Leveraging the performance of conventional spectroscopic techniques through data fusion approaches in high-quality edible oil adulteration analyses

Diego G. Much , Mirta R. Alcaraz , José M. Camiña , Héctor C. Goicoechea , Silvana M. Azcarate

PII: S2666-8319(24)00027-4
DOI: https://doi.org/10.1016/j.talo.2024.100313
Reference: TALO 100313

To appear in: *Talanta Open*

Received date: 2 February 2024
Revised date: 15 March 2024
Accepted date: 16 March 2024

Please cite this article as: Diego G. Much , Mirta R. Alcaraz , José M. Camiña , Héctor C. Goicoechea , Silvana M. Azcarate , Leveraging the performance of conventional spectroscopic techniques through data fusion approaches in high-quality edible oil adulteration analyses, *Talanta Open* (2024), doi: https://doi.org/10.1016/j.talo.2024.100313

TALO-D-24-00013

HIGHLIGHTS

- UV-vis NIR and fluorescence data to quantify adulterants in high-quality edible oils
- PLS regression models built with individual sources and fused data
- A comprehensive comparison between approaches to assess predictive performances
- The higher the chemical information, the better the predictive efficiency

Leveraging the performance of conventional spectroscopic techniques through data fusion approaches in high-quality edible oil adulteration analyses

Diego G. Much[a, d], Mirta R. Alcaraz[b, d,*], José M. Camiña[c,d], Héctor C. Goicoechea[b, d], Silvana M. Azcarate[c, d,**]

[a] Agencia de Investigación Científica (AIC) sede Santa Rosa, Corona Martínez 430, Santa Rosa, La Pampa, 6300, Argentina.

[b] Laboratorio de Desarrollo Analítico y Quimiometría (LADAQ) Cátedra de Química Analítica I, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional Del Litoral, Ciudad Universitaria, Santa Fe, 3000, Argentina

[c] Facultad de Ciencias Exactas y Naturales, Universidad Nacional de la Pampa, Instituto de Ciencias de la Tierra y Ambientales de la Pampa (INCITAP), Av. Uruguay 151, Santa Rosa, La Pampa, 6300, Argentina

[d] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Godoy Cruz 2290, CABA, 1425, Argentina

Corresponding authors:

*Mirta R. Alcaraz, e-mail: malcaraz@fbcb.unl.edu.ar

**Silvana M. Azcarate, e-mail: sazcarate@exactas.unlpam.edu.ar;

Abstract

The high demand, high cost, and low regulations surrounding high-quality edible oils (HQEO) make them a target for fraudulent actions, particularly adulteration with refined oils. Consequently, the authentication of this kind of oil is of great interest. This work assessed the adulteration degree of five HQEOs: sesame, flaxseed, chia, rapeseed, and extra virgin olive oils, using different chemometric strategies to enhance the detection capability of the analytical methodology. Refined oils used as adulterants were evaluated at low concentrations (2-15 % v/v). Three multidimensional spectroscopic techniques (UV-Visible, near-infrared, and excitation-emission matrix fluorescence) were used, and two data fusion strategies (low- and mid-level) were evaluated. Principal component analysis was applied as an exploratory analysis tool to visualise and interpret the information contained in the dataset. For the adulterant quantification, partial least squares regression analysis was used to build the sensitive predictive models. The results revealed that chemical information enhancement leverages the ability to attain reduced prediction compared to unidimensional signals. In scenarios with low sample variability, conventional unidimensional spectroscopy (UV-Visible or near-infrared) data was shown to be adequate to guarantee predictive efficiency. In contrast, when analysing predictive figures derived from models built using a dataset with high variability, e.g., brands, low-level data fusion approaches enhance predictive efficiency. The results showed that excitation-emission matrix-based or low-level data fusion approaches can be accurately implemented to guarantee the authenticity of edible oils even when a low content of adulterant oil is presented.

Keywords: food quality; high-quality edible oils; adulteration fraud; spectroscopic measurements; data fusion strategies; chemometric modelling

1. Introduction

High-quality edible oils (HQEO) play a crucial role in human nutrition and are widely used in cooking, baking, and food preparation. Unfortunately, due to their high cost, some manufacturers and traders commit fraud by adulterating expensive edible oils with cheaper alternatives or low-quality ingredients. The adulteration of HQEO is a global concern that seriously threatens public health and consumer rights [1]. Detecting food adulteration in these products is a complex and challenging task due to the diversity of potential adulterants and the sophistication of the techniques used to detect them.

In recent years, significant advancements in analytical techniques have led to the development of more accurate and reliable methods for detecting food adulteration in edible oils [2,3]. In this regard, spectroscopy has emerged as a powerful tool due to its speed, non-destructive nature, and low-cost analysis [4,5]. In addition, chemometrics proved to be of great importance due to its effectiveness in various food products and adulteration scenarios [6–9]. In this regard, it has been demonstrated that the combination of conventional spectroscopic techniques such as UV-Visible (UV-Vis), near-infrared (NIR), mid-infrared (MID), and fluorescence spectroscopy with chemometrics is an efficient strategy for the analysis of many complex matrices and valuable to detect and quantify adulterants in food products [10]. For instance, chemometrics approaches were utilised for the analysis of UV-Vis spectroscopy data to distinguish among different adulteration degrees in edible oils [11–14] and in combination with NIR spectroscopy to detect and quantify low-quality oils used as adulterants in high nutritional value oils, e.g., flaxseed (FSO), sesame palm, or rapeseed oils (RSO) [15–17]. On the other hand, excitation-emission fluorescence matrix (EEM)

spectroscopy associated with chemometric is particularly useful in detecting the presence of adulterants in high-quality oils, for example, the addition of pomace olive oil to extra virgin olive oil (EVOO) [18], or corn oil and soybean oil to sesame oil (SO) [16].

More recently, data fusion strategies have emerged as an appealing approach that exploits the synergic effect of individual techniques, improving the accuracy and reliability of the analysis [19–21]. In this context, three distinct data fusion approaches emerge: low-, mid- and high-level data fusions (LLDF, MLDF, and HLDF, respectively). These alternatives have garnered significant attention, particularly in predictive and classification fields, and several applications have been reported in the literature [22–24].

LLDF comprises modelling a data set built by concatenating raw data from different sources along the common mode. MLDF involves a two-step model in which data from various sources are individually modelled; then, the outputs are joined and used to construct the final model. In HLDF, the final decision is reached considering the outcomes obtained from the models built with data proceeding from individual sources [25].

This study investigates the potential of three conventional spectroscopic techniques (UV-Vis, NIR, and EEM) for detecting and quantifying adulterants in EVOO, FSO, SO, chia (CO), and RSO oils. Data fusion strategies were also applied to assess individual source synergistic/complementary information. Predictive analysis was assessed at a very low degree of adulteration (2-15 %), and a comprehensive comparative assessment of the predictive performance across various approaches was accomplished.

2. Materials and methods

2.1. Oil samples

High-quality and refined Argentinian edible oils were acquired from local markets. Five brands of EVOO, 4 brands of RSO, 2 brands of FSO, 1 brand of CO, and 1 brand of SO were analysed. Sunflower, corn, and soybean oil were used as adulterants.

2.2. Sample preparation

Artificial adulterated samples were built using an experimental design composed of two D-optimal designs for quaternary mixtures. For the first design, adulterations between 0 % and 15 % v/v were evaluated at five levels. The second design assessed adulterant oil at five concentration levels from 0 % to 7 % v/v. Table S1, Supplementary Material, shows detailed information about sample composition.

Five mL of adulterated samples were prepared by directly adding the aliquot of the refined oils to the target oil. After preparation, the samples were mechanically vortexed for 1 min and stored for 24 h in the dark at 4 °C before analysis to guarantee the homogenisation of the samples.

2.3. Instrumentation

UV-Vis spectroscopy measurements were performed on a 10 mm-path length quartz cell using a UV-Vis Ocean Optics CHEMUSB4 spectrophotometer with a linear diode array (LDA) detector. The UV-vis spectra were registered in the 200-800 nm range every 0.21 nm.

NIR spectra were collected between 900 nm and 1650 nm in transmittance mode (6 nm resolution) using a FLAME-NIR Ocean Optics spectrophotometer (Duiven, The Netherlands) equipped with a 10 mm-path length quartz cell. Dark body and empty

quartz cuvette, corresponding to 0 % and 100 % of transmittance, respectively, were used for the instrument calibration.

OceanView software (1.6.7 version; Ocean Optics, The Netherlands) was used to control the UV-vis and NIR spectrophotometers.

EEM matrices were recorded using an Agilent Cary Eclipse luminescence spectrophotometer (Agilent Technologies, Waldbronn, Alemania) equipped with a xenon flash lamp, using a 10×10 mm-path length quartz cell and Cary WinFLR software for instrument control and data acquisition. The EEMs were obtained by varying the excitation wavelength every 3 nm and recording the emission spectra every 2 nm according to the following parameters: 1) EEM_{EVOO} 280-580 nm/420-700 nm at 430 V PMT; 2) EEM_{RSO} 280-580 nm/380-700 nm at 500 V PMT; 3) EEM_{FSO}, EEM_{CO} and EEM_{SO} 280-500 nm/380-700 nm at 450 V PMT. The excitation and emission slits were set at 10 nm, and the scan rate was fixed to 9600 nm s^{-1} in all cases.

2.4. Data sets

First, non-informative spectral regions were avoided. Then, for UV-vis and NIR data sets (UVDS and NDS, respectively), 2-way arrays were built by staking the spectra of all analysed samples. On the other hand, the EEM data set (FDS) was constructed with the unfolded EEMs, which were then column-wise appended to build a 2-way array. In all cases, bidimensional samples×variables matrices were obtained.

UVDS, NDS, and FDS were row-wise concatenated to build the low-level data fusion dataset (LLDF). To construct the mid-level data fusion data set (MLDF), principal component analysis (PCA) was individually performed on the UVDS, NDS, and FDS, and the optimal principal component (PC) scores were combined in a single block. The size of each dataset is detailed in Table 1.

Table 1. Dataset dimensions

Dataset	Oil (sample×variable)				
	EVOO	*FSO*	*SO*	*CO*	*RSO*
UVDS	52×1719	31×289	16×289	12×1207	40×547
NDS	52×128	31×128	16×30	12×40	40×128
FDS	52×14241	31×12075	16×12075	12×12075	40×16261
LLDF	52×16088	31×12492	16×12394	12×13322	40×16936
MLDF	52×28	31×17	16×20	12×22	40×26

2.5. Software and data analysis

All calculations were done in MATLAB (The Mathworks, Natick, MA, USA). PCA was performed using PCA_toolbox, available at https://es.mathworks.com/matlabcentral/fileexchange/134751-pca-toolbox-for-matlab?s_tid=prof_contriblnk [26]. Quantification of adulterant oils was calculated by partial least square (PLS) regression analysis using the Regression_toolbox [27]. Both toolboxes are freely available at https://michem.unimib.it/download/matlab-toolboxes/. EEM_corr GUI, freely downloaded from https://fbcb.web1.unl.edu.ar/laboratorios/ladaq/download/, was utilised to subtract the Rayleigh and Raman scattering from the EEM [28]. All data sets were partitioned into calibration and validation sets using the Kennard-Stone (KS) algorithm [29].

3. Results and discussion

3.1. General considerations of spectroscopic features

Figures 1A and B show the UV-vis and NIR spectra of the raw and adulterated oils. Figure 1C depicts the EEM obtained for the raw samples. For more information about spectral profiles, the reader is referred to Figure S1, Supplementary Material.

In the UV-Vis spectra, the bands between 300-500 nm are mainly attributed to lignans, tocopherols, phytosterols, and carotenoids, which are natural antioxidants responsible for conferring the exceptional properties of high-quality edible oils. Similar behaviour is observed in the EEM, for which the same constituents show significant fluorescence intensities in the spectral region of 350-400 nm excitation (λ_{exc}) and 400-550 nm emission (λ_{em}) [17,30–32]. In particular, EVOO and blend oils exhibit a distinctive spectral band at 680 nm in the UV-Vis spectra given by the presence of chlorophylls and pheophytins [33], which is also observed in the EEM at 520-550 nm λ_{exc}/660-680 nm λ_{em}.

The significant UV-Vis absorbance band spanning 300-320 nm in SO is attributed to its high lignan content sourced from sesame seeds, recognised as rich sources of these compounds [34]. Similar spectral features are observed in RSO, showing a prominent band at lower wavelengths and minor bands within the 400-500 nm range.

Notably, differences in the UV-Vis absorbance profiles of EVOO between brands are observed (Figure S1, Supplementary Material). These differences can arise from variations in the locations of olive production. Different soil properties and varying climate conditions in these locations can lead to variability of aroma compound composition [35]. At the same time, the differences between FSO brands could be attributed to the production process declared by the manufacturer, wherein partial degreasing of the seeds is carried out. This process leads to a significant reduction in antioxidant compounds typically known to exhibit absorption within the range of 400-500 nm [33,36–38].

As can be appreciated, the higher the adulteration degree, the lower the UV-vis absorbance intensities of the significant bands, particularly those corresponding to carotenoids, tocopherols, and chlorophylls (400-700 nm). Moreover, it is worth noting that, in all cases, the fluorescence intensity at 320-400 nm λ_{exc}/400-550 nm λ_{em} increases as the adulterant concentration rises. This fact is attributable to fatty acid oxidation products from refined oils (Kongbonga et al., 2011). Also, a detriment in the fluorescence intensity at 500-550 nm λ_{exc}/660-690 nm λ_{em} (for EVOO) and at 350-430 nm λ_{exc}/ 660-690 nm λ_{em} for SO, CO, and RSO is noticed in adulterated samples as a consequence of the dilution effect.

NIR spectroscopy is a powerful technique for identifying distinct molecular groups based on their vibrational harmonics. Oils, rich in triacylglycerols and fatty acids, typically exhibit observable vibrational patterns, such as the stretching vibrational second harmonics of –CH bonds and the first stretching vibrational harmonics of $–CH_3$, $–CH_2$, and HC=CH groups found in unsaturated fatty acid chains [30,39,40]. Because of this, the NIR spectra of the different samples show a high degree of similarity, which becomes challenging for classification analysis.

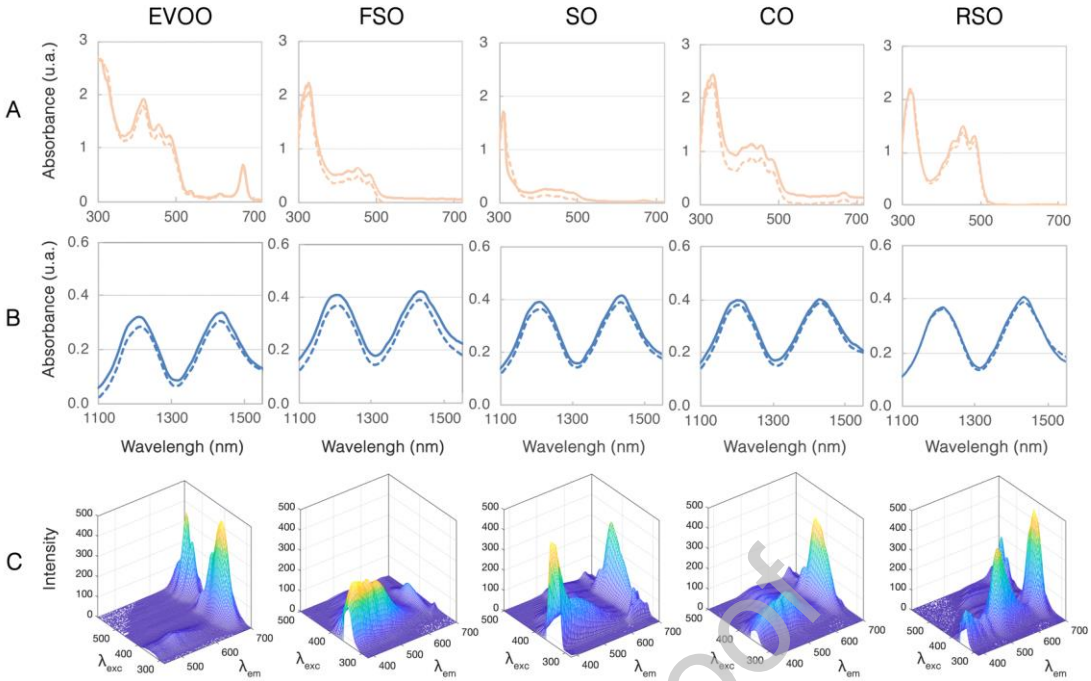


Figure 1. UV-Vis (**A**) and NIR (**B**) absorption spectra of unadulterated (solid line) and 15 % v/v adulterated (dashed line) samples. Excitation-emission matrices (EEMs) (**C**) for unadulterated oil samples.

3.2. Exploratory analysis through PCA

In a first attempt to investigate the adulteration effect on the target oils, an exploratory data analysis was performed to detect outliers and identify hidden patterns in sample distribution. In this line, PCA models were built, and a comparative, albeit exploratory, analysis among techniques was accomplished.

First, different pre-processing approaches were evaluated to implement the protocol that renders the best outcomes in terms of sample distribution by cluster formation. Hence, mean centring (MC), standard normal variate (SNV), and autoscaling, among others, were proven. Results demonstrated that implementing the MC approach to individual datasets led to a more reliable model regarding explained variance. On the other hand, the combination of SNV and MC approaches rendered the best performance figures for the LLDF-based strategy, while autoscale was also implemented for MLDF-based models.

Then, five PCA models were performed for each oil: three corresponding to the individual technique datasets (UVDS, NDS, FDS) and two for the fused data (LLDF, MLDF). Here, it is important to remark that MLDF was built with the PC scores obtained from the optimised PCA model of UVDS, NDS, and FDS. Table 2 summarises the optimal parameters obtained from the different PCA models of each oil and LLDF strategy, considering the number of PCs that minimise the root-mean-standard error for cross-validation (RMSECV).

Table 2. Optimal parameters obtained from PCA models

Oil	Strategy	Optimal PC number	Cumulative explained variance (%)
EVOO	UVDS	8	99.8
	NDS	11	99.8

	FDS	9	99.9
	LLDF	6	99.6
FSO	UVDS	6	99.9
	NDS	6	98.3
	FDS	5	99.9
	LLDF	4	99.9
SO	UVDS	6	99.8
	NDS	5	99.4
	FDS	9	99.5
	LLDF	4	98.4
CO	UVDS	6	99.7
	NDS	8	99.7
	FDS	8	99.0
	LLDF	4	95.8
RSO	UVDS	7	99.8
	NDS	13	99.4
	FDS	6	99.9
	LLDF	5	99.8

Notably, LLDF-based models attain the highest cumulative explained variance values with a low number of PCs, surpassing those selected for models obtained through individual technique datasets.

Figure 2 shows the 3D score plots for the first three PCs of the individual and fused data set-based PCA models and the ellipsoids of the confidence region at 95 %. Furthermore, the scores obtained from the raw oils were labelled to gain further insight into the effect of adulteration on the raw oils.

The first outcome accomplished from this analysis revealed that both individual and fused data sets provide enough chemical information to adequately discriminate samples among brands. In this regard, the models constructed with FDS achieved the best performance for EVOO, FSO, and RSO samples compared to those built with UVDS and NDS. At this point, it is worth highlighting that SO and CO samples were not subjected to brand analysis since only one brand was available for each oil.

Nevertheless, PCA models were built for further analysis. The PCA models using UVDS accounted for the highest sample variance for SO and CO. Similarly, this aligns with the previously mentioned outcomes of the optimal parameters obtained from PCA models (table 2), wherein the PCA models with the highest cumulative explained variance values are those obtained through FDS for EVOO, FSO and RSO samples and UVDS for SO and CO samples.

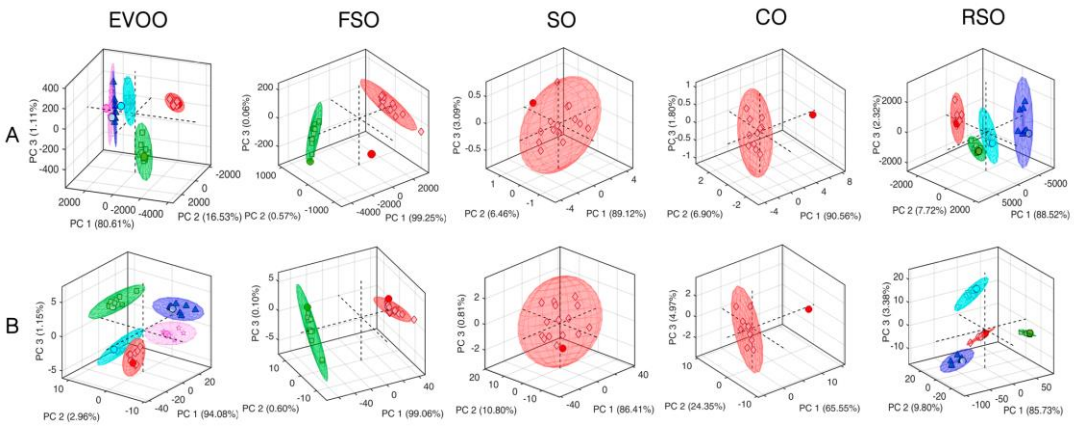


Figure 2. Scores plot of the first three PCs obtained by PCA according to the edible oil discriminated by brand. **A** shows PCAs obtained through individual data (FDS for EVOO, FSO, and RSO, and UVDS for SO and CO), while **B** presents PCAs based on fused data (LLDF for each oil). The ellipsoids show the confidence region at the 95 % level for each group according to the different oil brands. Dashed lines indicate the central axis of three-dimensional space. Non-adulterated oil samples (circles) exhibit different colors corresponding to their brand identities.

As can be appreciated, the sample variability of different brands is higher than that of adulterated samples of the same brand, and as many clusters as brands are discerned in each case. It becomes crucial to emphasise that, in all cases, the raw samples can be clearly distinguished from the adulterated samples since a clear separation from them is observable. Notably, the discrimination performances behave similarly among strategies, albeit smaller ellipsoids of confidence are obtained for fused data.

These outcomes evidenced the possibility of a binary distinction between raw and adulterated samples, even at low levels of adulteration. This observation holds significance considering the notable complexity of the studied system, primarily stemming from the variability of brands, the differences in the refined oils used as adulterants, and the low levels of adulteration. Nevertheless, to obtain more consistent results from a pattern recognition analysis, it is necessary to increase the number of samples to acquire more consistent groups of adulterated and unadulterated samples and to reach valid conclusions.

3.3. Multivariate predictive model development

Aiming to determine the degree of adulteration of the edible oils, PLS regression models were built. First, to establish the optimal approach that minimises the prediction and validation-associated errors, different data pre-processing strategies were evaluated, and models were compared regarding their predictive efficacy and the robustness of the statistical indicator.

For the model development, the samples were randomly selected by the Kennard-Stone algorithm [41] and split into calibration and validation sets. Each calibration set comprised 80 % of the samples used for the calibration and internal validation of the

models. The validation set comprised the remaining 20 % of the total samples and was used to evaluate the model's capability to detect the degree of adulteration.

The proper number of latent variables (LVs) for each model was assessed by the minimum error rate in cross-validation (Venetian blind, six splits). Once the number of total LVs was chosen, the prediction step was accomplished on the validation samples. To comprehensively interpret the model performance, cross-validation (RMSECV), and prediction (RMSEP), the determination coefficients (R^2) and the relative error of prediction were estimated (REP%). Table 2 summarises the parameters and the results obtained for the models built for each oil type.

The outcomes outlined in Table 3 demonstrate that, for EVOO, FSO, and RSO, an augmentation in the dimensionality of the data leads to a notable enhancement in the predictive efficacy of the models. The chemical information enhancement accomplished by acquiring multidimensional signals, i.e., EEM, enables the establishment of more sensitive models, leveraging the ability to attain reduced prediction compared to unidimensional signals (REP% higher than 25.8 % and 41.4 % for UV-Vis and NIR, respectively).

For SO and CO, satisfactory predictive figures were achieved for unidimensional UVDS-based models (REP% <15 % and % $\bar{R}$ c.a. 100 %). This outcome might stem from using a single brand for each oil type. This situation leads to a scenario with low sample variability; thus, conventional unidimensional spectroscopy (UV-Vis or NIR) data might be adequate to guarantee predictive efficiency.

Table 3. Statistical parameters of PLS regression models for prediction of adulteration using the different approaches

Oil	Strategy	Pre-processing*	LVs	R^2 cal	RMSECV (% v/v)	RMSEP (% v/v)	REP (%)	% $\bar{R}^b$
EVOO	UVDS	SNV	9	0.99	3.65	2.55	33.4	115.4
	NDS	MSC	10	0.83	6.30	3.16	41.4	88.5
	FDS	SC+MC	7	0.97	2.91	1.41	18.5	96.5
	LLDF	PoSS	10	0.99	2.41	0.97	12.7	98.5
	MLDF	PoSS+MC	8	0.72	3.70	1.52	19.9	102.0
FSO	UVDS	SNV+MC	6	0.96	5.78	2.17	25.8	109.6
	NDS	De+MSC+PoSS	7	0.86	4.37	2.96	35.2	125.3
	FDS	SC+MC	8	0.99	5.03	0.85	10.1	108.3
	LLDF	SNV	7	0.99	5.09	0.68	8.09	102.7
	MLDF	FD+MSC+PoSS	12	0.90	4.45	0.92	10.9	104.9
SO	UVDS	SNV	6	0.95	7.44	1.08	13.4	94.4
	NDS	FD+MSC+PoSS	7	0.80	8.52	3.58	44.5	76.8
	FDS	SC	5	0.99	2.63	1.12	13.9	107.0
	LLDF	PoSS	5	0.99	2.29	1.02	12.6	105.5
	MLDF	MC+PaSS	7	0.98	2.37	1.12	13.9	92.0
CO	UVDS	-	5	0.96	4.11	0.71	10.3	100.8
	NDS	FD +SNV	5	0.91	5.85	1.47	21.3	101.6
	FDS	FD+MC	3	0.99	3.36	1.22	17.7	100.0
	LLDF	FD+SNV	3	0.99	3.23	1.31	19.0	103.5
	MLDF	PaSS	4	0.99	2.98	1.80	26.2	92.1
RSO	UVDS	MC+PoSS	9	0.98	4.92	2.70	34.3	137.4
	NDS	FD+MSC+VSS	9	0.72	6.48	3.51	44.7	157.5
	FDS	SC	8	0.95	3.19	0.99	12.6	107.0
	LLDF	SNV	10	0.99	3.18	0.84	10.7	100.5
	MLDF	MC+PaSS	9	0.79	3.04	1.25	15.9	103.2

* De: Detrend, SNV: standard normal variate, MSC: multiplicative scattering correction, MC: mean centring, VSS: variance standard scaling, FD: first derivative, PaSS: Pareto standard scaling, PoSS: poison standard scaling, SC: Scattering correction.

[b]% $\bar{R}$= mean recovery in %

On the other hand, the unsatisfactory figures of merit obtained for all the NDS-based models imply their limited capability in accurately evaluating oil adulteration,

especially at low levels of adulteration. These observations align with results obtained in several works in which NIR data did not provide good statistical figures in the predictive evaluation of adulteration in edible oils. However, fused NIR data with MIR [40] or Raman [20]enhanced the performance of the models.

Additionally, when analysing predictive figures derived from models built using a dataset encompassing various brands such as EVOO, FSO, and RSO, it becomes evident that employing LLDF approaches enhances predictive efficiency. These methods yield REP values consistently below 12.7 %, showcasing notable improvements compared to using individual technique datasets.

Moreover, something that deserves attention is that, in most cases, MLDF strategies render poor prediction figures, not showing a distinctive enhancement of the performance in comparison to the individual technique datasets. This phenomenon could be attributed to the reduction in data dimensionality before treatment. In this case, although the previous step to the DF was based on extracting significant system characteristics through PCA, there is a potential loss of pertinent information for subsequent PLS analysis.

To gain further insights into the predictive capability of the models, the plot of nominal vs. predicted values and the elliptical joint of confidence region (EJCR) for the slope and the intercept of the predicted vs. nominal values were analysed. Figure 3 illustrates the results obtained for individual and fused data-based models and the EJCR plots at a 95 % confidence level.

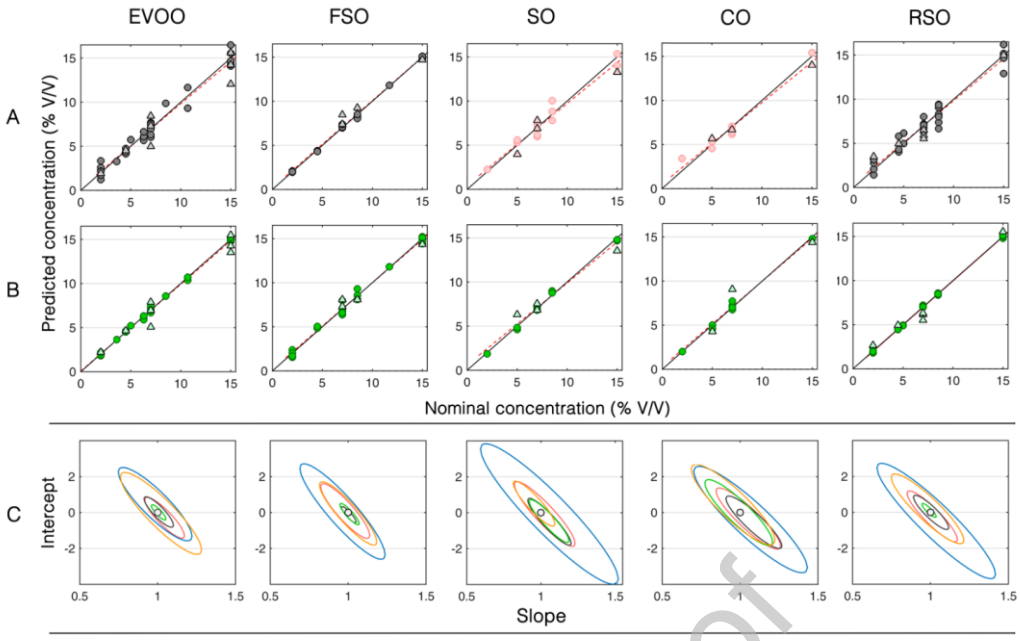


Figure 3. Nominal vs. predicted value plots derived from the information obtained using the best individual technique (**A**) or the best data fusion strategy applied (**B**) for EVOO FSO, SO, CO, and RSO. Elliptical joint confidence region (**C**) in the plane slope-intercept for a 95 % confidence level, indicating the ideal point [1, 0] (grey circle) for slope and intercept, respectively, for the data UVDS (pink), NDS (blue), FDS (grey), LLDF(green) and MLDF (orange). Full circles and triangles are calibration and validation samples, respectively.

As can be appreciated, in all cases, the predicted samples approximate the ideal regression line (Figure 3A and 3B). This observation is supported by the EJCR plots (Figure 3C), where the ideal points for slope and intercept (1 and 0, respectively) are included within the confidence region of the ellipses at a 95 % level. These outcomes shed light on the satisfactory accuracy achieved.

One achievement to highlight is that the PLS models allowed satisfactory prediction of those samples with the lowest level of adulteration (2 % v/v). This result is of paramount relevance since, to the best of our knowledge, this is the first time adulteration in seed oils using a mixture of refined oils as adulterants is assessed. In addition, it is noteworthy that most published works cover broad ranges of adulterant levels (typically between 10 % and 90 %) [5,13,40,42–45]. These ranges often surpass permissible limits in certain countries, raising pertinent concerns. For instance, according to the regulations outlined in the Argentine Food Code (CAA), adulterating EVOO is prohibited in Argentina [46]. Maintaining the integrity of EVOO samples is essential to ensure the authenticity and quality of the product. In this sense, investigations focusing on detecting low levels of adulterants in oil samples are particularly valuable due to the strict prohibition against adulteration, which contributes significantly to upholding the standards set and helping safeguard the genuineness of EVOO in the market.

Furthermore, while the CAA generally addresses FSO, SO, CO, and RSO, no specific regulation explicitly permits blending these oils with others of lower quality. Therefore, without established guidelines, such blending is not permitted. The lack of explicit permission prohibits mixing these high-quality oils with inferior ones, ensuring their quality standards are maintained and upheld.

4. Conclusions

This study shows a comprehensive comparative evaluation of data analysis approaches to analysing high-quality edible oil.

First, spectroscopic data obtained from conventional techniques (UV-Vis, NIR, and EEM) were subjected to PCA, and results regarding classificatory performance were compared. The results shed light on the fact that bidimensional signals such as EEM provide more chemical information than unidimensional signals (UV-Vis and NIR), enhancing the discrimination of samples according to the blend variability. Nevertheless, non-significant differences were encountered among techniques where blend variability does not occur.

The results obtained from modelling data proceeding from individual techniques were compared to those obtained from low-level and mid-level fused data. The study proved that low-level fused data render better results than mid-level fused data. These outcomes provide insights that the model performance is enhanced while increasing the chemical information by fusing data proceeding from different sources. These achievements align with the above observations that demonstrate that increasing the chemical information by acquiring bidimensional signals led to an improvement in the accomplishments of the method.

This study provides insights into the effectiveness of spectroscopy techniques for analysing edible oils and contributes to the development of rapid and non-destructive methods for the detection of adulteration of these kinds of samples.

Author contributions

Diego G. Much: Methodology, Investigation, Formal analysis, Writing – Original draft, Writing – review & editing. **Mirta R. Alcaraz**: Conceptualization, Supervision, Investigation, Visualization, Writing – Original draft, Writing – review & editing, Funding acquisition. **José M. Camiña**: Visualisation, Writing – review & editing. **Héctor C. Goicoechea**: Visualization, Writing – review & editing. **Silvana M. Azcarate**: Conceptualization, Supervision, Investigation, Visualization, Writing – Original draft, Writing – review & editing, Funding acquisition.

Acknowledgements

DGM acknowledge CONICET for his doctoral fellowship. MRA, HCG, JMC and SMA are permanent staff of CONICET. The authors are grateful to AGENCIA I+D+i (Agencia Nacional de Promoción Científica y Tecnológica Argentina, Project PICT-2020-00462, Project PICT 2020–0531) and CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, Project PIP 2021-2023 - 11220200100316CO; Project PIBAA 2022 - 2872021010-0925CO) for financial support.

References

[1] C.H. Tan, I. Kong, U. Irfan, M.I. Solihin, L.P. Pui, Edible Oils Adulteration: A Review on Regulatory Compliance and Its Detection Technologies, J Oleo Sci 70 (2021) ess21109. https://doi.org/10.5650/jos.ess21109.

[2] L. Han, M. Chen, Y. Li, S. Wu, L. Zhang, K. Tu, L. Pan, J. Wu, L. Song, Discrimination of different oil types and adulterated safflower seed oil based on electronic nose combined with gas chromatography-ion mobility spectrometry, Journal of Food Composition and Analysis 114 (2022) 104804. https://doi.org/10.1016/j.jfca.2022.104804.

[3] J.-J. Zhang, Y. Gao, M.-L. Zhao, X. Xu, B.-N. Xi, L.-K. Lin, J.-Y. Zheng, B. Chen, Y. Shu, C. Li, Y. Shen, Detection of walnut oil adulterated with high-linoleic acid vegetable oils using triacylglycerol pseudotargeted method based on SFC-QTOF-MS, Food Chem 416 (2023) 135837. https://doi.org/10.1016/j.foodchem.2023.135837.

[4] X. Li, D. Wang, F. Ma, L. Yu, J. Mao, W. Zhang, J. Jiang, L. Zhang, P. Li, Rapid detection of sesame oil multiple adulteration using a portable Raman spectrometer, Food Chem 405 (2023) 134884. https://doi.org/10.1016/j.foodchem.2022.134884.

[5] O. Uncu, B. Ozen, A comparative study of mid-infrared, UV–Visible and fluorescence spectroscopy in combination with chemometrics for the detection of adulteration of fresh olive oils with old olive oils, Food Control 105 (2019) 209–218. https://doi.org/10.1016/j.foodcont.2019.06.013.

[6] Demiati, W.T. Wahyuni, M. Rafi, B.R. Putra, The detection of goat milk adulteration with cow milk using a combination of voltammetric fingerprints and chemometrics analysis, Chemical Papers 77 (2023) 4307–4317. https://doi.org/10.1007/s11696-023-02780-w.

[7] M.A. Faqeerzada, S. Lohumi, R. Joshi, M.S. Kim, I. Baek, B.-K. Cho, Non-targeted detection of adulterants in almond powder using spectroscopic techniques combined with chemometrics, Foods 9 (2020) 876. https://doi.org/10.3390/foods9070876.

[8] P. Rana, S.-Y. Liaw, M.-S. Lee, S.-C. Sheu, Discrimination of four Cinnamomum species with physico-functional properties and chemometric techniques: application of PCA and MDA models, Foods 10 (2021) 2871. https://doi.org/10.3390/foods10112871.

[9] M. Roy, M. Doddappa, B.K. Yadav, R. Jaganmohan, V.R. Sinija, L. Manickam, S. Sarvanan, Detection of soybean oil adulteration in cow ghee (clarified milk fat): an ultrafast study using flash gas chromatography electronic nose coupled with multivariate chemometrics, J Sci Food Agric 102 (2022) 4097–4108. https://doi.org/10.1002/jsfa.11759.

[10] D. Caballero, R. Rios-Reina, J.M. Amigo, Chemometrics and food traceability, in: Comprehensive Foodomics, Elsevier, 2020: pp. 387–406. https://doi.org/10.1016/B978-0-08-100596-5.22859-X.

[11] A. Mannu, M. Poddighe, S. Garroni, L. Malfatti, Application of IR and UV–VIS spectroscopies and multivariate analysis for the classification of waste vegetable oils, Resour Conserv Recycl 178 (2022). https://doi.org/10.1016/j.resconrec.2021.106088.

[12] R. Ríos-Reina, S.M. Azcarate, How chemometrics revives the UV-Vis spectroscopy applications as an analytical sensor for spectralprint (nontargeted) analysis, Chemosensors 11 (2022) 8. https://doi.org/10.3390/chemosensors11010008.

[13] J. Tan, R. Li, Z.T. Jiang, M. Shi, Y.Q. Xiao, B. Jia, T.X. Lu, H. Wang, Detection of extra virgin olive oil adulteration with edible oils using front-face fluorescence and visible spectroscopies, JAOCS, Journal of the American Oil Chemists' Society 95 (2018) 535–546. https://doi.org/10.1002/aocs.12071.

[14] L. Valderrama, R.P. Gonçalves, P.H. Março, P. Valderrama, UV-Vis spectrum fingerprinting and chemometric method in the evaluation of extra virgin olive oil adulteration and fraud., Revista Brasileira de Pesquisa Em Alimentos 5 (2014). https://doi.org/10.14685/rebrapa.v5i2.170.

[15] O. Mba, P. Adewale, M.-J. Dumont, M. Ngadi, Application of near-infrared spectroscopy to characterize binary blends of palm and canola oils, Ind Crops Prod 61 (2014) 472–478. https://doi.org/10.1016/j.indcrop.2014.07.037.

[16] Y.Y. Yuan, S.T. Wang, J.Z. Wang, Q. Cheng, X.J. Wu, DM Kong, Rapid detection of the authenticity and adulteration of sesame oil using excitation-emission matrix fluorescence and chemometric methods, Food Control 112 (2020). https://doi.org/10.1016/j.foodcont.2020.107145.

[17] Z. Yuan, L. Zhang, D. Wang, J. Jiang, P. de B. Harrington, J. Mao, Q. Zhang, P. Li, Detection of flaxseed oil multiple adulteration by near-infrared spectroscopy and nonlinear one class partial least squares discriminant analysis, LWT 125 (2020) 109247. https://doi.org/10.1016/j.lwt.2020.109247.

[18] I. Durán Merás, J. Domínguez Manzano, D. Airado Rodríguez, A. Muñoz de la Peña, Detection and quantification of extra virgin olive oil adulteration by means of autofluorescence excitation-emission profiles combined with multi-way classification, Talanta 178 (2018) 751–762. https://doi.org/10.1016/j.talanta.2017.09.095.

[19] S. Wu, B. Tu, Y.-R. Yu, J. Wang, X. Zheng, D.-P. He, Quantitative analysis of peanut oil adulteration based on data fusion of multi-source spectra, DEStech Transactions on Environment, Energy and Earth Science (2017). https://doi.org/10.12783/dteees/sses/icfse2016/10699.

[20] Q. Xu, X. Yang, Z. Hu, Y. Li, X. Zheng, Identification of adulterated olive oil by fusion of near infrared and Raman spectroscopy, J Phys Conf Ser 1592 (2020) 012041. https://doi.org/10.1088/1742-6596/1592/1/012041.

[21] X. Zhou, X. Li, B. Zhao, X. Chen, Q. Zhang, Discriminant analysis of vegetable oils by thermogravimetric-gas chromatography/mass spectrometry combined with data fusion and chemometrics without sample pretreatment, LWT 161 (2022). https://doi.org/10.1016/j.lwt.2022.113403.

[22] H. Zhu, D. Zhu, J. Sun, Application of GC-IMS coupled with chemometric analysis for the classification and authentication of geographical indication agricultural products and food, Front Nutr 10 (2023). https://doi.org/10.3389/fnut.2023.1247695.

[23] M.K. Moro, E.V.R. de Castro, W. Romão, P.R. Filgueiras, Data fusion applied in near and mid infrared spectroscopy for crude oil classification, Fuel 340 (2023) 127580. https://doi.org/10.1016/j.fuel.2023.127580.

[24] Y. Hong, N. Birse, B. Quinn, Y. Li, W. Jia, P. McCarron, D. Wu, G.R. da Silva, L. Vanhaecke, S. van Ruth, C.T. Elliott, Data fusion and multivariate analysis for food authenticity analysis, Nat Commun 14 (2023) 3309. https://doi.org/10.1038/s41467-023-38382-z.

[25] M. Cocchi, Introduction: Ways and Means to Deal With Data From Multiple Sources, in: Data Handling in Science and Technology, Elsevier, 2019: pp. 1–26. https://doi.org/10.1016/B978-0-444-63984-4.00001-6.

[26] D. Ballabio, A MATLAB toolbox for principal component analysis and unsupervised exploration of data structure, Chemometrics and Intelligent Laboratory Systems 149 (2015) 1–9. https://doi.org/10.1016/j.chemolab.2015.10.003.

[27] V. Consonni, G. Baccolo, F. Gosetti, R. Todeschini, D. Ballabio, A MATLAB toolbox for multivariate regression coupled with variable selection, Chemometrics and Intelligent Laboratory Systems 213 (2021) 104313. https://doi.org/10.1016/j.chemolab.2021.104313.

[28] F.A. Chiappini, M.R. Alcaraz, H.C. Goicoechea, A.C. Olivieri, A graphical user interface as a new tool for scattering correction in fluorescence data, Chemometrics and Intelligent Laboratory Systems 193 (2019) 103810. https://doi.org/10.1016/j.chemolab.2019.07.009.

[29] R. Galvao, M. Araujo, G. Jose, M. Pontes, E. Silva, T. Saldhana, A method for calibration and validation subset partitioning, Talanta 67 (2005) 736–740. https://doi.org/10.1016/j.talanta.2005.03.025.

[30] M. Mburu, C. Komu, O. Paquet-Durand, B. Hitzmann, V. Zettel, Chia Oil Adulteration Detection Based on Spectroscopic Measurements, Foods 10 (2021) 1798. https://doi.org/10.3390/foods10081798.

[31] D. Rabiej, A. Szydłowska-Czerniak, Fluorescence and UV-VIS spectroscopy to determine the quality changes of rapeseed oil fortified with new antioxidants

after storage under various conditions, Food Anal Methods 13 (2020) 1973–1982. https://doi.org/10.1007/s12161-020-01804-5.

[32] J.-Y. Song, H.-W. Gu, Y. Wang, T. Geng, H.-N. Cui, Y. Pan, B. Ding, Z. Li, X.-L. Yin, Excitation-emission matrix fluorescence spectroscopy combined with multi-way chemometric methods for rapid qualitative and quantitative analyses of the authenticity of sesame oil, European Food Research and Technology (2023). https://doi.org/10.1007/s00217-023-04275-0.

[33] R. Tarakowski, A. Malanowski, R. Kościesza, R.M. Siegoczyński, VIS spectroscopy and pressure induced phase transitions – Chasing the olive oils quality, J Food Eng 122 (2014) 28–32. https://doi.org/10.1016/j.jfoodeng.2013.08.030.

[34] L.-K. Shi, R.-J. Liu, Q.-Z. Jin, X.-G. Wang, The contents of lignans in sesame seeds and commercial sesame oils of china, J Am Oil Chem Soc 94 (2017) 1035–1044. https://doi.org/10.1007/s11746-017-3018-7.

[35] Q. Zhou, S. Liu, Y. Liu, H. Song, Comparative analysis of volatiles of 15 brands of extra-virgin olive oils using solid-phase micro-extraction and solvent-assisted flavor evaporation, Molecules 24 (2019) 1512. https://doi.org/10.3390/molecules24081512.

[36] A. Abad, F. Shahidi, Compositional characteristics and oxidative stability of chia seed oil (Salvia hispanica L), Food Production, Processing and Nutrition 2 (2020) 9. https://doi.org/10.1186/s43014-020-00024-y.

[37] M.A. Khodasevich, D.A. Borisevich, Identification of flax oil by linear multivariate spectral analysis, J Appl Spectrosc 86 (2020) 996–999. https://doi.org/10.1007/s10812-020-00929-z.

[38] R. Upadhyay, R. Thirumdas, R.R. Deshmukh, U. Annapure, N.N. Misra, An exploration of the effects of low-pressure plasma discharge on the physicochemical properties of chia (Salvia hispanica L.) flour, Journal of Engineering & Processing Management 11 (2020). https://doi.org/10.7251/JEPM1902073U.

[39] H. Chen, Z. Lin, C. Tan, Fast quantitative detection of sesame oil adulteration by near-infrared spectroscopy and chemometric models, Vib Spectrosc 99 (2018) 178–183. https://doi.org/10.1016/j.vibspec.2018.10.003.

[40] Y. Li, Y. Xiong, S. Min, Data fusion strategy in quantitative analysis of spectroscopy relevant to olive oil adulteration, Vib Spectrosc 101 (2019) 20–27. https://doi.org/10.1016/j.vibspec.2018.12.009.

[41] R.W. Kennard, L.A. Stone, Computer Aided Design of Experiments, Technometrics 11 (1969) 137–148. https://doi.org/10.1080/00401706.1969.10490666.

[42] B. Elzey, D. Pollard, S.O. Fakayode, Determination of adulterated neem and flaxseed oil compositions by FTIR spectroscopy and multivariate regression

analysis, Food Control 68 (2016) 303–309. https://doi.org/10.1016/j.foodcont.2016.04.008.

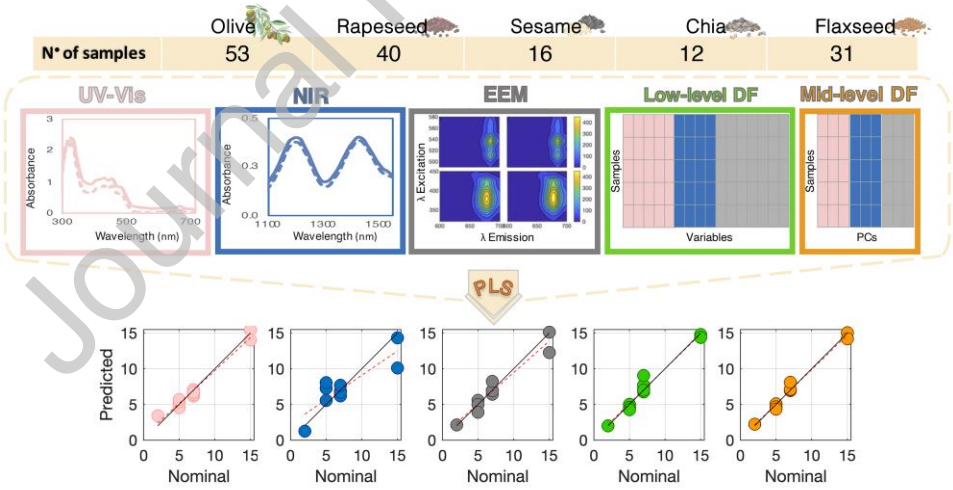
[43] H. Azizian, M.M. Mossoba, A.R. Fardin-Kia, P. Delmonte, S.R. Karunathilaka, J.K.G. Kramer, Novel, rapid identification, and quantification of adulterants in extra virgin olive oil using near-infrared spectroscopy and chemometrics, Lipids 50 (2015) 705–718. https://doi.org/10.1007/s11745-015-4038-4.

[44] E. Borràs, J. Ferré, R. Boqué, M. Mestres, L. Aceña, A. Calvo, O. Busto, Prediction of olive oil sensory descriptors using instrumental data fusion and partial least squares (PLS) regression, Talanta 155 (2016) 116–123. https://doi.org/10.1016/j.talanta.2016.04.040.

[45] A.M. Jiménez-Carvelo, V.A. Lozano, A.C. Olivieri, Comparative chemometric analysis of fluorescence and near infrared spectroscopies for authenticity confirmation and geographical origin of Argentinean extra virgin olive oils, Food Control 96 (2019) 22–28. https://doi.org/10.1016/j.foodcont.2018.08.024.

[46] CODEX, (2008). https://www.argentina.gob.ar/sites/default/files/anmat-capitulo_vii_grasosactualiz_2018-12.pdf (accessed January 23, 2024).

Graphical abstract



Declaration of competing 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 author is an Editorial Board Member/Editor-in-Chief/Associate Editor/Guest Editor for *[Journal name]* and was not involved in the editorial review or the decision to publish this article.

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