Title: Quality control of mint species based on UV-VIS and FTIR spectral data supported by chemometric tools

Author: Kamila Kucharska-Ambrożej, Agnieszka Martyna, Joanna Karpińska, Anna Kiełtyka-Dadasiewicz, Aleksandra Kubat-Sikorska

Citation style: Kucharska-Ambrożej Kamila, Martyna Agnieszka, Karpińska Joanna, Kiełtyka-Dadasiewicz Anna, Kubat-Sikorska Aleksandra. (2021). Quality control of mint species based on UV-VIS and FTIR spectral data supported by chemometric tools. "Food Control" (2021), vol. 129, art. no. 108228, s. 1-12. DOI: 10.1016/j.foodcont.2021.108228

Food Control 129 (2021) 108228

Contents lists available at ScienceDirect

Food Control

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

Quality control of mint species based on UV-VIS and FTIR spectral data supported by chemometric tools

Kamila Kucharska-Ambrożej [a,**], Agnieszka Martyna [b,*], Joanna Karpińska [a], Anna Kiełtyka-Dadasiewicz [c], Aleksandra Kubat-Sikorska [d]

[a] Environmental Chemistry Research Group, Institute of Chemistry, University of Bialystok, Ciolkowskiego 1K, 15-245 Bialystok, Poland
[b] Forensic Chemistry Research Group, Institute of Chemistry, Faculty of Science and Technology, University of Silesia in Katowice, Szkolna 9, 40-006 Katowice, Poland
[c] Department of Plant Production Technology and Commodities, University of Life Sciences, 15 Akademicka Str., 20-954 Lublin, Poland
[d] Garden of Cosmetic Plants and Raw Materials, Research and Science Innovation Center, Tarasowa 4/96 Str., 20-819 Lublin, Poland

ARTICLE INFO

Keywords:
Mentha, spectroscopy
Spectrophotometry
Chemometrics
Classification/discrimination

ABSTRACT

Mints are valued for their specific essential oil used in food, pharmaceutical and cosmetic industry. Chemical compounds differing between species, cause changes in medicinal/pharmacological properties, antioxidant activities or smell sensations. For this reason fast procedure for quality control of at least two most popular mint species, peppermint and spearmint, is the issue at hand. UV-VIS spectrophotometry and FTIR-ATR spectroscopy were used for recording the spectral fingerprints of a collection of more than 20 mint varieties harvested in three periods. Two-step chemometric approach for mints quality control involved SIMCA (Soft Independent Modeling of Class Analogy) to filter out species other than peppermint and spearmint. The samples suspected to be either peppermint or spearmint underwent final discrimination using adequate discrimination tools PLS-DA (Partial Least Squares-Discriminant Analysis) or SVM (Support Vector Machines). The model performance ranged between 60 and 80% depending on spectroscopic data used for model training and the harvest season.

1. Introduction

Mints are perennial herbal plants of the *Lamiaceae* L. family cultivated mainly due to their characteristic volatile oil. Taxonomy of plants of the genus *Mentha* L. is a complex problem, because they are characterized by the ability to natural, spontaneous hybridization. Fast classification and differentiation of mint samples is complicated due to the presence of many various chemical compounds. Additionally, they are also subjected to cross-fertilization in order to obtain new cultivars. The division of the genus *Mentha* into species is not unambiguous and their number is estimated at 13 to 18 depending on the classification (Gobert et al., 2002; Saric-Kundalic et al., 2009; Tucker & Naczi, 2007). For example Fejer et al. (2017) has reported that the most popular mint, of the greatest pharmaceutical importance, peppermint (*Mentha* × *piperita* L.), is a natural hybrid, formed by the crossbreeding of *M. aquatica* L. and *M. spicata* L. (Rita & Animesh, 2011) or by interbreeding of *M. silvestris* L., *M. longifolia* L., *M. viridis* L. and *M. aquatica* L. (Stanev & Zheljazkov, 2004; Sabboura et al., 2016). In addition, the chemical composition of oil plants can change during ontogenesis and depends on plant growth conditions (Marotti et al., 1994; Grulová et al., 2016).

Among the most popular mint species, peppermint and spearmint are cultivated in many countries because of their importance to the food, pharmaceutical and cosmetic industries (Tucker & Naczi, 2007) due to their specific flavour and presence of antioxidants (Freie et al., 2012; Saric-Kundalic et al., 2009; Abbas & Nisar, 2020). These two species are attracting considerable attention, leaving the remaining species less used. Mint species differ in chemical composition, which may cause changes in medicinal properties, antioxidant and antimicrobial activities or smell sensations. The characteristic scent of hybrid *Mentha* × *piperita* L. and its varieties is caused by the presence of menthol and its isomers, linalool and linalool acetate (Ludwiczuk et al., 2016; Stanev & Zheljazkov, 2004; Rita & Animesh, 2011). The active constituent found in cultivated spearmint is mainly carvon. (Hawryl et al., 2015; Ludwiczuk et al., 2016; Tucker & Naczi, 2007). The presence of different compounds translates into changes in mint scent and properties, not always easily perceived by consumers. It also necessitates the

* Corresponding author.
** Corresponding author.
E-mail addresses: k.kucharska@uwb.edu.pl (K. Kucharska-Ambrożej), agnieszka.martyna@us.edu.pl (A. Martyna), joasia@uwb.edu.pl (J. Karpińska), anna.kieltyka-dadasiewicz@up.lublin.pl (A. Kiełtyka-Dadasiewicz), a.kubat-sikorska@o2.pl (A. Kubat-Sikorska).

https://doi.org/10.1016/j.foodcont.2021.108228
Received 8 March 2021; Received in revised form 3 May 2021; Accepted 4 May 2021
Available online 7 May 2021

development of a procedure for quick and reliable quality control of certain types of mint to verify the purity of the material before it is used in a production, with a particular emphasis on the two most popular, peppermint and spearmint. Due to the taxonomical complexity of the genus *Mentha* it is important to define criteria of chemical profiles of mint plants for safety and efficacy of raw plant material applied in herbal medicines, food product, teas or diet supplements. The average consumer is not able to verify the authenticity of the plant raw material used in the case of adulteration of the product or distinguish between varieties. Conscious consumers, whose number has been growing steadily in recent years, are eager to know exactly what the bioactive components are included in table spices or herbs. Producers, especially in the food, herbal medicines, cosmetics and perfumery industries, need to know what species or varieties of mint are used due to their varied chemical composition and biological properties.

The quality control of plant material requires the use of techniques that will unequivocally assess the qualitative composition and purity of the examined part of materials. For this purpose morphological (Bezerra et al., 2019; Fialova et al., 2015), genetical (El-Sayed et al., 2013; Sabboura et al., 2016) as well as chemical (Anwar et al., 2017; Hawryl et al., 2015) methods have been used. Each of the mentioned approaches possesses its own advantages and limitations, e.g. the application of the first one gives the proper results only when the fresh material is available. In the case of dried and crushed plants, their identification is very problematic. The genetic or chemical identification typically using chromatographic techniques is time-consuming and expensive (Ercioglu et al., 2018; Jędrzejczyk & Rewers, 2018; Kiełtyka-Dadasiewicz et al., 2017; Ludwiczuk et al., 2016). Although the chromatographic analysis allows obtaining clear information on the quantitative or qualitative composition, final assay is preceded by a multi-stage sample preparation: extraction, purification of the obtained extracts and pre-concentration. These operations are very often burdened with high consumption of organic solvents which increases the total cost of the analysis and at the same time is harmful to the environment and the analyst. An additional disadvantage of chromatographic analysis is its long time. Generally, the use of advanced separation techniques is time consuming and laborious, which makes them difficult to use, especially when fast examination of large numbers of samples is required. So searching for methods that would be faster and more economical is still an open issue. Advances in spectroscopic techniques, as well as relatively lower costs of equipment and use, their green character, make spectroscopic tools a much more attractive alternative to chromatographic methods. The growing popularity of the methods based on chemometric analysis of spectra, especially FTIR spectra, is confirmed by the growing number of published articles, in which authors use them to solve various analytical problems. Each year, several dozen articles devoted to the use of a combination of spectral analysis and chemometric processing are published.

For this purpose Fourier Transform Infrared Spectroscopy (FTIR), Near Infrared Spectroscopy (NIR), Raman Spectroscopy and Nuclear Magnetic Resonance (NMR) techniques are most frequently applied (Kucharska-Ambrożej & Karpińska, 2020). The spectral characteristics of the tested materials (not only mints), and further chemometric processing permit the extraction of useful information that allows for distinguishing of different chemotypes. The most often applied chemometric techniques are principal component analysis (PCA) and hierarchical cluster analysis (HCA) for data exploration (unsupervised methods) and linear discriminant analysis (LDA), k-nearest neighbors (KNN) or orthogonal projection to latent structures (OPLS-DLA) for discrimination/classification (supervised methods) (Kucharska-Ambrożej & Karpińska, 2020). Procedures combining chemometric analysis of IR, NMR or UV-VIS spectra are usually used to assess the authenticity or quality control of the tested products (Petronijevic et al., 2017; Rachman & Muchtardi, 2018; Durazzo et al., 2018; Kiefer et al., 2019; Lucarini et al., 2020). However, careful analysis of available scientific data points that there are only a few articles devoted to application of spectral methods for evaluation, identification or discrimination of mints. Rosch et al. (2002) applied micro-Raman spectroscopy for investigation of different mint taxa: *M. piperita* L. nm. *citrata*, *M. piperita* L. var. *pallescens pallescens*, *M. piperita* L. var. *piperita* f. *piperita, M. spicata* L. ssp. *Crispata and M. spicata* L. ssp. *Spicata*. They found that hierarchical cluster analysis of recorded spectra can be used for distinguishing not only species but also subspecies and varieties. Additionally, they have discovered that Raman spectra of examined species can be applied for determination of stage of plant growth and maturation. The applicability of micro-Raman technique for taxonomic purposes was examined by Petry (Petry et al., 2003). The cluster analysis of second derivatives of the recorded spectra revealed that there are some spectral characteristic features which allowed distinction of individual taxa. The provided analysis has led to the conclusion that monitoring of subtle differences of Raman spectra allows to distinguish between subspecies and varieties as well as to evaluate seasonal variability of the content of essential oils of peppermint. This feature would be very useful for manufacturers of perfumes and aromatherapy products. Raman spectroscopy can be applied as a fast measurement method for the quality control of essential oils in plants alone or as a complement to IR or NIR techniques. FTIR spectroscopy was used to determine the presence of functional groups in the essential oil and to investigate the effect of pollution on the composition of essential oils from *Mentha arvensis* grown in the fields near motorways, railways or expressways (Prakash et al., 2013). The analysis of the recorded spectra allowed not only to identify the main components of the tested material, but also to indicate the differences between them. The subtle spectral differences were used for recognition of examined plants and for studies of influence of environmental factors on chemical composition of mints' essential oils. The usefulness of FTIR and Raman Spectroscopy for the quality control of the composition of essential oils in plants was proved. FTIR technique combined with canonical discriminant analysis (CDA) was applied for distinction of 70 samples of fresh mint (*M. pulegium*) harvested during flowering and growing in various places in Greece (Kanakis et al., 2012). The application of NIR spectroscopy for characterization of chemical composition and taste properties of mint tea (*Mentha haplocalyx* Brig.) was proposed by Dong (Dong et al., 2014). PCA and HCA of the spectral data allowed to distinguish mint samples from different geographical origins. NMR in combination with PCA was used for authentication of 21 mint samples cultivated in different geographical localizations (Manolache et al., 2018). The 1H NMR technique allowed for fast and direct profiling of the examined samples without prior sample processing.

Among a number of spectroscopic techniques, UV-VIS spectrophotometry is the least frequently used analytical technique for identifying plant material (Kucharska-Ambrożej & Karpińska, 2020). Complex rotary-vibrational-electronic nature of UV-VIS spectra causes difficulties in the visual recognition of individual bands and their use for taxonomic purposes. Due to this feature UV-VIS spectra are rarely used to identify and assess the authenticity of plant material (Kucharska-Ambrożej & Karpińska, 2020). Despite the fact that the UV-VIS technique is fast and available in almost every laboratory, which makes it a potential alternative for time-consuming GC-MS, there is viable shortage in its application for reliable examination of plant materials for classification purposes. To the best of our knowledge, the UV-VIS spectra of hexane extracts obtained from dried mint were not used to distinguish or identify individual species of mints (Kucharska-Ambrożej & Karpińska, 2020). There is also little research on the identification of dried mints using the FTIR-ATR technique. In order to fill the perceived gap, we made an attempt to evaluate the usefulness of the UV-VIS spectra of hexane extracts and FTIR-ATR spectra of dry ground mints materials for quality control of mint species. Their most important limitation is the provision of less direct information for identification than GC-MS, in which chromatographic profiles contain different peaks and thus provide clear differences between species. For extracting the proper information that will be useful for identification studies, chemometric tools are essential. In our research Soft Independent Modeling of Class

Analogy (SIMCA) was combined with partial least squares discriminant analysis (PLS-DA) or support vector machines (SVM). The research aimed at developing a procedure for quality control of mint samples described by UV-VIS or FTIR-ATR spectra, understood as creation of the model for assigning future mint samples of unknown species into two most popular species, i.e. peppermint, spearmint, or none of them.

2. Materials and methods

2.1. Plant material

The mint samples were obtained from the Garden of Cosmetic Plant and Raw Material collection, Research and Science Innovation Center located in Wola Zadybska (51° 44′49 "N 21° 50′38″ E) Lubelskie Region in Poland. The plants were grown on lessive soil which was slightly acidic (pH_{KCl} 6.1). Herbs were acquired twice in 2017 (June 19th – dataset I and August 12th, 2017 – dataset II) and once in 2018 (June 23rd - dataset III). Plants were harvested in the phase of forming flower buds, i.e. 51 BBCH scale. Each mint species was sampled in three possibly different locations of filed, marked with the letters A, B and C, such that: point A was on the extreme edge of the field from the north, point B in the central part of the field, surrounded on each side by other plants of the same variety, while the C was located on the edge of the field from the south side, i.e. the best insolated (Figure SM1). Such material acquisition allows to collect potentially the most chemically diverse material growing under the same agrotechnical conditions (on one plantation). Regrowth at the second harvest was cut precisely from the same plants. Each time, the whole shoots were cut at a height 10 cm above the ground. Freshly harvested plants were immediately dried (each sample separately) in a laboratory dryer with forced air circulation at 30-32 °C, and then the stalks were manually separated so that the leaves themselves were left to the test.

Each subset (A, B, C) contained a variety of mint plants, clustered further in three groups referring to the species: most numerous peppermint, spearmint and the remaining (denoted as mixed). Peppermint species contained *M.* × *piperita* 'Grapefruit', *M.* × *piperita* 'Granada', *M.* × *piperita* 'Almira', *M.* × *piperita* 'Chocolate', *M.* × *piperita* 'Swiss', *M.* × *piperita* 'Multimentha', *M.* × *piperita* 'Variegata', *M.* × *piperita* 'Citaro', *M.* × *piperita* 'Orangemint' varieties. *M. spicata* L., *M. spicata* 'Moroccan', *M. spicata* 'Crispa', *M. spicata* 'Cubana' varieties were considered within spearmint species. The remaining varieties *M. arvensis* L., *M. gracilis* 'Ginger', *M. rotundifolia, M. suaveolens* 'Variegata', *M.* 'Berries&Cream', *M. pulegium* 'Romana ', *M.* × *villosa* Huds., *M. crispata, M. rotundifolia* 'Apple Mint', *M. arvensis* 'Banana' were regarded in the mixed species group. Originally each species group within each subset contained all available varieties as listed above, each represented by a single sample. Thus there were more than 20 mint samples in each subset. Table SM1 provides the details of the group sizes (including replicates) for datasets (I, II, III) and subsets within them (A, B, and C). Differing group sizes arise from further outliers removal described in section 2.5. Thus not all varieties are represented in each subset.

2.2. Extraction procedure

1 g of fine ground dry plant material was weighed into a 16 ml vial. Next 12 ml of hexane was added and stirred vigorously for 30 min. Then the extract was removed to another vial and extraction was repeated with a new portion of solvent. The procedure was repeated three times. Subsequently, the extracts were combined and filtered through a paper filter. The resulting filtrate was concentrated to a volume of approximately 1.5 ml by evaporation of the solvent and then quantitatively transferred to a 2 ml vial.

2.3. Apparatus

The UV-VIS spectra of n-hexane extracts of plant material were recorded with n-hexane as a blank and using a Hitachi U-2800 A spectrophotometer (Hitachi High-Technologies Europe GmbH (Mannheim O_ce), Mannheim, Germany). The following working settings of the device were used: scan speed 1200 nm min^{-1} and spectral bandwidth 1.5 nm. The examined extracts were diluted with n-hexane in order to obtain values of absorbance <2 a. u.

Spectrometer Nicolet 6700 equipped with a diamond crystal was used for recording of FTIR spectra of dried plant material. Spectra of powdered leaves of plant were recorded in the range 645–4000 cm^{-1} after application onto the diamond crystal of the spectrometer.

The number of replicate spectra recorded for each sample is provided in Table SM1.

2.4. Software

Free software environment R was adopted for calculations (Core Team 2018). We have worked with home-written scripts that employed package rrcovHD (Todorov, 2016) for SIMCA, pls (Mevik, 2018) for PLS-DA and e1071 (Meyer, 2015, pp. 6–7) for SVM.

2.5. Chemometric analysis

Prior to the analysis UV-VIS and FTIR-ATR spectra were subjected to appropriate preprocessing. Its aim was to correct for any signal distortions, noise removal, or reduction of the "size effect" of the measured samples.

The spectral range of the UV-VIS spectra was truncated to 240–350 nm that covers the most informative region, completely acquired for all studied samples. The outlying samples were removed based on the inspection of the results of the robust principal component analysis (Hubert et al., 2004). Final number of samples that belonged to each dataset, species group or subset is given in Table SM1.

The signals demonstrated an increase of the noise with the signal magnitude (heteroscedastic noise). Square root transform has largely removed this unfavorable attribute of the noise which became homoscedastic afterwards. Then the first derivative was computed for each signal using Savitzky-Golay smoothing with third order polynomial fitted to smooth the signal in a window of five points. The transformed signals within each training or test set underwent probabilistic quotient normalisation (PQN) (Dieterle et al., 2006) for removing the differences in the absorbance arising from fluctuations of the samples concentrations, which are irrelevant for establishing their species. PQN scales each signal in the dataset using an individual scaling factor. This factor is computed as the most probable quotient, i.e. median, between the signals and the reference. The reference was taken to be the median signal in the dataset.

FTIR-ATR spectra were recorded in the range 645–4000 cm^{-1}. There were no outliers detected and Table SM1 provides the details of the groups sizes. Logarithm was taken to transform the heteroscedastic noise into homoscedastic. For this purpose 0 values (or slightly below 0) were substituted by 0.01. The transformed signals within each training or test set were normalised with PQN.

The identification of the mint species was conducted in two subsequent steps. The full rationale for using the two-stage procedure can be found in section 3.

Firstly, the samples were examined for whether they belonged to a common group of spearmint and peppermint, or neither. SIMCA modeling (Wold, 1976) was employed for this purpose, which uses principal component analysis (PCA) for modeling the space for each considered group. Here only a single group composed of spearmint and peppermint species together underwent modeling. SIMCA described each tested sample by the two characteristics, namely score (SD) and orthogonal distances (OD). Score distances are the Mahalanobis

distances of the observation in the PCA space from its center. Orthogonal distances are computed as Euclidean distances between the observation and its projection in the original space. Both parameters are actively used for deciding about the samples potential membership to the considered groups. The samples are assigned to the modeled groups if their SD and OD are below the established thresholds (Wold, 1976). If they go out of the cutoffs, it is concluded that they come from other groups, also these, which have not been modeled.

SIMCA modeling was conducted for each dataset (I-III) of samples described by either UV-VIS or FTIR-ATR spectra or both combined together. The concatenation was carried out after the preprocessing steps had been performed individually for each spectra type. To equalize the information potential of both spectroscopic techniques, preprocessed UV-VIS and FTIR-ATR spectra were divided by the square root of the variance averaged over all the variables measured by each technique respectively. In each case, SIMCA model was trained using one of the three available subsets of data collected within each dataset I-III. Then the model was applied to the next subset. The remaining subset was left for training of the discriminant variant of the partial least squares (PLS-DA) (Næs et al., 2002) method or support vector machines (SVM) (Belousov et al., 2002) for discrimination of the samples between spearmint and peppermint categories. Only the samples that were assigned in SIMCA to the common group of spearmint and peppermint underwent discrimination between these two groups. PLS-DA is a discriminant analogue of classical PLS regression, where the response variable is categorical, such as group membership. The task of PLS-DA is to predict the group membership based on a (usually large) set of predictors, e.g. analytical signals. The basis for modeling is to develop a limited number of new latent variables, which are linear combinations of the original predictors. They are generated to capture the maximum variance in the sets of responses and predictors, and the maximum covariance between them. The latter feature ensures that the performance of PLS in predicting the response is enhanced as the largest importance is assigned to the predictors with the highest ability to predict the response. SVM is a technique widely applied for regression and classification purposes. The concept of SVM, however, is to generate a hyperplane that best separates the investigated classes. The orientation of the hyperplane is set to maximize the margin, which is the greatest distance between the data of both classes. The position of the hyperplane is thus mostly governed by the closest points called support vectors.

There were six combinations of the subsets used for training and testing SIMCA and PLS-DA or SVM models:

- configuration ABC: A subset was used for testing the models, B subset was used for training SIMCA and C subset was used for training PLS-DA or SVM;
- configuration ACB: A subset was used for testing the models, C subset was used for training SIMCA and B subset was used for training PLS-DA or SVM;
- configuration BAC: B subset was used for testing the models, A subset was used for training SIMCA and C subset was used for training PLS-DA or SVM;
- configuration BCA: B subset was used for testing the models, C subset was used for training SIMCA and A subset was used for training PLS-DA or SVM;
- configuration CAB: C subset was used for testing the models, A subset was used for training SIMCA and B subset was used for training PLS-DA or SVM;
- configuration CBA: C subset was used for testing the models, B subset was used for training SIMCA and A subset was used for training PLS-DA or SVM.

The SIMCA models complexity (number of relevant principal components sufficient to characterize each group) was determined from the scree plots. The diagrams illustrated the amount of information each principal component carries. The principal components that were preserved were indicated by the "elbow point". Usually two or three components were preserved.

PLS-DA complexity (number of PLS components which minimize the misclassification error) was established for each PLS-DA model separately (i.e. for each datasets and each training/test sets configurations) using the leave-one sample-out cross validation protocol.

SVM used linear kernels for modeling.

3. Results and discussion

The obtained UV-VIS spectra (Figure SM2) are not specific and do not allow direct taxonomic classification of examined plants material. They all possess strong and intense bands in the range 190–210 nm which could be probably assigned to electron transfers $n\rightarrow\sigma^*$. Other much weaker bands at 220, 240, 270 or 290 nm could be results of $\pi\rightarrow\pi^*$ as well as $n\rightarrow\pi^*$ electron transfers. The location of the bands is not specific as it depends on the presence of auxochromic groups, coupling effects and interactions between the components and the solvent. All observed bands pointed the presence of such chromophores as systems of conjugate bonds, aromatic rings, carboxylic or carbonyl groups as well as interaction between components and solvent. The results provided by chromatographic analysis showed that the main components of extracts are mainly terpenoids: linalool, carvone, linalool acetate, piperitenone oxide, pulegone, menthone and menthol and others in different proportions depending on studied species.

The selected FTIR spectra of examined mints are presented in Figure SM3. Similarly to the UV-VIS spectra of the extracts, also the FTIR spectra of the tested mints are almost identical. The observed differences in the intensities of the bands and their position are not specific and do not allow the direct identification of the tested plants. The characteristic bands are a result of a variety of vibrations of bonds connected with the presence of functional groups of compounds present in surface of studied plant material but do not allow identification of parent compounds. The wavenumbers of recorded FTIR spectra and assigned to them functional groups or compounds are gathered in Table 1.

Even careful analysis shows that it is very difficult or impossible to distinguish individual mint species from each other on the basis of raw UV-VIS or FTIR spectra. Therefore, it was decided to check whether the chemometric processing of the spectral data would expose features which enable the correct classification of the tested objects.

3.1. Description of the data

Fig. 1(a) presents UV-VIS spectra transformed with the square root and truncated to 240–350 nm while Fig. 1(e) log-transformed FTIR-ATR spectra with the colour-coded datasets (i.e., mint plants harvested at different times). These spectra demonstrate the urge for normalisation to remove the size effect arising from various amounts of the measured samples. Thus Fig. 1(b) and (f) show the spectra which underwent PQN (section 2.5). PQN as a standardization strategy has proved to be useful in increasing the comparability of spectra and uncovering the features that differ between peppermint, spearmint and other mint species. This observation is particularly apparent in Fig. 1(c) and (g) which portray mean centred spectra (i.e., mean vector for variables was subtracted from each spectrum). Mean centred spectra, marked in colour depending on the dataset, clearly show that the largest differences between the spectra are mostly related to different datasets, I, II, or III. This makes sense, since datasets differ by the time (season) the mint plants were harvested. This, in turn, means different irrigation and weather conditions, sun exposure, and variability of other factors that become relevant for characterising the mint plants by their spectral fingerprints. For this reason each dataset should be examined individually. Principal component analysis (PCA) results plotted for first three components in Fig. 1(d) and (h) evidently confirm the previous findings. Especially for FTIR-ATR the datasets are quite well separated. For UV-VIS the separation is not that much manifested. However, it is still present as dataset

Table 1
The spectral data of recorded FTIR spectra of studied mints (Larkin P.J, 2018).

Wavenumber [cm^{-1}]	Type of vibration	Functional groups or compounds
670	deformation (γ) not in plain CAr-H	benzene-H
	deformation (γ) O–H not in plain	liquid alcohols or phenols
	deformation (γ) = C-A not in plain	cykloallenes polienes
1071	stretching (ν) C–O	saturated esters alcohols and phenols
	skeletal (ν) C–C	cycloalkanes
	stretching (ν) C–O–C	acid anhydrides
1247	stretching (ν) C–O	saturated esters - acetates dimers alcohols and phenols
	stretching (γ) C–C	alkanes
	deformation (γ) not in plain = C–H	lactones
1413	stretching (ν) –COO- deformation (δ) C–H	salts of symmetrical acids ketones lactones
1597	stretching (ν) –COO- stretching (ν) C═O	aromatic acids salts β-diketones (enolic forms)
1734	stretching (ν) C═O	formates α,β-unsaturated acids esters aromatic acids esters α-ketoesters β-ketoestry aromatic acids chlorides non-cyclic α, β-unsaturated anhydrides saturated aldehydes γ-diketones α-halogen acids non-cyclic α-halogen ketones
2850 i 2915	stretching (ν) C–H	-CH2-
3300	stretching (ν) O–H	free –OH group associated –OH group

I seems to be totally different since its cloud is practically disjoint in PCA from datasets II and III. For this reason the studies were conducted within each dataset.

Figs. 2 and 3 portray the three datasets individually in the spaces of the first three principal components. The colours refer to various mint species (p stands for peppermint, s for spearmint and m refers to other species) and point characters differ between the A, B or C subsets within each dataset. The most striking observation to emerge from the inspection of the plots is that the mint species are all usually mixed up and poorly separated. Only datasets I and III demonstrate some separation of the mixed species, but confuse peppermint and spearmint very often (see e.g. Fig. 3(a)). This may lead to unsatisfactory performance of the chemometric models in predicting the species. As Fig. 2 portrays for UV-VIS spectra, the first principal component does not capture the differences between species. Only the second explains this tiny part of variance (few %) that is associated directly to them. Score plots in Fig. 2 demonstrate that there are some samples from the species other than peppermint and spearmint (group m marked with black points), that may be suspected of being outliers. However, as belonging to mixed species group (m) they do not take part in any of the modeling (training), and therefore are used only for predictions. For this reason their outlyingness does not have an impact on the model development, but may affect (mostly positively due to their dissimilarity to p or s species) the overall performance. Their residuals, however, do not provide irrefutable evidence to remove them, especially that they represent the class of samples with a substantial variety which should not be ignored. Nevertheless, the plots yield strong evidence that the subsets A, B and C can be studied together since they do not form any clusters within the PCA spaces. This result has further strengthened our confidence in using different subsets as training and test sets for modeling and controlling the models performance (listed in section 2.5).

3.2. Species identification

Based on the above findings from data exploration using PCA, our research went in the direction described below. Datasets are composed of samples originating from two most common and well-defined species, i.e. peppermint (p) and spearmint (s), and some other mixed species (m). In practice the task of the chemometric modeling would be to filter out the species other than peppermint and spearmint (m) and conclude if the remaining are peppermint or spearmint. This is feasible with classification techniques (e.g. SIMCA), which establish whether an unknown sample is a member of well-defined classes (s, p) or none of them (m). They model only the known classes and if the sample does not fit them it is classified as coming from some other unmodeled class. However, for successful solutions, the task entails that the modeled classes should reveal some differences not only in relation to unmodeled data, but also between themselves. Regrettably, we were surprised to find that PCA, which is the basis for SIMCA modeling, is very often incapable of capturing the part of variance that distinguishes simultaneously between s and p and other species (m). Some expectations were laid only in datasets I and III, for which SIMCA does reveal at least partially separated clusters of mixed species, however, usually mixes s and p (see e.g. Fig. 3(a)). Thus although using SIMCA approach as a solution to classification of mint samples seems interesting, it suffers from PCA being an inadequate tool for finding the features that differentiate the most interesting species. This apparent lack of separation of the species can be attributed to the fact that the variability of data between species is not the primary one. In this sense the principal components focus on other sources of variance whilst the variance related to species is distributed between many components and hardly extractable. The evidence from this study points towards the idea of proceeding in two steps. Firstly, SIMCA as a classification tool will be used for filtering out the mixed species and leaving only these suspected to be peppermint or spearmint. In the second step these samples will undergo targeted discrimination analysis using partial least squares in its discriminant variant (PLS-DA) or support vector machines (SVM). The concept behind engaging the PLS-DA or SVM is to find new latent variables, which, contrary to principal components, best separate both species.

Fig. 4 illustrates the fraction of samples correctly assigned to each of the categories, i.e. into peppermint, spearmint or mixed, averaged across all categories, which is known as sensitivity of the classifier (see Supplementary Materials for more information) (Ballabio et al., 2018). Broadly speaking, Fig. 4 evidently points that the UV-VIS spectra of mint plants usually carry more information associated with their species than FTIR-ATR. The overall correct classification rate of the samples into peppermint, spearmint or neither using their UV-VIS spectra ranks at the level of 70–80% for dataset I and decreases to 50–60% for datasets II and III. This remains in compliance with the pervious findings illustrated in Fig. 2 stating that data clouds for different species are best isolated in dataset I. At the same time, FTIR-ATR spectra yield ca. 50–70% of correct classifications for dataset I, 30–70% for dataset II. However, this rate unexpectedly rises to more than 60–70% for dataset III, which is even superior to what UV-VIS spectra achieve for this dataset. This is because the clouds for the species are better resolved for dataset III in FTIR-ATR than UV-VIS, as Figs. 2 and 3 portray. Establishing the species of mint plants seems the toughest for dataset II, in which the signals for mixed species, peppermint or spearmint share very similar features.

Our study provides an evidence that two approaches based on PLS-DA and SVM for discrimination between spearmint and peppermint give comparable results. The rationale was to use two methods which, while focusing on linear differences, are based on different assumptions and principles. As evidenced from the research, neither one is better than the other. This confirmed that the results are valid since we arrived at the same conclusions using two distinct methods. The inability to decide which technique is better for recognizing the mint species led us

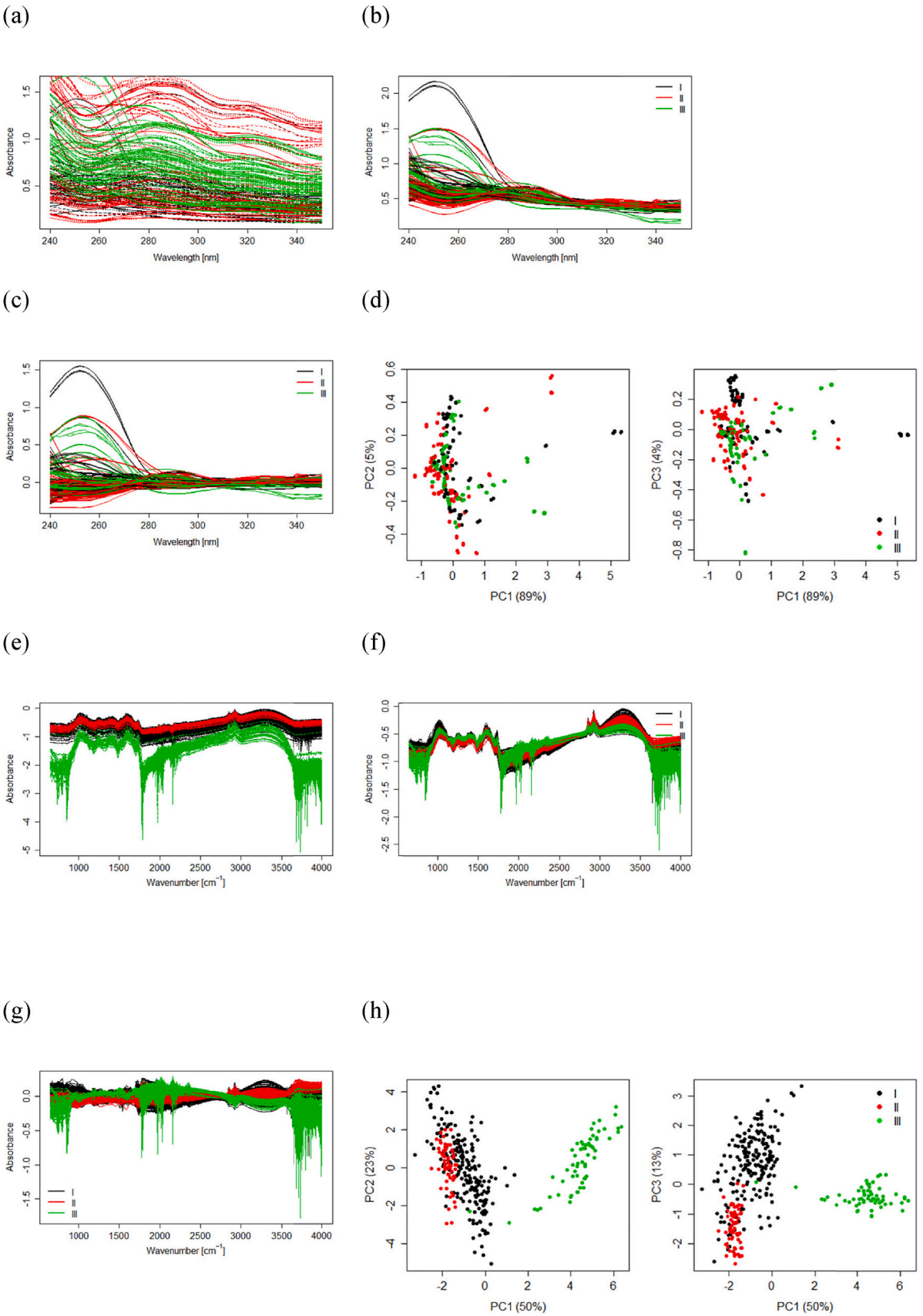

Fig. 1. (a) Square root-transformed UV-VIS spectra, (b) spectra normalised with PQN, (c) mean centred normalised UV-VIS spectra, (d) PCA of the normalised UV-VIS spectra, (e) log-transformed FTIR-ATR spectra, (f) spectra normalised with PQN, (g) mean centred normalised FTIR-ATR spectra, (h) PCA of the normalised FTIR-ATR spectra. Colours refer to different datasets (mint plants harvested at different times). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

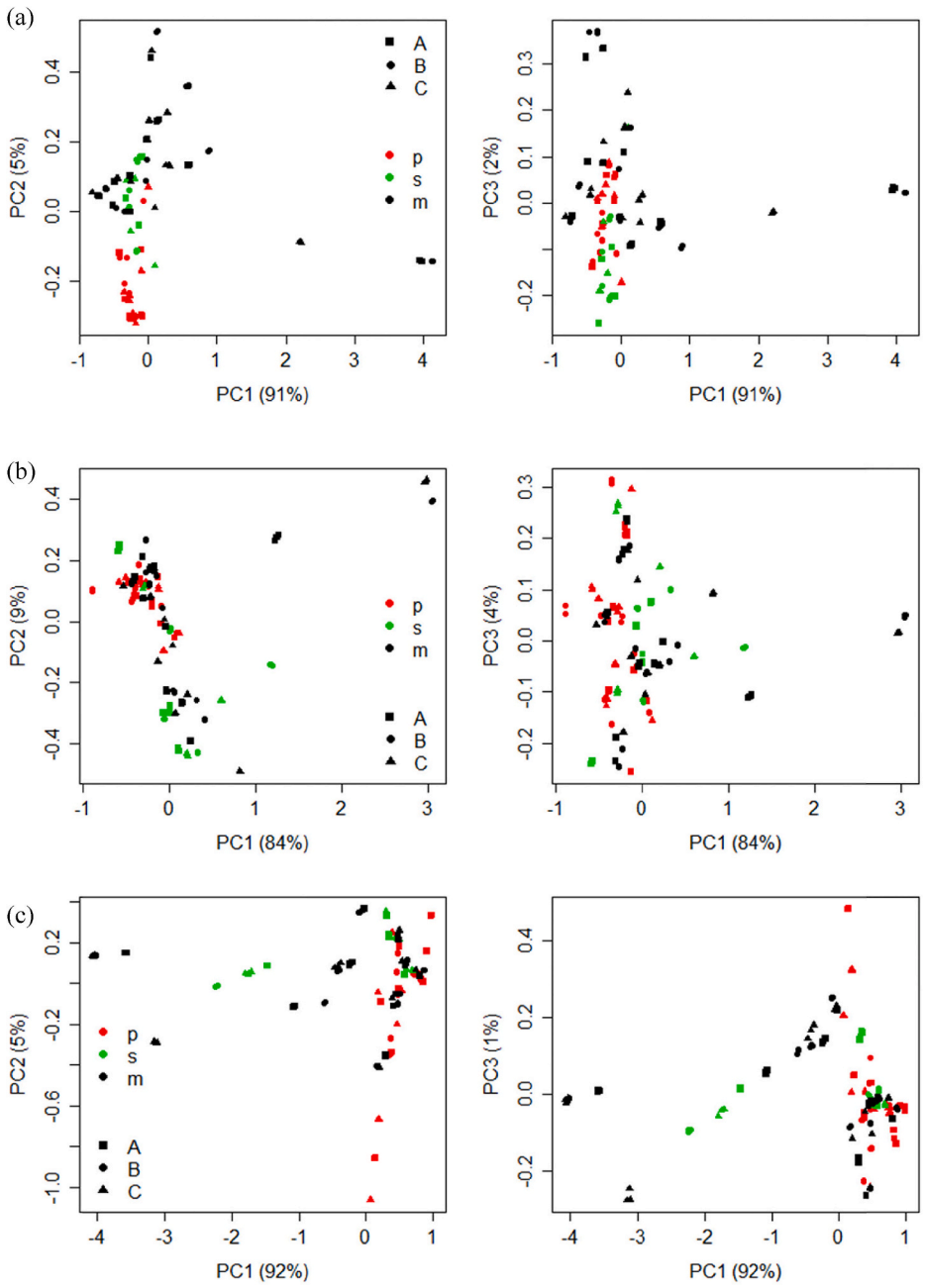

Fig. 2. UV-VIS spectra in the PCA space constituting (a) dataset I, (b) dataset II and (c) dataset III. The colours refer to various mint species (p stands for peppermint, s for spearmint and m refers to other species) and point characters differ between the A, B or C subsets within each dataset. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

to combine the information potential of both. The performance of the developed models applied to combined UV-VIS and FTIR-ATR data did not improve. The gain on efficiency resulting from the correct classification of samples that were misclassified using only a single technique was balanced by the erroneous assignments for these identified correctly previously. The results for the combined data are more like those for UV-VIS than for FTIR-ATR, suggesting that the features describing the species that are measured with UV-VIS are more manifested and rule classification. It is also very likely that the information that is relevant for discrimination between species is much more difficult to extract using a limited number of latent variables for fused data than for individual techniques. Worse performance of the developed models observed for the fused data of dataset III can be attributed also to some technical aspects of the proposed procedure. The number of significant PCA and PLS components for the models developed for the datasets before and after the data fusion differed and therefore spanned different quality and amount of information. Also, for the fused data, some outliers were detected which were not considered as such when single spectral methods were involved. For this reason the correct classification rates may be occasionally difficult to compare.

It is plausible that a number of issues could have influenced the results obtained. One of them is the need for establishing the complexity of PLS-DA, which is always a demanding activity for constructing reliable and not overfitted or underfitted models. The second issue is the need to tune the various parameters governing the performance of SVM.

With a few exceptions for dataset II, no significant differences were observed between the rates yielded for different configurations of test and training sets (listed in section 2.5). Despite obvious and natural

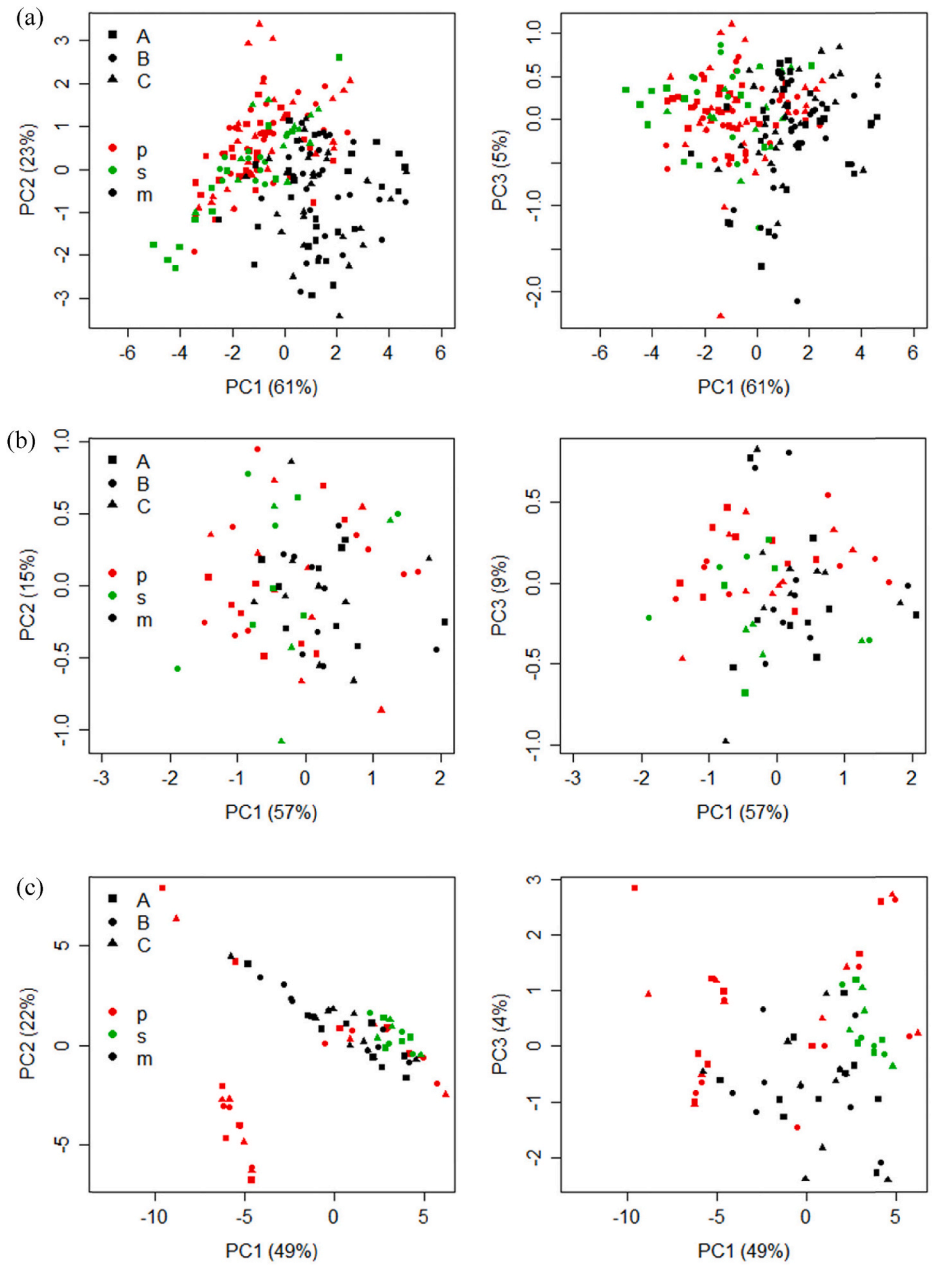

Fig. 3. FTIR-ATR spectra in the PCA space constituting (a) dataset I, (b) dataset II and (c) dataset III. The colours refer to various mint species (p stands for peppermint, s for spearmint and m refers to other species) and point characters differ between the A, B or C subsets within each dataset. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

variations between subsets A, B and C, the developed two-stage classification approach seems robust enough as it is virtually unaffected by the location of the crops of mint species harvested within the same season. However, it cannot be ruled out that some misclassifications may be attributed to the discrepancies arising from various locations of subsets. If some variations between subsets exist, then the models trained on one subset would have worse performance for a test set, which is obviously slightly different. The deliberately overfitted model trained and tested on the same set would help in recognizing if this variation negatively affects the performance of the developed procedure. Since the performance of such models does not reveal any remarkable improvement or deterioration in reference to the examined models employing all subsets to avoid overfitting, it becomes indisputable that the source of classification errors lies in the limited differences between the signals for peppermint, spearmint and mixed species rather than variations between subsets.

Further investigations suggest that insufficient differences between mixed and peppermint or spearmint species are mainly responsible for the final misclassification rate. These findings are also in line with the pictures in Figs. 2 and 3. The more remarkable the separation of mixed species from the peppermint and spearmint, the less the model performance is affected by the differences between subsets and vice versa. Fig. 5 briefly illustrates that the first step involving SIMCA may be blamed for the majority of classification errors. The diagrams show confusion tables, where the entries in the columns refer to the predictions of the membership of the samples originating from one specific species. Each single diagram in Fig. 5 refers to the classification within one dataset with training and test sets established according to one of the configurations listed in section 2.5. Each column sums up to 100% to demonstrate the fraction of the samples of a specific species that are correctly (diagonal) or mistakenly (off-diagonal) assigned. When the model behaves poorly either samples from mixed species are confused

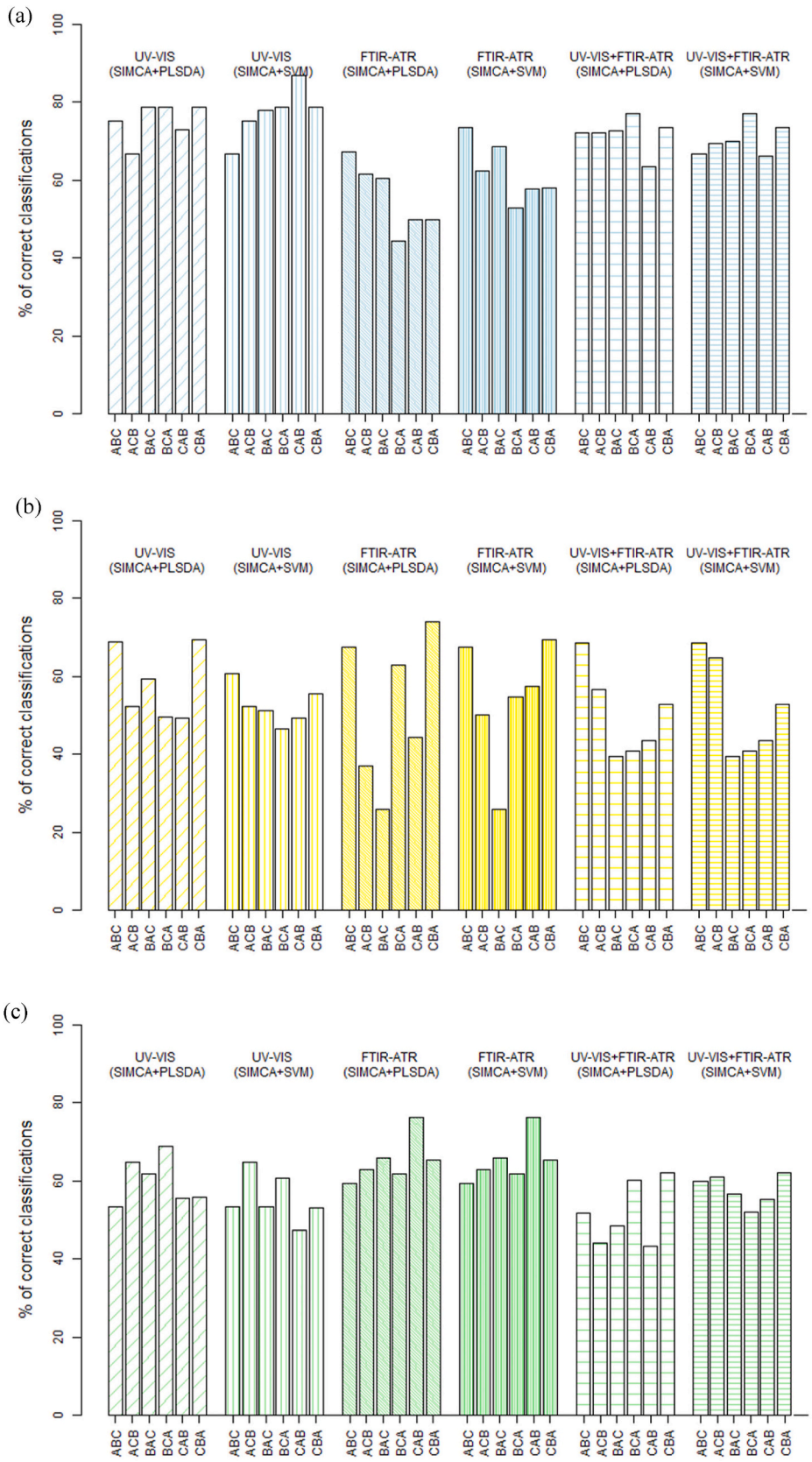

Fig. 4. Correct classification rates for (a) dataset I, (b) dataset II, (c) dataset III.

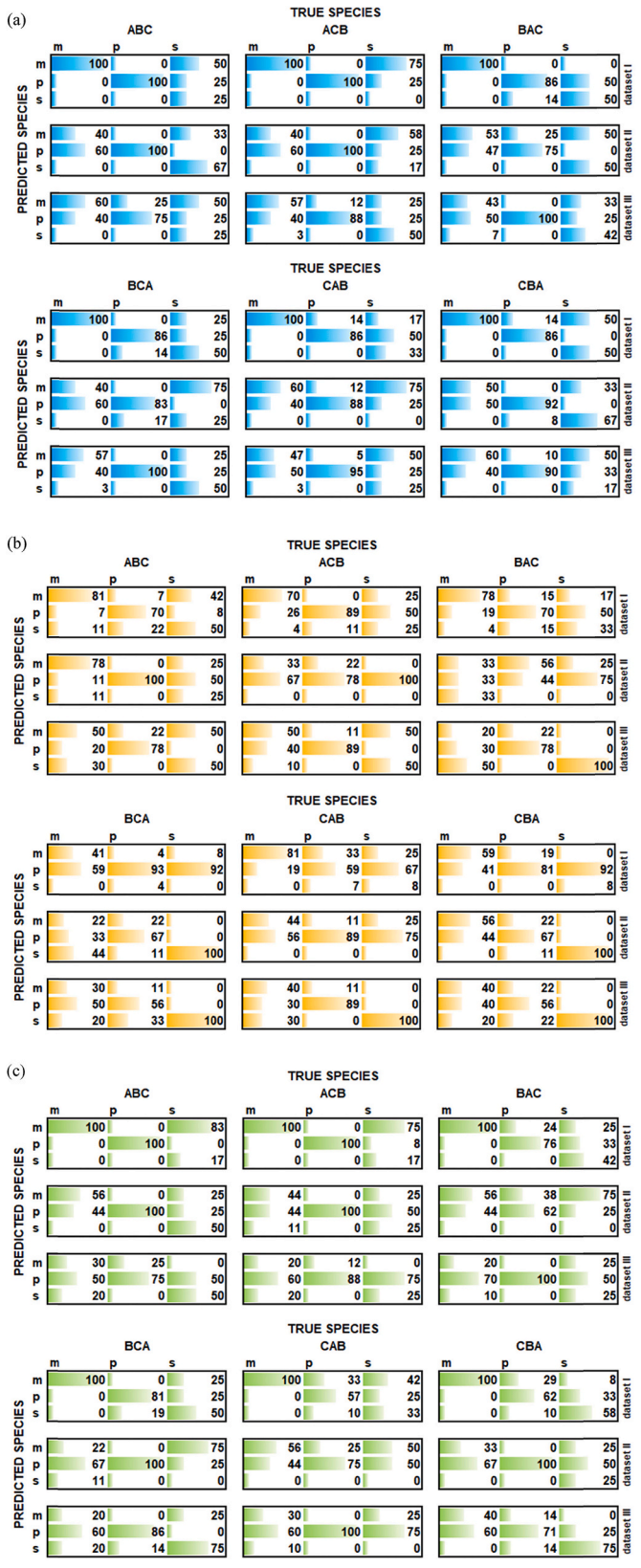

Fig. 5. The confusion tables for (a) SIMCA + PLS-DA model for UV-VIS spectra, (b) SIMCA + PLS-DA model for FTIR-ATR spectra and (c) SIMCA + PLS-DA model for combined UV-VIS and FTIR-ATR spectra (description of the symbols is provided in the text). The length of the bars illustrates the fraction of the samples from each species that are assigned as m, p or s.

with peppermint and spearmint combined group or true peppermint or spearmint samples are categorized as mixed species. Unfortunately, any misclassifications emerging at this stage are directly translated into decrease of efficiency of the overall model. Luckily, the performance of the PLS-DA or SVM for discrimination between spearmint and peppermint is rather appreciated with a few errors in erroneous assignment of spearmint samples as peppermint. In this way, the second step does not significantly deteriorate the overall performance.

The above findings are evidently confirmed by other classification performance metrics introduced and listed as Supplementary Materials in Tables SM2-SM7. When dealing with multiclass classification, i.e. with more than two classes, each classification task can be considered as a set of binary classifications (one for each class) in which samples are either classified into positive (particular considered class) or negative class (consisting of all other classes) (Ballabio et al., 2018). The performance of classification is then summarized with appropriate metrics for each class and globally for all of them on macro- and micro-levels (see Supplementary Materials).

As can be easily noted from the Tables SM2-SM7 listing the performance metrics for all developed models, the sensitivity and the corresponding false negative rate are the most disappointing. Low sensitivity and high false negative rate evidently confirm that the samples of particular species are easily misclassified and pretend to belong to other classes. This is mostly in regard to the spearmint or mixed species which are often mistaken for peppermint. The frequently occurring low precision also highlights these findings and indicates that the predicted classes are not entirely pure, i.e. contain incorrectly assigned samples truly belonging to other species. However, very poor sensitivity for spearmint class is occasionally compensated by very high precision pointing that only true spearmint samples (but not all of them) are predicted to spearmint class. Limited precision and sensitivity for the mixed species class suggests that mixed species are either misclassified to be the members of peppermint and spearmint or peppermint and spearmint are regarded as mixed species. The same can be concluded for peppermint samples. Specificity, accuracy, error rate and false positive rate count the items assigned to the negative class in binary cases (see Supplementary Materials). For multiclass cases, they tend to overestimate the performance as assigning a sample into a negative class does not always mean that correct class was predicted for this sample. It only means that the class was any other than positive in a particular binary case. For this reason, their levels will not be described further to summarize the effectiveness of the identification of mint species.

4. Conclusions

We have outlined studies on the applicability of UV-VIS and FTIR-ATR spectral analysis to distinguish between different mint species, with an emphasis on spearmint and peppermint, as the most popular and widely used in food, pharmaceutical, or cosmetic industries. The problems of discrimination of mint species, especially peppermint, may be caused by their ability to create natural hybrids. Specific requirements for mint products necessitate the development of mint samples quality control strategy understood as the identification of mint species into peppermint, spearmint and other less relevant species before they are used in the production. For this purpose UV-VIS spectra of hexane extracts and FTIR-ATR spectra of finely ground dried plants material were recorded. The concept of our procedure for quality control of mint samples assumed using firstly SIMCA method for cleansing the data from any other species than peppermint and spearmint. Observable similarity between spearmint and peppermint species forced to use more adequate tools for their effective differentiation. This step was accomplished by PLS-DA and SVM techniques, which substantially reinforced the performance of the overall model and contributed to the reduction of the misclassification rate into these two species in relation to SIMCA only.

Our work has led us to conclude that UV-VIS spectrophotometry of mint plants seems to deliver much more adequate information related to

their species and thus chemometric models based on its outcomes are more reliable and appreciated than using FTIR-ATR. We also provided further evidence that the use of FTIR-ATR does not contribute to the efficiency of the UV-VIS only models. Our comprehensive results prove that the quality control is equally successful using either PLS-DA or SVM as discrimination techniques.

We have succeeded in identifying the bottleneck of our quality control procedure, which may be blamed for the majority of final classification errors. The most important limitation of our procedure relates to the filtering step using SIMCA, in which mixed species are very frequently confused with peppermint or spearmint and peppermint or spearmint are mistaken for mixed species. The second step, however, accomplished with PLS-DA or SVM, scores much higher and provides efficient discrimination between peppermint and spearmint. The upshot of this is the possibility to identify the peppermint and spearmint species with high accuracy if the sample under investigation is a priori known to be one of the two and not the other species. Thus skipping the stage of verifying whether or not the sample belongs to the mixed species category translates into much better results since the first filtration step is practically entirely responsible for errors occurring in the second step.

Another shortfall of the developed procedure is related to the fact that in order to achieve acceptable models performance, it has to be trained on the samples harvested within the same season, wherein location of the crops is irrelevant. It might, however, generate some inconvenience in practice, when unknown mint sample will undergo quality control studies using our procedure.

In our view these results constitute still not perfect but an encouraging initial step toward routine identification of the mint species based on their spectral fingerprints, supported by chemometric tools, acting as a likely alternative for time and money consuming GC-MS.

CRediT authorship contribution statement

Kamila Kucharska-Ambrożej: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing, Visualization. **Agnieszka Martyna:** Conceptualization, Methodology, Software, Validation, Formal analysis, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Joanna Karpińska:** Writing – original draft, Writing – review & editing, Project administration, Management and coordination responsibility for the research activity planning and execution. **Anna Kiełtyka-Dadasiewicz:** Conceptualization, Resources, Writing – original draft. **Aleksandra Kubat-Sikorska:** Resources, Writing – review & editing.

Declaration of competing interest

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodcont.2021.108228.

References

Abbas, S. R., & Nisar, M. (2020). Herbal significance of horse mint (*Mentha longifolia*). *J. Biotech Sci. V-, 8*, 138–143.

Anwar, F., Alkharfy, K. M., Najeeb-ur-Rehman, Adam, E. H. K., & Gilani, A. H. (2017). Chemo-geographical variations in the botanical attributes of *Mentha longfolia* (L.) essential oils from Saudi arabia. *International journal of Pharmacolog, 13*, 408–424.

Ballabio, D., Grisoni, F., & Todeschini, R. (2018). Multivariate comparison of classification performance measures. *Chemometrics and Intelligent Laboratory Systems, 174*, 33–44.

Belousov, A. I., Verzakov, S. A., & von Frese, J. (2002). Applicational aspects of support vector machines. *Journal of Chemometrics, 16*, 482–489.

Bezerra, A. C., da Silva Barbarosa, L., Zuza, J. F. C., de Oliveira, A. M. F., & de Azevedo, C. F. (2019). Structural characterization of mint (*Mentha* x villosa Huds) steam and leaf. *Journal of Experimental Agriculture International, 36*, 1–6.

Core R Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL https://www.R-project.org/.

Dieterle, F., Ross, A., Schlotterbeck, G., & Senn, H. (2006). Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Anayltical Chemistry, 78*, 4281–4290.

Dong, W., Ni, Y., & Kokot, S. (2014). A Novel near-infrared spectroscopy and chemometrics method for rapid analysis of several chemical components and antioxidant activity of mint (*Mentha haplocalyx* Brig.) samples. *Applied Spectroscopy, 68*, 245–254.

Durazzo, A., Kiefer, J., Lucarini, M., Camilli, E., Marconi, S., Gabrielli, P., Aguzzi, A., Gambelli, L., Lisciani, S., & Marletta, L. (2018). Qualitative analysis of traditional Italian dishes: FTIR approach. *Sustainability, 10*, 4112.

El-Kashourty, E.-S. A., El-Askary, H. I., & Kandil, Z. A. (2013). Botanical and genetic characterization of Mentha suaveolens Ehrh. cultivated in Egypt. *Pharmacognosy Journal, 5*, 228–237. https://doi.org/10.1016/j.phcgj.2013.10.002

Ercioglu, E., Velioglu, H. M., & Boyaci, I. H. (2018). Determination of terpenoid contents of aromatic plants using NIRS. *Talanta, 178*, 716–721. https://doi.org/10.1016/j.talanta.2017.10.017

Fejér, J., Grulová, D., & De Feo, V. (2017). Biomass production and essential oil in a new bred cultivar of peppermint (*Mentha×piperita* L.). *Industrial Crops and Products, 109*, 812–817.

Fialova, S., Halusova, L., Tekelova, D., & Grancai, D. (2015). The exploitation of micromorphological parameters for identification in section *Mentha*. *Acta Facultatis Pharmaceuticae Universitatis Comenianae, 62*, 9–14. https://doi.org/10.1515/afpuc-2015-0004

Freire, M. M., Jham, G. N., Dhingra, O. D., Jardim, C. M., Barcelos, R. C., & Valente, V. M. M. (2012). Composition, antifungal activity and main fungitoxic components of the essential oil of *Mentha piperita* L. *Journal of Food Safety, 32*, 29–36. http://10.1111/j.1745-4565.2011.00341.x.

Gobert, V., Moja, S., Colson, M., & Taberlet, P. (2002). Hybridization in the section *Mentha* (*Lamiaceae)* inferred from AFLP markers. *American Journal of Botany, 98*, 2017–2023.

Grulová, D., De Martino, L., Mancinib, E., Tkáčiková, L., Šalamon, I., Fejer, J., & De Feo, V. (2016). Phytotoxic and antibacterial activity of essential oil of new peppermint cultivar. *Natural Product. Communications, 11*, 1631–1774.

Hawrył, M. A., Skalicka-Woźniak, K., Świeboda, R., Niemiec, M., Stępak, K., Waksmundzka-Hajnos, M., Hawrył, A., & Szymczak, G. (2015). GC-MS fingerprints of mint essential oils. *Open Chemistry, 13*, 1326–1332.

Hubert, M., & Engelen, S. (2004). Robust PCA and classification in bioscience. *Bioinformatics, 20*(11), 1728–1736.

Jędrzejczyk, I., & Rewers, M. (2018). Genome size and ISSR markers for *Mentha* L. (*Lamiaceae*) genetic diversity assessment and species identification. *Industrial Crops and Products, 120*, 171–179.

Kanakis, C. D., Petrakis, E. A., Kimbaris, A. C., Pappas, C., Tarantilis, P. A., & Polissiou, M. G. (2012). Classification of Greek *Mentha pulegium* L. (Pennyroyal) samples, according to geographical location by fourier transform infrared spectroscopy. *Phytochemical Analysis, 23*, 34–43. https://doi.org/10.1002/pca.1322

Kiefer, J., Lampe, A. I., Nicoli, S. F., Lucarini, M., & Durazzo, A. (2019). Identification of passion fruit oil adulteration by chemometric analysis of FTIR spectra. *Molecules, 24*, 3219.

Kiełtyka-Dadasiewicz, A., Okoń, S., Ociepa, T., & Król, B. (2017). Morphological and genetic diversity among peppermint (*Mentha* × *piperita* L.) cultivars. *ACTA Scientarum Polonorum Hortorum Cultus, 16*(3), 151–161. https://doi.org/10.24326/asphc.2017.3.15

Kucharska-Ambrożej, K., & Karpińska, J. (2020). The application of spectroscopic techniques in combination with chemometrics for detection adulteration of some herbs and spices. *Microchemical Journal, 153*, 104278. https://doi.org/10.1016/j.microc.2019.104278

Larkin, P. J. (2018). *Infrared and Raman spectroscopy principles and spectral interpretation* (Elsevier ed). Second Edition.

Lucarini, M., Durazzo, A., Kiefer, J., Santini, A., Lombard-Boccia, G., Souto, E. B., Romani, A., Lampe, A., Nicoli, S. F., Gabrielli, P., Bevilacqua, N., Campo, M., Morassut, M., & Cecchini, F. (2020). Grape seeds: Chromatographic profile of fatty acids and phenolic compounds and qualitative analysis FTIR-ATR spectroscopy. *Foods, 9*, 10.

Ludwiczuk, A., Kiełtyka-Dadasiewicz, A., Sawicki, R., Golus, J., & Ginalska, G. (2016). Essential oils of some *Mentha* species and cultivars, their chemistry and bacteriostatic activity. *Natural Product Communication, 1*, 1015–1018.

Manolache, F.-A., Todaşcă, C., Ionescu, V., Iorga, E., Beldie, A. A., & Matei, M. A. (2018). Authentication of wild *Mentha aquatica*, from Danube delta, using 1H-NMR spectroscopy. *UPB Scientific Bulletin, Series B, 80*, 127–134.

Marotti, M., Piccaglia, R., Giovanelli, E., Deans, S. G., & Eaglesham, E. (1994). Effects of planting time and mineral fertilization on peppermint (*Mentha* × *piperita* L.) essential oil composition and its biological activity. *Flavour and Fragrance Journal, 9*, 125–129.

Mevik, B.-H., Wehrens, R., & Liland, K. H. (2018). *pls: Partial least squares and principal component regression. R package version 2.7-0*. https://CRAN.R-project.org/package=pls.

e1071 Meyer, D. D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2015). Probability theory group (formerly: E1071), TU wien. R package version 1. In *Misc functions of the department of statistics*.

Næs, T., Isaksson, T., Fearn, T., & Davies, T. (2002). *Multivariate calibration and classification*. Chichester: NIR Publications.

October) Petronijevic, R. B., Velebit, B., & Baltic, T. (2017). Shedding light on food fraud: Spectrophotometric and spectroscopic methods as a tool against economically motivated adulteration of food. *59th International Meat Industry Conference MEATCON 2017; IOP Conference Series: Earth and Environmental Science, 85*, Article 012024. https://doi.org/10.1088/1755-1315/85/1/012024.

Petry, R., Schmitt, M., & Popp, J. (2003). Raman spectroscopy–A prospective tool in the life sciences. *ChemPhysChem, 4*, 14–30.

Prakash, N., & Yunus, M. (2013). Fourier Transform Infrared Spectroscopy analysis of oil of *Mentha arvensis* grown at sites varying with vehicular traffic loads in Lucknow City, India. *International Journal of Environment, 2*, 16–25.

Rachman, Z., An, U., & Muchtardi, M. (2018). Analysis of quality and quantity control of herbal compounds using Fourier –transform infrared spectroscopy. *A review journal. Drug Invention Today, 10*, 2734–2742.

Rita, P., & Animesh, D. K. (2011). An updated overview on peppermint (*Mentha piperita* L.). *International Research Journal of. Pharmacy, 2*, 1–10.

Rosch, P., Kieffer, W., & Popp, J. (2002). Chemotaxonomy of mints of genus *Mentha* by applying Raman spectroscopy. *Biopolymers, 67*, 358–361.

Sabboura, D., Yacoub, R., & Lawand, S. (2016). Assessment of genetic relationships among Mint Species. *International Journal of ChemTech Research, 9*, 462–468.

Saric-Kundalic, B., Fialova, S., Dobes, C., Olzant, S., Tekelova, D., Granca, D., Reznicek, G., & Saukel, J. (2009). Multivariate numerical taxonomy of *Mentha* species, hybrids, varieties and cultivars. *Scientia Pharmaceutica, 77*, 851–876.

Stanev, S., & Zheljazkov, V. D. (2004). Study on essential oil and free menthol accumulation in 19 cultivars, populations, and clones of peppermint (*Mentha* × *piperita*). *Acta Horticultirae, 629*, 149–152.

Todorov, V. (2016). *rrcovHD: Robust multivariate methods for high dimensional data. R package version 0.2-5*. https://CRAN.R-project.org/package=rrcovHD.

Tucker, A. O., & Naczi, R. F. C. (2007). Mint: The genus Mentha. In B. M. Lawrence (Ed.), *Mentha: An overview of its classification and relationships* (pp. 1–39). Boca Raton, FL: CRC Press.

Wold, S. (1976). Pattern recognition by means of disjoint principal components models. *Pattern Recognition, 8*, 127–139.