Cabrera, J., Siche, R., Behrens, J. H., Selani, M. M., de Almeida, …, Contreras-Castillo, C. J. (2019). Relationship between volatile compounds and consumer-based sensory characteristics of bacon smoked with different Brazilian woods. Food Research International, 119(September 2018), 839–849. https://doi.org/10.1016/j.foodres.2018.10.067.

Sales, C., Cervera, M. I., Gil, R., Portolés, T., Pitarch, E., & Beltran, J. (2017). Quality classification of Spanish olive oils by untargeted gas chromatography coupled to hybrid quadrupole-time of flight mass spectrometry with atmospheric pressure chemical ionization and metabolomics-based statistical approach. *Food Chemistry, 216*, 365–373. https://doi.org/10.1016/j.foodchem.2016.08.033.

Sales, C., Portolés, T., Johnsen, L. G., Danielsen, M., & Beltran, J. (2019). Olive oil quality classification and measurement of its organoleptic attributes by untargeted GC–MS and multivariate statistical-based approach. *Food Chemistry, 271*, 488–496. https://doi.org/10.1016/j.foodchem.2018.07.200.

Sérot, T., Baron, R., Knockaert, C., & Vallet, J. L. (2004). Effect of smoking processes on the contents of 10 major phenolic compounds in smoked fillets of herring (*Cuplea harengus*). *Food Chemistry, 85*(1), 111–120. https://doi.org/10.1016/j.foodchem.2003.06.011.

Soria, Ana C., García-Sarrió, M. J., Ruiz-Matute, A. I., & Sanz, M. L. (2017). Headspace techniques for volatile sampling. *Comprehensive Analytical Chemistry, 76*, 255–278. https://doi.org/10.1016/bs.coac.2017.02.001.

Soria, Ana Cristina, Martínez-Castro, I., & Sanz, J. (2008). Some aspects of dynamic headspace analysis of volatile components in honey. *Food Research International, 41*(8), 838–848. https://doi.org/10.1016/j.foodres.2008.07.010.

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., … Viant, M. R. (2007). Proposed minimum reporting standards for chemical analysis. *Metabolomics, 3*(3), 211–221. https://doi.org/10.1007/s11306-007-0082-2.

Thomsen, B. R., Yesiltas, B., Sørensen, A. D. M., Hermund, D. B., Glastrup, J., & Jacobsen, C. (2016). Comparison of three methods for extraction of volatile lipid oxidation products from food matrices for GC–MS analysis. *JAOCS, Journal of the American Oil Chemists' Society, 93*(7), 929–942. https://doi.org/10.1007/s11746-016-2837-2.

Tomita, S., Nakamura, T., & Okada, S. (2018). NMR- and GC/MS-based metabolomic characterization of sunki, an unsalted fermented pickle of turnip leaves. *Food Chemistry, 258*, 25–34. https://doi.org/10.1016/j.foodchem.2018.03.038.

van den Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K., & van der Werf, M. J. (2006). Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics, 7*. https://doi.org/10.1186/1471-2164-7-142.

Varlet, V., Serot, T., & Prost, C. (2009). Smoke flavoring technology in seafood. In *Handbook of seafood and seafood products analysis* (pp. 233–254). CRC Press. https://doi.org/10.1201/9781420046359-c15.

Vidal, N. P., Goicoechea, E., Manzanos, M. J., & Guillén, M. D. (2017). Effect of smoking using smoke flavorings on several characteristics of farmed sea bass (*Dicentrarchus labrax*) fillets and on their evolution during vacuum-packed storage at refrigeration temperature. *Journal of Food Processing and Preservation, 41*(2), Article e12800. https://doi.org/10.1111/jfpp.12800.