Journal Pre-proof

Rapid screening of tuna samples for food safety issues related to histamine content using Fourier-transform mid-infrared (FT-MIR) and chemometrics.

Mónica Sánchez-Parra, Juan Antonio Fernández Pierna, Vincent Baeten, José Manuel Muñoz-Redondo, José Luis Ordóñez-Díaz, José Manuel Moreno-Rojas

PII: S0260-8774(24)00195-X

DOI: https://doi.org/10.1016/j.jfoodeng.2024.112129

Reference: JFOE 112129

To appear in: *Journal of Food Engineering*

Received Date: 2 April 2024

Revised Date: 7 May 2024

Accepted Date: 9 May 2024

Please cite this article as: Sánchez-Parra, M., Fernández Pierna, J.A., Baeten, V., Muñoz-Redondo, J.M., Ordóñez-Díaz, J.L., Moreno-Rojas, J.M., Rapid screening of tuna samples for food safety issues related to histamine content using Fourier-transform mid-infrared (FT-MIR) and chemometrics., *Journal of Food Engineering*, https://doi.org/10.1016/j.jfoodeng.2024.112129.

Rapid screening of tuna samples for food safety issues related to histamine content using Fourier-transform mid-infrared (FT-MIR) and chemometrics.

Mónica Sánchez-Parra[1,2], Juan Antonio Fernández Pierna[3], Vincent Baeten[3], José Manuel Muñoz-Redondo[1], José Luis Ordóñez-Díaz[1], José Manuel Moreno-Rojas[1*]

[1]Department of Agroindustry and Food Quality, Andalusian Institute of Agricultural and Fisheries Research and Training (IFAPA), Alameda del Obispo, Avda. Menéndez–Pidal, s/n., 14004 Córdoba, Spain

[2]PhD program Ingeniería agraria, alimentaria, forestal y de desarrollo rural sostenible. Universidad de Córdoba, Córdoba, Spain.

[3]Quality and Authentication of Products Unit. Knowledge and Valorization of Agricultural Products Department, Walloon Agricultural Research Centre (CRA-W), Chaussée de Namur 24, 5030 Gembloux, Belgium

*Corresponding author. E–mail: josem.moreno.rojas@juntadeandalucia.es (J.M.M.-R).

ABSTRACT

Biogenic amines (BAs) generally result from the decarboxylation reaction of free amino acids as a result of the activity of different microorganisms. A build-up of these compounds can result in food being spoilt. Therefore, the rapid and precise detection of BAs like histamine is an important task for food safety. This research aimed to explore the potential of Fourier-Transform Mid-Infrared (FT-MIR) spectroscopy combined with chemometric methods to assess histamine in fresh tuna quantitatively. Based on the FT-MIR data, partial least squares regression models for the prediction of histamine were successfully constructed with R^2>0.90. Machine learning algorithms (partial least squares-discrimination analysis, k-nearest neighbours, and support vector machine) were applied, and excellent discrimination results were achieved based on the limits specified in two different legislations (EU and FDA). The results support the use of a rapid, economic and reliable approach for the discrimination of samples that could pose a health risk to consumers.

Keywords: Food safety; Histamine; Tuna; FT-MIR; HPLC; Machine learning

1. INTRODUCTION

Yellowfin tuna (*Thunnus albacares, YFT*) is one of the most important fish species belonging to the *Scombridae* family, constituting approximately 30% of the global tuna harvest (FAO, 2011). Tuna is considered to be of high nutritional value and can play a very important role in a balanced human diet. This species provides many essential nutrients and health benefits, being a source of high-quality proteins, vitamins, amino acids and n-3 polyunsaturated fatty acids (Khalili and Sampels, 2018; Salvador et al., 2019).

With regard to food safety, fish and its derived products are considered one of the most perishable products. Many extrinsic and intrinsic factors make these products highly susceptible to chemical and microbiological contamination (Herpandi et al., 2011; Xie et al., 2020). Fish handling during storage and transport may enhance the potential human health risks associated with its consumption (Papageorgiou et al., 2018). The most widely known hazard associated with tuna is the presence of high levels of histamine, which constitutes an issue for the food industry (Feng et al., 2016; Prester, 2011). Histamine, a biogenic amine (BA), is produced in the flesh from histidine due to the activity of the bacterial enzyme histidine-decarboxylase (Lehane and Olley, 2000; Ordóñez and Callejón, 2019). Although histamine is essential to many key functions in humans and animals, a high intake of histamine, known as scombroid food poisoning, may cause adverse toxicological effects, such as neurological disorders, gastrointestinal diseases, headaches and urticaria, among others (Hungerford, 2010; McLauchlin et al., 2006), the effects of which will depend on the sensibility of each person. The production of histamine is intricately linked to microbial growth, and other BAs like cadaverine and putrescine can be generated concurrently (Sánchez-Parra et al., 2022; Shakila et al., 2003).

Histamine is the only BA appearing in legislation to which regulatory limits have been applied. Commission Regulation (EC) No. 2073/2005 on microbiological criteria for foodstuffs established a maximum histamine level of 200 mg/kg as acceptable in fresh fish. In application of this legislation, in a sample of nine randomly collected units, only two may contain between 100 – 200 mg/kg of histamine and none may be above the limit of 200 mg/kg in fish species associated with a high amount of histidine such as the *Scombridae, Clupeidae, Coryphaenidae, Engraulidae, Pomatomidae,* and

Scomberesocidae families. Nevertheless, the Food and Drug Administration (FDA, 2011) set a maximum limit of 50 mg/kg. Other countries such as Canada, Finland and Switzerland established 200 mg/kg as a maximum recommended limit for this BA (DeBeer et al., 2021). No regulations have been established for cadaverine (Sánchez-Parra et al., 2023), since the information regarding the toxicity of this BA is limited, as only a few studies are available in animals (Omer et al., 2021) and no studies are available that analyse dose response in humans.

In the context of foodborne outbreaks of Scombroid poisoning, ensuring control over histamine levels in tuna emerges as critically significant. In the Rapid Alert System for Food and Feed (RASFF) organized among the member states of the European Union, histamine represents one of the most common notifications. During the periods from 2000 to 2010 and 2011 to 2021, a total of 314 and 383 notifications were reported for histamine in fish and fish products, respectively. Numerous studies have been published focusing on the assessment of incidents related to histamine poisoning (Colombo et al., 2018; Leuschner et al., 2013; Visciano et al., 2020). Hence, the development of analytical techniques for the determination of BA levels is essential in assessing food toxicity as it serves as a quality marker of freshness, adherence to good manufacturing practices, and product preservation status (Peng et al., 2008; Shakila et al., 2003). Presently, high-performance liquid chromatography (HPLC) stands as the reference analytical method in the European Union (EU) as governed by the Commission Regulation (EC) No. 1441/2007, amending Regulation (EC) No. 2073/2005, and by Commission Regulation (EU) No. 1019/2013, amending Annex I to Regulation No. 2073/2005. However, in the literature, several alternative techniques have been used for the analysis of BAs, such as thin-layer chromatography (Bajc and Gačnik, 2009; Lapa-Guimarães and Pickova, 2004; Tao et al., 2011), gas chromatography-mass spectrometry (GC-MS) (Awan, 2008; Huang et al., 2016; Kamankesh et al., 2019), the enzyme-linked immunosorbent assay (Köse et al., 2011), fluorimetric methods (Muscarella et al., 2013), ion-mobility spectrometry (Cohen et al., 2015) or real-time mass spectrometry (Nei et al., 2017). These techniques have been widely used to assess the freshness of fish due to their proven performance and accuracy (Cheng et al., 2013). Nevertheless, these techniques have several drawbacks since they not only need to use expensive and contaminant/toxic solvents but are also time consuming and involve laborious sample preparation. Besides, the analysis of samples using chromatographic techniques poses additional difficulties due to limited

absorption properties within the visible, ultraviolet, or fluorescence wavelength ranges, necessitating derivatization for detection (Ordóñez et al., 2016). In this context, in recent years, simple, quick, precise, inexpensive and non-destructive methods have been proposed to complement or replace the traditional techniques. As an alternative, vibrational techniques (near-infrared, Raman spectroscopy, hyperspectral imaging) can offer sensitive, swift, and unique chemical insights for assessing fish quality and safety (Rodriguez-Saona et al., 2016). As a large amount of data is obtained after the analysis of the samples using vibrational techniques, chemometric tools are used to build mathematical models that allow samples to be differentiated (Almoujahed et al., 2023; Cárdenas-Escudero et al., 2023; Massart et al., 1997; Peris-Díaz and Krezel, 2021).

In the literature, vibrational spectroscopic techniques coupled with machine learning have been used to assess food (Abdel-Nour et al., 2011; Magwaza et al., 2012; Nguyen et al., 2022; Qi et al., 2022) and fish freshness (Franceschelli et al., 2020; Wang et al., 2019). Furthermore, the quantification of histamine in fish matrices to implement quick and efficient tools to facilitate verification within the industry is necessary. In their research, Ghidini et al. (2021) investigated the application of NIR spectroscopy (1000-2500 nm) for estimating the histamine levels in both raw and processed tuna fish. They employed Orthogonal PLS regression to establish a correlation between the spectral data and the histamine concentrations determined through the reference HPLC method, ranging from 10 to 1000 mg/kg. Moreover, Asghari et al. (2022) proposed attenuated total reflectance-Fourier transform infrared (ATR-FTIR) spectroscopy as a non-invasive, robust and rapid method to measure histamine levels in tuna fish samples. They utilized a PLS regression alongside two wavelength selection techniques: interval-partial least squares and the generic algorithm (GA). The authors reported a lower quantification limit (4.68 mg/kg) for GA-PLS obtained in a histamine range between 5 and 100 mg/kg. The two studies mentioned above showed promising results when measuring histamine content in fresh and processed tuna fillets, but they presented drawbacks that made both approaches more time-consuming than expected, since pre-processing steps such as sample extraction were included and therefore involved a destructive analysis of the sample.

The purpose of this research is to evaluate the use of FT-MIR spectroscopy in combination with chemometric methods for two main objectives. The first one corresponds to the quantitative measurement of different histamine levels in raw tuna

fillets, without any other sample pretreatment. The second objective, which we believe is novel, consists of the integration of discrimination algorithms with FT-MIR spectroscopic analysis to distinguish tuna samples according to their histamine concentration, in accordance with European Commission and FDA regulations.

2. Material and methods

2.1. Chemicals

All reagents utilized were of analytical-chemical grade. Panreac (Barcelona, Spain) supplied toluene, perchloric acid, methanol, acetonitrile, toluene, deionized water, and sodium carbonate. Histamine, dansyl chloride, L–proline and 1,7– diaminoheptane (internal standard) were obtained from Sigma–Aldrich (Steinheim, Germany).

2.2. Fish Samples and Experimental Design

A total of 66 samples of Yellowfin tuna (Thunnus albacares) fillets from six different tuna batches were collected during the period between July 2020 and December 2021 from five different companies in Andalusia, Spain (FAO area 34, Eastern Atlantic Ocean). The samples were kept on ice during both sampling and transportation to the laboratory. Then, in the laboratory, the samples were filleted with an area of 3.5 x 3.5 cm in freezing conditions, individually packed in plastic bags, divided into randomly six groups of 11 samples and kept at -20ºC. The experimental design consisted of a control group (coded as 0), and five groups stored at 22 ± 2°C for 1, 3, 5, 7, and 10 days, respectively, to potentially vary histamine levels.

2.3. HPLC-DAD analysis

The content of histamine was analysed following the official reference method mandated by European Regulation No. 2073/2005, as per the procedure proposed by Duflos et al. (1999) and modified by Duflos et al. (2019). In summary, 5 g of fresh tuna samples were weighed and mixed with 10 mL of 0.2 M perchloric acid solution and 100 µL of 1,7– diaminoheptane (internal standard, 6.4 mg/L) in a centrifuge tube. This mixture underwent homogenization using an Ultra-turrax homogenizer (Ultraturrax®, Stauten, Germany) within an ice bath and was subsequently centrifuged for 10 min at 15,000 rpm at 4 ºC. For the derivatization step, 0.100 mL of supernatant was combined with 0.300 mL of sodium carbonate solution and 0.400 mL of dansyl chloride solution

(7.5 mg/mL) in an Eppendorf tube. After vortexing, the tube was incubated at 60 °C for 5 minutes in a stirred water bath (Unitronic Reciprocating Shaking Bath, model 6032011, J.P. Selecta, Barcelona, Spain). The excess of the derivatization reagent was then neutralized by adding 0.1 mL of L-proline solution (0.1 mg/L) and left in the dark for 15 minutes, followed by another 15 minutes of darkness after vortexing.

Following this, 0.5 mL of toluene was added, and the tubes were vortexed and frozen to separate the organic phase containing the histamine dansyl derivate of the aqueous phase at –80 °C for 15 minutes. The isolated organic phase underwent concentration through drying using a Speedvac concentrator (Eppendorf, Hamburg, Germany) and the resulting dry residue was resuspended in 0.2 mL of a solution comprising acetonitrile:water (6:4;v/v), followed by vortexing and centrifugation.

The histamine was determined using a liquid chromatograph system coupled to a diode array detector (HPLC-DAD, PerkinElmer PE200, Waltham, MA, USA) equipped with an autosampler. The reverse-phase column employed was a Luna C18 (5µm, 250 x 4.6 mm) with a C18 pre-column, 4.0 x 3.0 mm, from Analytical Phenomenex (Torrance, CA, USA). The detection was performed at 254 nm. The histamine was quantified using the individual standard curve (Table S1). All measurements were conducted three times for each sample.

2.4. Fourier transform mid-infrared (FT-MIR) spectroscopy

The instrumentation utilized was a FT-MIR Vertex 70 spectrometer (Bruker Optics, Ettlingen, Germany) equipped with a Globar source and a room temperature deuterated lanthanum α-alanine-doped triglycine sulfate (DLaTGS) detector. For analysis, each tuna flesh (3.5 x 3.5 cm) sample was stored in a refrigeration chamber at 4°C for 15 hours, until completely thawed. Then, the analyses were replicated three times placing the tuna samples directly onto the attenuated total reflectance (ATR) crystal, ensuring complete coverage of the crystal surface and optimal contact between the tuna samples and the crystal. To mitigate external interferences, the ATR crystal underwent cleaning with distilled water and 70% ethanol, followed by drying with wiping paper after each sample scan. Prior to each instance of scanning, the spectra against air (mainly H_2O and CO_2) were recorded. Spectra were obtained at a resolution of 4 cm^{-1}, spanning the spectral range of 4000 to 600 cm^{-1} (1763 variables), with an average of 64 scans. All data

analyses were performed at a room temperature of 25 ºC. The data were exported using Opus 7.2 software package (Bruker Optics Inc., Billerica, MA, USA).

2.5. Chemometric methods

In the chemometric analyses, the dependent variable (Y) was the concentration of histamine determined by the reference methods previously indicated, while the FT-MIR spectra were used as the independent variables (X). The chemometric techniques applied in this research, as follows: 1) Principal component analysis (PCA) to discern interrelationships among samples (clusters) and identify outliers in the X dataset, and 2) Partial least squares (PLS) regression. Furthermore, three discriminant analysis methods were applied to develop models for classifying samples under the limits set forth by FDA and European legislation: i) Partial least squares-discriminant analysis (PLS-DA); ii) K-nearest neighbors (KNN); and iii) Support Vector Machine (SVM) (Berrueta et al., 2007; Borràs et al., 2015; Cozzolino et al., 2011; Djuris et al., 2013). These multivariate analyses were carried out using PLS toolbox 7.0 (v.7.0, Eigenvector USA) under Matlab 2017bR (Mathworks, USA).

2.5.1. Data pre-processing

The average of the three replicates for each sample was computed to derive a final representative spectrum (Figure 1). To remove baseline offset, light scattering effects, signal noise, optical path shift or universal intensity changes, among others, spectral pre-processing was applied to the raw data (Mishra et al., 2020; Yang et al., 2021). In this work, ten different spectral preprocessing methods were applied to improve recognition accuracy prior to modeling, including First-Derivative (Saviztky-Golay algorithm; 9 and 15-point window), Second-Derivative (Saviztky-Golay algorithm; 9 and 15-point window) (Savitzky and Golay, 1964), Standard Normal Variate (SNV) (Barnes et al., 1989), Multiplicative Scattering Correction (MSC) (Geladi et al., 1985), and combined pretreatment as First Derivative-SNV (FD-SNV), Second Derivative-SNV (SD-SNV), SNV-First Derivative (SNV-FD) and SNV-Second Derivate (SNV-SD). Once the different combinations were checked, the best spectra pre-processed observed were the SNV combined with Savitzky-Golay first derivate (quadratic polynomial fit using a 15-point window).

2.5.2. Principal component analysis

PCA is an unsupervised pattern recognition technique widely used to categorize samples based on spectral differences. PCA reduces the dimensionality of complex datasets by extracting essential information based on the spectral attributes of the samples examined. The creation of new uncorrelated variables from the original set of variables is called principal components (PC) (Liu et al., 2020; Shao et al., 2022). As a result, the interrelationships between the original variables and the samples, in the new PC space, can be identified as clusters in the maps of loadings and scores, respectively. In this study, PCA was applied to investigate the clustering of samples and to detect outliers by applying the Hotelling T^2 statistic (95% confidence interval) (Nieuwoudt et al., 2004; Vermeulen et al., 2021).

2.5.3. Quantitative Models for Histamine

Partial least squares regression (PLSR) models were developed to establish a linear correlation between the FT-MIR spectral data of samples (X) and the different variables to be predicted (Y), namely histamine (Lee et al., 2011; Lutz et al., 2006; Ramadan et al., 2006). PLSR, an advanced technique that combines the features of PCA and regression, effectively addressed the main drawbacks associated with spectral data analysis, such as collinearity and overlapping bands. This technique is advantageous when the number of independent variables exceeds the number of dependent variables (Iqbal et al., 2013; Mahanti et al., 2020). PLSR aims to maximize covariance, capturing variance and establishing correlations within the data (Xiaobo et al., 2010). The fundamental principle underlying PLSR is to extract latent variables (LVs) that account for as much of the spectral variance as possible while modelling the variables to be predicted (Roy et al., 2015).

To avoid bias in the subset division, all samples were initially sorted in ascending order based on their reference histamine levels. Subsequently, the dataset was randomly divided into three subsets:

- Subset 1: samples with concentrations <100 mg/kg (number of samples 20)
- Subset 2: samples with concentrations >300 mg/kg (number of samples 23)
- Subset 3: samples with concentrations between 100 and 300 mg/kg (number of samples 23)

In this regard, the models were constructed as follows:

(A) Model 1: calibration set comprising Subset 2 and 3 (N=46) and validation set including Subset 1 (N=20).
(B) Model 2: calibration set comprising Subset 1 and 3 (N=43) and validation set including Subset 2 (N=23).
(C) Model 3: calibration set comprising Subset 1 and 2 (N=43) and validation set including Subset 3 (N=23).

To validate the PLSR models, a 10-fold cross-validation was employed. This cross-validation approach was employed to guarantee that the model was trained to accommodate the variability present in the calibration dataset throughout the training phase. The performance of the calibration model was assessed using the number of LVs and the root mean squares error of cross-validation (RMSECV) as an internal indicator of the predictive ability of the models (Equation 1):

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad [1]$$

where $\hat{y}_i$ is the concentration of histamine predicted by the model, y_i is the concentration reference of histamine and n is the number of samples in the calibration set.

Once the final models were established, the validation set was used to evaluate the expected error when the model was applied to predict new samples. Root mean squared errors of prediction (RMSEP) were calculated using Equation 2:

$$RMSEP = \sqrt{\frac{\sum_{i=l}^{n_t}(y_{t,i} - \hat{y}_{t,i})^2}{n_t}} \quad [2]$$

where $\widehat{y_{t,i}}$ is the concentration of histamine predicted by the model, $y_{t,i}$ is the concentration reference of histamine and n_t is the number of samples in the validation set.

The criteria for selecting the most suitable model focused on minimizing the RMSECV and RMSEP, optimizing the number of LVs and maximizing R^2 values (Wilkerson et al., 2013).

2.5.4. Discriminant Analysis Algorithms

Three different multivariate data analysis techniques — namely, Partial Least Squares-Discriminant Analysis (PLS-DA), the K-nearest neighbors (KNN) technique and Support Vector Machines (SVM) — were employed to build models for classifying

samples as falling below or above the limits specified by the European legislation (100 mg/kg) or the FDA regulation (50 mg/kg) to achieve a correct classification based on the histamine concentration limits.

The confusion matrix is a concept from machine learning, to evaluate the performance of a classification model. It offers a concise overview of the model's predictions, categorizing instances into four groups: true positives (TP, correctly predicted positive instances), true negatives (TN, correctly predicted negative instances), false positives (FP, incorrectly predicted as positive), and false negatives (FN, incorrectly predicted as negative). In this study, the performance of the models was evaluated in terms of sensitivity (SEN) and specificity (SPE) defined as follows:

$$\mathrm{SEN} = \frac{\mathrm{TP}}{(\mathrm{TP+FN})} \qquad [3]$$

$$\mathrm{SPE} = \frac{\mathrm{TN}}{(\mathrm{TN+FP})} \qquad [4]$$

where SEN demonstrated its capability to identify samples belonging to the target class and SPE indicated the model's capacity to distinguish and reject samples from other classes. The percentage of correct predictions (true positive and true negative) and its corresponding confusion matrix served as the ultimate parameters for assessing the goodness of the model.

Validating discriminant models holds great importance in assessing the performance of classification models. Due to the influence of the choice of segmentation on the sensitivity of validated misclassification rates (Kjeldahl and Bro, 2010), drawing conclusions about the real dependency of data predictions on a random artefact of a simple structure becomes challenging (Ojala and Garriga, 2010). The Kennard-Stone algorithm (KS) (Saptoro et al., 2012) considers all samples as potential candidates for the training set and, in turn, selects specific samples to form the validation set based on a 2:1 ratio. The main objective of the KS algorithm is to maximize the minimum Euclidean distances between the already selected samples and the remaining ones (Claeys et al., 2010), as defined by Equation [5].

$$d_x(p,q) = \sqrt{\sum_{j=1}^{N} [x_p(j) - x_q(j)]^2};\ p,q \in [1,N] \qquad [5]$$

where N is the number of spectral wave points of the sample, and x_p and x_q represent two different samples.

Initially, KS identifies the two samples with the maximum Euclidean distance, forming the starting point of the selection process. Subsequently, for each remaining sample, the algorithm stores the nearest Euclidean distances in a distance list along with the corresponding sample number. From this list, the sample with the maximum distance is chosen, and this iterative procedure continues until the desired number of samples is achieved. In our modelling approach, two-thirds of the samples (Ncalibration = 45) were allocated to the calibration group, and the remaining one-third of the sample set was selected as a validation group to obtain predictions using the model (Nvalidation = 21).

As we all know, ensuring an effective evaluation is crucial to check the performance of each classifier. This enables us to ascertain whether a specific classification approach is sufficiently adapt for particular predictive tasks. However, this critical evaluation step is seldom addressed in other spectroscopy research publications. For this purpose, to optimize the model parameters in both PLS-DA, KNN, and SVM, a 10-fold cross-validation technique was employed. During this cross-validation process, each subset was employed once as a validation set, with the remaining nine subsets utilized as the calibration set. This process was repeated ten times, employing distinct subsets for evaluation on each iteration, and the outcomes were averaged to gauge model performance. This approach provides a more objective examination of how the model might perform with unknown samples (Stone, 1974). Hence, five different datasets were randomly constructed, and the three models were reconstructed. Moreover, for the validation of the classification model, a permutation test was proposed (Westerhuis et al., 2008).

In summary, in a permutation test, the class labels are permuted and randomly reassigned 'incorrectly' to the samples. The model is then constructed using these samples with the wrong class assignments, compelling the model to generate both false negative and false positive outcomes (Golland et al., 2005; de Souza et al., 2020). This procedure enables the assessment of the probability that predictions occur by random chance, aiding in verifying whether the optimized parameters of a model are susceptible to overfitting or not (Liu et al., 2006). In this study, a permutation test with 200 iterations was conducted, generating a random dataset under a null hypothesis H_0 (no difference between classes). For the results obtained from the non-permuted sample set to be considered significant,

they must fall outside the 95 or 99% confidence limits of the H_0 distribution derived from the permuted classifications. Subsequently, the significance of the models was evaluated using the Wilcoxon test (Pratt, 1959), the sign test (Thomas, 2003), and the t-random test (van der Voet, 1994), all at a significance level of 95% ($\alpha = 0.05$).

2.5.4.1. Partial Least-squares Discriminant Analysis (PLS-DA)

PLS-DA is a supervised method based on the partial least squares regression, which transforms the regression method into a technique for discriminating multivariate chemical data (Gromski et al., 2015). Its primary objective is to construct classification models applicable to future predictions (de Santana et al., 2016). This is achieved through the utilization of a dummy matrix (Y), an N x F, matrix with N rows (total number of samples) and F columns (number of classes), encoding class membership through a binary system. In a two-class scenario, as discussed in this study, the dummy vector Y contains 1s in rows corresponding to Class 1 and 0s in the remaining rows (Class 2). The PLS-DA model is then computed by estimating Equation 6:

$$y = Xb + e \quad [6]$$

where X represents the data matrix, and b and e are vectors of regression coefficients and residuals, respectively.

When applying the model to a new set of measures (X_{new}), the predicted Y_{new} comprises continuous values, so a rule is needed to classify the samples. In a two-class scenario, a common practice is to set a threshold at 0.5, despite potential classification errors due to the method's general nature. However, various literature approaches aim to refine this choice (Barker and Rayens, 2003; Indahl et al., 2007; Pérez et al., 2009). In this study, a more refined threshold was determined using a Bayesian algorithm, initially estimating probabilities, and subsequently discriminating samples. The objective was to identify a threshold where FP and FN are minimized (Tormena et al., 2019). Values exceeding this threshold signify that the samples pertain to the modeled class, while lower values suggest samples that do not belong to this class (Valderrama et al., 2022).

2.5.4.2. K-Nearest Neighbors (KNN)

The KNN is a linear and non-parametric supervised pattern recognition method (Clarke et al., 1974). The principle of this method is based on proximity - it classifies an unknown sample of the validation set based on the majority of its K-nearest neighbours

in the calibration set (Berrueta et al., 2007). Commonly, similarity is measured by the Euclidean distance between spectra, and classification is performed on the group to which most of the k objects belong, with ties broken by the sums of the relevant distances. The parameter K has a large impact on the classification model; and it is optimized by calculating the prediction ability at different values of k. It is often advisable to choose lower k values, such as 3 or 5, when employing this algorithm (Chen et al., 2011).

The KNN method offers several advantages, among other, its mathematical simplicity, allowing it to produce classification outcomes potentially superior to other more intricate pattern recognition techniques. Additionally, its efficiency remains consistent regardless of the spatial distribution of classes. However, it's worth noting that KNN may struggle when significant disparities exist in the sample sizes of each class, as this can lead to excessively slow computations (Berrueta et al., 2007; Jiang et al., 2007).

2.5.4.3. Support vector machines (SVM)

The theory of SVM has been extensively described in literature (Fernández-Pierna et al., 2004; Fernández-Pierna et al., 2005). SVM is a non-linear supervised statistical learning method developed by Vapnik and co-workers (Cortes and Vapnik, 1995; Vapnik, 1995).

The concept of SVM stems from the classification of binary problems, aiming to identify a hyperplane that effectively separates two data sets. In case the linear boundary in the low-dimension input space is insufficient for the proper separation of two classes, it is possible to create a hyperplane that allows a linear separation in a higher-dimensional feature space. This transformation is achieved through a conversion function that maps data from the original input space to a higher-dimensional feature space, making it linearly separable. This transformation is facilitated by a kernel function (Mammone et al., 2009). Through an appropriate selection of a kernel function, any consistent training set can be made separable. In this study, a Gaussian kernel function was selected as it is the simplest and quickest to calculate (Berrueta et al., 2007). Its structure is the radial basic function (RBF) Equation 7:

$$K(x_i, x_j) = \exp\left(-\frac{||x_i - x_j||^2}{2\sigma^2}\right) \qquad [7]$$

where σ is the bandwidth of the RBF function (kernel parameter) and it reflects the degree of generalization.

To obtain a good performance of SVM model, some parameters need to be optimized by grid search (GS) (Fayed and Atiya 2019). These parameters include: C (the penalty factor) and σ (the radial width of the kernel function). C minimizes both the fitting error and the model's complexity, whereas σ determines the non-linear mapping from the input space to the high-dimensional feature space (Li et al., 2019). When C is decreased, more emphasis is placed on maximizing margin and enhancing generalization. In addition, generalization can also be improved by increasing the value of σ in the Gaussian function.

3. Results and discussion

3.1. Histamine analysis

In order to identify possible Y data (histamine concentration) outliers during the 10-day incubation period at room temperature (22 ± 2 ºC), a boxplot analysis was carried out and is presented in Figure 2. In this graph, the median was represented as the central mark, the 25th and 75th percentiles marked the edges of the box, and the whiskers extended to the most extreme data points, excluding outliers, as outlined by Quintelas et al. (2019).

Regarding the data, the samples were distributed based on their concentration (Table 1). In this sense a total of 16 samples were identified with a histamine concentration below 50 mg/kg (the limit established by the FDA). After being caught, tuna is frozen aboard fishing vessels using a brine immersion freezer set at temperatures below -8 ºC. The protective influence of salt is credited with reducing the likelihood of microbial contamination (Barbosa et al., 2018). This explains why low histamine levels were found in those samples. For the range established between 50 and 100 mg/kg, only five samples were detected, while 13 samples presented concentrations in the range between 100 and 200 mg/kg. The levels of histamine in the samples at room temperature showed a significant increase over the incubation period (Figure 2), reaching concentrations close to 1000 mg/kg. This phenomenon can be attributed to the elevated bacterial growth and enzyme activity observed at this temperature, as reported by Ekici and Omer (2020), which accelerates the decarboxylation of the amino acid histidine into histamine (Altieri et al., 2016).

The fluctuations in histamine levels observed over the storage time of the tuna could be attributed to variances in the microbial levels among the collected samples. Additionally,

prior studies have indicated that the production of this amine is influenced by factors such as the individual fish, the sampled fish part, time, temperature, and the type and number of bacterial species present (Economou et al., 2007; Sánchez-Parra et al., 2022). The interplay of these factors may contribute to the variation in the histamine levels within individual lots of fish and even among individual fish.

3.2. Principal Component Analysis (PCA)

PCA was performed using pre-processed FT-MIR spectra to identify the capacity of samples clustering by SNV + first derivate Savitzky-Golay. Figure 3a shows the scores plot for the first two PCs. The total variance explained by the first two components was 84.52% (PC1 = 72.37% and PC2 = 12.15%). It is possible to observe a pattern clearly demonstrates the grouping of tuna samples into seven categories, based on their histamine levels (Fig. 3a). Moreover, the PCA suggests the presence of two clusters. The samples with a higher concentration of histamine are on the positive quadrant of the PC1 (histamine levels ranging from 200 to 1000 mg/kg) and samples with lower histamine levels (<LOD (limit of detection) to 200 mg/kg) tended to move towards the PC1 negative quadrant. These findings suggest that the second primary component correlates with the spectral variability among samples, stemming from differences between individual fish and batches, while the first primary component is linked to the level of tuna flesh decomposition (histamine content).

To further investigate the bands with the greatest influence on potential sample discrimination, we analyzed the graph of the loadings corresponding to PC1 and PC2 (Fig. 3b). Typically, the MIR spectrum encompasses four identifiable regions: the double-bond region (2000–1500 cm^{-1}), the fingerprint region (1500–600 cm^{-1}), the X–H stretching region (4000–2500 cm^{-1}), and the triple-bond region (2500–2000 cm^{-1}) (Karouiet al., 2010). The higher loading of PC1 and PC2 are linked to the most important regions of the tuna samples spectrum. In our study, specific peaks stood out, specifically at 1634 and 1659 cm^{-1} (weights positively in PC1 and PC2, respectively). These peaks correspond to the Amide I band, which represents the most intense absorption band in proteins. The highlighted peaks resulted from the stretching vibrations of the C=O (70–85% of the potential energy) and C-N groups (10–20%) (Bandekar, 1992; Karoui et al., 2010). This frequency typically falls within the range of 1600 to 1700 cm^{-1}. Next to the Amide I group, another prominent peak emerges at 1564 cm^{-1}, attributed to Amide II group. This region is more intricate than Amide I and mainly stems from in-plane N-H

bending (40–60%), with the remaining potential energy assigned to C-N (18-40%) and C-C stretching vibrations (approximately 10%) (Venyaminov and Kalnin, 1990). The increase in the signal intensity could be explained as a direct increase of free amino acids and peptides resulting from proteolysis during the tuna decomposition period. Additionally, the loadings revealed a contribution from bands belonging to the region between 1400 to 1200 cm^{-1}, corresponding to Amide III (1381 and 1377 cm^{-1}). This region encompasses complex bands influenced by the force field details, the nature of the side chain and the hydrogen bonding (Karoui et al., 2010). Moreover, absorption bands were observed in the 950-1200 cm^{-1} range, corresponding to the C-N stretch of histamine compound (1160 cm^{-1}). Lastly, the peak at 3300 cm^{-1} is due to the vibration of the O–H bond in water and the amide A of proteins (3270 cm^{-1}). These results indicated that the negative variation found by PCA analysis, regarding PC1, could be associated with the decrease in histamine content present in the tuna samples.

3.3. PLS models for quantitative predictions

The PLSR models for histamine quantification were developed utilizing various preprocessing techniques outlined above. Various parameters were employed to evaluate the performance of these PLSR models. Accuracy was assessed through coefficients of determination for calibration (R^2_C) and prediction (R^2_P). A coefficient of determination approaching 1 indicates a strong correlation between the predicted and measured values in both calibration and prediction sets. The RMSECV based on contiguous cross-validation procedure was utilized to evaluate the modeling capacity of the PLSR model using calibration set. In this study, ten different spectral preprocessing methods, namely, FD (9 and 15 points window), SD (9 and 15 points window), SNV, MSC, FD-SNV, SD-SNV, SNV-FD and SNV-SD, were compared to investigate their influences on the performance of the PLSR models. The results of the optimization of the spectral pretreatment methods are shown in Tables S2, S3 and S4.

Three PLSR models were developed (Figure 4). The best performance of the PLSR model 1 (Subsets 2 and 3) for the quantitation of histamine was obtained based on the preprocessing of SNV+FD (Savitzky-Golay algorithm, quadratic polynomial fit using a 15 points window) (Savitzky and Golay, 1964), with the latent variables of 5, the R_{cv}^2 and R_p^2 of 0.991 and 0.978, and the RMSECV and RMSEP of 21.7875, and 5.8435 mg/kg, respectively (Table S2). With respect to PLSR models 2 and 3 (Subsets 1 and 3, and Subsets 1 and 2, respectively), the optimal quantification models were developed

after the same pretreatment of SNV+FD, with the latent variable of 6. The optimum latent variables used in the model development determined the lowest error in cross-validation (Fig. 4) and therefore avoided overfitting (Prieto et al., 2014). The RMSECV was 22.7376 and 15.6790 mg/kg, respectively. Moreover, the R_p^2 and RMSEP were 0.970 and 9.28 mg/kg for the PLSR model 2 (Table S3) and 0.967 and 37.0270 mg/kg for the PLSR model 3 (Table S4).

The values of the R^2 in all models indicated the excellent quality of the fit. When utilizing spectroscopy methods to predict chemical characteristics and quantify components in food, achieving an R^2 value surpassing 0.95 is considered an excellent indicator of the model's quality (Shenk and Westerhaus,1996). However, the probability of attaining such optimal outcomes rises proportionally with the quantity of the specific component to be quantified. Despite histamine in fish being present at parts per million (ppm) levels, the reliability of obtaining an $R^2 > 0.90$ can be linked to the suitability of pre-processing techniques. These techniques are known to enhance the linear relationship between spectral signals and analyte concentrations (Rinnan et al., 2009). This improvement has been validated for various food contaminants found at comparable or lower concentration ranges than those of biogenic amines, like total volatile basic nitrogen (TVBN) or K value (Ding et al., 2014; Liu et al., 2022; Yan et al., 2023). Although the RMSECV obtained in the models using the 10-fold cross-validation were higher than the calibration, a well-dispersed and random distribution of the residuals was obtained (data no shown). This fact can be explained by the number of spectra used in the validation. The parameter RMSEP expresses the average error expected when the calibration model is applied to unknown tuna samples in future predictions. Low values indicate that these models were reliable and robust. Hence, the prediction models derived from FT-MIR spectra fitted well with experimental measurements and hold significant promise for advancing the development of a novel methodology for quality control and food safety applications.

3.4. Qualitative analysis of histamine in tuna samples with classification purposes based on existing (EU and FDA) legislations

Recent research suggests that a promising avenue for further exploration involves employing the European and FDA thresholds to construct classification models for discerning fish samples (tuna, sardines) based on their histamine content (Asghari et al., 2022; Ghidini et al., 2021). In most food-related quality control problems, higher

sensitivity rates are more critical than specificity. Based on the data regarding the notifications of histamine in fish and fish products, we considered that the availability of rapid techniques would be of interest to the industries, as those techniques could be applied in the reception of tuna and along the production chain as a means of discriminating between tuna samples.

The three supervised methods used, PLS-DA, KNN and SVM, were developed using two sets (calibration and prediction) by similarity grouped according to the Kennard–Stone algorithm. The calibration set was composed of 45 samples whereas the test set consisted of the remaining 21 samples. Only the calibration data set was used to build the classification model, while the prediction data set was used to test its ability to classify new samples. This process was carried out 5 times to obtain robust and non-random discrimination models.

For the European regulation models, the samples were divided into two classes: Class 1 (tuna samples with HIS concentration < 100 mg/kg) and Class 2 (tuna samples with HIS concentration > 100 mg/kg). Different pre-treatments were applied on the calibration set. In particular, the SNV in combination with the first derivative Savitzky-Golay with a 15 points window and quadratic polynomial fit was the best pre-treatment to the correct classification in a 10-fold-cross-validation. Table 2 shows the results obtained for PLS-DA, KNN and SVM. Sensitivity and specificity were used in order to evaluate the classification models.

In machine learning algorithms, adjusting parameters is the algorithm's learning process. Consequently, identifying appropriate parameters is crucial for the outcome of each algorithm (Xia et al., 2023). The optimal number of latent variables in the PLS-DA model is determined by 10-fold cross-validation. However, when the number of latent variables reaches 10 or higher, there is no improvement in the classification accuracy of the PLS-DA model. At this point, when the LV was 4, the error rate was lowest. The PLS-DA model provided a sensitivity and a specificity for both classes of 1, namely 100% classification accuracy (Table 2). The threshold for each class was obtained by using the Bayesian theorem and respective data. By means of five different models, it is correctly classified 14.6/15 samples of Class 1 and 29.8/30 samples of Class 2 (Table S5), while in external validation, 94.3% of the samples with a concentration of histamine below 100 mg/kg were correctly classified (Class 1). In the SVM models, the radial basis function (RBF) was chosen as the core function where only two parameters need to be tuned: the

cost function and the kernel parameter. To optimize these two hyperparameters, the most accurate values in the 10-fold cross-validation were selected to avoid overfitting and to obtain a good accuracy rate. For the five different models, the cost function value was 100, while the kernel parameters were 1, 0.01, 0.032, 0.01 and 3.16. As seen in Table 2, all samples were correctly classified (100%) by SVM in the calibration set. On the other hand, in the 10-fold cross-validation, 97.1 % of samples below 100 mg/kg were well-classified. For external validation, the sensitivity and specificity values were 92.9% and 98.2%, respectively. The classification of the Class 2 samples was worse using this model, where only 28.6/30 samples were classified correctly, as confirmed by the confusion matrix (Table S5). Thirdly, modelling using the KNN approach was carried out. In this study, 10-fold cross-validation was employed to ensure the K value. When the K value was 3, the highest accuracy was ensured. In terms of sensitivity and specificity for cross-validation, the model showed values of 97.1% and 98.8%, respectively (Table 2). The best classification results in the training set were obtained using KNN model for FT-MIR data, giving 100 % sensitivity for the two different histamine concentrations (Table 2), indicating reliability and good generalization of the model data. Although the sensitivity of PLS-DA model was 94.3 % for the samples below 100 mg/kg of histamine (Class 1), it showed good predictability for unknown samples since none of the 15 samples that exceeded the 100 mg/kg threshold were misclassified by the confusion matrix (Table S5).

The FDA regulation sets a stringent limit of 50 mg/kg for histamine in fish. To accommodate this stricter regulation, we categorised the samples into two classes for the classification models: Class 1 (comprising tuna samples with HIS concentration < 50 mg/kg) and Class 2 (consisting of tuna samples with HIS concentration > 50 mg/kg). The optimal number of latent variables in the PLS-DA model was 3 and the optimum cost function value for the five SVM models was 10, while the kernel parameters were 0.01, 0.01, 0.032, 0.01 and 1. In KNN, the highest accuracy was ensured with a value of K=3. Table 3 shows the data obtained for the three calibration models. Regarding sensitivity and specificity in 10-fold cross-validation, the following results were obtained: 94% and 93.7% for the PLS-DA model; 90% and 98.3% for the SVM model; and 92% and 95.3% for the KNN model for Class 1. On the other hand, the prediction results obtained for the PLS-DA and KNN models indicated that 5.4 out of 6 samples were correctly classified as samples with histamine values below 50 mg/kg. In contrast, for the SVM model, only 4.6 out of 6 samples were correctly classified (Table S6).

Furthermore, due to the validation set having few samples, it is essential to determine whether the model's predictions stem from a genuine reliance on the spectral data or are merely the outcome of random chance. To address this, we employed a permutation test with 200 iterations to validate the significance of the three models' predictions for self-prediction. As indicated by the Wilcoxon, Sign Test, and Rand t-test values, the models exhibited statistically significant differences ($p < 0.05$) (Table S7 and S8). Consequently, we can confidently conclude that the models yield dependable predictions that are inherently linked to the underlying structure of the data, rather than being a product of chance (Brereton, 2006; Kjeldahl and Bro, 2010). The models examined in this research effectively categorised tuna samples that fail to meet the maximum histamine limits set by EU and US-FDA regulations in a quick and non-destructive manner. They can be used to guarantee that enterprises acquire tuna of the required quality, for both producers and consumers from a food safety perspective.

In brief, the predictive performance of the KNN model was superior to that of PLS-DA and SVM for both regulations. This arises from the nature of the KNN algorithm, which operates as a supervised learning method, incorporating all data during each training phase to determine the most effective model. Typically, this method yields to good results in scenarios where there are small differences among samples within the same group (Zhao et al., 2010). Additionally, KNN is robust against data variability and has few hyperparameters to optimize, unlike SVM.

4. Conclusions

In this paper, the histamine content in yellowfin tuna was analyzed by HPLC-DAD. This amine showed an increasing trend with the increase in the incubation period, attributed to higher bacterial growth and enzyme activity that accelerate the decarboxylation of amino acids. The current research highlights the great potential to improve the estimation of the histamine level using FT-MIR spectroscopy, particularly when employing appropriate spectral pre-processing techniques and chemometric methods. In general, the PLSR models proved to be robust. Better results were achieved by employing SNV with First Derivate Savitzky-Golay pre-processing, leading to decreased errors in predicting histamine content.

The regulation of histamine levels in tuna is highly relevant, especially in the context of food poisoning outbreaks. To address this concern, three classification models (PLS-DA, SVM, and KNN) were applied to distinguish tuna samples surpassing threshold limits established by EU and US-FDA regulations. The results obtained underscore the feasibility of using FT-MIR spectroscopy combined with multivariate analysis for rapid and non-destructive safety inspections. This technology serves as a vital and expeditious complement to the reference HPLC-DAD method in the industry. Future research should concentrate on validating the transferability of models to portable devices for on-site and real-time screening. Additionally, collecting more representative samples annually to update the database and enhance model robustness should be a focus of future work. The results supported that machine learning models could enhance the prediction performance compared to traditional modelling. These machine learning approaches were validated, including internal 10- fold cross-validation and external independent validation. Following cross-validation, the KNN model yielded the highest classification, achieving a 100% classification accuracy during external validation, according to EU Regulation.

Subsequent works should focus on confirming the adaptability of the models for on-site and real-time screening on portable devices. Moreover, there is a need to collect more diverse samples annually to continually update the database and strengthen the models' reliability.

Acknowledgements

This research was supporting by project PR.PEI.IDF2019.003 funded by the Andalusian Institute of Agricultural and Fisheries Research and Training (IFAPA, Spain), and by the European Rural Development Fund (ERDF, EU). J.L.O.D. received support through a postdoctoral contract from the Junta de Andalucía under the PAIDI 2020 program (POSTDOC_21_00914). M.S.P. received funding through a research contract from IFAPA, as well as support from the National Youth Guarantee System, funded by the European Social Fund, European Union (ESF), and the Youth Employment Initiative, Spain (YEI). The authors thanks to Mr. Quentin Arnould of the CRA-W for his technical assistance with the FT-MIR.

REFERENCES

Abdel-Nour, N., Ngadi, M., Prasher, S., & Karimi, Y. (2011). Prediction of Egg Freshness and Albumen Quality Using Visible/Near Infrared Spectroscopy. *Food and*

Bioprocess Technology, *4*(5), 731–736. https://doi.org/10.1007/s11947-009-0265-0

Almoujahed, M. B., Rangarajan, A. K., Whetton, R. L., Vincke, D., Eylenbosch, D., Vermeulen, P., & Mouazen, A. M. (2023). Non-destructive detection of fusarium head blight in wheat kernels and flour using visible near-infrared and mid-infrared spectroscopy. *Chemometrics and Intelligent Laboratory Systems*, *245*(October 2023), 105050. https://doi.org/10.1016/j.chemolab.2023.105050

Altieri, I., Semeraro, A., Scalise, F., Calderari, I., & Stacchini, P. (2016). European official control of food: Determination of histamine in fish products by a HPLC-UV-DAD method. *Food Chemistry*, *211*, 694–699. https://doi.org/10.1016/j.foodchem.2016.05.111

Asghari, A., Haj Hosseini, A., & Ghajarbeygi, P. (2022). Fast and non-destructive determination of histamine in tuna fish by ATR-FTIR spectroscopy combined with PLS calibration method. *Infrared Physics and Technology*, *123*(2), 104093. https://doi.org/10.1016/j.infrared.2022.104093

Awan, M. A. (2008). Development of non-aqueous single stage derivatisation method for the determination of putrescine and cadaverine using GC-MS. *Central European Journal of Chemistry*, *6*(2), 229–236. https://doi.org/10.2478/s11532-008-0007-6

Bajc, Z., & Gačnik, K. Š. (2009). Densitometric TLC analysis of histamine in fish and fishery products. *Journal of Planar Chromatography - Modern TLC*, *22*(1), 15–17. https://doi.org/10.1556/JPC.22.2009.1.3

Bandekar, J. (1992). Amide modes and protein conformation. *Biochimica et Biophysica Acta (BBA)/Protein Structure and Molecular*, *1120*(2), 123–143. https://doi.org/10.1016/0167-4838(92)90261-B

Barbosa, R. G., Gonzaga, L. V., Lodetti, E., Olivo, G., Costa, A. C. O., Aubourg, S. P., & Fett, R. (2018). Biogenic amines assessment during different stages of the canning process of skipjack tuna (Katsuwonus pelamis). *International Journal of Food Science and Technology*, *53*(5), 1236–1245. https://doi.org/10.1111/ijfs.13703

Barker, M., & Rayens, W. (2003). Partial least squares for discrimination. *Journal of Chemometrics*, *17*(3), 166–173. https://doi.org/10.1002/cem.785

Barnes, R. J., Dhanoa, M. S., & Lister, S. J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Applied spectroscopy, 43(5), 772-777. https://doi.org/10.1366/0003702894202201

Berrueta, L. A., Alonso-Salces, R. M., & Héberger, K. (2007). Supervised pattern recognition in food analysis. *Journal of Chromatography A*, *1158*(1–2), 196–214. https://doi.org/10.1016/j.chroma.2007.05.024

Borràs, E., Ferré, J., Boqué, R., Mestres, M., Aceña, L., & Busto, O. (2015). Data fusion methodologies for food and beverage authentication and quality assessment - A review. Analytica Chimica Acta, 891, 1–14. https://doi.org/10.1016/j.aca.2015.04.042

Brereton, R. G. (2006). Consequences of sample size, variable selection, and model validation and optimisation, for predicting classification ability from analytical data. *TrAC - Trends in Analytical Chemistry*, *25*(11), 1103–1111. https://doi.org/10.1016/j.trac.2006.10.005

Cárdenas-Escudero, J., Galán-Madruga, D., & Cáceres, J. O. (2023). Rapid, reliable and easy-to-perform chemometric-less method for rice syrup adulterated honey detection using FTIR-ATR. *Talanta*, *253*. https://doi.org/10.1016/j.talanta.2022.123961

Chen, Q., Zhao, J., Chen, Z., Lin, H., & Zhao, D. A. (2011). Discrimination of green tea quality using the electronic nose technique and the human panel test, comparison of linear and nonlinear classification tools. *Sensors and Actuators, B: Chemical*, *159*(1), 294–300. https://doi.org/10.1016/j.snb.2011.07.009

Cheng, J. H., Dai, Q., Sun, D. W., Zeng, X. A., Liu, D., & Pu, H. Bin. (2013). Applications of non-destructive spectroscopic techniques for fish quality and safety evaluation and inspection. *Trends in Food Science and Technology*, *34*(1), 18–31. https://doi.org/10.1016/j.tifs.2013.08.005

Claeys, D. D., Verstraelen, T., Pauwels, E., Stevens, C. V., Waroquier, M., & Speybroeck, V. Van. (2010). Conformational sampling of macrocyclic alkenes using a kennard-stone-based algorithm. *Journal of Physical Chemistry A*, *114*(25), 6879–6887. https://doi.org/10.1021/jp1022778

Clarke, M. R. B., Duda, R. O., & Hart, P. E. (1974). Pattern Classification and Scene Analysis. *Journal of the Royal Statistical Society. Series A (General)*, *137*(3), 442. https://doi.org/10.2307/2344977

Cohen, G., Rudnik, D. D., Laloush, M., Yakir, D., & Karpas, Z. (2015). A Novel Method for Determination of Histamine in Tuna Fish by Ion Mobility Spectrometry. *Food Analytical Methods*, *8*(9), 2376–2382. https://doi.org/10.1007/s12161-015-0129-3

Colombo, F. M., Cattaneo, P., Confalonieri, E., & Bernardi, C. (2018). Histamine food poisonings: A systematic review and meta-analysis. *Critical Reviews in Food Science and Nutrition*, *58*(7), 1131–1151. https://doi.org/10.1080/10408398.2016.1242476

Cozzolino, D., Cynkar, W. U., Shah, N., & Smith, P. (2011). Multivariate data analysis applied to spectroscopy: Potential application to juice and fruit quality. Food Research International, 44(7), 1888–1896. https://doi.org/10.1016/j.foodres.2011.01.041

Debeer, J., Bell, J. W., Nolte, F., Arcieri, J., & Correa, G. (2021). Histamine limits by country: A survey and review. *Journal of Food Protection*, *84*(9), 1610–1628. https://doi.org/10.4315/JFP-21-129

de Santana, F. B., Gontijo, L. C., Mitsutake, H., Mazivila, S. J., De Souza, L. M., & Borges Neto, W. (2016). Non-destructive fraud detection in rosehip oil by MIR spectroscopy and chemometrics. *Food Chemistry*, *209*, 228–233. https://doi.org/10.1016/j.foodchem.2016.04.051

de Souza, S. J., Valderrama, P., Consolin Filho, N., Pilau, E. J., Coelho Tanamati, A. A., Wentzell, P. D., & Março, P. H. (2020). Partial least squares discrimination applied to a few samples dataset: A case for predicting the presence of pesticide in lettuce. *Journal of Chemometrics*, *34*(12), 1–9. https://doi.org/10.1002/cem.3299

Ding, R., Huang, X., Han, F., Dai, H., Teye, E., & Xu, F. (2014). Rapid and nondestructive evaluation of fish freshness by near infrared reflectance spectroscopy combined with chemometrics analysis. *Analytical Methods*, *6*(24), 9675–9683. https://doi.org/10.1039/c4ay01839g

Djuris, J., Ibric, S., & Djuric, Z. (2013). Chemometric methods application in pharmaceutical products and processes analysis and control. In J. Djuris (Ed.), Woodhead publishing series in biomedicine, Computer-aided applications in pharmaceutical technology (pp. 57–90). Woodhead Publishing. ISBN 9781907568275. https://doi.org/10.1533/9781908818324.57

Duflos, G., Dervin, C., Malle, P., & Bouquelet, S. (1999). Relevance of matrix effect in determination of bigoenic amines in plaice (*Pleuronectes platessa*) and withing (*Merlangus merlangus*). *Journal of AOAC International, 82,* 1097–1101.

Duflos, G., Inglebert, G., Himber, C., Degremont, S., Lombard, B., & Brisabois, A. (2019). Validation of standard method EN ISO 19343 for the detection and quantification of histamine in fish and fishery products using high-performance liquid chromatography. International Journal of Food Microbiology, 288(June 2018), 97–101. https://doi.org/10.1016/j.ijfoodmicro.2018.07.023Economou, V., Brett, M. M., Papadopoulou, C., Frillingos, S., & Nichols, T. (2007). Changes in histamine and microbiological analyses in fresh and frozen tuna muscle during temperature abuse. *Food Additives and Contaminants*, *24*(8), 820–832. https://doi.org/10.1080/02652030701278321

European Comission - RASFF Portal. Available online: https://webgate.ex.europa.eu/rasff-window/portal/ (accessed on 14/09/2023)

Ekici, K., & Omer, A. K. (2020). Biogenic amines formation and their importance in fermented foods. *BIO Web of Conferences*, *17*, 00232. https://doi.org/10.1051/bioconf/20201700232

FAO (2011). Global food losses and food waste-extent, causes and prevention. Italy: FAO Rome.

Fayed, H. A., & Atiya, A. F. (2019). Speed up grid-search for parameter selection of support vector machines. *Applied Soft Computing Journal*, *80*, 202–210. https://doi.org/10.1016/j.asoc.2019.03.037

Feng, C., Teuber, S., & Gershwin, M. E. (2016). Histamine (Scombroid) Fish Poisoning: a Comprehensive Review. *Clinical Reviews in Allergy and Immunology*, *50*(1), 64–69. https://doi.org/10.1007/s12016-015-8467-x

Fernández Pierna, J. A., Baeten, V., Renier, A. M., Cogdill, R. P., & Dardenne, P. (2004). Combination of support vector machines (SVM) and near-infrared (NIR) imaging spectroscopy for the detection of meat and bone meal (MBM) in compound feeds. *Journal of Chemometrics*, *18*(7–8), 341–349. https://doi.org/10.1002/cem.877

Fernández Pierna, J. A., Volery, P., Besson, R., Baeten, V., & Dardenne, P. (2005). Classification of modified starches by fourier transform infrared spectroscopy using support vector machines. *Journal of Agricultural and Food Chemistry*, *53*(17), 6581–6585. https://doi.org/10.1021/jf0501544

Food and Drug Administration (FDA) (2011). Fish and fisheries products hazards and controls guide. Chap 13 (4th ed.). (Accessed on 20.03.22).

Franceschelli, L., Marconi, G., & Mater, A. (2020). *Vis / NIR hyperspectral imaging to assess freshness of sardines (Sardina pilchardus).* 124–128.

826 Geladi, P., MacDougall, D., & Martens, H. (1985). Linearization and scatter-correction
827 for near-infrared reflectance spectra of meat. Applied spectroscopy, 39(3), 491-500.
828 https://doi.org/10.1366/0003702854248656

829 Ghidini, S., Chiesa, L. M., Panseri, S., Varrà, M. O., Ianieri, A., Pessina, D., & Zanardi,
830 E. (2021). Histamine control in raw and processed tuna: A rapid tool based on nir
831 spectroscopy. *Foods*, *10*(4), 1–16. https://doi.org/10.3390/foods10040885

832 Golland, P., Liang, F., Mukherjee, S., Panchenko, D. (2005). Permutation Tests for
833 Classification. In: Auer, P., Meir, R. (eds) Learning Theory. Vol 3559. Springer,
834 Berlin, Heidelberg. https://doi.org/10.1007/11503415_34

835 Gromski, P. S., Muhamadali, H., Ellis, D. I., Xu, Y., Correa, E., Turner, M. L., &
836 Goodacre, R. (2015). A tutorial review: Metabolomics and partial least squares-
837 discriminant analysis - a marriage of convenience or a shotgun wedding. *Analytica*
838 *Chimica Acta*, *879*, 10–23. https://doi.org/10.1016/j.aca.2015.02.012

839 Herpandi, Huda, N., Rosma, A., & Wan Nadiah, W. A. (2011). The Tuna Fishing
840 Industry: A New Outlook on Fish Protein Hydrolysates. *Comprehensive Reviews in*
841 *Food Science and Food Safety*, *10*(4), 195–207. https://doi.org/10.1111/j.1541-
842 4337.2011.00155.x

843 Huang, J., Gan, N., Lv, F., Cao, Y., Ou, C., & Tang, H. (2016). Environmentally friendly
844 solid-phase microextraction coupled with gas chromatography and mass
845 spectrometry for the determination of biogenic amines in fish samples. *Journal of*
846 *Separation Science*, *39*(22), 4384–4390. https://doi.org/10.1002/jssc.201600893

847 Hungerford, J. M. (2010). Scombroid poisoning: A review. *Toxicon*, *56*(2), 231–243.
848 https://doi.org/10.1016/j.toxicon.2010.02.006

849 Indahl, U. G., Martens, H., & Næs, T. (2007). From dummy regression to prior
850 probabilities in PLS-DA EXTRACTION AND CLASSIFICATION. *Journal of*
851 *Chemometrics*, *21*(August), 529–536. https://doi.org/10.1002/cem

852 Iqbal, A., Sun, D. W., & Allen, P. (2013). Prediction of moisture, color and pH in cooked,
853 pre-sliced turkey hams by NIR hyperspectral imaging system. *Journal of Food*
854 *Engineering*, *117*(1), 42–51. https://doi.org/10.1016/j.jfoodeng.2013.02.001

855 Jiang, L., Cai, Z., Wang, D., & Jiang, S. (2007). Survey of improving K-nearest-neighbor
856 for classification. *Proceedings - Fourth International Conference on Fuzzy Systems*
857 *and Knowledge Discovery, FSKD*, *1*, 679–683.
858 https://doi.org/10.1109/FSKD.2007.552

859 Kamankesh, M., Mohammadi, A., Mollahosseini, A., & Seidi, S. (2019). Application of
860 a novel electromembrane extraction and microextraction method followed by gas
861 chromatography-mass spectrometry to determine biogenic amines in canned fish.
862 *Analytical Methods*, *11*(14), 1898–1907. https://doi.org/10.1039/c9ay00224c

863 Karoui, R., Downey, G., & Blecker, C. (2010). Mid-infrared spectroscopy coupled with
864 chemometrics: A tool for the analysis of intact food systems and the exploration of
865 their molecular structure-quality relationships-A review. *Chemical Reviews*,
866 *110*(10), 6144–6168. https://doi.org/10.1021/cr100090k

867 Khalili Tilami, S., & Sampels, S. (2018). Nutritional Value of Fish: Lipids, Proteins,
868 Vitamins, and Minerals. *Reviews in Fisheries Science and Aquaculture*, *26*(2), 243–

253. https://doi.org/10.1080/23308249.2017.1399104

Kjeldahl, K., & Bro, R. (2010). Some common misunderstandings in chemometrics. *Journal of Chemometrics*, *24*(7–8), 558–564. https://doi.org/10.1002/cem.1346

Köse, S., Kaklikkaya, N., Koral, S., Tufan, B., Buruk, K. C., & Aydin, F. (2011). Commercial test kits and the determination of histamine in traditional (ethnic) fish products-evaluation against an EU accepted HPLC method. *Food Chemistry*, *125*(4), 1490–1497. https://doi.org/10.1016/j.foodchem.2010.10.069

Lapa-Guimarães, J., & Pickova, J. (2004). New solvent systems for thin-layer chromatographic determination of nine biogenic amines in fish and squid. *Journal of Chromatography A*, *1045*(1–2), 223–232. https://doi.org/10.1016/j.chroma.2004.06.014

Lee, L., Petter, S., Fayard, D., & Robinson, S. (2011). On the use of partial least squares path modeling in accounting research. *International Journal of Accounting Information Systems*, *12*(4), 305–328. https://doi.org/10.1016/j.accinf.2011.05.002

Lehane, L., & Olley, J. (2000). Histamine fish poisoning revisited. *International Journal of Food Microbiology*, 58, 1–37.

Leuschner, R., Rogers, D. S., & Charvet, F. F. (2013). A meta-analysis of supply chain integration and firm performance. *Journal of Supply Chain Management*, *49*(2), 34–57. https://doi.org/10.1111/jscm.12013

Li, L., Xie, S., Ning, J., Chen, Q., & Zhang, Z. (2019). Evaluating green tea quality based on multisensor data fusion combining hyperspectral imaging and olfactory visualization systems. *Journal of the Science of Food and Agriculture*, *99*(4), 1787–1794. https://doi.org/10.1002/jsfa.9371

Liu, H., Papa, E., & Gramatica, P. (2006). QSAR prediction of estrogen activity for a large set of diverse chemicals under the guidance of OECD principles. *Chemical Research in Toxicology*, *19*(11), 1540–1548. https://doi.org/10.1021/tx0601509

Liu, L., Zuo, Z. tian, Wang, Y. zhong, & Xu, F. rong. (2020). A fast multi-source information fusion strategy based on FTIR spectroscopy for geographical authentication of wild Gentiana rigescens. *Microchemical Journal*, *159*, 105360. https://doi.org/10.1016/j.microc.2020.105360

Liu, Y., Wang, C., Xia, Z., Wang, Q., & Duan, S. (2022). Monitoring freshness of crayfish (Prokaryophyllus clarkii) through the combination of near-infrared spectroscopy and chemometric method. *Journal of Food Measurement and Characterization*, *16*(5), 3438–3450. https://doi.org/10.1007/s11694-022-01451-w

Lutz, U., Lutz, R. W., & Lutz, W. K. (2006). Metabolic profiling of glucuronides in human urine by LC-MS/MS and partial least-squares discriminant analysis for classification and prediction of gender. *Analytical Chemistry*, *78*(13), 4564–4571. https://doi.org/10.1021/ac0522299

Magwaza, L. S., Opara, U. L., Nieuwoudt, H., Cronje, P. J. R., Saeys, W., & Nicolaï, B. (2012). NIR Spectroscopy Applications for Internal and External Quality Analysis of Citrus Fruit-A Review. *Food and Bioprocess Technology*, *5*(2), 425–444. https://doi.org/10.1007/s11947-011-0697-1

Mahanti, N. K., Chakraborty, S. K., Kotwaliwale, N., & Vishwakarma, A. K. (2020).

912 Chemometric strategies for nondestructive and rapid assessment of nitrate content
913 in harvested spinach using Vis-NIR spectroscopy. *Journal of Food Science*,
914 *85*(10), 3653–3662. https://doi.org/10.1111/1750-3841.15420

915 Mammone, A., Turchi, M., & Cristianini, N. (2009). Support vector machines. Wiley
916 Interdisciplinary Reviews: Computational Statistics, 1(3), 283–289.
917 https://doi.org/10.1002/wics.49

918 Massart, D. L., Vandeginste, B. G. M., Buydens, L. M. C., De Jong, S., Lewi, P. J., &
919 Smeyers-Verbeke, J. (1997). Handbook of Chemometrics and Qualimetrics: Part A:
920 Elservier Science B. Amsterdam, The Netherlands.

921 McLauchlin, J., Little, C. L., Grant, K. A., & Mithani, V. (2006). Scombrotoxic fish
922 poisoning. *Journal of Public Health*, *28*(1), 61–62.
923 https://doi.org/10.1093/pubmed/fdi063

924 Mishra, P., Biancolillo, A., Roger, J. M., Marini, F., & Rutledge, D. N. (2020). New data
925 preprocessing trends based on ensemble of multiple preprocessing techniques. *TrAC*
926 *- Trends in Analytical Chemistry*, *132*, 116045.
927 https://doi.org/10.1016/j.trac.2020.116045

928 Muscarella, M., Lo Magro, S., Campaniello, M., Armentano, A., & Stacchini, P. (2013).
929 Survey of histamine levels in fresh fish and fish products collected in Puglia (Italy)
930 by ELISA and HPLC with fluorimetric detection. *Food Control*, *31*(1), 211–217.
931 https://doi.org/10.1016/j.foodcont.2012.09.013

932 Nei, D., Nakamura, N., Ishihara, K., Kimura, M., & Satomi, M. (2017). A rapid screening
933 of histamine concentration in fish fillet by direct analysis in real-time mass
934 spectrometry (DART-MS). *Food Control*, *75*, 181–186.
935 https://doi.org/10.1016/j.foodcont.2016.12.001

936 Nguyen, D. T., Pissard, A., Pierna, J. A. F., Rogez, H., Souza, J., Dortu, F., Goel, S.,
937 Hernandez, Y., & Baeten, V. (2022). A method for non-destructive determination of
938 cocoa bean fermentation levels based on terahertz hyperspectral imaging.
939 *International Journal of Food Microbiology*, 365.
940 https://doi.org/10.1016/j.ijfoodmicro.2022.109537

941 Nieuwoudt, H. H., Prior, B. A., Pretorius, I. S., Manley, M., & Bauer, F. F. (2004).
942 Principal component analysis applied to Fourier transform infrared spectroscopy for
943 the design of calibration sets for glycerol prediction models in wine and for the
944 detection and classification of outlier samples. *Journal of Agricultural and Food*
945 *Chemistry*, *52*(12), 3726–3735. https://doi.org/10.1021/jf035431q

946 Ojala, M., & Garriga, G. C. (2010). Permutation tests for studying classifier performance.
947 *Journal of Machine Learning Research*, *11*, 1833–1863.

948 Omer, A. K., Mohammed, R. R., Ameen, P. S. M., Abas, Z. A., & Ekici, K. (2021).
949 Presence of biogenic amines in food and their public health implications: A
950 review. *Journal of food protection*, *84*(9), 1539-1548.

951 Ordóñez, J. L., Troncoso, A.M., García–Parrilla, M.D.C., & Callejón, R.M. (2016).
952 Recent trends in the determination of biogenic amines in fermented beverages. A
953 review. *Analytica Chimica Acta*, *939*, 10–25.

954 Ordóñez, J. L., & Callejón, R. (2019). Biogenic amines in non–fermented food. In

Biogenic Amines in Food: Analysis, Occurrence and Toxicity, 76–102, Royal Society of Chemistry, London.

Papageorgiou, M., Lambropoulou, D., Morrison, C., Kłodzińska, E., Namieśnik, J., & Płotka-Wasylka, J. (2018). Literature update of analytical methods for biogenic amines determination in food and beverages. *TrAC - Trends in Analytical Chemistry*, *98*, 128–142. https://doi.org/10.1016/j.trac.2017.11.001

Peng, J. feng, Fang, K. teng, Xie, D. hua, Ding, B., Yin, J. Y., Cui, X. mei, Zhang, Y., & Liu, J. fu. (2008). Development of an automated on-line pre-column derivatization procedure for sensitive determination of histamine in food with high-performance liquid chromatography-fluorescence detection. *Journal of Chromatography A*, *1209*(1–2), 70–75. https://doi.org/10.1016/j.chroma.2008.09.028

Pérez, N. F., Ferré, J., & Boqué, R. (2009). Calculation of the reliability of classification in discriminant partial least-squares binary classification. *Chemometrics and Intelligent Laboratory Systems*, *95*(2), 122–128. https://doi.org/10.1016/j.chemolab.2008.09.005

Peris-Díaz, M. D., & Krężel, A. (2021). A guide to good practice in chemometric methods for vibrational spectroscopy, electrochemistry, and hyphenated mass spectrometry. *TrAC - Trends in Analytical Chemistry*, *135*. https://doi.org/10.1016/j.trac.2020.116157

Pratt, J. W. (1959). Remarks on Zeros and Ties in the Wilcoxon Signed Rank Procedures. *Journal of the American Statistical Association*, *54*(287), 655–667. https://doi.org/10.1080/01621459.1959.10501526

Prester, L. (2011). Biogenic amines in fish, fish products and shellfish: a review. In *Food Additives and Contaminants - Part A Chemistry, Analysis, Control, Exposure and Risk Assessment* (Vol. 28, Issue 11, pp. 1547–1560). https://doi.org/10.1080/19440049.2011.600728

Prieto, N., López-Campos, Ó., Aalhus, J. L., Dugan, M. E. R., Juárez, M., & Uttaro, B. (2014). Use of near infrared spectroscopy for estimating meat chemical composition, quality traits and fatty acid content from cattle fed sunflower or flaxseed. *Meat Science*, *98*(2), 279–288. https://doi.org/10.1016/j.meatsci.2014.06.005

Qi, W., Tian, Y., Lu, D., & Chen, B. (2022). Research Progress of Applying Infrared Spectroscopy Technology for Detection of Toxic and Harmful Substances in Food. *Foods*, *11*(7). https://doi.org/10.3390/foods11070930

Quintelas, C., Mesquita, D. P., Ferreira, E. C., & Amaral, A. L. (2019). Quantification of pharmaceutical compounds in wastewater samples by near infrared spectroscopy (NIR). *Talanta*, *194*, 507–513. https://doi.org/10.1016/j.talanta.2018.10.076

Ramadan, Z., Jacobs, D., Grigorov, M., & Kochhar, S. (2006). Metabolic profiling using principal component analysis, discriminant partial least squares, and genetic algorithms. *Talanta*, *68*(5), 1683–1691. https://doi.org/10.1016/j.talanta.2005.08.042

Regulation (EU) No. 2073/ 2005 of the European Parliament and of the Council of 15 November 2005 on microbiological criteria for foodstuffs. In Official Journal of the European Union, L338, Eur-Lex: Brussels, Belgium, 2005; p 1.

Regulation (EU) No. 1441/2007 of 5 December 2007 amending Regulation (EC) No 2073/2005 on microbiological criteria for foodstuffs. In Official Journal of the European Union, *322*, 12-29.

Regulation (EU) No 1019/2013 of 23 October 2013 Amending Annex I to Regulation (EC) No 2073/2005 as Regards Histamine in Fishery Products. In Official Journal of the European Union, *282*, 46-47.

Rinnan, Å., Berg, F. van den, & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC - Trends in Analytical Chemistry*, *28*(10), 1201–1222. https://doi.org/10.1016/j.trac.2009.07.007

Rodriguez-Saona, L. E., Giusti, M. M., & Shotts, M. (2016). Advances in infrared spectroscopy for food authenticity testing. In *Advances in Food Authenticity Testing*. Elsevier Ltd. https://doi.org/10.1016/B978-0-08-100220-9.00004-7

Roy, K., Kar, S., & Das, R. N. (2015). Selected statistical methods in QSAR, 1st edition, chapter-6. In *K. Roy, S. Kar, & R. N. Das (Eds.), Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment* (pp. 191–228). Cambridge, Massachusetts: United States of America. https://doi.org/10.1016/C2014-0-00286- 9.

Salvador, A. M., García-Maldonado, E., Gallego-Narbón, A., Zapatera, B., & Vaquero, M. P. (2019). Fatty acid profile and cardiometabolic markers in relation with diet type and omega-3 supplementation in Spanish vegetarians. *Nutrients*, *11*(7). https://doi.org/10.3390/nu11071659

Sánchez-Parra, M., Lopez, A., Muñoz-Redondo, J. M., Montenegro-Gómez, J. C., Pérez-Aparicio, J., Pereira-Caro, G., Rodríguez-Solana, R., Moreno-Rojas, J. M., & Ordóñez-Díaz, J. L. (2022). Study of the influence of the fishing season and the storage temperature in the fishing vessel on the biogenic amine and volatile profiles in fresh yellowfin tuna (*Thunnus albacares*) and dry-cured mojama. *Journal of Food Composition and Analysis*, *114*, 104845. https://doi.org/10.1016/J.JFCA.2022.104845

Sánchez-Parra, M., Lopez, A., Ordóñez-Díaz, J. L., Rodríguez-Solana, R., Montenegro-Gómez, J. C., Pérez-Aparicio, J., & Moreno-Rojas, J. M. (2023). Evaluation of Biogenic Amine and Free Fatty Acid Profiles During the Manufacturing Process of Traditional Dry-Cured Tuna. *Food and Bioprocess Technology*. https://doi.org/10.1007/s11947-023-03134-w

Saptoro, A., Tadé, M. O., & Vuthaluru, H. (2012). A modified Kennard-Stone algorithm for optimal division of data for developing artificial neural network models. *Chemical Product and Process Modeling*, *7*(1). https://doi.org/10.1515/1934-2659.1645

Savitzky, A. & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, *36*, 1627–1639.

Shakila, R. J., Vijayalakshmi, K., & Jeyasekaran, G. (2003). Changes in histamine and volatile amines in six commercially important species of fish of the Thoothukkudi coast of Tamil Nadu, India stored at ambient temperature. *Food Chemistry*, *82*(3), 347–352. https://doi.org/10.1016/S0308-8146(02)00552-6

Shao, Y., Shi, Y., Qin, Y., Xuan, G., Li, J., Li, Q., Yang, F., & Hu, Z. (2022). A new quantitative index for the assessment of tomato quality using Vis-NIR hyperspectral imaging. *Food Chemistry*, *386*(November 2021), 132864. https://doi.org/10.1016/j.foodchem.2022.132864

Shenk, J. S. & Westerhaus, M. O. (1996). Calibration the ISI way. In: Davies, A.M.C., Williams, P.C. (Eds.), Near Infrared Spectroscopy: the Future Waves. NIR Publications, Chichester, UK, 198–202.

Stone, M. (1974) Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, *36*, 111-147.

Tao, Z., Sato, M., Han, Y., Tan, Z., Yamaguchi, T., & Nakano, T. (2011). A simple and rapid method for histamine analysis in fish and fishery products by TLC determination. *Food Control*, *22*(8), 1154–1157. https://doi.org/10.1016/j.foodcont.2010.12.014

Thomas, E. V. (2003). Non-parametric statistical methods for multivariate calibration model selection and comparison. *Journal of Chemometrics*, *17*(12), 653–659. https://doi.org/10.1002/cem.833

Tormena, C. D., Marcheafave, G. G., Pauli, E. D., Bruns, R. E., & Scarminio, I. S. (2019). Potential biomonitoring of atmospheric carbon dioxide in Coffea arabica leaves using near-infrared spectroscopy and partial least squares discriminant analysis. *Environmental Science and Pollution Research*, *26*(29), 30356–30364. https://doi.org/10.1007/s11356-019-06163-1

Valderrama, L., Valderrama, P., & Carasek, E. (2022). A semi-quantitative model through PLS-DA in the evaluation of carbendazim in grape juices. *Food Chemistry*, *368*, 130742. https://doi.org/10.1016/j.foodchem.2021.130742

van der Voet, H. (1994). Comparing the predictive accuracy of models using a simple randomization test. *Chemometrics and Intelligent Laboratory Systems*, *25*(2), 313–323. https://doi.org/10.1016/0169-7439(94)85050-X

Venyaminov, S. Y., & Kalnin, N. N. (1990). Quantitative IR spectrophotometry of peptide compounds in water (H2O) solutions. I. Spectral parameters of amino acid residue absorption bands. *Biopolymers*, *30*(13–14), 1243–1257. https://doi.org/10.1002/bip.360301309

Vermeulen, M., Smith, K., Eremin, K., Rayner, G., & Walton, M. (2021). Application of Uniform Manifold Approximation and Projection (UMAP) in spectral imaging of artworks. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, *252*, 119547. https://doi.org/10.1016/j.saa.2021.119547

Visciano, P., Schirone, M., & Paparella, A. (2020). An overview of histamine and other biogenic amines in fish and fish products. *Foods*, *9*(12). https://doi.org/10.3390/foods9121795

Wang, X., Shan, J., Han, S., Zhao, J., & Zhang, Y. (2019). Optimization of Fish Quality by Evaluation of Total Volatile Basic Nitrogen (TVB-N) and Texture Profile Analysis (TPA) by Near-Infrared (NIR) Hyperspectral Imaging. *Analytical Letters*, *52*(12), 1845–1859. https://doi.org/10.1080/00032719.2019.1571077

Westerhuis, J. A., Hoefsloot, H. C. J., Smit, S., Vis, D. J., Smilde, A. K., Velzen, E. J. J., Duijnhoven, J. P. M., & Dorsten, F. A. (2008). Assessment of PLS-DA cross validation. *Metabolomics*, *4*(1), 81–89. https://doi.org/10.1007/s11306-007-0099-6

Wilkerson, E. D., Anthon, G. E., Barrett, D. M., Sayajon, G. F. G., Santos, A. M., & Rodriguez-Saona, L. E. (2013). Rapid assessment of quality parameters in processing tomatoes using hand-held and benchtop infrared spectrometers and multivariate analysis. *Journal of Agricultural and Food Chemistry*, *61*(9), 2088–2095. https://doi.org/10.1021/jf304968f

Xia, J., Min, S., & Li, J. (2023). Rapid analysis the type of customs paper using Micro-NIR spectrometers and machine learning algorithms. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, *290*, 122272. https://doi.org/10.1016/j.saa.2022.122272

Xiaobo, Z., Jiewen, Z., Povey, M. J. W., Holmes, M., & Hanpin, M. (2010). Variables selection methods in near-infrared spectroscopy. *Analytica Chimica Acta*, *667*(1–2), 14–32. https://doi.org/10.1016/j.aca.2010.03.048

Xie, J., Bian, Z., Lin, T., Tao, L., Wu, Q., & Chu, M. (2020). Global occurrence, bioaccumulation factors and toxic effects of polychlorinated biphenyls in tuna: A review. *Emerging Contaminants*, *6*, 388–395. https://doi.org/10.1016/j.emcon.2020.11.003

Yan, W., Yao, J., Yue, Z., Lin, H., Wang, L., Wang, K., & Li, J. (2023). Non-destructive monitoring the freshness of sea bass fillets using Raman spectroscopy with orthogonal signal correction and multivariate analysis. *Microchemical Journal*, *191*(March), 108859. https://doi.org/10.1016/j.microc.2023.108859

Yang, C., Guang, P., Li, L., Song, H., Huang, F., Li, Y., Wang, L., & Hu, J. (2021). Early rapid diagnosis of Alzheimer's disease based on fusion of near- and mid-infrared spectral features combined with PLS-DA. Optik, 241(January), 1–9. https://doi.org/10.1016/j.ijleo.2021.166485

Zhao, J., Lin, H., Chen, Q., Huang, X., Sun, Z., & Zhou, F. (2010). Identification of egg's freshness using NIR and support vector data description. *Journal of Food Engineering*, *98*(4), 408–414. https://doi.org/10.1016/j.jfoodeng.2010.01.018

Figure Captions

Figure 1. FT-MIR spectra of tuna fish samples without any pre-treatment

Figure 2. Boxplot analysis of the histamine levels (mg/kg) in tuna fillets during the incubation period (10 days) at 22 ± 2 ºC. Data are expressed as mean ± standard deviation.

Figure 3. Principal Component Analyses of the FT-MIR spectra of tuna samples: (a) Score plot; (b) Loading plot

Figure 4. Partial Least Square Regression (PLSR) cross-validation models for histamine levels, including (a) Model 1, (b) Model 2 and (c) Model 3.

Table 1. Histamine content of the different tuna samples analyzed by HPLC-DAD in the ranges and models established.

ID	Histamine level (mg/kg)	Nº of samples	Min	Max
1	< 50	16	n.d.	40.89
2	50 – 100	5	50.51	68.92
3	100 – 200	13	100.68	194.75
4	200 – 300	9	207.34	275.02
5	300 – 500	11	307.14	475.09
6	500 – 700	8	506.03	687.03
7	700 – 1000	4	761.60	879.73
FDA model	< 50	16	n.d.	40.89
	> 50	50	50.51	879.73
EU model	< 100	21	n.d.	95.85
	> 100	45	100.68	879.73

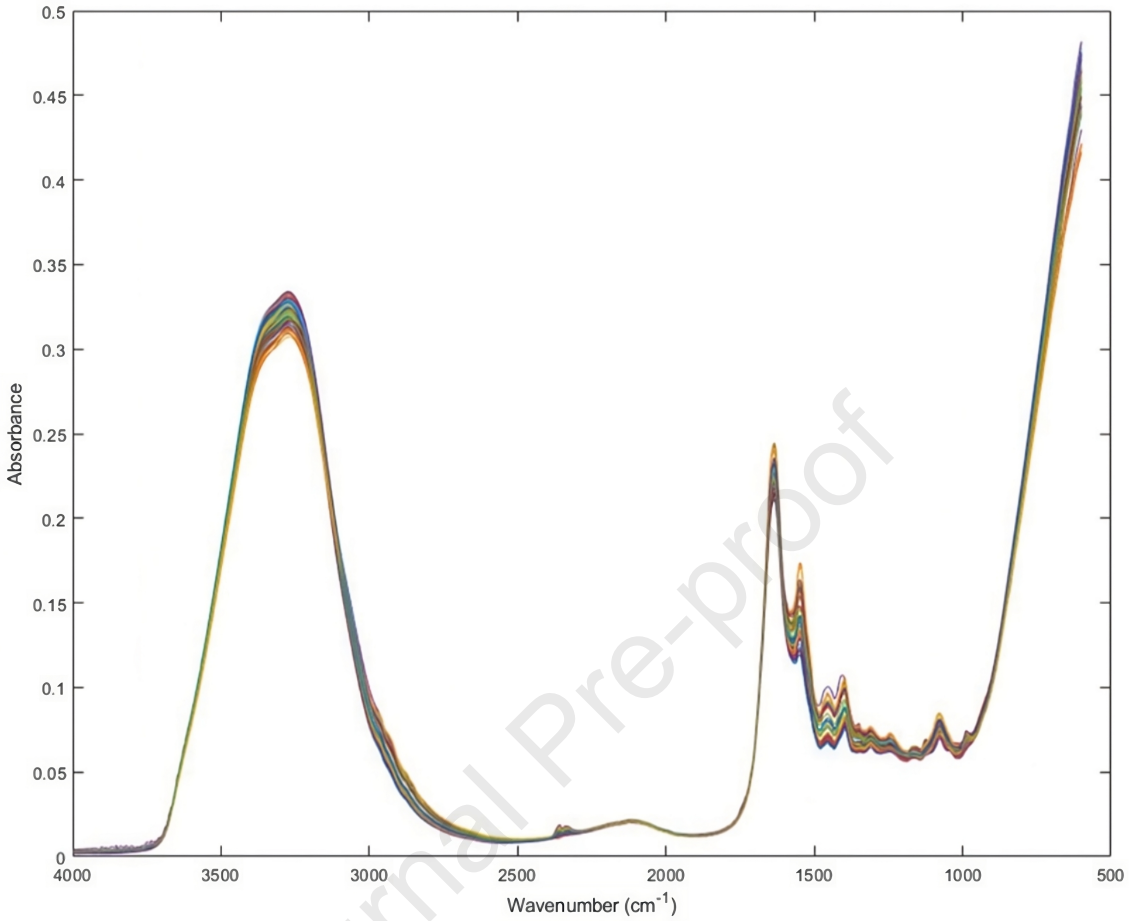
Min=minimum; Max= maximum; n.d. =not detected.

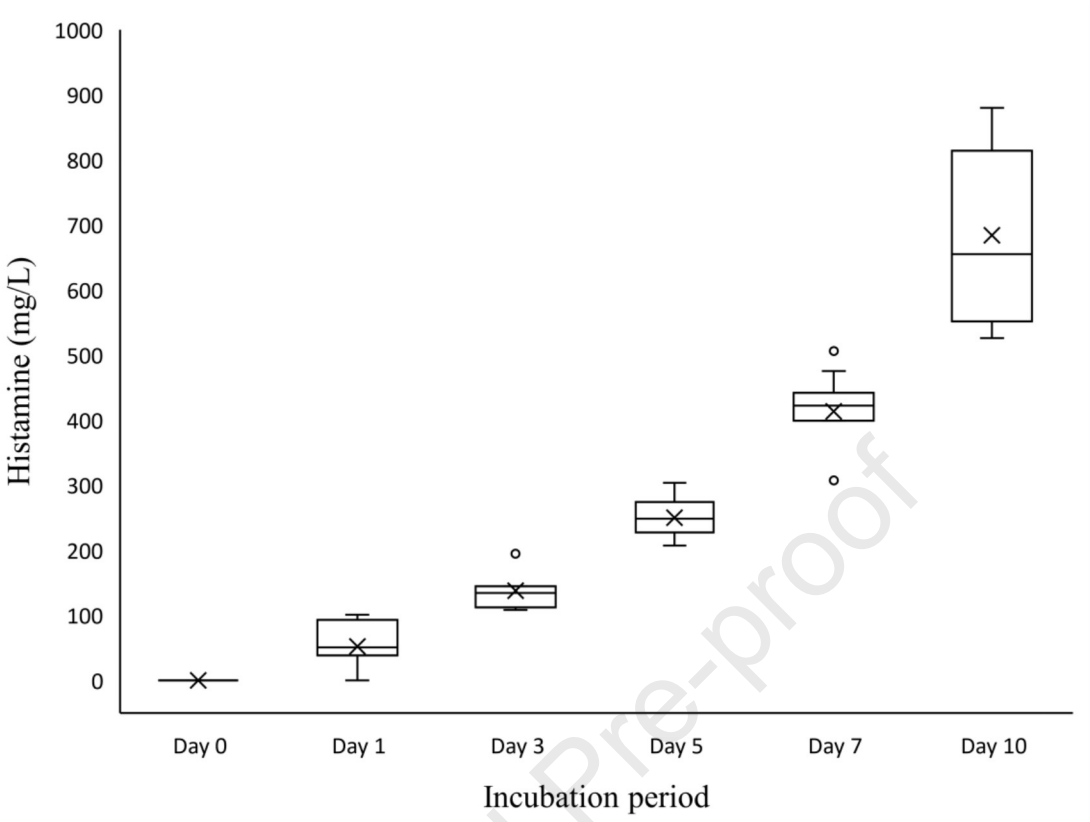
Table 2. Sensitivity and specificity (%) for the classification models in calibration, cross-validation and prediction of histamine.

EU Regulation		**PLS-DA**		**SVM**		**KNN**	
		< 100 mg/kg	**> 100 mg/kg**	**< 100 mg/kg**	**> 100 mg/kg**	**< 100 mg/kg**	**> 100 mg/kg**
Calibration	Sensitivity (%)	100 ± 0.0	100 ± 0.0	100 ± 0.0	100 ± 0.0	100 ± 0.0	100 ± 0.0
	Specificity (%)	100 ± 0.0	100 ± 0.0	100 ± 0.0	100 ± 0.0	100 ± 0.0	100 ± 0.0
Cross-Validation	Sensitivity (%)	97.1 ± 1.8	98.8 ± 1.2	97.1 ± 1.8	95.3 ± 1.7	97.1 ± 1.8	98.8 ± 1.2
	Specificity (%)	98.8 ± 1.2	97.1 ± 1.8	95.3 ± 1.7	97.1 ± 1.8	98.8 ± 1.2	97.1 ± 1.8
Prediction	Sensitivity (%)	94.3 ± 0.6	100 ± 0.0	92.9 ± 6.4	98.2 ± 1.6	100 ± 0.0	100 ± 0.0
	Specificity (%)	100 ± 0.0	94.3 ± 0.6	98.2 ± 1.6	92.9 ± 6.4	100 ± 0.0	100 ± 0.0

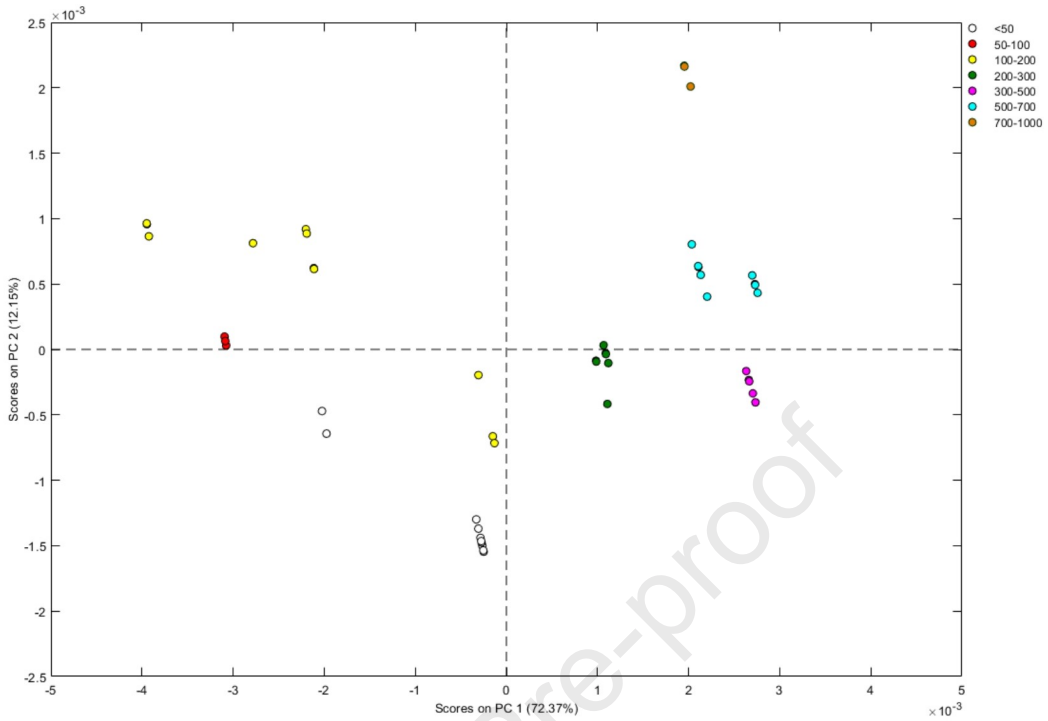
Table 3. Sensitivity and specificity (%) for the classification models in calibration, cross-validation and prediction of histamine.

FDA regulation		**PLS-DA**		**SVM**		**KNN**	
		< 50 mg/kg	**> 50 mg/kg**	**< 50 mg/kg**	**> 50 mg/kg**	**< 50 mg/kg**	**> 50 mg/kg**
Calibration	Sensitivity (%)	100 ± 0.0	93.0 ± 1.1	84.1 ± 4.5	99.4 ± 0.6	100 ± 0.0	100 ± 0.0
	Specificity (%)	93.0 ± 1.1	100 ± 0.0	99.4 ± 0.6	84.1 ± 4.5	100 ± 0.0	100 ± 0.0
Cross-Validation	Sensitivity (%)	94.0 ± 2.4	93.7 ± 1.7	90.0 ± 4.5	98.3 ± 0.7	92.0 ± 3.7	95.3 ± 0.8
	Specificity (%)	93.7 ± 1.7	94.0 ± 2.4	98.3 ± 0.7	90.0 ± 4.5	96.5 ± 0.8	90.8 ± 3.7
Prediction	Sensitivity (%)	89.9 ± 6.6	95.9 ± 2.7	76.6 ± 4.1	98.7 ± 1.3	89.9 ± 4.1	97.3 ± 2.7
	Specificity (%)	95.9 ± 2.7	89.9 ± 6.6	98.7 ± 1.3	76.6 ± 4.1	97.3 ± 2.7	89.9 ± 4.1

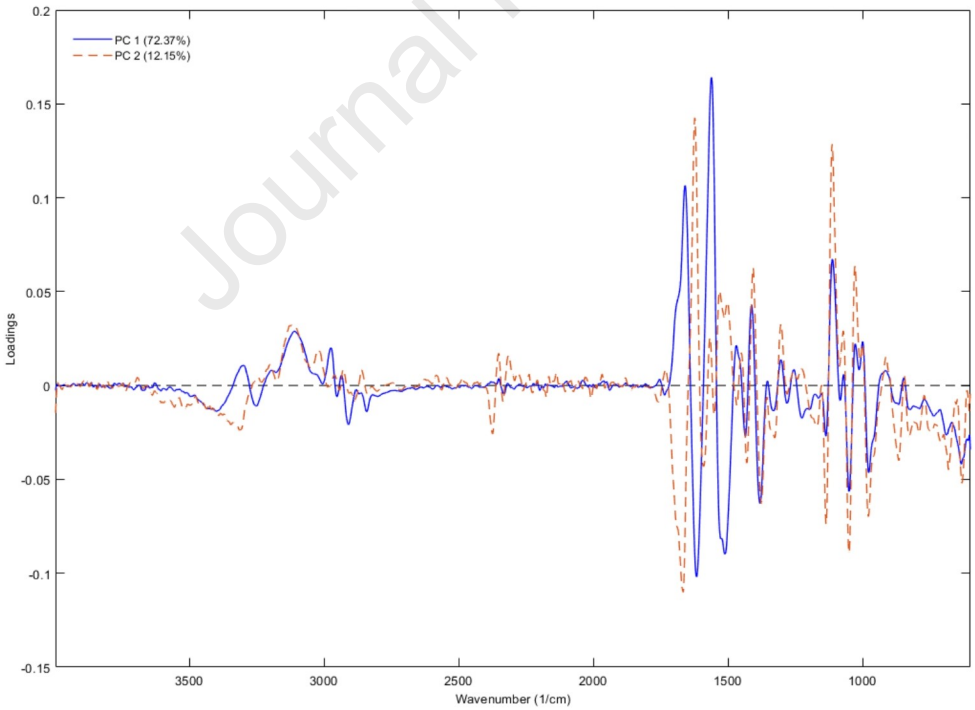




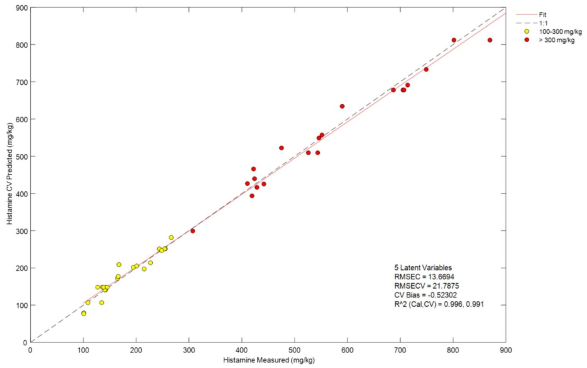
a



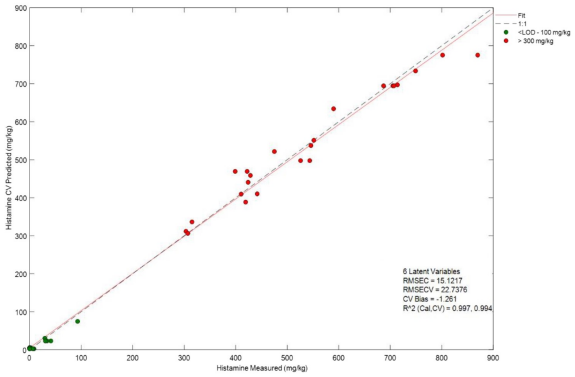
b



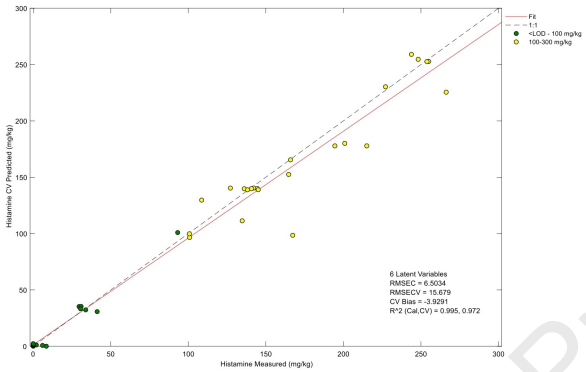
a



b



c



HIGHLIGHTS

- Models created analyzing FT-MIR spectra with machine learning algorithms.
- The best spectral pre-processing technique was the combination of SNV + Savitzky-Golay derivative.
- FT-MIR was used to discriminate tuna samples according to their histamine concentration.
- Contribution to improving quality control and safety inspections in the industry.

Declaration of interests

☒ 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.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: