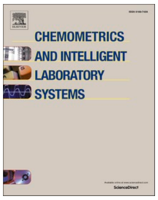
Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemometrics



Addressing adulteration challenges of dried oregano leaves by NIR HyperSpectral Imaging

Veronica Ferrari [a], Rosalba Calvini [a,b,*], Camilla Menozzi [a], Alessandro Ulrici [a,b], Marco Bragolusi [c], Roberto Piro [c], Alessandra Tata [c], Michele Suman [d,e], Giorgia Foca [a,b]

[a] Dipartimento di Scienze della Vita, Università di Modena e Reggio Emilia, Padiglione Besta, Via Amendola, 2, 42122, Reggio Emilia, Italy
[b] Centro Interdipartimentale BIOGEST-SITEIA, Università degli Studi di Modena e Reggio Emilia, Piazzale Europa, 1, 42122, Reggio Emilia, Italy
[c] Istituto Zooprofilattico Sperimentale Delle Venezie, Laboratorio di Chimica Sperimentale, Viale Fiume 78, 36100, Vicenza, Italy
[d] Analytical Food Science, Barilla G. e R. Fratelli S.p.A., Via Mantova, 166, 43122, Parma, Italy
[e] Department for Sustainable Food Process, Catholic University Sacred Heart, Piacenza, Italy

ARTICLE INFO

Keywords:
Oregano adulteration
Hyperspectral imaging
NIR
Classification
SIMCA
Soft PLS-DA

ABSTRACT

Dried oregano leaves are particularly prone to adulteration because of their widespread distribution and their easy mixing with leaves of other plants of lower commercial value, such as olive, myrtle, strawberry tree, or sumac. To reveal the presence of adulteration, in this study we considered an untargeted analytical approach, which instead of involving the *a priori* selection of specific compounds of interest is focused on defining the characteristic spectral signature of authentic oregano with respect to its most frequent adulterants. NIR HyperSpectral Imaging (NIR-HSI) represents a state-of-the-art, rapid and non-destructive technique, allowing for the collection of both spectral and spatial information from the sample, making it particularly suitable for characterizing visually heterogeneous samples.

Authentication issues are typically assessed through class modelling techniques and Soft Independent Modelling of class Analogy (SIMCA) is one of the most used algorithms in this scenario. However, the high variability and heterogeneity within the authentic oregano class resulted in poor outcomes when SIMCA was applied. As an alternative, Soft Partial Least Squares Discriminant Analysis (Soft PLS-DA) algorithm was applied to differentiate authentic oregano samples from pure adulterants. Soft PLS-DA represents a hybrid approach that combines the advantages of both discriminant and class modelling techniques. The resultant classification model has indeed led to promising results, achieving a prediction efficiency of 92.9 %. Finally, based on the percentage of pixels predicted as oregano in the Soft-PLSDA prediction images, a threshold value of 10 % was established, serving as a detection limit of NIR-HSI to distinguish authentic oregano samples from adulterated ones.

1. Introduction

The global market of herbs and spices has experienced unprecedented growth in recent years, driven by an increasing demand for culinary diversity, natural flavour enhancers, and the perceived health benefits of herbal products [1]. While this growth presents excellent opportunities for the industry, it also raises concerns about authenticity and integrity of these food products. An alarming consequence of this flourishing market is the increasing risk of adulteration, which is defined by European Spice Association as "the deliberate and intentional inclusion in herbs and spices of substances whose presence is not legally declared, is not permitted or is present in form which might mislead or confuse the customer, leading to an imitated food and/or product of reduced value" [2].

In this context, we are referring to Economically Motivated Adulteration (EMA), which involves the deliberate act of altering products, particularly in the food industry, with the aim of gaining a financial advantage. This practice entails substituting expensive ingredients with lower-quality alternatives to reduce costs [3,4]. In the case of herbs like oregano, EMA often includes adding lower-cost plant materials, which may encompass different botanical species. This unethical activity is facilitated by complexity of the supply chains, spanned in multiple stages occurring in countries that are different from that of the final sale [5]. Beyond the economic harm to consumers and the damaged

[*] Corresponding author. Dipartimento di Scienze della Vita, Università di Modena e Reggio Emilia, Padiglione Besta, Via Amendola, 2 – 42122 Reggio Emilia, Italy.
E-mail address: rosalba.calvini@unimore.it (R. Calvini).

https://doi.org/10.1016/j.chemolab.2024.105133
Received 26 February 2024; Received in revised form 3 April 2024; Accepted 22 April 2024
Available online 23 April 2024

reputation of honest producers and distributors, adulteration is a practice that can pose also significant health risks. Indeed, adulterated products may contain potential allergens that are not declared in the food labels, or the adulteration process may involve the introduction of different pesticides, which can accumulate dangerously in the adulterated product [6,7].

According to a technical report by the European Commission's science and knowledge service (JCR) [5], oregano – the herb used to flavour many foods such as pizza and responsible for its characteristic "Italian" aroma – has emerged as the most adulterated herb, with 48 % of samples suspected of adulteration. The botanical species introduced as substitutes typically include olive leaves, sumac, cistus, strawberry tree and myrtle, whose dried and ground leaves are visually indistinguishable from oregano. Since unintentional contamination can occur during processing, the presence of extraneous matter is still tolerated within 2 % [8,9].

While the need for a proper authentication of dried oregano is evident, the selection of the most appropriate analytical technique is not straightforward. Various methodologies have been employed to face this authentication issue, each with its own advantages and limitations. In this context, we can delineate two primary approaches that have been employed by various research groups: targeted and untargeted approaches.

Targeted approaches are based on the identification and quantification of specific chemical markers of adulteration or authenticity for a specific product. For instance, Dabrova and colleagues [6] employed advanced mass spectrometry methods to analyse 400 pesticides in both genuine and adulterated oregano samples, identifying a number of compounds that were exclusively present in those adulterated. Other studies employed liquid chromatography coupled with mass spectrometry to detect additional biomarkers of adulteration [10,11]. Cottened and coauthors [12] used a DNA metabarcoding approach to analyse commercial samples of spices and herbs; in 22 % of the examined samples, they identified undeclared species in complex mixtures containing down to 1 % of adulterants. On the other hand, Pages-Rebull and coauthors [13] utilized HPLC-UV technique to determine phenolic compounds in various spices and aromatic herbs. Their study revealed that the presence and quantity of six of these compounds define their typical profile, making them suitable markers of authenticity. Targeted methodologies are highly selective and capable of detecting even very low levels of adulteration. However, they come with significant drawbacks, including high costs for the analysis, the requirement for highly specialized personnel, and the need for prior knowledge of the specific markers, which may not always be available. Additionally, the extraction and identification of markers often involve a lengthy and labour-intensive sample preparation process, particularly when dealing with complex matrices such as food samples.

Conversely, untargeted approaches entail a comprehensive examination of a sample distinctive chemical profile, without prior selection of specific compounds of interest. This approach facilitates a deep comprehension of the sample composition, essentially "recognizing" the profile of an authentic sample, much like its fingerprint, against which adulterated samples exhibit differences. Untargeted methods, therefore, necessitate the processing of analytical results using multivariate chemometric techniques, essential for extracting pertinent information from the fingerprint [14].

Among the recent untargeted techniques employed for oregano authentication, notable examples include nuclear magnetic resonance (NMR) spectroscopy, which has been effectively utilized for initial fingerprinting to discern oregano types, geographical origins, and the presence of other plant additives [15]. In other instances, mass spectrometry has been employed, using instrumental setups that offer rapid analyses with minimal or no sample preparation, such as Proton-Transfer Reaction Time-of-Flight Mass Spectrometry (PTR-TOF-MS), enabling real-time detection of volatile organic compounds [16]. Additionally, various applications of Ambient Mass Spectrometry, such as Direct Analysis in Real Time (DART-MS) [17–19] or Atmospheric Solid Analysis Probe (ASAP-MS) [19], have been utilized. While in these cases sample preparation may not be as time-consuming, challenges persist regarding instrument costs and the high level of technical expertise required of analysts. Mid- and Near-InfraRed spectroscopic fingerprinting techniques can thus overcome these limitations, simplifying and cost-effectively enhancing the analysis process. For these reasons, in the very recent years, several studies have emerged regarding the use of FTIR (Fourier-Transform InfraRed) and NIR (Near-InfraRed) spectroscopies for the authentication of both oregano [16,20–22] and other herbs and spices [4,23–25]. The last relatively unexplored frontier of infrared spectroscopic analysis for herbs and spices authentication is NIR HyperSpectral Imaging (NIR-HSI) [16,26,27].

NIR-HSI is a cutting-edge, rapid and non-destructive technique that allows the collection of both spectral and spatial information of the sample [28–30]. Indeed, each pixel of a NIR hyperspectral image contains a complete NIR spectrum, which represents a sort of chemical fingerprint at the corresponding sample position. In this manner, it is possible to obtain the so-called chemical maps of acquired samples, i.e., to characterize sample chemical composition and evaluate how it varies on sample surface. This method is particularly suitable for characterizing heterogeneous food matrices, such as ground herbs. Note that the ground herbs and spices contaminated with extraneous plant species may exhibit a variety of fragments, that potentially differ in terms of chemical composition.

In this study we used NIR-HSI to analyse authentic oregano samples, pure adulterants, and oregano samples adulterated with various types and amounts of spiked adulterants. The collected images were first explored by Principal Component Analysis (PCA) to assess spectral differences between pure oregano samples and adulterants. Afterwards, multivariate classification methods were used to obtain predictive models able to distinguish between authentic oregano samples and adulterated ones.

Ideally, authentication issues, such as the one considered in this study, should be assessed using Class Modelling (CM) classification approaches [31–33], which construct individual class models based on similarities among samples belonging to the same target class (i.e., authentic oregano). Consequently, a new sample can be assigned to one or more of the modelled classes, or to none of them. However, CM methods usually provide poor results when the variability within the target class (i.e., authentic oregano) is greater than the variability between target and non-target classes (i.e., pure adulterants), resulting in strong overlapping of the classes.

In contrast to CM, discriminant analysis (DA) methods maximise the differences among the studied classes, even if these differences are subtle, thus providing better results when dealing with overlapping classes. One of the most used algorithms for this purpose is Partial Least Squares Discriminant Analysis (PLS-DA), a modified version of the PLS statistical regression method [34,35]. Classical DA methods are recommended when the classes of interest are well defined as they force the class assignment of a new sample to one of the modelled classes. Therefore, they are not suitable for authentication issues where new unknown samples may belong to none of the considered classes.

To benefit from the advantages of both CM and DA methods, soft discriminant approaches may represent a valid alternative. These algorithms can be considered hybrid classification methods that combine CM and DA as they enable the classification of samples based on differences among the considered classes, while simultaneously identifying samples belonging to none of them [36–39].

Therefore, classification models were calculated on a training set of authentic oregano and of pure adulterants using both a CM approach, namely Soft Independent Modelling of Class Analogies (SIMCA) [31], and a soft discriminant method, namely Soft Partial Least Squares Discriminant Analysis (Soft PLS-DA) [39]. Essentially, Soft PLS-DA is the same as PLS-DA, however class assignment is subjected to some additional rules involving the calculation of further thresholds based on Q residuals and on *y* predictions. These additional thresholds do not constrain an unknown sample to belong to one of the modelled classes, thus facilitating effective handling of samples adulterated with extraneous herbs or materials not previously accounted for model calculation [36].

Finally, the SIMCA and Soft PLS-DA models were applied at the pixel level to all the acquired images, including those of oregano samples adulterated with different percentages of adulterants. The resulting prediction images were used to provide a final classification of the samples into genuine or adulterated oregano. Moreover, the percentage of pixels predicted as oregano from the Soft PLS-DA model allowed to define a sort of detection limit of NIR-HSI in this context.

2. Materials and methods

2.1. Oregano samples

The sample set included forty-nine samples: in detail, we analysed twenty-six authentic oregano samples (*Origanum vulgare*, *Origanum onites*, *Coleus amboinicus* and *Origanum vulgare* subsp. *Viridulum*, also known as Sicilian oregano), including two samples containing inflorescences and other two samples certified from the FAPAS proficiency tests 2985A-C. Four samples of pure adulterants including myrtle leaves (*Lagerstroemia indica*), sumac leaves (*Rhus coriaria*), strawberry tree leaves (*Arbutus unedo*) and olive leaves (*Olea europaea*) were also investigated. Moreover, nineteen adulterated oregano samples intentionally mixed with different percentages of myrtle leaves (*Lagerstroemia indica*), sumac leaves (*Rhus coriaria*), strawberry tree leaves (*Arbutus unedo*) and olive leaves (*Olea europaea*) and unintentionally polluted during various steps of the supply chain (with rosemary, cistus, hazelnut and sumac leaves) were tested. Among them, one certified oregano sample adulterated with olive leaves from FAPAS proficiency tests 2985A-C was included. The authentic samples originated from Italy, France, Turkey and Albany, and were harvested between 2019 and 2022. The percentages of adulterations ranged between 1.5 and 60 % (*see* Table S1 for the details of each sample).

2.2. Image acquisition and elaboration

Three random aliquots of each sample, ranging between 0.2 g and 1.0 g, were placed inside a glass Petri dish of 6.0 cm diameter and acquired as an individual image using a HSI line-scan system. Such system was composed of a desktop NIR Spectral Scanner (DV Optic, Padova, Italy) embedding a Specim N17E reflectance imaging spectrometer, coupled to a Xenics XEVA 1.7–320 camera (320 × 256 pixels) embedding Specim Oles 31 f/2.0 optical lens and covering the spectral range from 900 to 1700 nm (5 nm resolution, 150 spectral channels). Due to low S/N values, the wavelengths at the extremes of the spectral range were excluded: the final hyperspectral images, covering the spectral range between 980 and 1660 nm (137 wavelengths), were considered for further analysis.

To evaluate the system's stability over time, a setup composed of a silicon carbide sandpaper as sample background – characterized by a very low and constant reflectance spectrum [40] – a 99 % reflectance standard and two ceramic tiles with different grayscale tones and intermediate reflectance values, were used for the acquisition of all the images. The raw data were then converted into reflectance values by applying the instrument calibration procedure, which involved measuring the high-reflectance standard reference and the dark current. As a first step of image elaboration, an additional internal calibration was performed to minimize any residual variability among the images over time [41]. In total, 147 hyperspectral images were acquired (=49 samples × 3 replicates).

The corrected images were then cropped to a size of 248 × 199 pixels, in order to consider only the sample area. Subsequently, the pixels associated with the black sandpaper background and the glass Petri dish were removed from each image using a fast-thresholding procedure: all the pixels with reflectance values lower than 0.50 reflectance units measured at 980 nm were ascribable to the background and removed. Finally, a morphological erosion procedure, using a disk-shaped structuring element with a radius of 2 pixels, was performed to remove the pixels placed at the edges of the samples, which were affected by scattering phenomena and specular reflections of the glass Petri dish [29].

These image elaboration steps were performed using routines written *ad hoc* in MATLAB language (R2020b, The MathWorks Inc., USA).

2.3. Data analysis

2.3.1. Exploratory analysis

A preliminary exploratory analysis of the images was performed by means of PCA both at the *pixel-level* and at the *image-level*. In both cases, PCA was performed using linear detrend and mean center as spectral preprocessing methods.

For *pixel-level* exploratory analysis, some representative images of authentic oregano, pure adulterants (myrtle leaves, sumac leaves, strawberry tree leaves, and olive leaves) and adulterated oregano samples (mixtures of oregano and adulterants) were selected and merged together. PCA was applied to the merged images in order to have a preliminary evaluation of the spectral differences between oregano and the considered adulterants.

Subsequently, in order to have a global evaluation of the whole dataset structure at the *image-level*, as well as to gain an overall understanding of sample characteristics and behaviour, the average spectrum was calculated from each image and PCA was applied to the average spectra dataset.

2.3.2. Classification

Classification was carried out using two classification methods: SIMCA as a class modelling technique [31] and the soft discriminant method Soft PLS-DA [39].

For both classification methods the ability to distinguish authentic and adulterated oregano samples was evaluated in two steps. Firstly, *pixel level* models able to classify genuine oregano and pure adulterants (single class including leaves of myrtle, strawberry tree, olive and sumac) were calculated. Then, both models were applied to all the acquired hyperspectral images, including those of adulterated oregano samples (i.e., mixtures of oregano and adulterant), and for each image the corresponding percentage of pixels predicted as oregano (PPO%) was calculated. Based on this value, a threshold was set in order to identify each sample as authentic or adulterated oregano (see **Section 2.3.2.4**). As done for the previous exploratory analysis step, the spectra were preprocessed by applying linear-detrend followed by mean centering.

The classification performances of the models were evaluated by cross-validation (CV) and prediction of the external test set (TS) by calculating the statistical parameters sensitivity, specificity and efficiency [42], where.

- sensitivity (SENS), also known as *True Positive Rate*, measures the classifier's ability to correctly identify samples belonging to a considered class. SENS is calculated as the ratio between objects correctly assigned to the modelled class (*true positives*, TP) and all objects belonging to the class.
- specificity (SPEC), also known as *True Negative Rate*, evaluates the classifier's ability to reject samples belonging to other classes. SPEC is calculated as the ratio between objects correctly rejected by the modelled class (*true negatives*, TN) and all objects that do not belong to the considered class.
- efficiency (EFF), defined as the geometric mean of SENS and SPEC, provides an overall assessment of classification performance.

These classification performances were assessed by initially applying the models to a dataset comprising spectra references of authentic oregano and pure adulterants, and subsequently, to all the acquired hyperspectral images (see **Sections 2.3.2.1** and **2.3.2.4**, respectively).

2.3.2.1. Dataset structure. Firstly, we developed *pixel-level* models able to classify genuine oregano and pure adulterants. To this aim, we built-up a dataset of representative spectra belonging to both classes. This phase is crucial as it determines the representativeness of spectra references for the two classes, thus affecting the robustness and reliability of the classification models.

To this aim, a PCA model was calculated on the mean centered spectra of each image considering only the pixels belonging to the sample and selecting 3 principal components (PCs). Then, the pixels outside the 99.9 % confidence limit on both Hotelling T^2 and Q residuals values were excluded. Indeed, a deeper investigation these few pixels allowed to observe that they were ascribable to specular reflections or to small portions of the glass Petri dish that were not removed during background segmentation and erosion. Finally, a new PCA model was calculated, and Kennard-Stone algorithm [43] was applied in the PC space to select a representative number of pixel spectra. In particular, 600 spectra were selected from each image of pure adulterants, resulting in a total of 7200 representative spectra collected for the pure adulterants class (=600 spectra × 12 hyperspectral images), while 100 pixel spectra were selected from each image of authentic oregano samples, for a total of 7200 representative spectra (=100 spectra × 72 hyperspectral images). Note that two authentic oregano samples, listed as # 9 and # 10 in Table S1, were excluded in this phase due to the presence of branch fragments, but they were used for the final validation of the classification models (*see* **Section 2.3.2.4**). This dataset of representative spectra included therefore an overall number of 14400 spectra, and it was then used to develop the classification models.

Before calculating the classification models, the dataset was split into a training (TR) set used for model calculation and a test (TS) set used for model validation. The samples of authentic oregano were randomly subdivided in training and test samples with a ratio of 2/3 and 1/3, respectively, and the corresponding spectra were then assigned to the TR or TS dataset accordingly. Considering the pure adulterants, the subdivision into TR and TS sets was based on acquisition replicates: for each pure adulterant sample, the spectra of two of the three aliquots were included in the TR and one in the TS dataset. Therefore, the composition of TR and TS datasets can be summarised as follows.

- TR: 9600 spectra in total, 4800 spectra selected from 48 images of authentic oregano samples and 4800 spectra selected from 8 images of pure adulterants;
- TS: 4800 spectra in total, 2400 spectra selected form 24 images of authentic oregano samples and 2400 spectra selected from 4 images of pure adulterants.

Fig. 1 reports the mean spectrum of the selected pixel spectra of authentic oregano belonging to the TR set together with the mean spectra of the different pure adulterants considered in this study.

2.3.2.2. Spectra classification by SIMCA. Under a CM perspective, the authentication issue of this study can be considered a one-class classification problem. In fact we are interested in defining the boundaries of a single target class, i.e., authentic oregano, and predicting if a new sample belongs or not to the target class. Therefore, we developed a one-class SIMCA model considering only the spectra of authentic oregano of the TR set, while the spectra of pure adulterants of the TR set were used during cross-validation following a compliant approach [44].

SIMCA algorithm models the similarities among samples of the target class (i.e., authentic oregano), assuming that the main features of the target class can be represented by a Principal Component (PC) space of adequate dimensionality, commonly known as class subspace. Class assignment of new observations is carried out by calculating two statistical metrics accounting for the distance between the new observation and the target class subspace: the Orthogonal Distance (OD) and the Score Distance (SD). OD is the squared Euclidean distance of each new observation from its projection into the PCA model, while SD is defined as the squared Mahalanobis distance between the projection of the sample into the PCA subspace and the origin of the PCs [45].

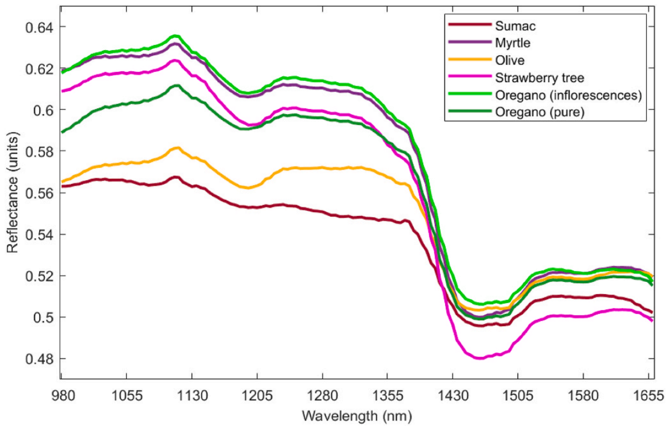


Fig. 1. Average spectrum of selected pixel spectra belonging to the TR set of authentic oregano class together with the average spectrum of each pure adulterant.

Afterwards, OD and SD values are usually compared with the corresponding critical limits (OD_{crit} and SD_{crit}, respectively) at a defined confidence level to perform the final assignment of the new observation. Different versions of SIMCA algorithm have been developed based on how OD and SD metrics as well as the corresponding confidence limits are used to perform class assignment [31].

In this study, we used Alternative SIMCA algorithm (Alt-SIMCA), which combines OD and SD values in a single statistical parameter, *d*, that defines the limits of the acceptance subregion (Eq. (1)) [31,46]:

$$d = \sqrt{\left(\frac{OD}{OD_{crit}}\right)^2 + \left(\frac{SD}{SD_{crit}}\right)^2} \qquad (1)$$

where OD_{crit} and SD_{crit} values correspond to the critical limits at 95 % confidence level. Only the observations with $d \leq \sqrt{2}$ are assigned to the target class, while those not meeting this decision rule are rejected and defined as non-target samples.

The optimal number of PCs was selected by maximizing the cross-validation efficiency (as defined in **Section 2.3.2**) following the compliant strategy for one-class modelling, which consists in using also samples not belonging to the target class for model optimisation. This strategy is generally recommended when dealing with overlapping classes [44]. In particular, the authentic oregano samples of the TR set were randomly split into two deletion groups and, based on this subdivision, the corresponding spectra were assigned to the two groups; in this manner, spectra selected from replicate images of the same oregano sample were kept in the same deletion group. Conversely, the TR set spectra belonging to pure adulterants were divided into the two deletion groups based on replicates., i.e., the spectra of one of the two replicate images of the TR set were assigned to one deletion group, while the spectra of the other replicate image were assigned to the other deletion group.

Afterwards, the classification performance of the Alt-SIMCA model was assessed though external validation using the TS set, which contains spectra belonging to both authentic oregano and pure adulterants.

Alt-SIMCA model was calculated using routines written *ad hoc* in MATLAB environment (ver. 2020b, The MathWorks, USA) based on

PLS_Toolbox functions (ver. 8.5, Eigenvector Research Inc., USA). The reader is referred to Vitale et al. [31] for an in-depth description of the Alt-SIMCA algorithm.

2.3.2.3. Spectra classification by soft PLS-DA. The TR set was used to calculate a classification model to discriminate between genuine oregano and pure adulterants by means of Soft PLS-DA. Soft PLS-DA is a soft discriminant algorithm that combines the advantages of discriminant analysis and class modelling approaches. Its configuration allows for increased flexibility and robustness in classification models: by applying additional constraints for class assignment, it effectively identifies possible outliers. In this manner, Soft PLS-DA overcomes the limitations of PLS-DA in handling new objects not belonging to the target classes, maximizing at the same time the discrimination between the classes of interest [36,39].

In details, a new sample is assigned to a defined class according to the following criteria.

- it must have Q residuals values falling within the 99.9 % confidence limit of the model. This limit has been chosen to set boundaries wide enough to consider as much as possible within classes variability, but allowing at the same time to exclude samples with a very poor fit to the model;
- it must have *y* predicted values falling within an acceptability range for the considered class. The lower limit is defined by the PLS-DA threshold value for the class under investigation, while the upper limit allows for the rejection of objects located at the extremes of the Gaussian probability density function;
- for multiclass classification, the samples must be unambiguously assigned to only one class.

The samples that do not match all the three criteria defined by the Soft PLS-DA decision rules are not assigned to any class and automatically labelled as “not assigned” samples (NA).

Soft PLS-DA model was optimized by using the same custom cross-validation scheme and samples splitting criterion previously mentioned in **Section 2.3.2.2**. Furthermore, external validation was performed by predicting class assignment of the samples belonging to the TS set.

Soft PLS-DA model was calculated using routines written *ad hoc* in MATLAB environment (ver. 2020b, The MathWorks, USA). The MATLAB routine to run Soft PLS-DA algorithm [39] is freely downloadable from http://www.chimslab.unimore.it/downloads/. The reader is referred to Calvini et al. [39] for a detailed description of the Soft PLS-DA algorithm.

2.3.2.4. Validation on external images. Alt-SIMCA and Soft PLS-DA models, obtained as previously described in **Section 2.3.2.2** and **Section 2.3.2.3**, were applied to all the acquired hyperspectral images, including the images of adulterated oregano samples. The resulting prediction images, in which each pixel is coloured according to the class assignment of the corresponding spectrum, were used to directly visualize the prediction performance on the images and obtain a quantitative evaluation of the classification performances on the entire set of images. To this aim, the percentage of pixels predicted as authentic oregano (PPO%) was calculated for each prediction image.

Finally, to assess the overall ability of the classification model to differentiate between authentic and adulterated oregano, we defined a threshold value based on the percentage of pixels predicted as pure oregano in each image. Samples whose images had PPO% values higher than the threshold were considered as authentic oregano samples, while samples whose images had PPO% values lower than the threshold were considered as adulterated. This threshold value was calculated as the minimum PPO% value obtained for the images of pure oregano belonging to the training set (*see* **Section 3.4**).

3. Results and discussion

3.1. Exploratory analysis at the pixel-level

An exploratory analysis at the *pixel-level* was performed to evaluate the spectral differences between authentic oregano, pure adulterants (myrtle leaves, sumac leaves, strawberry tree leaves, and olive leaves) and adulterated oregano. To this aim, for each adulterant type a unique hyperspectral image was obtained by merging together one image of the pure adulterant, one image of an authentic oregano sample and two images of oregano adulterated with the corresponding adulterant at different percentages. The merged hyperspectral images were then analysed by PCA.

Fig. 2 reports the results of the PCA model calculated on the merged image of an authentic oregano sample (47_04, 0 % myrtle leaves), of two oregano samples adulterated with myrtle leaves (48_29, 10 % myrtle leaves, and 48_04, 60 % myrtle leaves) and of the pure adulterant (A_05, 100 % myrtle leaves).

Fig. 2A shows the PC1-PC2 score plot, where the first two principal components account for 87.69 % and 6.67 % of total variance, respectively, whereas Fig. 2B reports the corresponding loading vectors. In the score plot each object represents a single pixel and it is coloured according to pixel density, i.e., red colour represents a region of the PC1-PC2 score space with a high density of pixels, while blue colour corresponds to low pixel density. From this score plot it is possible to observe the presence of two clusters of pixels, separated along PC2. The PC2 score image reported in Fig. 2C shows that the differences observed along PC2 are ascribable to the spectral differences between authentic oregano and myrtle leaves. Indeed, the pixels of the authentic oregano sample (0 % of adulteration) are mainly characterised by low PC2 score values, while the pixels of the myrtle sample (100 % of adulteration) show generally high PC2 score values. Note that the adulterated samples show an intermediate behaviour somehow proportional to the percentage of adulteration. In fact, the oregano sample adulterated with 10 % of myrtle leaves has PC2 score values comparable to that of authentic oregano, whereas the oregano sample with 60 % of adulteration generally presents positive PC2 score values, slightly lower than those of the image of the myrtle leaves.

Similar results were obtained from the PCA model calculated on the merged hyperspectral image for strawberry tree leaves as adulterant (Fig. S1 of Supplementary Material). Also in this case, PC2 allows separating the pixel spectra of the authentic oregano sample from those of the sample with only strawberry tree leaves, and the oregano samples adulterated with 20 % and 30 % of strawberry tree leaves show an intermediate behaviour. The PCA model calculated on the merged hyperspectral image considering sumac as adulterant (Fig. S2 of Supplementary Material) provides comparable results to those previously discussed for myrtle and strawberry tree leaves. Finally, the results of the investigation regarding olive leaves as adulterant are reported in Fig. S3 of Supplementary Material. In this case, the direction which better reflects the differences between the images according to the percentage of adulterant is represented by PC3.

Therefore, the PCA models calculated on the merged images allowed capturing the presence of detectable spectral differences between authentic oregano and the adulterants investigated in this study. Considering the loading vectors of the relevant PCs for this separation (mainly PC2 for myrtle leaves, strawberry tree leaves and sumac, and PC3 for olive leaves as adulterants), it is possible to identify some common spectral regions that contribute to these findings (Fig. 2B and Figs. S1B–S3B of Supplementary Material).

These spectral regions fall into the 980–1080 nm spectral range, corresponding to O–H third overtone and C–H second overtone, associated with polyphenol content, in the 1150–1200 nm spectral range corresponding to asymmetric stretching of C–H second overtone, ascribable to alcohols, in the 1420–1450 nm, ascribable to O–H stretch first overtone, C=O stretch third overtone and N–H stretch first

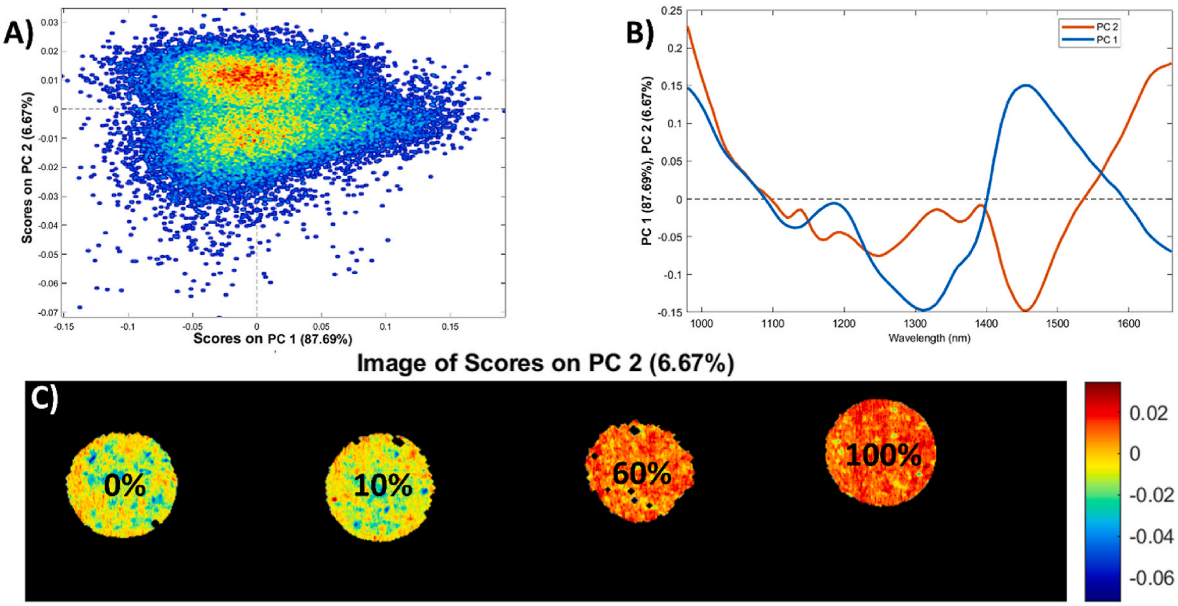


Fig. 2. Principal component analysis (PCA) results of the merged hyperspectral image containing one authentic oregano sample (0 % of adulteration), two samples adulterated with different percentages of myrtle leaves (10 % and 60 % of adulteration) and one sample of pure myrtle leaves (100 % of adulteration). In (A) PC1-PC2 score plot; in (B) PC1 and PC2 loading vectors and in (C) PC2 score image.

overtone which could be ascribable to terpenoids and cellulose content, and in the 1620–1660 nm region, ascribable to the stretching of aromatic C–H first overtone [21,47–51]. Additional spectral regions that codify for myrtle, strawberry tree and sumac can be found around 1250 nm and 1380 nm, related to alcohol and methyl groups [49].

3.2. *Exploratory analysis at the image-level*

As mentioned in **Section 2.3.1**, the whole image dataset was also evaluated at the *image-level* to gain an insight on sample characteristics and behaviour. To this aim, the average spectrum was obtained from each image and a global PCA model was calculated on the average spectra dataset using linear detrend and mean center as preprocessing methods.

Since in the previous evaluation performed at the *pixel-level* (*see* **Section 3.1**) we observed detectable differences between the spectral signatures of authentic oregano and pure adulterants, the *image-level* analysis was focused on identifying possible trends due to adulterant type and amount. Therefore, a PCA model was calculated on the average spectra of the sole authentic and adulterated oregano. The resulting PC1-PC2 score plot is reported in Fig. 3, accounting for 92.06 % of explained variance. In Fig. 3A the samples in the score plot are coloured according to authentic or adulterated class. It is worth noting that the oregano samples have a wide chemical variability and morphological heterogeneity, probably due to the different geographical origins and multiple harvest years. Furthermore, the two classes of authentic and adulterated oregano samples are partly overlapped. Only a limited separation of some adulterated samples, characterised by extreme (both positive and negative) PC2 score values was observed. A more in-depth investigation, based on adulterant type and percentage (Fig. 3B), revealed that the adulterated samples showing more marked differences from pure oregano were those characterized by adulteration percentages equal or higher than 60 %, 30 % and 20 % with myrtle leaves, olive leaves and strawberry tree leaves, respectively.

These findings confirm the results of the PCA models calculated at the *pixel-level*, where adulterated samples with percentages lower than 20 % showed similar behaviour to authentic oregano samples.

3.3. *Classification between authentic oregano and pure adulterants*

The first step in the identification of adulterated oregano samples consists in the development of *pixel-level* classification models able to distinguish genuine oregano from pure adulterants. Table 1 reports the results obtained in calibration (CAL), cross-validation (CV) and validation of the external test set (TS) of the Alt-SIMCA and Soft PLS-DA models. The classification performances were evaluated by calculating SENS, SPEC and EFF values: for Alt-SIMCA, the results refer to authentic oregano, which is the target class of the model, while for Soft PLS-DA the results of both pure adulterants and authentic oregano classes are reported. The results in Table 1 clearly show that the Soft PLS-DA model achieved good classification performances, with SENS and SPEC values for both classes higher than 90 %.

Interestingly, about 39 % of misclassified pixel spectra of authentic oregano class (i.e., spectra of authentic oregano class but predicted as pure adulterants by Soft PLS-DA model) in cross-validation belong to genuine oregano with inflorescences. This finding suggests that the presence of inflorescences may negatively affect the classification performances.

Conversely, despite the Alt-SIMCA model reached excellent SENS values both in cross-validation and TS set prediction, its ability to correctly reject not-authentic oregano samples is unsatisfactory, as indicated by SPEC values around 50 %. The poor classification performances obtained with Alt-SIMCA can be explained considering the significant within-class variability both for authentic oregano and for pure adulterants, resulting in partial overlap between the two classes. In this context, class modelling approaches are generally not effective.

A more in-depth evaluation of the classification results was also performed, based on adulterant type. For this reason, considering Alt-SIMCA algorithm, Table 2 reports for each adulterant type the percentage of spectra accepted or rejected by the authentic oregano class model. Similarly, Table 2 also reports the results obtained for Soft PLS-DA, expressed as percentage of spectra assigned by the model to authentic oregano class, pure adulterants class and not assigned spectra. For both models, the results reported in Table 2 are referred to cross-validation and prediction of the TS set, while for Soft PLS-DA model the results obtained also in calibration are reported in Table S2 of Supplementary Material.

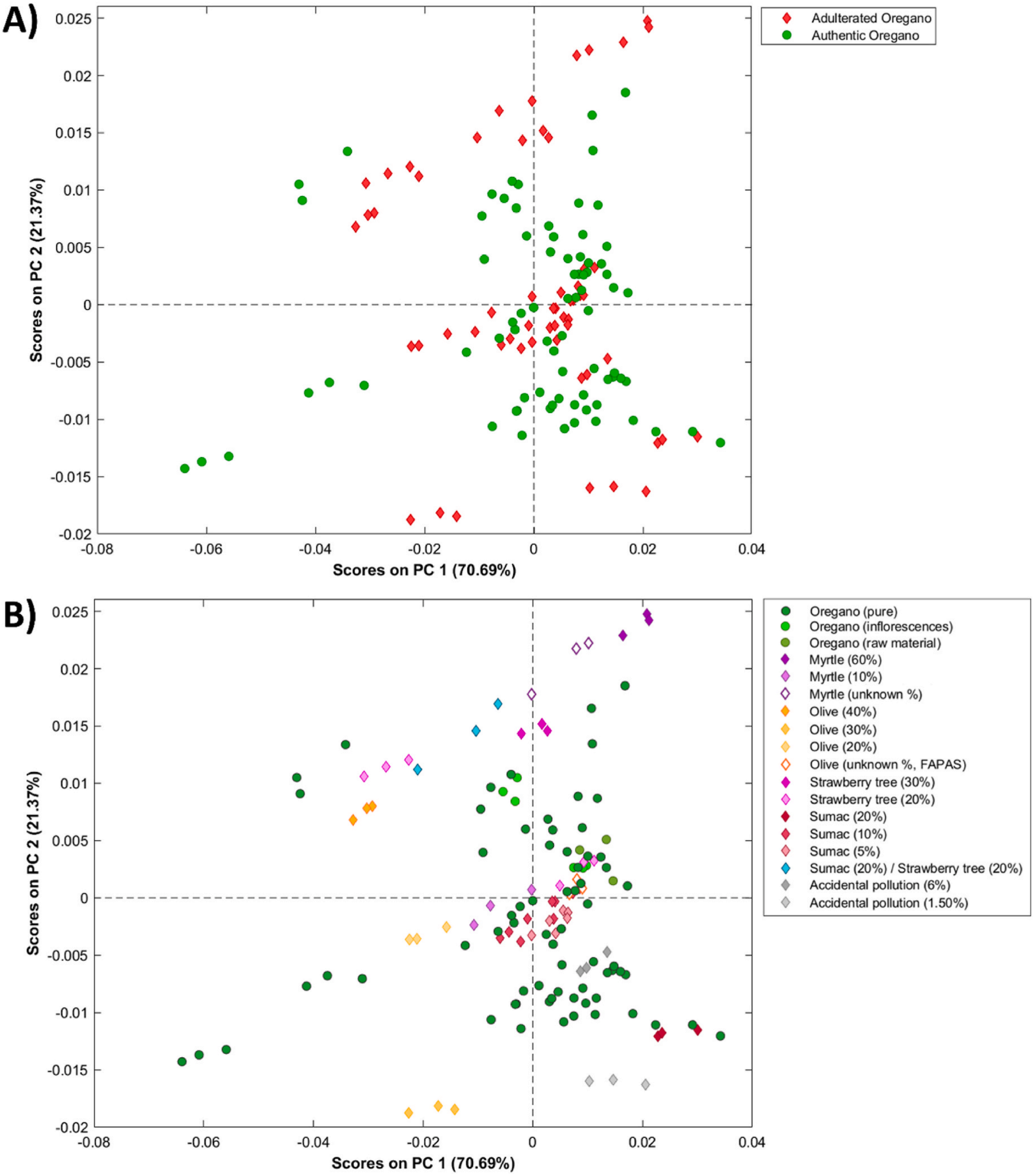


Fig. 3. PC1-PC2 score plot obtained by calculating a PCA model considering the authentic oregano (circle) and adulterated oregano (rhombus) samples average spectra. In (**A**) samples are coloured according to authentic and adulterated class; in (**B**) samples are coloured based on adulterant type and percentage.

Alt-SIMCA provided satisfactory classification performances only for olive leaves, achieving a percentage of correctly rejected spectra of 79.7 % for TS set prediction. Conversely, overall poor classification performances were obtained for the other adulterant types. Specifically, approximately half of the spectra belonging to strawberry tree leaves and sumac were wrongly accepted by the authentic oregano class model, and the same applies to the vast majority of myrtle spectra.

Concerning Soft PLS-DA, the best performances were obtained for strawberry tree leaves and sumac, with a percentage equal to or less than 1.7 % of spectra misclassified as genuine oregano both in cross-validation and prediction of the test set. Conversely, higher misclassifications were obtained for myrtle and olive leaves, where the percentage of correctly assigned spectra ranged between 82.7 % and 89.1 %. Therefore, the Soft PLS-DA model better recognized sumac and strawberry tree leaves as adulterants; however, the results obtained for myrtle and olive leaves can still be considered satisfactory.

In order to evaluate the spectral regions that contribute the most to the identification of authentic oregano, Fig. 4 reports the Variable Importance in Projection (VIP) scores of the Soft PLS-DA model. In particular, the spectral variables with VIP scores higher than 1 (red dashed line in Fig. 4) are those with higher relevance for the classification model. These variables fall into the intervals at 980–1010 nm (O–H third overtone and C–H second overtone), 1130–1150 nm (C–H second overtone and stretching of C=O fourth overtone), 1180–1225 nm (asymmetric stretching of C–H second overtone), 1390–1420 nm (stretching of O–H first overtone for ROH and ArOH), 1445–1470 nm (stretching of O–H first overtone for water, stretching of C=O third overtone, stretching of N–H first overtone) and 1640–1650 nm

Table 1
Classification performances of Alt-SIMCA and Soft PLS-DA models in calibration (CAL), cross-validation (CV) and prediction of the external test set (TS).

		Alt-SIMCA	Soft PLS-DA	
		Authentic Oregano	Pure Adulterants	Authentic Oregano
	PCs/ LVs	6	7	
Calibration (CAL)	**SENS (%)**	94.9	92.4	93.5
	SPEC (%)	–	94.3	93.8
	EFF (%)	**-**	**93.4**	**93.7**
	NA (%)	–	1.4	0.8
Cross-validation (CV)	**SENS (%)**	90.9	91.7	90.4
	SPEC (%)	50.8	91.5	93.7
	EFF (%)	**67.9**	**91.4**	**92.0**
	NA (%)	–	2.4	1.1
Prediction (TS)	**SENS (%)**	96.9	90.8	93.6
	SPEC (%)	47.1	94.1	92.3
	EFF (%)	**67.5**	**92.4**	**92.9**
	NA (%)	–	1.5	0.5

(stretching of aromatic C–H first overtone). The wavebands related to C–H and C═O absorption could be associated to polyphenols and terpenoids, whereas the wavebands related to O–H and N–H absorption can be associated to water, cellulose, hemicellulose and lignin. In addition, the relevance of O–H and aromatic C–H can be associated with hydroxyl and aromatic groups, particularly present also in polyphenols. Therefore, the discriminant wavebands can be associated to differences in the aromatic content, in terms of polyphenols, terpenoids, esters and alcoholic groups, and to cellulose, hemicellulose and lignin, which are ubiquitous in plant tissues [21,47–51].

3.4. Validation on external images and identification of adulterated samples

Alt-SIMCA and Soft PLS-DA classification models were applied to all the acquired hyperspectral images. For Soft PLS-DA, some representative prediction images are reported in Fig. 5, where the pixel spectra predicted as authentic oregano are represented in green colour, the pixel spectra predicted as pure adulterant are reported in red colour while the not assigned pixels are reported in grey colour.

Specifically, the first row of images in Fig. 5 reports the prediction images obtained from five authentic oregano samples, including one sample with inflorescences. Note that almost all the pixels belonging to the pure oregano samples were correctly predicted as authentic oregano and a few misclassifications are ascribable to an intrinsic error of the classification model. The oregano sample containing inflorescences represents an exception, since it has a high number of misclassified pixels; this fact confirms what already observed in **Section 3.3**, where a relevant number of misclassified spectra of genuine oregano belonged to samples with inflorescences.

In the second row of Fig. 5 six prediction images, calculated on adulterated oregano samples, are shown: from left to right, the images are reported at decreasing concentrations of adulterants. In this case, the samples with adulterant concentrations equal to or higher than 10 % have a relevant number of pixels predicted as adulterants and their amount is roughly proportional to the adulterant concentration. On the other hand, the prediction images of oregano samples adulterated at percentages lower than 10 % have a number of pixels predicted as adulterant comparable to or even smaller than the number of misclassified pixels of those of the pure oregano.

In order to perform a global evaluation of the prediction ability of Alt-SIMCA and Soft PLS-DA classification models, the percentage of pixels predicted as oregano class (PPO%) was calculated for each prediction image obtained with both methods. The scatter plot in Fig. 6 shows the relationship between the actual oregano percentage in the analysed sample aliquots and the PPO% values obtained from the corresponding prediction images, calculated by applying Alt-SIMCA (Fig. 6A) and Soft PLS-DA (Fig. 6B) models.

The plot with Alt-SIMCA results (Fig. 6A) reveals that images of authentic oregano correctly present PPO% values higher than 90 %. However, also all the images of adulterated samples have PPO% values around 90 %, regardless of the actual adulterant concentration. High PPO% values are also evident for the pure adulterants, particularly for myrtle which has PPO% values around 90 %. These results confirm the low specificity of Alt-SIMCA model, i.e., its poor ability of correctly

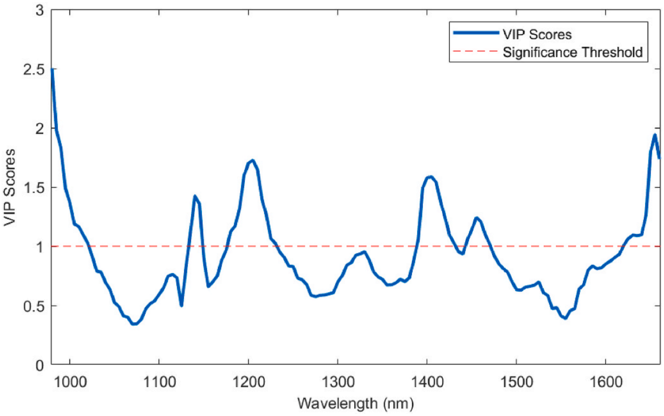


Fig. 4. VIP scores of the Soft PLS-DA model for the discrimination of authentic oregano and pure adulterants.

Table 2
Classification performances of pure adulterants in cross-validation (CV) and prediction of the test set (TS). Alt-SIMCA: for each adulterant type the percentage of spectra accepted or rejected by the authentic oregano class model is reported. Soft PLS-DA: for each adulterant type the percentage of spectra predicted as authentic oregano, pure adulterants or not assigned (NA) is reported.

			Myrtle	Olive	Strawberry tree	Sumac
Alt-SIMCA	CV	Authentic oregano (%)	83.9	20.2	45.7	47.1
		Not Authentic oregano (%)	16.1	79.8	54.3	52.9
	TS	**Authentic oregano (%)**	89.0	20.3	52.0	50.3
		Not Authentic oregano (%)	**11.0**	**79.7**	**48.0**	**49.7**
Soft PLS-DA	**CV**	**Authentic oregano (%)**	14.7	7.9	1.1	1.7
		Pure adulterants (%)	**84.8**	**89.1**	**97.1**	**94.2**
		NA (%)	0.6	3.0	1.8	4.2
	TS	**Authentic oregano (%)**	17.2	11.7	0.7	1.5
		Pure adulterants (%)	**82.7**	**83.8**	**98.7**	**97.8**
		NA (%)	0.2	4.5	0.7	0.7

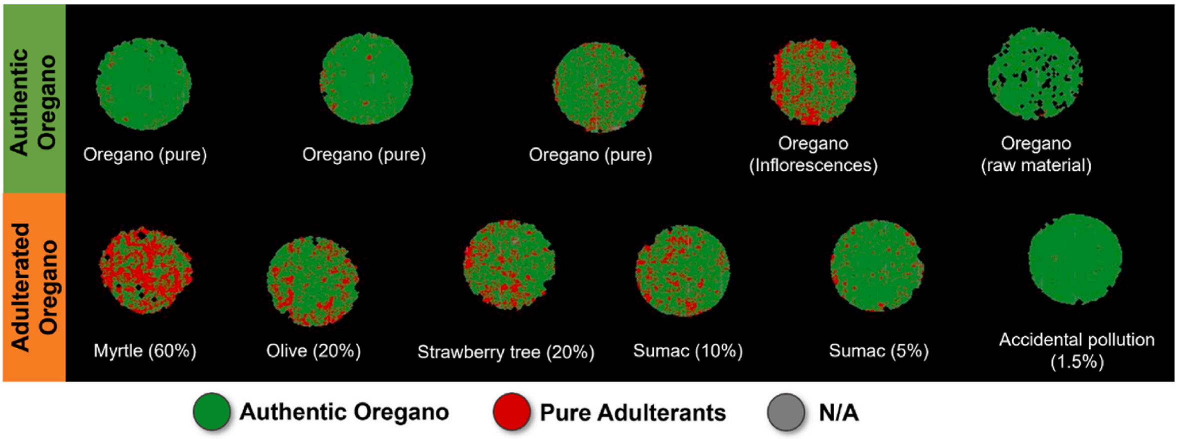


Fig. 5. Prediction images obtained by applying the Soft PLS-DA model to hyperspectral images of authentic and adulterated oregano samples. Prediction images of pure oreganos, one oregano sample with inflorescences and one sample of raw oregano are reported in the first row, whereas the prediction images of adulterated oregano samples at decreasing percentages of adulteration (from left to right) are reported in the second row.

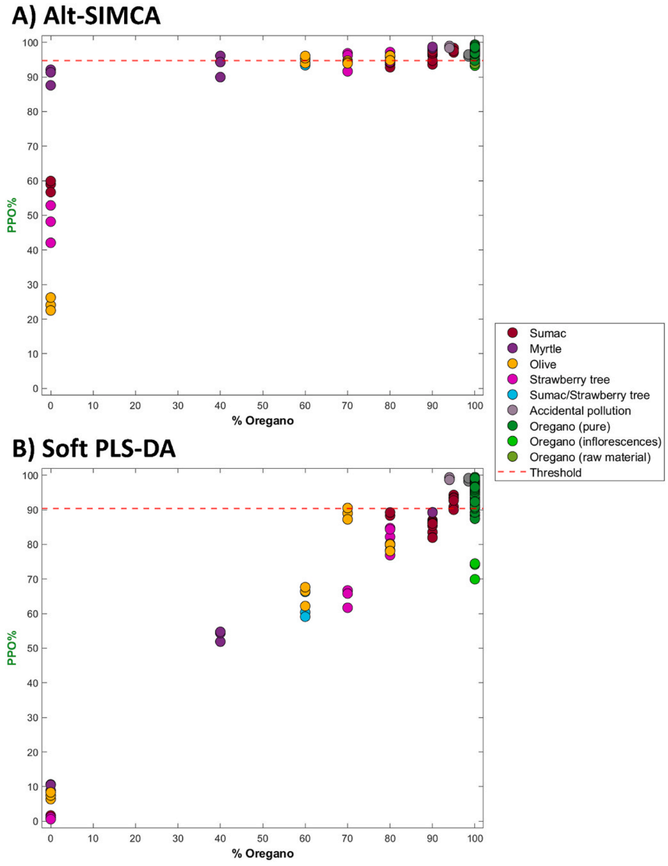


Fig. 6. Actual oregano concentration vs. percentage of pixels predicted as oregano (PPO%) by the Alt-SIMCA (A) and Soft PLS-DA (B) models. The red dashed line represents the threshold value used to discriminate images of authentic oregano from images of adulterated oregano. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

rejecting samples not belonging to the authentic oregano class.

Conversely, the plot related to Soft PLS-DA results (Fig. 6B) shows a discrete correlation between actual oregano content and the percentage of pixels predicted as authentic oregano extracted from the prediction images. Concerning the pure adulterants, the PPO% values were found to be in the 0–10 % range. Except for sample 47_08 and sample 47_15 containing inflorescences (*see* Table S1 of Supplementary Material), the images of authentic oregano showed PPO% values around 90 % or higher. In this case, the adulterated samples have PPO% values that are generally proportional to the actual concentration of pure oregano. Except for one image of a sample adulterated with 30 % olive leaves, only the samples adulterated with less than 10 % of adulterant show PPO % values higher than 90 %, comparable to those of authentic oregano.

Since the final goal of this study was the identification of authentic and adulterated oregano samples, we decided to define a threshold value based on PPO% value to assign the samples to one of the two classes. This threshold value was defined as the minimum PPO% value obtained for the hyperspectral images of authentic oregano samples belonging to the training set (*see* **Section 2.3.2.1**). The threshold was set at PPO% values equal to 94.74 % and 90.41 % for Alt-SIMCA and Soft PLS-DA, respectively. Therefore, the hyperspectral images, whose prediction images have PPO% values equal or higher than the threshold values, were assigned to the authentic oregano class whereas those with a lower value were classified as adulterated.

The resulting outcomes are reported as confusion matrix (Table 3), where the columns represent the actual classes, and the rows are the assigned classes. Table 3 shows the results obtained only for the images used as external validation, i.e., the authentic oregano images used as test images and all the images of adulterated samples. For ease of interpretation, the results of the images of adulterated oregano were reported after splitting them in three categories based on adulterant concentration: images of samples with adulterant concentration lower

Table 3
Classification results of the images used for external validation into authentic and adulterated oregano classes based on the threshold considering PPO% values, obtained by applying Alt-SIMCA and Soft PLS-DA. For the images of adulterated oregano, the results are reported by splitting the samples according to adulterant concentration: adulterant concentration lower than 10 % (adult. < 10 %), adulterant concentration equal to or higher than 10 % (adult. ≥ 10 %) and unknown concentration.

		Actual class			
		Authentic oregano	Adulterated oregano		
			adult.< 10 %	adult.≥ 10 %	unknown
Assigned class (Alt-SIMCA)	**Authentic**	**29**	9	26	6
	Adulterated	4	**3**	**13**	**0**
Assigned class (Soft PLS-DA)	**Authentic**	**27**	11	1	3
	Adulterated	6	**1**	**38**	**3**

than 10 % (adult. < 10 %), with adulterant concentration equal to or higher than 10 % (adult. ≥ 10 %) and with unknown concentration.

Concerning the ability of the models in correctly recognizing authentic oregano samples, the performances are comparable: Alt-SIMCA correctly classified 29 images of authentic oregano out of 33, which corresponds to a SENS value of 87.9 %, while Soft PLS-DA correctly classified 27 images out of 33, which coincide with a SENS value of 81.8 %. In both cases, the misclassified images of authentic oregano include 3 replicates of two oregano samples with inflorescences, confirming that the presence of inflorescences can negatively affect the classification performances.

Conversely, the two models have a different ability to identify adulterated oregano samples. Concerning Alt-SIMCA, only 16 adulterated oregano images out of 57 (28.1 %) were correctly recognized.

On the other hand, Soft PLS-DA correctly attributed 42 adulterated oregano images out of 57 to the corresponding class (73.7 %). Furthermore, among adulterated oregano there is a clear difference in the classification performances based on adulterant concentration. As expected, 11 out of 12 images with adulterant concentration lower than 10 % were erroneously assigned to the authentic class, while 38 out of 39 images with adulterant concentration equal to or higher than 10 % were correctly classified as adulterated oregano. The misclassified image is a replicate of one adulterated sample whose remainder replicates were correctly classified. Supposing a sample-based classification by majority voting of the assignments done on the three replicated images of each sample, we can state that all the samples with adulterant concentration equal to or higher than 10 % were correctly identified by Soft PLS-DA.

The image-level classification based on the Soft PLS-DA results on the one hand confirmed the preliminary findings obtained by exploratory data analysis, and on the other hand allowed to better identify the minimal spectral differences between pure adulterants and authentic oregano. In accordance with previous studies [16], NIR-HSI is affected by a detection limit of 10 % adulteration. According to the European Spice association [9], 2 % of extraneous matter is tolerated; however, a detection limit of 10 % may still be considered acceptable, since oregano adulteration levels are generally higher that this value to lead to a concrete economic advantage.

4. Conclusions

The aim of the present study was to evaluate NIR-HSI as a rapid, non-destructive and untargeted method to authenticate oregano samples which, due to their heterogeneity, can benefit from the coupling of spectral and spatial information.

The initial exploratory analysis performed both at the *pixel level* on some representative images and at the *image level* on the average spectra dataset allowed to identify the presence of spectral differences between authentic oregano and pure adulterants, to point out a remarkable heterogeneity among different genuine oregano samples and to highlight the need of accounting for spatial variation of sample composition to authenticate the samples.

Based on these considerations, Alt-SIMCA and Soft PLS-DA algorithms were used to build classification models able to differentiate authentic oregano and its most frequent adulterants, i.e., myrtle, olive leaves, strawberry tree leaves and sumac.

Due to classes overlapping and heterogeneity of the authentic oregano class, Alt-SIMCA led to overall poor classification performances. Conversely, Soft PLS-DA achieved satisfactory outcomes, with efficiency values in classification higher than 91 % in calibration, cross-validation and validation of the external test set. In this case, the spectra of pure strawberry tree and sumac leaves were easier to distinguish from authentic oregano, while pure myrtle and olive leaves presented higher misclassifications.

To obtain a final assignment of the acquired oregano samples into authentic and adulterated classes, both classification models were applied to all the acquired hyperspectral images and from each image the corresponding percentage of pixels predicted as oregano (PPO%) was calculated. Once defined a PPO% threshold value to differentiate authentic oregano samples from adulterated ones, it was possible to reach SENS values for authentic class equal to 87.9 % and 81.8 % for Alt-SIMCA and Soft PLS-DA, respectively. In both cases, the misclassifications of authentic oregano were mainly due to samples containing inflorescences.

The main differences in classification performances were encountered in the ability of correctly differentiating adulterated oregano samples. Indeed, while Alt-SIMCA was unable to correctly identify most of the adulterated samples, Soft PLS-DA successfully distinguished all adulterated oregano samples with adulterant concentrations equal to or greater than 10 %. These results confirm that soft discriminant approaches like Soft PLS-DA are an effective and powerful alternative to CM and DA methods when dealing with authentication problems.

Furthermore, according to the results obtained with Soft PLS-DA, we can consider the 10 % of adulteration as a sort of limit of detection of NIR-HSI to identify adulterated oregano samples. Considering that the percentage of adulteration detected on market oregano samples has been found very often at much higher levels, these results seem rather satisfactory to corroborate NIR-HSI potentialities as a screening technique able to face adulteration issues, also considering the possibility of performing the analysis in a fast and non-destructive manner.

Further developments may involve expanding the dataset to include additional authentic samples, aiming to better represent the intrinsic variability of oregano matrices and diverse types of adulterants at different percentages of adulteration as well. From the results obtained from the Soft PLS-DA model, a discernible correlation between actual oregano percentages and PPO% values emerged, suggesting the potential for developing a quantitative model with proper sampling.

Moreover, the application of spectral variable selection methods could enhance the robustness and flexibility of the model, while also offering a monitoring system that is easier to apply. Indeed, the selected wavebands can be used to implement multispectral imaging systems, which are more suitable for industrial applications in terms of computational time, durability and lower costs of the optical components.

CRediT authorship contribution statement

Veronica Ferrari: Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation. **Rosalba Calvini:** Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **Camilla Menozzi:** Writing – review & editing, Investigation. **Alessandro Ulrici:** Writing – review & editing, Supervision. **Marco Bragolusi:** Writing – review & editing, Resources. **Roberto Piro:** Writing – review & editing, Supervision, Resources. **Alessandra Tata:** Writing – original draft, Resources. **Michele Suman:** Writing – review & editing, Supervision, Resources. **Giorgia Foca:** Writing – original draft, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgements

The authors would like to thank the funding programme "FAR_DIP2022", Department of Life Sciences – University of Modena and Reggio Emilia.

Rosalba Calvini would like to thank the Italian funding programme Fondo sociale europeo REACT-EU - PON "Ricerca e Innovazione" 2014–2020 – Azione IV.6 Contratti di ricerca su tematiche Green (D.M. 1062 del 10/08/2021) for supporting her research (CUP: E95F21002330001; contract number 17-G-13884–4).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemolab.2024.105133.

References

[1] P. Galvin-King, S.A. Haughey, C.T. Elliott, Herb and spice fraud; the drivers, challenges and detection, Food Control 88 (2018) 85–97, https://doi.org/10.1016/j.foodcont.2017.12.031.

[2] European Spice Association, Adulteration Awareness Document, 2018. https://www.esa-spices.org/download/esa-adulteration-awareness-document2. (Accessed 9 October 2023).

[3] P.F. Ndlovu, L.S. Magwaza, S.Z. Tesfay, R.R. Mphahlele, Destructive and rapid non-invasive methods used to detect adulteration of dried powdered horticultural products: a review, Food Res. Int. 157 (2022) 111198, https://doi.org/10.1016/j.foodres.2022.111198.

[4] M. Shannon, J.L. Lafeuille, A. Frégière-Salomon, S. Lefevre, P. Galvin-King, S. A. Haughey, D.T. Burns, X. Shen, A. Kapil, T.F. McGrath, C.T. Elliott, The detection and determination of adulterants in turmeric using fourier-transform infrared (FTIR) spectroscopy coupled to chemometric analysis and micro-FTIR imaging, Food Control 139 (2022) 109093, https://doi.org/10.1016/j.foodcont.2022.109093.

[5] A. Maquet, A. Lievens, V. Paracchini, G. Kaklamanos, B. de la Calle, L. Garlant, S. Papoci, D. Pietretti, T. Zdiniakova, A. Breidbach, J. Omar Onaindia, A. Boix Sanfeliu, T. Dimitrova, F. Ulberth, Results of an EU Wide Coordinated Control Plan to Establish the Prevalence of Fraudulent Practices in the Marketing of Herbs and Spices, EUR30877EN, Publications Office of the European Union, 2021. JRC126785, https://doi:10.2760/309557.

[6] L. Drabova, G. Alvarez-Rivera, M. Suchanova, D. Schusterova, J. Pulkrabova, M. Tomaniova, V. Kocoureka, O. Chevallier, C. Elliott, J. Hajslova, Food fraud in oregano: pesticide residues as adulteration markers, Food Chem. 276 (2019) 726–734, https://doi.org/10.1016/j.foodchem.2018.09.143.

[7] S. Schaarschmidt, Public and private standards for dried culinary herbs and spices – Part I: standards defining the physical and chemical product quality and safety, Food Control 70 (2016) 339–349, https://doi.org/10.1016/j.foodcont.2016.06.004.

[8] FAO and WHO, Standard for dried oregano, codex alimentarius standard, No. CXS 342-2021, Codex Alimentarius Commission (2021).

[9] European Spice Association, European Spice Association Quality Minima Document, 2018. https://www.esa-spices.org/download/esa-qmd-rev-5-update-as-per-esa-tc-26-03-18.pdf. (Accessed 15 January 2024).

[10] C. Black, S.A. Haughey, O.P. Chevallier, P. Galvin-King, C.T. Elliott, A comprehensive strategy to detect the fraudulent adulteration of herbs: the oregano approach, Food Chem. 210 (2016) 551–557, https://doi.org/10.1016/j.foodchem.2016.05.004.

[11] E. Wielogorska, O. Chevallier, C. Black, P. Galvin-King, M. Delêtre, C.T. Kelleher, S. A. Haughey, C.T. Elliott, Development of a comprehensive analytical platform for the detection and quantitation of food fraud using a biomarker approach. The oregano adulteration case study, Food Chem. 239 (2018) 32–39, https://doi.org/10.1016/j.foodchem.2017.06.083.

[12] G. Cottenet, C. Cavin, C. Blancpain, P.F. Chuah, R. Pellesi, M. Suman, S. Nogueira, M. Gadanho, A DNA metabarcoding workflow to identify species in spices and herbs, J. AOAC Int. 106 (1) (2022) 65–72, https://doi.org/10.1093/jaoacint/qsac099.

[13] J. Pages-Rebull, C. Pérez-Ràfols, N. Serrano, M. del Valle, J.M. Díaz-Cruz, Classification and authentication of spices and aromatic herbs by means of HPLC-UV and chemometrics, Food Biosci. 52 (2023) 102401, https://doi.org/10.1016/j.fbio.2023.102401.

[14] K. Kucharska-Ambrożej, J. Karpinska, The application of spectroscopic techniques in combination with chemometrics for detection adulteration of some herbs and spices, Microchem. J. 153 (2020) 104278, https://doi.org/10.1016/j.microc.2019.104278.

[15] F. Flügge, T. Kerkow, P. Kowalski, J. Bornhöft, E. Seemann, M. Creydt, B. Schütze, U.L. Günther, Qualitative and quantitative food authentication of oregano using NGS and NMR with chemometrics, Food Chem. 145 (2023) 109497, https://doi.org/10.1016/j.foodcont.2022.109497.

[16] J. Van De Steene, J. Ruyssinck, J.A. Fernandez-Pierna, L. Vandermeersch, A. Maes, H. Van Langenhove, C. Walgraeve, K. Demeestere, B. De Meulenaer, L. Jacxsens, B. Miserez, Authenticity analysis of oregano: development, validation and fitness for use of several food fingerprinting techniques, Food Res. Int. 162 (2022) 111962, https://doi.org/10.1016/j.foodres.2022.111962.

[17] A. Massaro, A. Negro, M. Bragolusi, B. Miano, A. Tata, M. Suman, R. Piro, Oregano authentication by mid-level data fusion of chemical fingerprint signatures acquired by ambient mass spectrometry, Food Control 126 (2021) 108058, https://doi.org/10.1016/j.foodcont.2021.108058.

[18] C. Zacometti, A. Massaro, T. di Gioia, S. Lefevre, A. Frégière-Salomon, J. L. Lafeuille, I. Fiordaliso Candalino, M. Suman, R. Piro, A. Tata, Thermal desorption direct analysis in real-time high-resolution mass spectrometry and machine learning allow the rapid authentication of ground black pepper and dried oregano: a proof-of-concept study, J. Mass Spectrom. 58 (10) (2023) e4953, https://doi.org/10.1002/jms.4953.

[19] T. Damiani, N. Dreolin, S. Stead, C. Dall'Asta, Critical evaluation of ambient mass spectrometry coupled with chemometrics for the early detection of adulteration scenarios in *Origanum vulgare* L, Talanta 227 (2021) 122116, https://doi.org/10.1016/j.talanta.2021.122116.

[20] G. Sammarco, M. Alinovi, L. Fiorani, M. Rinaldi, M. Suman, A. Lai, A. Puiu, L. Giardina, F. Pollastrone, Oregano herb adulteration detection through rapid spectroscopic approaches: fourier transform-near infrared and laser photoacoustic spectroscopy facilities, J. Food Compos. Anal. 124 (2023) 105672, https://doi.org/10.1016/j.jfca.2023.105672.

[21] C. McVey, T.F. McGrath, S.A. Haughey, C.T. Elliott, A rapid food chain approach for authenticity screening: the development, validation and transferability of a chemometric model using two handheld near infrared spectroscopy (NIRS) devices, Talanta 222 (2021) 121533, https://doi.org/10.1016/j.talanta.2020.121533.

[22] O.Ye Rodionova, A.L. Pomerantsev, Chemometric tools for food fraud detection: the role of target class in nontargeted analysis, Food Chem. 317 (2020) 126448, https://doi.org/10.1016/j.foodchem.2020.126448.

[23] R. Khodabakhshian, M.R. Bayati, B. Emadi, Adulteration detection of Sudan Red and metanil yellow in turmeric powder by NIR spectroscopy and chemometrics: the role of preprocessing methods in analysis, Vib. Spectrosc. 120 (2022) 103372, https://doi.org/10.1016/j.vibspec.2022.103372.

[24] A.M. Elfiky, E. Shawky, A.R. Khattab, R.S. Ibrahim, Integration of NIR spectroscopy and chemometrics for authentication and quantitation of adulteration in sweet marjoram (Origanum majorana L.), Microchem. J. 183 (2022) 108125, https://doi.org/10.1016/j.microc.2022.108125.

[25] A. Massaro, M. Bragolusi, A. Tata, C. Zacometti, S. Lefevre, A. Frégière-Salomon, J. L. Lafeuille, G. Sammarco, I. Fiordaliso Candalino, M. Suman, R. Piro, Non-targeted authentication of black pepper using a local web platform: development, validation and post-analytical challenges of a combined NIR spectroscopy and LASSO method, Food Control 145 (2023) 109477, https://doi.org/10.1016/j.foodcont.2022.109477.

[26] J.P. Cruz-Tirado, Y. Lima Brasil, A. Freitas Lima, H. Alva Pretel, H. Teixeira Godoy, D. Barbin, R. Siche, Rapid and non-destructive cinnamon authentication by NIR-hyperspectral imaging and classification chemometrics tools, Spectrochim. Acta Mol. Biomol. Spectrosc. 289 (2023) 122226, https://doi.org/10.1016/j.saa.2022.122226.

[27] Z. Jiang, A. Lv, L. Zhong, J. Yang, X. Xu, Y. Li, Y. Liu, Q. Fan, Q. Shao, A. Zhang, Rapid prediction of adulteration content in *Atractylodis rhizoma* based on data and image features fusions from near-infrared spectroscopy and hyperspectral imaging techniques, Foods 12 (2023) 2904, https://doi.org/10.3390/foods12152904.

[28] R. Calvini, A. Ulrici, J.M. Amigo, Growing applications of hyperspectral and multispectral imaging, in: J.M. Amigo (Ed.), Data Handling in Science and Technology – Vol. 32 Hyperspectral Imaging, Elsevier, Amsterdam, 2019, pp. 605–629.

[29] R. Calvini, S. Michelini, V. Pizzamiglio, G. Foca, A. Ulrici, Exploring the potential of NIR hyperspectral imaging for automated quantification of rind amount in grated Parmigiano Reggiano cheese, Food Control 112 (2020) 107111, https://doi.org/10.1016/j.foodcont.2020.107111.

[30] V. Ferrari, R. Calvini, B. Boom, C. Menozzi, A.K. Rangarajan, L. Maistrello, P. Offermans, A. Ulrici, Evaluation of the potential of near infrared hyperspectral imaging for monitoring the invasive brown marmorated stink bug, Chemometr. Intell. Lab. Syst. 234 (2023) 104751, https://doi.org/10.1016/j.chemolab.2023.104751.

[31] R. Vitale, M. Cocchi, A. Biancolillo, C. Ruckebusch, F. Marini, Class modelling by soft independent modelling of class analogy: why, when, how? A tutorial, Anal. Chim. Acta 1270 (2023) 341304, https://doi.org/10.1016/j.aca.2023.341304.

[32] P. Oliveri, Class-modelling in food analytical chemistry: development, sampling, optimisation and validation issues–a tutorial, Anal. Chim. Acta 982 (2017) 9–19, https://doi.org/10.1016/j.aca.2017.05.013.

[33] O.Y. Rodionova, A.V. Titova, A.L. Pomerantsev, Discriminant analysis is an inappropriate method of authentication, Trends Anal. Chem. 78 (2016) 17–22, https://doi.org/10.1016/j.trac.2016.01.010.

[34] M. Barker, W. Rayens, Partial least squares for discrimination, J. Chemom. 17 (2003) 166–173, https://doi.org/10.1002/cem.785.

[35] D. Ballabio, V. Consonni, Classification tools in chemistry. Part 1: linear models. PLS-DA, Anal. Methods 5 (2013) 3790–3798, https://doi.org/10.1039/C3AY40582F.

[36] Z. Małyjurek, D. de Beer, E. Joubert, B. Walczak, Combining class-modelling and discriminant methods for improvement of products authentication, Chemometr. Intell. Lab. Syst. 228 (2022) 104620, https://doi.org/10.1016/j.chemolab.2022.104620.

[37] Z. Małyjurek, D. de Beer, H. van Schoor, J. Colling, E. Joubert, B. Walczak, Class-modelling of overlapping classes. A two-step authentication approach, Anal. Chim. Acta 1191 (2022) 339284, https://doi.org/10.1016/j.aca.2021.339284.

[38] A.L. Pomerantsev, O.Y. Rodionova, Multiclass partial least squares discriminant analysis: taking the right way—a critical tutorial, J. Chemom. 32 (2018) e3030, https://doi.org/10.1002/cem.3030.

[39] R. Calvini, G. Orlandi, G. Foca, A. Ulrici, Development of a classification algorithm for efficient handling of multiple classes in sorting systems based on hyperspectral imaging, J. Spectr. Imaging 7 (2018) a13, https://doi.org/10.1255/jsi.2018.a13.

[40] J. Burger, P. Geladi, Hyperspectral NIR image regression part II: dataset preprocessing diagnostics, J. Chemom. 20 (3–4) (2006) 106–119, https://doi.org/10.1002/cem.986.

[41] A. Ulrici, S. Serranti, C. Ferrari, D. Cesare, G. Foca, G. Bonifazi, Efficient chemometric strategies for PET-PLA discrimination in recycling plants using hyperspectral imaging, Chemometr. Intell. Lab. Syst. 122 (2013) 31–39, https://doi.org/10.1016/j.chemolab.2013.01.001.

[42] D. Ballabio, F. Grisoni, R. Todeschini, Multivariate comparison of classification performance measures, Chemometr. Intell. Lab. Syst. 174 (2018) 33–44, https://doi.org/10.1016/j.chemolab.2017.12.004.

[43] R.W. Kennard, L.A. Stone, Computer aided design of experiments, Technometrics 11 (1969) 137–148, https://doi.org/10.1080/00401706.1969.10490666.

[44] O.Y. Rodionova, P. Oliveri, A.L. Pomerantsev, Rigorous and compliant approaches to one-class classification, Chemometr. Intell. Lab. Syst. 159 (2016) 89–96, https://doi.org/10.1016/j.chemolab.2016.10.002.

[45] S. Wold, Pattern recognition by means of disjoint principal components models, Pattern Recogn. 8 (1976) 127–139, https://doi.org/10.1016/0031-3203(76)90014-5.

[46] A. Biancolillo, R. Bucci, A.L. Magrì, A.D. Magrì, F. Marini, Data-fusion for multiplatform characterization of an Italian craft beer aimed at its authentication, Anal. Chim. Acta 820 (2014) 23–31, https://doi.org/10.1016/j.aca.2014.02.024.

[47] T.J. Bruno, P.D.N. Svoronos, CRC Handbook of Fundamental Spectroscopic Correlation Charts, CRC Press, Boca Raton, 2005 (chapter 2).

[48] J.S. Shenk, J.J. Workman Jr., M.O. Westerhaus, Application of NIR spectroscopy to agricultural products, in: D.A. Burns, E.W. E.W. Ciurczak (Eds.), Handbook of Near-Infrared Analysis, CRC Press, Boca Raton, 2007, pp. 356–365.

[49] J.J. Workman Jr., L. Weyer, Practical Guide to Interpretive Near-Infrared Spectroscopy, CRC Press, Boca Raton, 2007. Chapters 5-6.

[50] J.J. Workman Jr., The Handbook of Organic Compounds, Three-Volume Set: NIR, IR, and UV-Vis Spectra Featuring Polymers and Surfactants –, vol. 2, Academic Press, Cambridge, 2000, pp. 160–165.

[51] I. Oniga, C. Pușcaș, R. Silaghi-Dumitrescu, N.K. Olah, B. Sevastre, R. Marica, I. Marcus, A.C. Sevastre-Berghian, D. Benedec, C.E. Pop, D. Hanganu, *Origanum vulgare ssp. vulgare*: chemical composition and biological studies, Molecules 23 (8) (2018) 2077, https://doi.org/10.3390/molecules23082077.