Talanta Open 7 (2023) 100175

Contents lists available at ScienceDirect

Talanta Open

journal homepage: www.sciencedirect.com/journal/talanta-open

Guidelines to build PLS-DA chemometric classification models using a GC-IMS method: Dry-cured ham as a case of study

Andrés Martín-Gómez [a,b], Pablo Rodríguez-Hernández [c], María José Cardador [a,*], Belén Vega-Márquez [d], Vicente Rodríguez-Estévez [c], Lourdes Arce [a]

[a] Department of Analytical Chemistry, Institute of Fine Chemistry and Nanochemistry, Marie Curie Annex Building, University of Córdoba, Campus de Rabanales 14014 Córdoba, Spain
[b] COVAP. S. Coop. And. Ganadera del Valle de los Pedroches, Calle Mayor, 56, 14400 Pozoblanco, Córdoba, Spain
[c] Department of Animal Production, International Agrifood Campus of Excellence (ceiA3), University of Córdoba, Campus de Rabanales 14014 Córdoba, Spain
[d] Department of Computer Languages and Systems, University of Seville, 41012 Seville, Spain

ARTICLE INFO

Keywords:
GC-IMS
PLS-DA
Iberian ham
Multivariate analysis
Chemometrics

ABSTRACT

The number of representative samples to build a calibration model plays a major role in the success of chemometric models for class discrimination; therefore, knowing which samples should be used for the calibration of prediction models is essential. The aim of this work is to design a basic guideline for the training of partial least squares discriminant analysis (PLS-DA) models to classify complex samples analysed by Gas Chromatography (GC) coupled to Ion Mobility Spectrometry (IMS) using dry-cured Iberian ham as an example. The effect of the number, proportion and class of samples for training and validation and the use of two data types (spectral fingerprint or pre-selected markers) has been assessed by analysing with GC-IMS nearly 1000 dry-cured Iberian ham samples obtained from 7 different curing plants. Subsequently, these were classified with PLS-DA according to the pig's feeding regime (acorn-fed vs. feed-fed) and it has been demonstrated that 450 out of 997 samples are enough for model training to achieve a maximum average prediction accuracy rate. Furthermore, the use of pre-selected GC-IMS markers provides slightly better prediction results than the use of the complete spectral fingerprint. In summary, these results represent a tentative guide for the classification of samples in an industrial setting using GC-IMS and PLS-DA. This methodology would allow authorities and producers to ensure the quality of the agri-food products put on the market as is proven in this study.

1. Introduction

Gas Chromatography-Ion Mobility Spectrometry (GC-IMS) is a suitable technique for the analysis of complex matrices using the information of volatile organic compounds (VOCs). The potential of GC-IMS has been widely explored in different fields, food analysis being one of the main ones in which it is being used. Due to the huge amount of data that is contained in GC-IMS spectral plots, non-targeted approaches such as fingerprinting are required when volatilome analysis is used for classification purposes. Several authors have developed analytical methods for the characterisation and classification of food products using non-targeted GC-IMS data analysis of VOCs [1]. These products include fresh [2], cooked [3] smoked [4] and cured [5–7] meat, cheese [8], honey [9,10], olive oil and other vegetal oils [11,12], fish [13,14], vegetables and cultivars [15,16] and drinks [17,18]. The use of chemometrics plays a major role in the success of sample discrimination. Partial Least Squares Discriminant Analysis (PLS-DA) is one of the most widely employed chemometric techniques. The research group of the authors of this article was pioneer in the use of GC-IMS data combined with PLS-DA for the classification of food samples, whose usefulness for the classification of samples has been demonstrated in previous studies with Iberian ham [5,19,20] and olive oil [11,21,22]. Subsequently, many authors have developed similar methodologies [1,23,24]. Nevertheless, no case has described the minimum number of samples required to build a qualitative calibration model in order to classify them and draw conclusions, while guidelines for quantitative models have been indeed developed for infrared (IR) data [25] PLS-DA may construct overly complex models that force a separation between samples based on class membership [26,27] and therefore require external validation to ensure reliability [28,29]. In addition, the class and number of

* Corresponding author.
E-mail address: q22cadum@uco.es (M.J. Cardador).

https://doi.org/10.1016/j.talo.2022.100175
Received 14 October 2022; Received in revised form 6 December 2022; Accepted 14 December 2022
Available online 15 December 2022

training samples employed has a fundamental impact on the prediction power of PLS-DA models [30].

In the present work, guidelines for the best sample selection approach for PLS-DA calibration with GC-IMS data are discussed. As an example, nearly 1000 dry-cured Iberian ham samples obtained by non-destructive means in different curing plants have been used to build PLS-DA models in order to classify them according to the pig's feeding regime (acorn-fed vs. feed-fed). The main objective of this work is to perform a tentative assessment of how many samples are needed to calibrate a PLS-DA model for the classification of blind samples. The best training samples selection approach is also determined, which has never been attempted before with GC-IMS data. The consecution of this objective will serve as a tentative guide to build a basic database for the classification of blind samples. The second objective is to discuss the advantages and disadvantages of each data type (the complete spectral fingerprint or the intensity of pre-selected features) for classification purposes.

2. Materials and method

2.1. Samples and standards

997 subcutaneous fat samples were extracted from Iberian pig dry-cured hams with two feeding regimes during the finishing phase (500 from acorn-fed pigs and 497 from feed-fed pigs). The sampling procedure was non-destructive and consisted of punching a disposable stainless-steel 2.1 × 60 mm sterile needle (*Bovivet-Kruuse*, Langeskov, Denmark) in the rump of the ham to be impregnated with cured fat, as described in a previous study [19]. The metallic part of the impregnated needle was cut off with pliers and placed in a 20 mL glass vial closed with a silicone septum. Sampling was carried out in 7 different curing plants of 2 different regions of Spain and all samples had full traceability. In summary, the following list of samples was used:

- Region A:
  - Curing plant 1: 220 samples (175 acorn-fed and 45 feed-fed pigs).
  - Curing plant 2: 103 samples (31 acorn-fed and 72 feed-fed pigs).
  - Curing plant 3: 46 samples from feed-fed pigs.
  - Curing plant 4: 326 samples (165 acorn-fed and 161 feed-fed pigs).
- Region B:
  - Curing plant 5: 75 samples (50 acorn-fed and 25 feed-fed pigs).
  - Curing plant 6: 94 samples (44 acorn-fed and 50 feed-fed pigs).
  - Curing plant 7: 133 samples (34 acorn-fed and 99 feed-fed pigs).

In addition, 65 individual standards of 1 mg/L were prepared in refined oil supplied by *Sovena S.A.* (Brenes, Spain). The selection of standards was based on lists of VOCs of dry-cured ham described by a previous review [31]. The list included 18 alcohols: (E)-hex-2-en-1-ol, (Z)-pent-2-en-1-ol, 1-penten-3-ol, 2-methylbutan-1-ol, 2-phenylethan-1-ol, 3,3-dimethylbutan-1-ol, 3-methylbutan-1-ol, butan-1-ol, butane-1,3-diol, ethanol, heptan-1-ol, heptan-2-ol, hexan-1-ol, oct-1-en-3-ol, octan-1-ol, pentan-1-ol, phenylmethanol and propan-2-ol; 16 aldehydes: (E)-decen-2-al, (2E)-hepten-2-al, (E)-hexen-1-al, (E)-nonen-2-al, (E)-octen-2-al, 2-methylbutyraldehyde, 2-methylpropanal, 3-methylbutyraldehyde, benzaldehyde, butyraldehyde, decanal, heptanal, hexanal, nonanal, octanal and pentanal: 8 ketones: 6-methyl-5-hepten-2-one, butan-2-one, cyclohexanone, heptan-2-one, hexan-2-one, nonan-2-one, oct-1-en-3-one and pentan-2-one; 5 acids: 3-methylbutanoic, acetic, butanoic, pentanoic and propionic: esters, ethyl esters: 2-methylbutyrate, 2-methylpropanoate, 3-methylbutyrate, acetate, heptanoate, hexanoate, octanoate, pentanoate and propanoate: aromatics: p-xylene, m-xylene, 2,6-dimethylpyrazine and 2-methylpyrazine: terpenes: (*R*)-limonene and α-pinene: and others: dimethyl disulfide, octane and γ-caprolactone.

A quality control mixture was prepared dissolving six high purity (≥ 99%) ketones (nonan-2-one, octan-2-one, heptan-2-one, hexan-2-one, pentan-2-one and butan-2-one) at 0.5 mg·L^{-1} in ultrapure water (Milli-Q Plus, *Millipore Bedford*, MA, USA). This working solution was analysed along with every sample batch for quality control of the analyses performed with the GC-IMS instrument.

2.2. Instrumental method

The GC-IMS device used had an integrated Agilent 7697A headspace sampler connected by a transfer line to an Agilent 8860 gas chromatograph (*Agilent*, Santa Clara, CA, US) and a standalone ion mobility spectrometer (*G.A.S. Gesellschaft für analytische Sensorsysteme mbH*, Dortmund, Germany) with a ^{3}H ionisation source and a 10 cm drift tube. The analysis method was similar to that of a previous paper [32], which consisted of a sample incubation of 15 min at 60 °C. Subsequently, the vial was pressurised at 14 psi and 1 mL (loop) of headspace volume was injected in split mode 1:5. The loop and transfer line were heated at 100 and 110 °C, respectively. GC separation was performed on a 30 m HP-5 (5%-phenyl)-methylpolysiloxane non-polar column with an internal diameter of 0.32 mm and a 0.5 μm film (*Agilent*, Santa Clara, CA, US). Helium (*Abelló Linde*, Sevilla, Spain) was used as the carrier gas at a constant flow rate of 1 mL/min. The temperature ramp consisted of 3 min at 40 °C from the start of the analysis, an increase to 100 °C at a rate of 5 °C/min, an increase to 130 °C at a rate of 15 °C/min, and a 130 °C plateau until the end of the analysis at 27 min. After separation in the GC column, VOCs entered the ionisation chamber of the IMS module, whose detector was working at positive polarity. Nitrogen was used as drift gas at a flow of 150 mL/min. IMS parameters were 150 μs of injection pulse width, 45 °C drift tube temperature, signal averaging each 32 spectra, a repetition rate of 30 ms, and drift, blocking and injection voltages of 237 V, 40 V and 2500 V, respectively.

The 997 fat samples and the 65 individual standards of VOCs were analysed with the GC-IMS method described.

2.3. Data treatment and chemometrics

After sample analysis, GC-IMS data was exported with VOCal 1.0.0. software (*G.A.S. Gesellschaft für analytische Sensorsysteme mbH*, Dortmund, Germany). Then two different data treatment approaches were carried out: approach A, using the intensities (mV) of pre-selected markers, and approach B, using all data available (spectral fingerprint). For approach A, the intensity (mV) of the individual peak of 279 pre-selected features (61 of them corresponding to tentatively identified VOCs and 218 unidentified) were extracted from the GC-IMS spectral plot. Thus, each one of the 279 features had a unique retention time (s) in the GC column and a drift time (ms) in the IMS drift tube. The 61 tentatively identified VOCs were identified comparing their retention and drift times with that of the standards described in Section 2.1.; in order to be completely sure that the feature corresponds to a VOC, analysis with a high-resolution MS would be required. Subsequently, data was imported into Matlab R2016a software (*Mathworks*) with PLS Toolbox plug-in (*Eigenvector*) for pre-treatment, which consisted of normalization with the RIP intensity and autoscaling (mean-centring and scaling to unit variance). For approach B, the complete GC-IMS spectral fingerprint was extracted as a matrix. Subsequently, data was imported into Matlab for pre-treatment and in this case the pre-treatment consisted of normalization with the RIP and a transformation of the data matrix into a single data row with 27,178 variables, constituted by the concatenated spectra. In addition, a mean-centring of the concatenated spectra was applied. The difference between data types in approaches A and B can be seen in Supplementary Figure 1.

Next, and in order to detect possible sample outliers, a detection process was carried out. Outliers are those observations that differ strongly from the other data points within sampled population. The methodologies described in the literature for outlier detection are based on determining certain thresholds of closeness between samples and

then establishing a value to act as a reference. In this case the method used was a Principal Components Analysis (PCA) followed by a classification of samples in two classes (acorn-fed vs. feed-fed Iberian ham) carried out through PLS-DA. The first step for PLS-DA multivariate analysis was the selection of samples to constitute the calibration and validation sets. In the present case, an automatic random selection of samples was performed taking into account the distribution of classes: In each random calibration and validation subset the proportion between classes was always maintained as the original set (≈50%), unless otherwise stated. Several PLS-DA models were built with different number and types of samples for training and validation. For comparison purposes, each model was built with its optimum number of latent variables (LVs) or sets of components made from linear combinations of variables. The selection of the optimal number of LVs for PLS-DA models was performed automatically considering the average cross-validation classification error (CVCE) [33]. In this procedure, the number of LVs is increased until the CVCE reaches a plateau.

The PLS-DA models were trained and cross-validated using training sets with a specific number of samples while the rest were used to validate the performance of our proposal. In all cases, cross-validation was performed employing randomly-selected training samples with 5 data splits (20% of the calibration samples each) and 20 iterations; in order to detect a possible overfitting, permutation tests were performed on all the developed PLS models. The tests involved repeatedly and randomly reordering the validation set and rebuilding the model after each iteration while assigning samples to wrong classes. Then, the test searches for a chance correlation between the calibration and validation blocks. After each permutation, the predictions for each sample from cross-validation, RMSEC (root-mean-standard error for calibration) and RMSECV are recorded. The shuffling of classes is repeated multiple times. The cross-validated and self-prediction values should be relatively close to each other, but should be significantly less than the results for the non-permuted validation block. The further away the unpermuted results are from the mean, the more unlikely it is the original model is over-fit. In the present case, all the models developed were suitable for prediction.

Prediction was performed with the samples of the validation set. The evaluation of the prediction ability of the models was done through two parameters: the accuracy (number of samples correctly classified, or true positives plus true negatives vs. the number of total samples) and the Area Under the Receiver Operating Characteristic (AUROC). Accuracy and AUROC are the most suitable parameters to assess the quality of PLS-DA models that discriminate between two classes [33], such as in the present case. If minimizing false negative or false positive errors was a priority, an optimization based on sensitivity and specificity is also recommended. Furthermore, in approach A, based on the intensity of preselected markers as data, the Variable Importance for the Projection (VIP) of each variable was calculated.

3. Results and discussion

3.1. Univariate analysis

After a visual study of the GC-IMS spectral plots of 500 acorn-fed and 497 feed-fed Iberian ham samples, 279 markers were located, of which 61 were identified, corresponding to 36 VOCs. The main criteria for feature selection were that their intensity was be over the baseline and the absence of overlapping between them. The features that correspond to identified VOCs were located after the analysis of standards, if the position of the features of these standards matched with that of the features of the samples. In approach A, the markers' intensity was used to classify the samples according to their class, as described in Section 2.3. Furthermore, a univariate sample classification through analysis of variance (ANOVA) was attempted using these markers, whose aim was to identify relevant markers (identified or unidentified) of each feeding class for sample discrimination.

23 markers of the 279 originally selected from the GC-IMS spectral plot, were considered the most relevant for acorn-fed vs. feed-fed discrimination after a point-biserial correlation test, showing an absolute value of Pearson correlation coefficient (PCC) >0.25 with the feeding regime and a p-value <0.001. Of those 23 features, 4 were tentatively identified: the protonated monomer of hexan-2-one and the proton-bound dimers of nonan-2-one, 3-methylbutan-1-ol and octan-2-one. As an example, the signal of the branched alcohol 3-methylbutanol showed higher intensity in acorn-fed samples, which is in accordance with previous studies [32,34,35]. The intensity of the 6 features with the highest PCC ("247″, the protonated monomer of hexan-2-one, the proton-bound dimer of nonan-2-one, "118″, "267″ and "49″) is represented as violin plots in Fig. 1. As can be seen in this figure, upper intensity limits could be determined for these 6 features. The upper part of the rectangle represents the third quartile (Q3): 75% of the intensity values (mV) are equal or lower than this value. As an example, marker "247″ showed an intensity >75 mV in 25% of acorn-fed samples, while this only happened with 1.4% of feed-fed samples (see Fig. 1a). On the contrary, the marker "207″ (protonated monomer of hexan-2-one) showed an intensity >263 mV in 25% of feed-fed samples, while this only happened in 7.6% of acorn-fed samples (see Fig. 1b). Even so, in some cases the differences between classes might be more related to the curing plant of origin rather than the feeding regime itself. As can be seen in Supplementary Figure 2, the differentiation between acorn-fed and feed-fed classes was mostly determined by the unusually high intensity of marker "247″ in feed-fed samples of curing plant 7, which could be due to the differences and evolution of the yeast population of the curing plant during manufacturing and ripening of dry-cured hams [36,37]. This might give a false impression of a high correlation of that marker with an acorn-fed regime, which proves that univariate analysis is not suitable in the case of dry-cured Iberian ham discrimination.

3.2. Multivariate analysis

Due to the unfeasibility of obtaining a feeding regime differentiation using individual markers (univariate analysis), multivariate PLS-DA was employed to classify the 997 samples. The evaluation of the prediction ability of the models was done through two parameters: Accuracy and AUROC. In the present approach, several experiments were carried out to assess the influence of the following factors in the prediction results:

1) Use of discrepant samples as a validation set.
2) Proportion of samples used for training and validation.
3) Minimum number of training samples required.
4) Proportion of samples of each class in training and validation sets.
5) Use of the whole spectral fingerprint vs. the intensity of pre-selected features.

3.2.1. Use of discrepant samples as a validation set

The objective of this experiment was to assess the influence of the variability of samples within each class in the accuracy. For this purpose, two PCAs of the samples of each class (499 acorn-fed and 498 feed-fed, 997 in total) were performed. The scores of 36 acorn-fed and 38 feed-fed samples (74 in total) that were outside the Hotelling's T^2 ellipse were considered discrepant. In this case, 39% of the 74 discrepant samples came from the same curing plant. These might have been just different from the rest of the samples due to their processing conditions or superficial fungi [36], amongst other factors, not necessarily being outliers. After a careful study, none were considered outliers: the discrepant samples were correctly labelled and analysed according to quality control procedure (described in Section 2.1.), and their GC-IMS spectral plots were not so different compared to the other samples.

Subsequently, the influence of the uniformity of the calibration set on the results was studied using three different calibration sets: a homogeneous set, a heterogeneous set and a combination of both. The

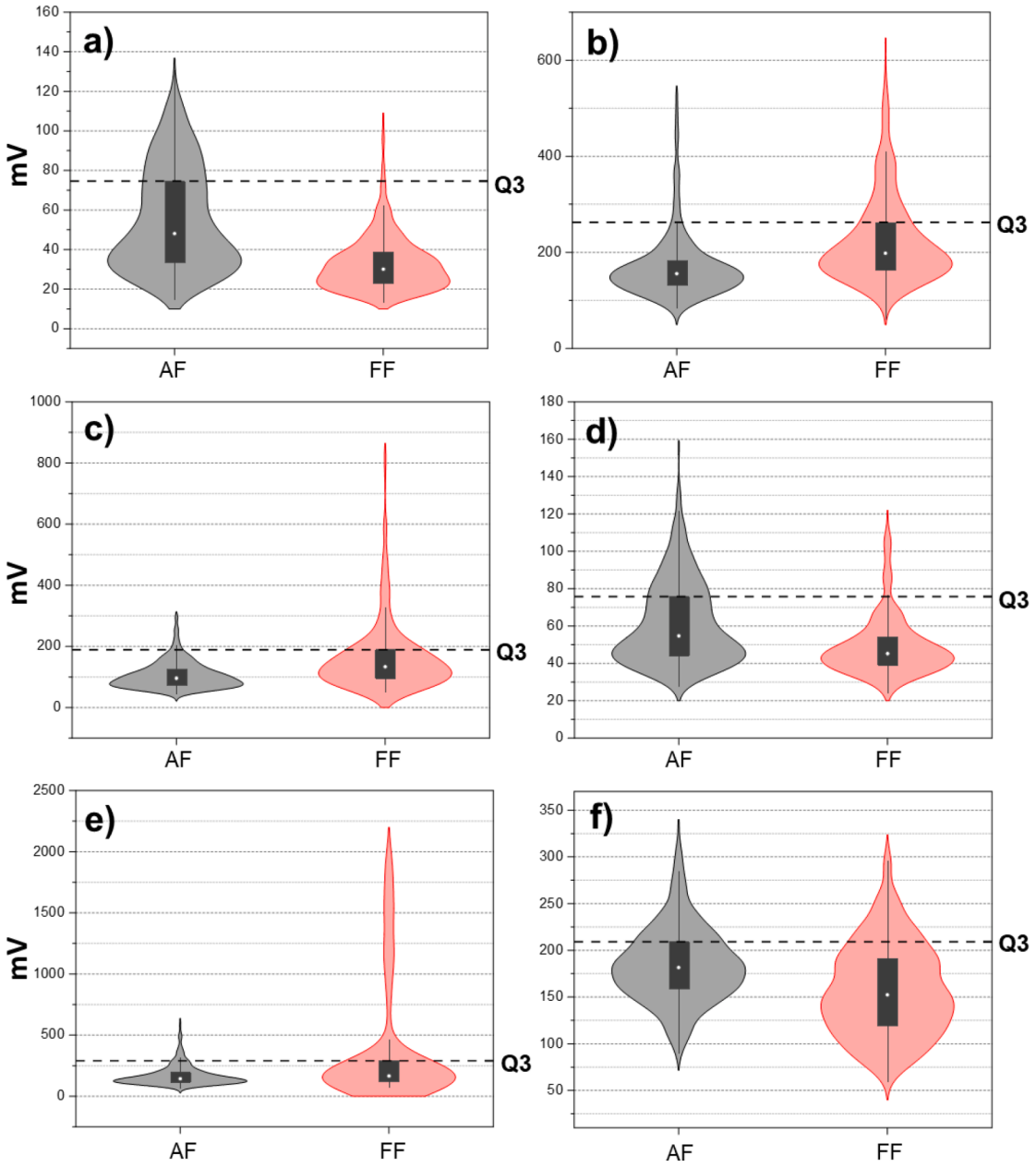

Fig. 1. Violin plots of the signal intensity (mV) of features a) 247, b) protonated monomer of hexan-2-one, c) proton-bound dimer of nonan-2-one, d) 118, e) 267 and f) 49 in acorn-fed samples and feed-fed samples. AF: Acorn-fed, FF: Feed-fed, *Q3*: Third quartile.

homogeneous set was constituted by very similar samples whose scores had low Hotelling's T^2 values in the PCAs. Therefore, the samples used for training were from acorn-fed and feed-fed classes located in the middle of PCA, whereas the heterogeneous set was constituted by samples that were very different within each class and whose scores showed high T^2 values, which were segregated in the score plot. The mixed calibration set, used as a reference, consisted of a mix of the homogeneous and heterogeneous sets; all three training sets contained 100 samples each; this number was selected so that the effect of the different types of samples would be more apparent. The 74 discrepant samples previously detected were included in the heterogeneous set. A single validation set was used in all cases, which also consisted of 100 samples. The results of this experiment are shown in Table 1, and as can be seen, the best results were obtained training the model with a mixed approach that includes very different samples that covered the maximum space within a PCA score plot. In fact, training with similar samples within each class or outliers implied a poorer prediction result.

Table 1
Results of PLS-DA models using, within each class, very similar samples (lowest T^2 value), very different samples (highest T^2 value) and mixed samples for training while predicting a single set of 100 blind samples.

Training set ($n = 100$)	% Success	AUROC
Using very similar samples (homogeneous set)	76	0.87
Using very different samples (heterogeneous set)	52	0.76
Mixed approach	80	0.88

AUROC: Area Under the Receiver Operating Characteristic.

Therefore, trying to predict the natural feeding regime of free-range Iberian pigs [38], which give rise to dry-cured hams with slight differences in the curing process, requires flexibility when training the PLS-DA model. Thus, the analysis of a wide range of pieces of each class is recommended for model training, in order to capture those natural differences.

3.2.2. Proportion of samples used for training and validation

Since the multivariate approach can be sensitive to the proportion of samples used in calibration and validation sets, in this experiment different proportions of samples were used for PLS-DA model training and validation (from a training/validation samples proportion of 0.3 to 6.5), which was done to assess the need to increase the size of the training set, depending on the number of blind samples whose class needed to be predicted. For this aim, a regression was performed using the proportion of calibration samples vs. the accuracy (shown in Fig. 2) and a significant positive correlation was found ($p<0.001$). In the present case, an average accuracy of ≥85% was obtained in a PLS-DA model with a minimum proportion of training samples 1.8 times higher than the number of blind samples. Thus, the number of training samples should be adapted to the number of blind samples following this suggestion. The complete results are shown in Supplementary Table 1.

3.2.3. Minimum number of training samples required

For this experiment, three validation sets were fixed with 100, 200 and 300 samples (10%, 20% and 30% of the total samples, respectively). Subsequently, the number of samples in the corresponding training sets was progressively and randomly decreased from 650 to 100 samples. The training and validation sets contained a similar proportion of acorn-fed and feed-fed samples (≈50%). The objective of this approach was to assess how many known samples are necessary to train a PLS-DA model in order to predict the class of blind samples from several curing plants, or just from a single one. An average accuracy of 85%, which represented a maximum, was obtained using 450 samples for calibration (see Fig. 3). The complete obtained results are shown in Supplementary Table 2: a calibration set with ≈450 samples from several origins was enough to achieve a PLS-DA classification model for the prediction of the class of 100, 200 or 300 blind samples from different curing plants.

The same approach was applied to study the results using samples of individual curing plants. The objective was to assess the lowest number of training samples necessary to train a PLS-DA classification model for a single company, which would be useful for producers that want to ensure the quality of their own products. As can be drawn from Supplementary Tables 3 and 4, curing plants 4 and 6 reached a maximum average prediction accuracy when the proportion of training samples vs. blind samples was at least ≈1.2. Nevertheless, with that same proportion, the accuracy was different depending on the curing plant. Prediction accuracy in curing plant 4, with 326 samples, was ≥0.95 and in curing plant 6, with 94 samples, was ≥0.85, which can be explained by the different homogeneity of the pieces in each curing plant. PLS-DA models of plants where the cured pieces sampled show more differences might need a more exhaustive training.

3.2.4. Proportion of samples of each class in training and validation sets

For this experiment, whose objective was to verify if the number of samples of each class should be equilibrated in the training set, three training sets with 400 samples were used to predict the class of a single validation set with 200 samples. The training sets were constituted by A) 50% acorn-fed and 50% feed-fed samples, B) 75% acorn-fed and 25% feed-fed samples, and C) 25% acorn-fed and 75% feed-fed samples. The results are shown in Table 2, in which the general accuracy in the classification using the training sets A, B and C (82±3%, 77±0% and 82 ±2%, respectively) was similar. However, the accuracy obtained with sets B and C was severely biased depending on the class of the training samples. As can be seen in Table 2, the best prediction results of acorn-fed and feed-fed samples took place when a higher proportion of samples of those classes was used for training the PLS-DA model. Likewise, an equal proportion of samples of both classes (50/50) provided a global result that is equal or better than that with the use of biased training sets, preventing low accuracy rates for one of the classes.

Therefore, according to the results of experiments previously described in Sections 3.2.1. and 3.2.4., the classes in the training set must be balanced (approximately half of them must belong to each class) and, within both classes, samples should be diverse. Under the "acorn-fed" class a high variability of feed intake may exist, as the feeding

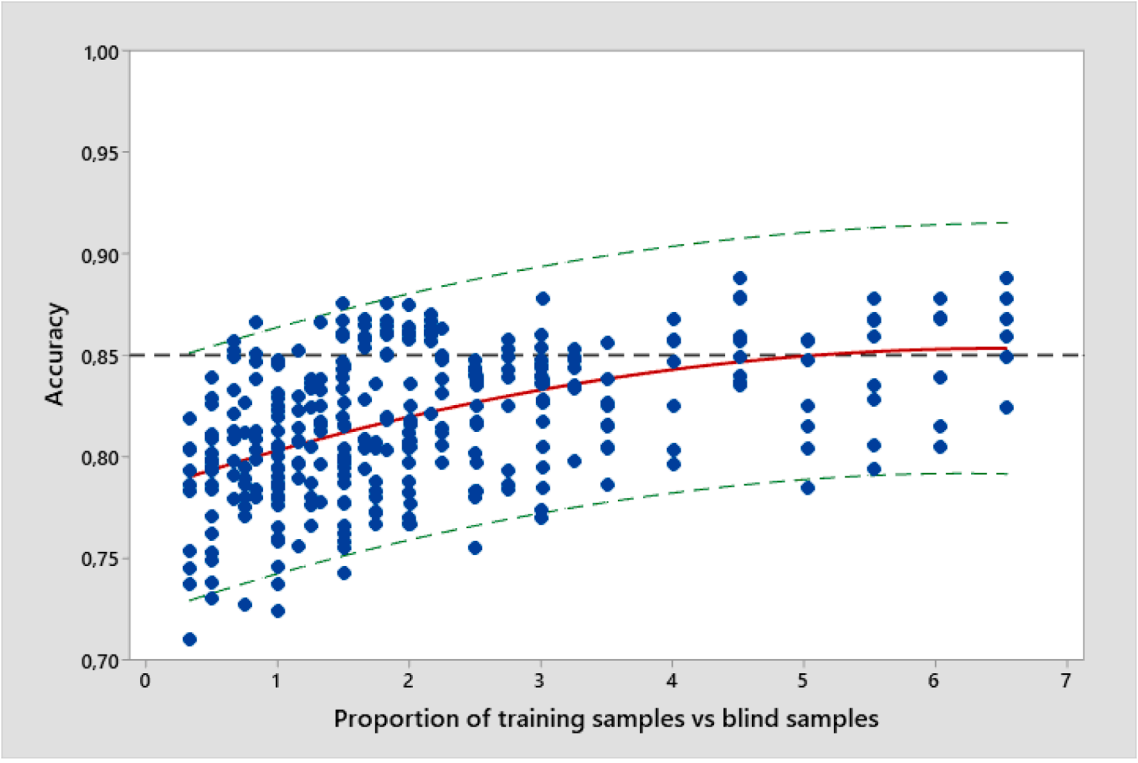

Fig. 2. Polynomic regression of the proportion of training/blind samples for PLS-DA model (X) vs the accuracy of the prediction (Y). $p<0.001$. $R^2 = 0.26$. Equation: $y = 0.783 + 0.021x - 0.002\times^2$.

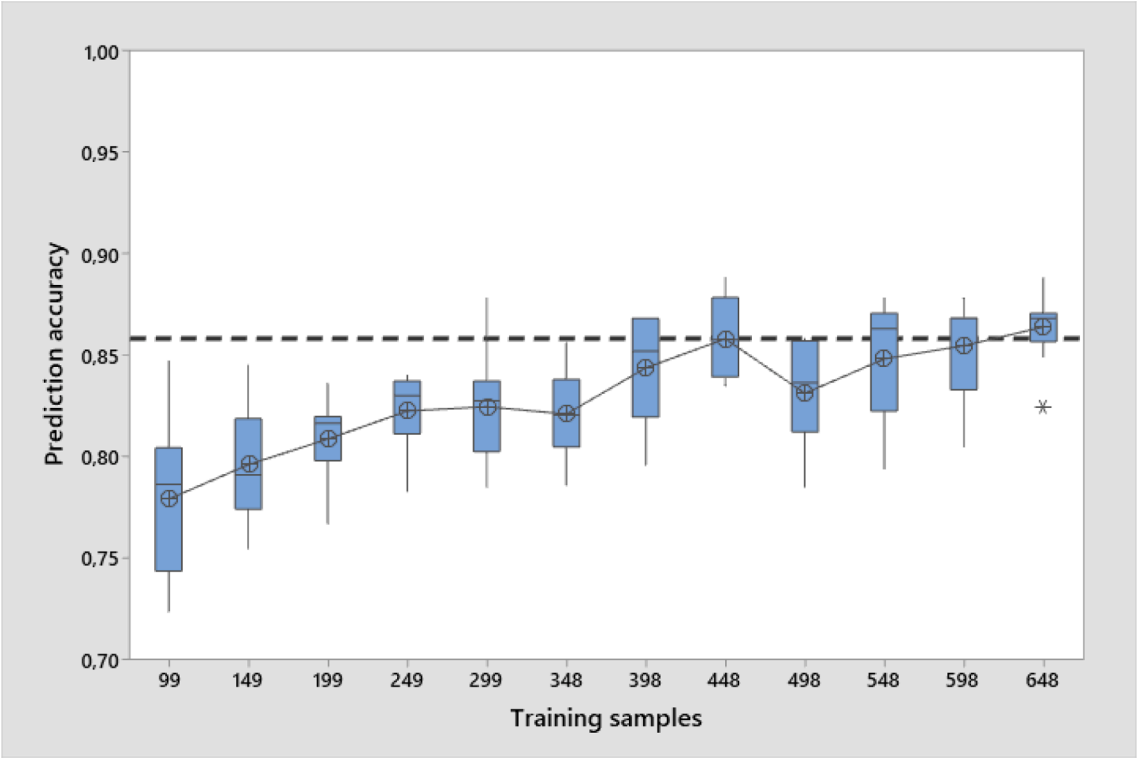

Fig. 3. Box-plot of the accuracy of PLS-DA models predicting the class (acorn-fed or feed-fed) of fixed sets of 100, 200 and 300 blind samples using trainings sets of different size.

Table 2
Results of PLS-DA models using different proportions of samples of each class for the prediction of a single set of 200 blind samples using 400 training samples. AF: Acorn-fed, FF: Feed-fed.

Proportion AF/FF in training set ($n = 400$)	Total% Success	Acorn-fed% Success	Feed-fed% Success	AUROC
50/50	82 ± 3^{ab}	89 ± 5^{a}	75 ± 9^{b}	0.93 ± 0.01^{a}
75/25	77 ± 0^{b}	98 ± 0^{a}	56 ± 0^{c}	0.91 ± 0.00^{a}
25/75	82 ± 2^{a}	73 ± 6^{b}	92 ± 2^{a}	0.92 ± 0.01^{a}

a and b represent different groups of classification according to Tukey test.

regime is mainly based on a free-range grazing of pasture and acorns [38]. Furthermore, it has been demonstrated that microbial populations on the surface of Iberian hams produce differences in their volatile profile [36]. Similarly, a previous study with olive oil [11] demonstrated that a mix of training samples of different campaigns (different years) improved the accuracy in the prediction, while when training samples from a single campaign were used to predict blind samples from another campaign, the results were much poorer [11].

In addition, the training samples should belong to different industries (curing plants) to capture as much variability as possible. In order to demonstrate this hypothesis several PLS-DA models were built excluding the samples from each curing plant from the training set and at the same time these samples were used as the validation set. This greatly lowered the accuracy of their classification by feeding regime, which on average fell to nearly 51±22%. Therefore, an already trained PLS-DA model must be retrained if samples of a new origin need to be predicted.

3.2.5. *Marker selection for PLS-DA models*

In addition, in order to simplify the data treatment of samples, a marker selection for PLS-DA models was attempted. A model built with 650 known samples for training and 200 blind samples for prediction was selected as a starting point, after which new models with progressively less variables were built and their discrimination power was assessed. The procedure was performed by the evaluation of VIP scores and on each iteration, markers with the highest VIP (at least over the significancy threshold, which had a value of 1) were selected for subsequent models, while the rest of the variables (markers) were discarded. The classification results with different number of variables are shown in Fig. 4. As can be seen, the accuracy in the classification and the AUROC reached a maximum with 35 variables (equivalent to an 87% reduction of variables). Therefore, the reduction of variables when using the intensity of pre-selected markers as data implies the improvement of classification accuracy, although in this case reducing more variables would imply a decline of the prediction accuracy, that falls below a 75% rate.

3.2.6. *Use of the whole spectral fingerprint vs. pre-selected features*

For chemometric treatments, the use of the whole spectral fingerprint might appear faster than the use of pre-selected features due to its complete untargeted nature. However, the use of the whole fingerprint implies that features of contaminants have a much higher impact in the classification results. In addition, the much larger amount of data requires considerable computation power thus data treatment is more time-consuming, whereas the use of pre-selected features requires a visual selection of signals present in the GC-IMS spectral plot. Despite this step of feature selection being time consuming at the beginning, once it is finished the features' template can be saved and transferred as needed. In addition, the lower amount of data implies that modest computers are able to carry out its processing. Furthermore, variable selection is possible using VIP scores or other techniques when the intensity of pre-selected features is used as data, thus simplifying subsequent targeted

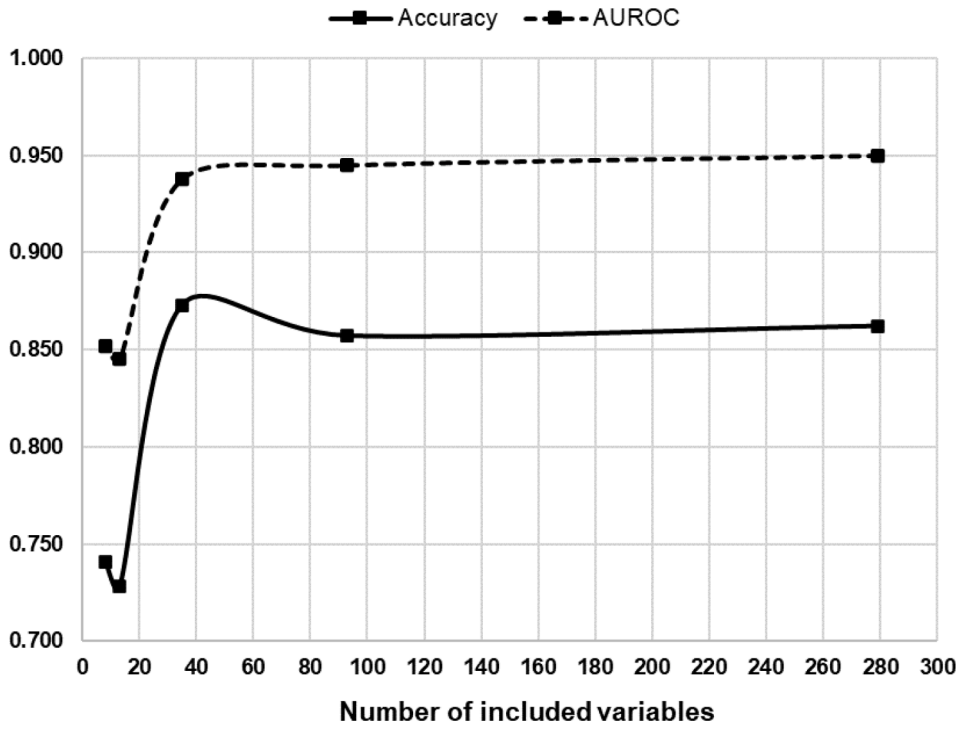

Fig. 4. Results of PLS-DA models that include different number of variables with a fixed size of calibration (n = 650) and validation (n = 300) sets.

analysis if required.

In order to assess the performance of the approaches based on pre-selected markers (approach A) and the whole spectral fingerprint (approach B), PLS-DA models were built with the same training and validation sets for both data types. Subsequently, the classification results were compared, which are shown in Fig. 5. The difference of accuracy using both approaches was statistically significant (p<0.03) although the values were similar: average accuracy values were 0.85 and 0.86 using the whole spectral fingerprint and pre-selected markers, respectively. On the other hand, the median of the prediction accuracy was also higher using markers vs the spectral fingerprint (0.87 vs 0.85). Therefore, the use of markers provides more slightly better prediction results than the use of the fingerprint. Therefore, the first approach may be more advantageous. The complete results are included in Supplementary Table 5.

In summary, the high validated classification rates obtained in both

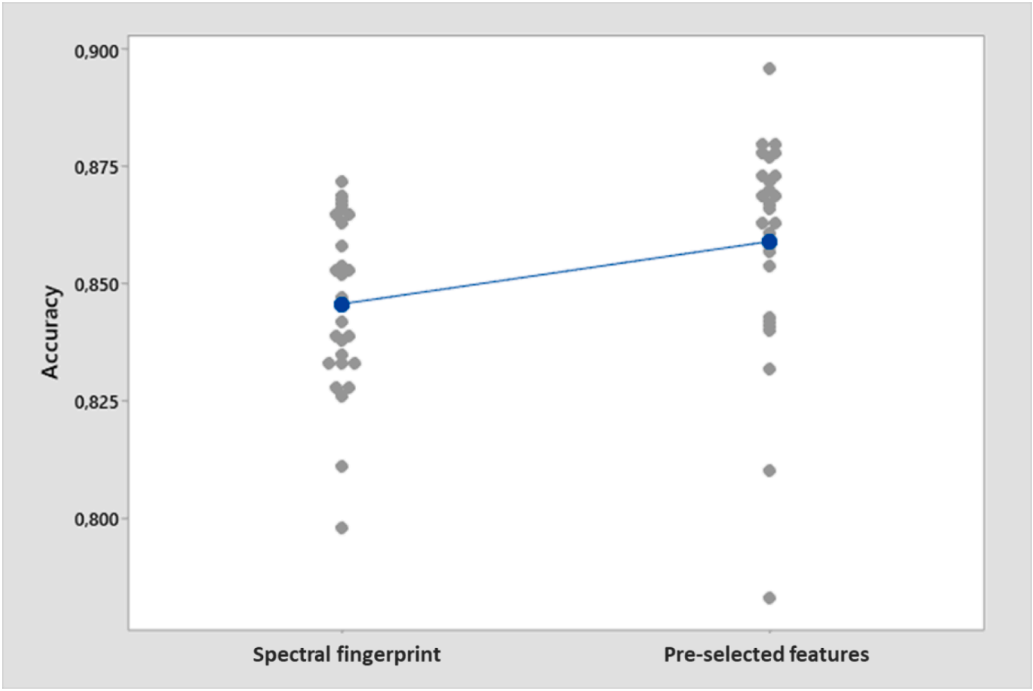

Fig. 5. Comparison of accuracy of PLS-DA models using GC-IMS pre-selected markers or the complete spectral fingerprint with different proportion of samples for training and prediction.

approaches (using the spectral fingerprint or pre-selected markers) demonstrate the suitability of PLS-DA for the classification of complex agri-food matrices, such as dry-cured Iberian ham fat. The influence of the geographical origin and curing time in the capability of the model to discriminate the samples by feeding regime is yet to be assessed.

3.3. Issues found during the analysis period using GC-IMS

Analysing samples for a prolonged period of time is required to build classification models that can be applied in an industrial context [39]. However, dealignment of the plots and reactant ion peak (RIP) value fluctuations might take place when analysing ≈1000 samples during a year with GC-IMS [40]. The impact of the dealignments is greater when using the whole spectral fingerprint approach because a smaller amount of data is discarded, thus weakening the results of classification. In the present case, these changes were assessed through the periodic analysis of a mix of ketones acting as external standards (ES). If required, a manual alignment of the samples was carried out following the procedure described in a previous study [40]. The results obtained show an adequate robustness of the method for a period of 12 months (see Table 3). The GC column used was able to withstand the analysis of ≈1000 samples using the previously described method, and so for the purpose of a similar study a change of GC column should not be needed, at least within that period of time. Besides that, VOCs might naturally degrade inside of vials. This evolution is caused by the equilibrium between the headspace and the solid sample inside the vial as it has been demonstrated that fat deteriorates due to lipolytic enzymes that remain active at freezing temperatures [41]. In the present case, this can happen after freezing due to lipid oxidation and so samples should be analysed within 1–2 weeks after being frozen at −18 °C.

The variety, quality and traceability of the samples used for the training of chemometric models is more important for the accuracy of classification than their number [39]. This was demonstrated by Jurado-Campos et al. [11], who attempted an olive oil classification in three categories and whose prediction results improved using a training set based on a mix of two seasons instead of a set with samples of a single season. Therefore, knowing the origin and season of those samples was equally important than knowing just their category: increasing the number of samples with limited information about their traceability does not necessarily mean a better accuracy of classification. In any case, outliers can appear due to experimental errors or due to an incorrect labelling of the samples. In the present case, the 997 samples analysed were directly obtained in the cellars of each company while the dry-curing process of the hams was taking place. Each piece was unequivocally identified by a unique code fitted to it from the start of its processing, ensuring traceability, and when experimental errors took place e.g., the analysis of vials with badly fitted caps, the samples involved were discarded.

Table 3
Values of% RSD of retention and drift times of the proton-bound dimers of the 500 mg/L ketone mix after the analysis of 56 aliquots during a period of 6 months.

Proton-bound dimer	Average Retention time (s)	Retention time RSD (%)	Average Drift time (ms)	Drift time RSD (%)
butan-2-one	108.89±1.20	1.10	10.24±0.06	0.58
pentan-2-one	164.29±1.51	0.92	11.26±0.06	0.54
hexan-2-one	297.93±2.12	0.71	12.34±0.07	0.53
heptan-2-one	488.05±2.80	0.57	13.39±0.07	0.52
octan-2-one	694.07±3.49	0.50	14.41±0.07	0.51
nonan-2-one	899.89±8.16	0.91	15.44±0.18	1.15

4. Conclusion

In the case of dry-cured Iberian ham fat, classifying samples with univariate methods is not possible. As described in Section 3.1., only limits of intensity of individual markers can be obtained, which constrain the ability to discriminate samples through targeted analysis. In order to calibrate a PLS-DA model based on GC-IMS data to assign the correct category to an unknown sample, the following guidelines should be followed:

- The influence of the geographical origin and curing time in the capability of PLS-DA models to discriminate Iberian ham samples by the pig's feeding regime is yet to be assessed.
- A statistically reliable number of samples representative of the problem under study is required. In the present case, it was confirmed that for the classification of 100, 200 or 300 blind samples of 7 different origins, 450 samples for PLS-DA model training were enough to achieve an average prediction accuracy rate of 85%.
- If the samples whose class will be predicted come from the same curing plant, a proportion of ≈1.2 between training and validation samples is enough to reach a maximum in the average prediction accuracy rate. However, the value of this maximum could be different depending on the heterogeneity of the samples in the curing plant.
- Training samples must have good traceability: Fewer well-identified samples are preferable over many samples if some of them are erroneously labelled, because a previous lack of traceability may imply prediction mistakes. Samples certified are very useful for this purpose (for instance, certified by a Denomination of Protected Origin, such as the samples studied).
- Analysing the same number of known samples of each class in order to equilibrate the training set is recommended. In addition, sampling a wide range of samples within each class improves prediction results. The samples of the present study were obtained in several curing plants, ensuring variability in order to enrich the training sets. The periodic enrichment of the model with enough variety of samples within each class is desirable. Any classification attempt using GC-IMS should take this into account.
- In the present case, a reduction of 87% of the original number of variables slightly improved classification results.
- During a 12-month analysis period, no significant problems were detected with the GC-IMS device using the method described.

In summary, it has been demonstrated that GC-IMS combined with PLS-DA is a useful tool for the ham industry to perform a control of their products. The use of the intensity of pre-selected GC-IMS markers is recommended as data for chemometric treatment over the complete spectral fingerprint. Nevertheless, the accuracy in the classification of blind samples still depends on the curing plant of origin where calibration samples are obtained, and therefore individual calibration models should be built for each curing plant.

Funding

This work was supported by *Programa operativo FEDER Andalucía 2014–2020*, 2018 call, project code 1261925-R, by *Consejería de Economía, Conocimiento, Empresas y Universidad*, 2017 call (*Ayudas a actividades de transferencia de conocimiento entre los agentes del sistema andaluz del conocimiento y el tejido productivo*), project code 1155447, and by *Ministerio de Ciencia e Innovación*, 2017 call (*Doctorados Industriales*), project code DI-17–09645.

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.

Data Availability

The data that has been used is confidential.

Acknowledgements

Andrés Martín-Gómez would like to thank *Ministerio de Ciencia e Innovación* for a pre-doctoral grant and *Sociedad Cooperativa Andaluza Ganadera del Valle de los Pedroches (COVAP)* for its collaboration and an employment contract. The authors would also like to thank *Los Pedroches* and *Jabugo* Protected Designations of Origin for their collaboration and the industries included in the study, whose personnel allowed sampling of the Iberian hams subject to analysis: *Sociedad Cooperativa Andaluza Ganadera del Valle de los Pedroches (COVAP), Comercial Jabu S.L., Dehesa de Campo Alto S.L., Dehesas Reunidas S.L., Embutidos Camilo Ríos, Hijos de Onofre Sánchez Martín and Sierra de Jabugo S.L.*

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.talo.2022.100175.

References

[1] S. Wang, H. Chen, B. Sun, Recent progress in food flavor analysis using gas chromatography–ion mobility spectrometry (GC–IMS), Food Chem. 315 (2020), 126158.
[2] F. Wang, Y. Gao, H. Wang, B. Xi, X. He, X. Yang, W. Li, Analysis of volatile compounds and flavor fingerprint in Jingyuan lamb of different ages using gas chromatography–ion mobility spectrometry (GC–IMS), Meat Sci. 175 (2021), 108449.
[3] W. Yao, Y. Cai, D. Liu, Y. Chen, J. Li, M. Zhang, N. Chen, H. Zhang, Analysis of flavor formation during production of Dezhou braised chicken using headspace-gas chromatography-ion mobility spec-trometry (HS-GC-IMS), Food Chem. 370 (2022), 130989.
[4] J. Zhang, W. Zhang, L. Zhou, R. Zhang, Study on the influences of ultrasound on the flavor profile of unsmoked bacon and its underlying metabolic mechanism by using HS-GC-IMS, Ultrason. Sonochem. 80 (2021), 105807.
[5] A. Martín-Gómez, M.P. Segura-Borrego, R. Ríos-Reina, M.J. Cardador, R. M. Callejón, M.L. Morales, V. Rodríguez-Estévez, L. Arce, Discrimination of defective dry-cured Iberian ham determining volatile compounds by non-destructive sampling and gas chromatography, LWT - Food Sci. Tech. 154 (2021), 112785.
[6] J.H. Aheto, X. Huang, X. Tian, R. Lv, C. Dai, E. Bonah, X. Chang, Evaluation of lipid oxidation and volatile compounds of traditional dry-cured pork belly: The hyperspectral imaging and multi-gas-sensory approaches, J. Food Process Eng. 43 (2019) e13092.
[7] W. Li, Y.P. Chen, I. Blank, F. Li, C. Li, Y. Liu, GC × GC-ToF-MS and GC-IMS based volatile profile characterization of the Chinese dry-cured hams from different regions, Food Res. Int. 142 (2021), 110222.
[8] J. Gallegos, R. Garrido-Delgado, L. Arce, L.M. Medina, Volatile metabolites of goat cheeses determined by ion mobility spectrometry. Potential applications in quality control, Food Anal. Methods 8 (2014) 1699–1709.
[9] N. Arroyo-Manzanares, M. García-Nicolás, A. Castell, N. Campillo, P. Viñas, I. López-García, M. Hernández-Córdoba, Untargeted headspace gas chromatography - Ion mobility spectrometry analysis for detection of adulterated honey, Talanta 205 (2019), 120123.
[10] X. Wang, K.M. Rogers, Y. Li, S. Yang, L. Chen, J. Zhou, Untargeted and Targeted Discrimination of Honey Collected by Apis cerana and Apis mellifera Based on Volatiles Using HS-GC-IMS and HS-SPME-GC–MS, J. Agric. Food Chem. 67 (2019) 12144–12152.
[11] M.M. Contreras, N. Jurado-Campos, L. Arce, N. Arroyo-Manzanares, A robustness study of calibration models for olive oil classification: targeted and non-targeted fingerprint approaches based on GC–IMS, Food Chem. 288 (2019) 315–324.
[12] A. Tata, A. Massaro, T. Damiani, R. Piro, C. Dall'Asta, M. Suman, Detection of soft-refined oils in extra virgin olive oil using data fusion approaches for LC-MS, GC-IMS and FGC-Enose techniques: The winning synergy of GC-IMS and FGC-Enose, Food Control 133 (2021), 108645.
[13] J. Chen, L. Tao, T. Zhang, J. Zhang, T. Wu, D. Luan, L. Ni, X. Wang, J. Zhong, Effect of four types of thermal processing methods on the aroma profiles of acidity regulator-treated tilapia muscles using E-nose, HS-SPME-GC-MS, and HS-GC-IMS, LWT - Food Sci. Tech. 147 (2021), 111585.
[14] S. Jia, Y. Li, S. Zhuang, X. Sun, L. Zhang, J. Shi, H. Hong, Y. Luo, Biochemical changes induced by dominant bacteria in chill-stored silver carp (Hypophthalmichthys molitrix) and GC-IMS identification of volatile organic compounds, Food Microbiol. 84 (2019), 103248.
[15] Y. Yang, B. Wang, Y. Fu, Y.-g. Shi, F.-l. Chen, H.-n. Guan, L.-l. Liu, C.-y. Zhang, P.-y. Zhu, Y. Liu, N. Zhang, HS-GC-IMS with PCA to analyze volatile flavor compounds across different production stages of fermented soybean whey tofu, Food Chem. 346 (2020), 128880.
[16] X. Fan, X. Jiao, J. Liu, M. Jia, C. Blanchard, Z. Zhou, Characterizing the Volatile Compounds of Different Sorghum Cultivars by Both GC-MS and HS-GC-IMS, Food Res. Int. 140 (2020), 109975.
[17] R. Brendel, S. Schwolow, S. Rohn, P. Weller, Volatilomic Profiling of Citrus Juices by Dual-Detection HS-GC-MS-IMS and Machine Learning-An Alternative Authentication Approach, J. Agric. Food Chem. 69 (2021) 1727–1738.
[18] R. Zhou, X. Chen, Y. Xia, M. Chen, Y. Zhang, Q. Li, D. Zhen, S. Fang, Research on the application of liquid-liquid extraction-gas chromatography-mass spectrometry (LLE-GC-MS) and headspace-gas chromatography-ion mobility spectrometry (HS-GC-IMS) in distinguishing the Baiyunbian aged liquors, Int. J. Food Eng. 17 (2020) 83–96.
[19] A. Martín-Gómez, N. Arroyo-Manzanares, V. Rodríguez-Estévez, L. Arce, Use of a non-destructive sampling method for characterization of Iberian cured ham breed and feeding regime using GC-IMS, Meat Sci. 152 (2019) 146–154.
[20] N. Arroyo-Manzanares, A. Martín-Gómez, N. Jurado-Campos, R. Garrido-Delgado, C. Arce, L. Arce, Target vs spectral fingerprint data analysis of Iberian ham samples for avoiding labelling fraud using headspace - gas chromatography-ion mobility spectrometry, Food Chem. 246 (2018) 65–73.
[21] E. Borràs, J. Ferré, R. Boqué, M. Mestres, L. Acena, A. Calvo, O. Busto, Olive oil sensory defects classification with data fusion of instrumental techniques and multivariate analysis (PLS-DA), Food Chem. 203 (2016) 314–322.
[22] M.M. Contreras, N. Arroyo-Manzanares, C. Arce, L. Arce, HS-GC-IMS and chemometric data treatment for food authenticity assessment: Olive oil mapping and classification through two different devices as an example, Food Control 98 (2018) 82–93.
[23] S. Gu, J. Zhang, J. Wang, X. Wang, D. Du, Recent development of HS-GC-IMS technology in rapid and non-destructive detection of quality and contamination in agri-food products, TrAC Trends Anal. Chem. 144 (2021), 116435.
[24] C. Capitain, P. Weller, Non-Targeted Screening Approaches for Profiling of Volatile Organic Compounds Based on Gas Chromatography-Ion Mobility Spectroscopy (GC-IMS) and Machine Learning, Molecules 26 (2021) 5457.
[25] Standard Practices for Infrared Multivariate Quantitative Analysis - E1655-05, ASTM International, West Conshohocken, Pennsylvania, USA, 2012.
[26] B. Worley, R. Powers, PCA as a practical indicator of OPLS-DA model reliability, Curr. Metabolom. 4 (2016) 97–103.
[27] R.G. Brereton, G.R. Lloyd, Partial least squares discriminant analysis: taking the magic away, J. Chemom. 28 (2014) 213–225.
[28] D. Ruiz-Pérez, H. Guan, P. Madhivanan, K. Mathee, G. Narasimhan, So you think you can PLS-DA? BMC Bioinf. 21 (2020) 1–10.
[29] J.A. Westerhuis, H.C.J. Hoefsloot, S. Smit, D.J. Vis, A.K. Smilde, E.J.J. van Velzen, J.P.M. van Duijnhoven, F.A. van Dorsten, Assessment of PLSDA cross validation, Metabolomics 4 (2008) 81–89.
[30] E. Saccenti, M.E. Timmerman, Approaches to Sample Size Determination for Multivariate Data: Applications to PCA and PLS-DA of Omics Data, J. Proteome Res. 15 (2016) 2379–2393.
[31] M. Narváez-Rivas, E. Gallardo, M. León-Camacho, Analysis of volatile compounds from Iberian hams: a review, Grasas Aceites 63 (2012) 432–454.
[32] M.P. Segura-Borrego, A. Martín-Gómez, R. Ríos-Reina, M.J. Cardador, M. L. Morales, L. Arce, R.M. Callejón, A non-destructive sampling method for food authentication using gas chromatography coupled to mass spectrometry or ion mobility spectrometry, Food Chem. 373 (2021), 131540.
[33] E. Szymanska, E. Saccenti, A.K. Smilde, J.A. Westerhuis, Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies, Metabolomics 8 (2012) 3–16.
[34] J. Ruiz, J. Ventanas, R. Cava, A. Andrés, C. García, Volatile compounds of dry-cured Iberian ham as affected by the length of the curing process, Meat Sci. 52 (1999) 19–27.
[35] A.I. Andrés, R. Cava, J. Ventanas, E. Muriel, J. Ruiz, Lipid oxidative changes throughout the ripening of dry-cured Iberian hams with different salt contents and processing conditions, Food Chem. 84 (2004) 375–381.
[36] A. Martín, J.J. Córdoba, F. Núñez, M.a.J. Benito, M.A. Asensio, Contribution of a selected fungal population to proteolysis on dry-cured ham, Int. J. Food Microbiol. 94 (2004) 55–66.
[37] C. Alapont, P.V. Martínez-Culebras, M.C. López-Mendoza, Determination of lipolytic and proteolytic activities of mycoflora isolated from dry-cured teruel ham, J. Food Sci. Technol. 52 (2015) 5250–5256.
[38] V. Rodríguez-Estévez, A. García, F. Peña, A.G. Gómez, Foraging of Iberian fattening pigs grazing natural pasture in the dehesa, Livestock Sci. 120 (2009) 135–143.
[39] B. Debus, H. Parastar, P. Harrington, D. Kirsanov, Deep learning in analytical chemistry, TrAC Trends Anal. Chem. 145 (2021), 116459.
[40] N. Jurado-Campos, A. Martín-Gómez, D. Saavedra, L. Arce, Usage considerations for headspace-gas chromatography-ion mobility spectrometry as a suitable technique for qualitative analysis in a routine lab, J. Chromatogr. A 1640 (2021), 461937.
[41] P. Hernández, J.L. Navarro, F. Toldrá, Effect of frozen storage on lipids and lipolytic activities in the longissimus dorsi muscle of the pig, Z Lebensm Unters Forsch 208 (1999) 110–115.