DATA FUSION APPROACHES IN SPECTROSCOPIC CHARACTERIZATION AND CLASSIFICATION OF PDO WINE VINEGARS

Rocío Ríos-Reina[a,*], Raquel M Callejón[a], Francesco Savorani[b], José M Amigo[c], Marina Cocchi[d,*]

[a] *Dpto. de Nutrición y Bromatología, Toxicología y Medicina Legal, Facultad de Farmacia, Universidad de Sevilla, C/P. García González n°2, E-41012 Sevilla, Spain*

[b] *Department of Applied Science and Technology (DISAT), Polytechnic University of Turin – Corso Duca degli Abruzzi 24, 10129 Torino (TO), Italy*

[c] *Chemometrics and Analytical Techniques, Department of Food Science, University of Copenhagen, Rolighedsvej 26, 1958 Frederiksberg C, Denmark*

[d] *Dipartimento di Scienze Chimiche e Geologiche, Università di Modena e Reggio Emilia, Via Campi 103, 41125 Modena, Italy*

*corresponding authors: rrios5@us.es; marina.cocchi@unimore.it

Abstract

Spain is one of the major producers of high-quality wine vinegars having three protected designations of origin (a.k.a. PDOs): “Vinagre de Jerez”, “Vinagre de Condado de Huelva” and “Vinagre de Montilla-Moriles”. Their high prices due to their high quality and their high production costs explain the need for developing an adequate quality control technique and the interest in extensive characterization in order to capture the identity of each denomination. In this framework, methodologies based on non-targeted techniques, such as spectroscopies, are becoming popular in food authentication. Thus, for improving vinegar quality assessment, fusion of data blocks obtained from the same samples but different analytical techniques could be a good strategy, since the quantity and quality of sample knowledge could be enhanced providing new insights into the differentiation of vinegars. Therefore, the aim of this manuscript is the development of a multi-platform methodology and a model able to classify the Spanish wine vinegar PDOs. Sixty-five PDO wine vinegars were analyzed by four spectroscopic techniques: Fourier-transform mid-infrared spectroscopy (MIR), near infrared spectroscopy (NIR), multidimensional fluorescence spectroscopy (EEM) and proton nuclear magnetic resonance (^{1}H-NMR). Two different data fusion strategies were evaluated: Mid-level data fusion with different preprocessing, and Common Component and Specific Weights analysis multiblock method. Exploratory and classification analysis on the data from individual techniques were also performed and compared with data fusion models. The data fusion models improved the classification, providing a more efficient differentiation, than the models based on single methods, and supporting the approach to combine these methods to achieve synergies for an optimized PDO differentiation.

Keywords: wine vinegars, food authentication, spectroscopy, classification, data fusion, P-Comdim.

1. INTRODUCTION

Nowadays, there is a growing consumer's demand for high quality food products. The term "quality" in food is directly related to a known origin and specific chemical composition, adequate and satisfactory physical and sensory properties, as well as meeting safety and health requirements [1,2]. Protected Designation of Origin (PDO) indication is one of the label adopted by the European Community as recognition of some specific food quality attributes. A product with a PDO registration must be produced, processed and prepared in a given geographical area using a recognized know-how [3]. The PDO denomination confers to these products a high added value, consequently there is also an increasing of deceptive practices aiming at counterfeiting them, such as mislabeling of geographical origin, disregarding the production protocol or adulteration of the product. In this respect, assessing the authenticity of traditional food is a complex issue because it has to encompass several aspects going from assessing the compliance to the legal requirements stated in the product label, i.e. controlling the geographical origin and the respect of the traditional protocols, to detecting fraudulent processing practices or adulteration.

Among the PDO products with high demand there are the high-quality vinegars. In particular, in addition to the well-known "Aceto Balsamico Tradizionale di Modena" from Italy [4], Spain is also one of the major producers of high-quality wine vinegars. Thus, three important Spanish wine vinegars have gained the PDO label because of their unique characteristics and traditional production, namely: "Vinagre de Jerez" (also known as "Sherry wine vinegar"), "Vinagre de Condado de Huelva" and the most recently "Vinagre de Montilla-Moriles". Furthermore, within each PDO, there are different categories according to their time and method of aging (''criaderas and solera" or ''añada" systems) in wood barrels as well as the sweetness. The high quality of these wine vinegars is linked to the raw material used (i.e. high quality wines, also protected by the corresponding PDO), the traditional production protocol and method of aging in wooden barrels. Therefore, the high prices of these vinegars, due to their high quality, the long aging time and hence, the high cost of their production, explain the need of proper characterization in order to provide an adequate quality control to defend their identity [4–9].

Due to the traditional making procedure, the raw material used and the aging process, these wine vinegars are very complex multi-component mixtures from the chemical point of view, thus different analytical techniques have been applied to obtain an extensive characterization in order to assess their authenticity [6,7,9–11]. Spectroscopic techniques, based on infrared (IR), fluorescence or nuclear magnetic resonance (NMR) spectroscopy, are the most commonly used food fingerprinting techniques in untargeted approach. In particular, these spectroscopic techniques share the advantage of requiring minimal sample preparation, moreover IR is non-destructive and chip, while NMR may allow quantification of a wide range of compounds. Good results were obtained by spectroscopic analysis of the three Spanish PDO wine vinegars in terms of assessing their aging and sweet categories or characterizing each PDO separately [6,7,12]. However, the possibility of discriminating these three wine vinegars PDOs, regardless of the presence of different ageing or sweetness features, within each distinct PDO, has been less considered in the literature [7,9,11].

In order to gather more detailed knowledge about the specificity of each PDOs and aiming at improving their quality assessment and differentiation, the combination and fusion of the data acquired by several analytical platforms could be useful [2,13,14]. Data fusion methodologies have demonstrated to be a powerful tools for obtaining more reliable authentication models with respect to the results obtained by each technique separately [2,13,15–17]. In fact, the fusion of the different information obtained can enhance the quantity and quality of knowledge about the distinctive features among samples/categories. Moreover, the integration of the different data types into a single model also allows assessing the correlation and the similar/different information content among the different techniques.

Data fusion may be accomplished at different levels (i.e. low-, mid- and high-level data fusion), depending on the objective, number and type of data sets to combine [2,18–20]. The low-level fusion is a conceptually simple method: raw data from more than one source are directly fused (concatenated) after preprocessing issues are addressed. This level of data fusion has been widely applied for the authentication and quality control of many food and beverages [2]. The main limitations are a high data volume and the possible predominance of one data source over the others and possible discontinuities regions when spectral data are fused. This is partially overcome by the mid-level fusion, in which a previous extraction of some relevant features from each single data source is performed and then, these features are concatenated into a single

array. Moreover, this type of fusion enables an easy interpretation of the results, since the contribution of each individual block can be visualized. The main parameters to take into account are the number of features to retain from each model, the method to be used for data reduction, and the type of scaling to apply, however this last issue is less severe than in low-level data fusion, considering that data reduction has already been applied. Mid-level data fusion has been also applied in authentication and quality control of food and beverages [2,17].

On the other hand, other approaches based on multiblock analysis are also suitable in data fusion context, such as the Common Components and Specific Weights Analysis (CCSWA, also referred to as ComDim, which is as well the name of the algorithmic implementation) [21–23], which has been recently revised and extended to the supervised context (P-ComDim) [24], *i.e*. to deal with the case where one of the blocks (Y block) holds responses that are to be predicted on the basis of the information provided by the other blocks. The main purpose of the ComDim algorithm is to provide the common sources of information shared by each data block, i.e. the common components, at the same time assigning to each single block a specific weight (or salience) associated to each dimension of the common space [24,25]. This method has been recently applied to the analysis of several food products in order to differentiate e.g. an organically or conventional production [26,27], or cheese products obtained by different manufacturing or ripening [28] as well as it has been applied to predict sensory attributes [21].

A major general advantage of ComDim approach, compared to the low and mid-level data fusion approaches, is that it provides information about the relation between individual data blocks (i.e. common variables) and their contribution to each common component. Thus, ComDim can be applied in order to study the complementarity, and also the differences, of the various spectroscopic techniques. In particular, the study of the saliences (weights of each data block in the common model) could be particularly interesting due to the fact that if a dimension has close saliences for two or more techniques, this may be due to a physical phenomenon that is described in a similar way for both methods. On the other hand, if there is an important difference between the saliences for a given dimension, it could mean that this dimension reveals a phenomenon only visible by one technique and not by the others. This could be used for focusing the selectivity of the spectroscopic techniques studied in this work.

Moreover, used in the predictive context, i.e. P-ComDim, we could infer and assess which information in the different data blocks is relevant for the discrimination of the different categories, which is shared and which is peculiar to each of them [24,28,29]

Taking this background into consideration, the aim of this work was to perform a multiplatform characterization and develop classification models for the different Spanish wine vinegar PDOs by assessing different data fusion approaches, as well as to study the synergy/complementarity among the techniques considered for that purpose. To this aim, the same wine vinegar samples were measured by four spectroscopic techniques: Fourier-transform infrared spectroscopy (i.e. mid infrared, MIR), near infrared spectroscopy (NIR), multidimensional fluorescence spectroscopy (EEM) and proton nuclear magnetic resonance (^{1}H-NMR). These techniques were selected due to the individual efficacy in the characterization of PDO wine vinegars as previously reported [6,7,12] , as well as because they have gained wide acceptance in foods characterization, authenticity and classification purposes [15, 30–34].

The main contribution of this study is to comparatively discuss the different data fusion strategies, in term of capability to improve discrimination of the three PDO's vinegars and to highlight the role of each spectroscopic technique. In fact, although they can share some repeated pieces of information, they are mostly complementary.

2. MATERIALS AND METHODS

2.1. Samples

Sixty-five PDO wine vinegar samples were provided by several local wineries through the Council Regulation of each PDO. Twenty-one samples belonging to the PDO "Vinagre de Condado de Huelva", twenty-eight to "Vinagre de Jerez" PDO and sixteen to the most recently designed PDO "Vinagre de Montilla-Moriles" were analyzed by the four analytical techniques which are described below. Furthermore, within each PDO, samples from the different commercialized categories (aged and sweet) were included in the analysis. Samples were analyzed in duplicate. More information about the samples is presented in Table 1.

2.2. Instrumental analysis

2.1.1. Mid-infrared spectroscopy (MIR)

Samples were analyzed, according to the method reported in [6], by using a Bruker Vertex 70 FTIR spectrometer equipped with a DGTS detector (Bruker Optics, Ettlingen,

Germany) and a multi-reflection attenuated total reflectance accessory (ATR, six bounces, Specac, Orpington, U.K.). Samples were directly analyzed without sample pre-treatment, recording the spectra at the same temperature (22 ± 0.05 ˚C) in the region of 4000-600 cm^{-1} (by an average of 50 scans at a resolution of 4 cm^{-1}) and were examined using OPUS version 7.0 (Bruker Optics, Ettlingen, Germany) and manipulated with OMNIC software. The raw MIR spectra are shown in Fig.I. Supplementary Material.

2.1.2. Near-infrared spectroscopy (NIR)

NIR spectra were collected following the method published in [12], by using an ABB Bomen IR spectrometer (Q-interline, X, Denmark), equipped with a 1 mm path length cuvette. Spectral data were collected in the range of 12000–4000 cm^{-1}, resolution of 8 cm^{-1}, and 64 scans for both backgrounds and samples. Samples were directly analyzed without sample pre-treatment in a random sequence at room temperature (21± 2 °C) by pipetting them into 1 mL shell vial, 40x80 mm transparent (Skandinaviska Genetec AB, Lund, Sweden) before measurement. The spectrometer was interfaced to a computer with GRAMS/AI™ Spectroscopy Software (Thermo Fisher Scientific software) for spectral acquisition and exportation. The raw NIR spectra are shown in Fig.I. Supplementary Material.

2.1.3. Excitation-Emission Multidimensional Fluorescence (EFM)

Wine vinegar samples were directly analyzed without sample pre-treatment an at the same temperature (25.00 ± 0.05 ˚C) by a Varian Cary-Eclipse fluorescence spectrophotometer (Varian Iberica, Madrid, Spain), equipped with two Czerny-Turner monochromators, and a Xenon discharge lamp pulsed at 80 Hz with a half peak height of ≈2 μs, according to the method reported in [7]. Cary-Eclipse software was used for spectral acquisition and exportation. The fluorescence Excitation-Emission Matrices (EEMs) were obtained by varying the excitation wavelength (λex) between 250 and 700 nm (every 5 nm) and recording the emission spectra (λem) from 300 to 800 nm (every 2 nm), with excitation and emission slits set at 5 nm and the scan rate fixed to 1200 nm min^{-1}. EEMs were preprocessed in order to avoid noisy and non-informative areas by selecting shorter spectral ranges (λex from 250 to 680 nm, and λem from 310 to 800 nm). The EEM landscape of a vinegar is shown in Fig.I. Supplementary Material as an example.

2.1.4. ^{1}H-Nuclear Magnetic Resonance (^{1}H-NMR)

Samples were prepared by adding 100 µL of 0.16% of 3-(Trimethylsilyl) propionic-2,2,3,3-d4 acid sodium salt (TMSP-2,2,3,3-d4) in D2O (99.97%) dissolution, to 600 µL of each wine vinegar. TMSP was used as both a chemical shift reference (δ=0) and internal standard. ^{1}H-NMR spectra have been acquired at 300 K of temperature on a Bruker AVIII 700 spectrometer (Bruker Biospin GmbH Rheinstetten, Karlsruhe, Germany) operating at 700.25 MHz. The ^{1}H-NMR data were acquired using the Bruker spin−echo sequence "cpmgpr.fb" (Carr–Purcell–Meiboom–Gill, Bruker Library) with water presaturation, applied to suppress broad resonance signals. FIDs' have been recorded as the sum of 64 scans of 7.4 s each covering a spectral width of 11.0 ppm with 1s between each consecutive scan. Data acquisition was carried out using the "baseopt" Bruker sequence to optimize the baseline after Fourier Transform. The raw ^{1}H-NMR spectra are shown in Fig.I. Supplementary Material.

2.3. Data analysis

Since four different instrumental fingerprints were recorded for each sample, each one with different data structures, several chemometric algorithms were employed in order to extract and merge the information presents in each data set.

The data analysis workflow included: i) building separate models: both exploratory analysis and classification were performed on the data obtained from the individual analytical techniques; ii) in order to take advantage of the multiplatform characterization of the samples, the data of different sources were processed by means of different data fusion (DF) strategies. The objectives were to assess common and specific information pertaining to each analytical platform and obtaining improved classification results. A schematization of the global data analysis flow is presented in Fig.1.

Figure 1 to be inserted about here

2.3.1. Data sets

In total sixty-five samples were analyzed by each spectroscopic technique. In order to validate the models, the samples were split in a training set of forty-seven samples (fifteen "Vinagre de Condado de Huelva", twenty "Vinagre de Jerez" and twelve "Vinagre de Montilla-Moriles" PDO samples) and a test set of eighteen samples (six "Vinagre de Condado de Huelva", eight "Vinagre de Jerez" and four "Vinagre de Montilla-Moriles" PDO samples) using the Duplex algorithm [35]. This algorithm ensures a representative spanning of the whole data domain for both calibration and

validation sets, we also checked for a balanced representation of each category in both sets. Moreover, since the number of samples is rather limited, the splitting was repeated five times (always checking by exploratory data analysis, that both sets spanned the whole variability domain and balanced category representation was achieved) hence five classification models were calculated for each analyzed data set (NIR+MIR, NMR, EEM, mid-level Data Fused, P-Comdim raw data and P-Comdim extracted features). In the results the average classification errors are reported.

2.3.2. Decomposition methods

As summarized in Fig. 1, different decomposition methods were applied, according to the type of dataset, for exploratory data analysis as well as for data reduction to obtain the features which were then used for the data fusion models, *i.e*. mid-level DF and features-based P-ComDim.

MIR and NIR individual data sets, as detailed in Section 2.3.5.1, were concatenated at low-level DF and the obtained dataset was compressed by principal component analysis (PCA).

The EEM data array, after Rayleigh and Raman scattering correction [7], was decomposed by PARAllel FACtor analysis (PARAFAC) [36,37] in order to extract the relevant features (fluorophores).

Finally, for ^{1}H-NMR dataset, after proper alignment and baseline correction, multivariate curve resolution (MCR) [38,39] was used to resolve the chemical components. The peak areas of the resolved components were then used as features.

PARAFAC and MCR decomposition methods have been widely described in the literature. Applied constrains and preprocessing details for each data block are reported in Section 2.3.4.

2.3.3. Classification Analysis.

Partial least squares-discriminant analysis (PLS-DA) is a classification technique based on partial least squares (PLS) algorithm with a so-called dummy matrix reporting class membership as Y block [40]. In our study, three different Spanish PDO were considered, therefore, the size of the Y dummy matrix was $n°\ samples \times 3$ (one column for each one of the classes) and codification was 1/0 (belonging/not belonging to the category).

In the case of EEM data set, which is a three-way array, N-way Partial least squares-discriminant analysis (NPLS-DA) [41] based on multilinear PLS (NPLS) [42] has been used and codification of the Y block is the same as for PLS-DA.

In both cases, classification was achieved by applying linear discriminant analysis (LDA) on the X-scores calculated by PLS-DA/NPLS-DA [43]. The minimum classification error rate in cross-validation (venetian blind, seven splits) was used to assess the number of latent variables, i.e. components of the PLS-DA/NPLS-DA models.

2.3.4. Preprocessing and analysis of individual data blocks

2.3.4.1. MIR and NIR datasets

Concerning the MIR data, as is described in our previous work [6], no preprocessing was needed and the raw spectra were just mean centered. Moreover, only the region between 1500 and 900 cm^{-1} was included in the analysis [6] in order to discard the uninformative variables with excessive noise.

With regards to NIR data, different preprocessing methods were evaluated prior to data analysis as was contemplated in a previous work [12] The best pre-processing approach resulted to be smoothing (Savitzky-Golay filter, 7 points window and second order polynomial degree) to reduce random noise, followed by standard normal variate (SNV) [44] to correct additive scattering. In addition, the spectra were always mean centered prior to any analysis. As mentioned before, based on previous expertise or literature [12,45,46], two segments of the spectrum were removed from the whole acquired wavenumber range: the first one (4000-5430 cm^{-1}) because of low signal/noise ratio and the second one due to the strong combination band of O-H from water (7200-6400 cm^{-1}).

2.3.4.2. EEM dataset

EEM data were preprocessed in order to avoid noisy and non-informative areas by selecting shorter spectral ranges, according to the preprocessing steps described in [7]. Thus, the emission over 680 nm and the excitation below 310 nm were cut. Then, EEM data were corrected for Rayleigh and Raman scattering [47], removing and replacing the scattering areas with interpolated values [47]. After this correction, EEM data was decomposed by PARAFAC [37]. A model based on five factors, constrained for non-negativity in all modes (both concentration and spectral profiles), was built. The proper number of factors was determined by taking into account the CORe CONsistency DIAgnostic test (COR- CONDIA) [48], the explained variance and the visual inspection of the recovered spectral profiles and residuals. The PARAFAC scores (first mode loadings) for these factors were used as features to build the mid-level fused dataset.

2.3.4.3. ^{1}H-NMR dataset

Prior to data analysis, several preprocessing steps were applied to NMR spectra. The regions below 0.84 ppm and over 9.8 ppm were discarded because they were uninformative. Also the region between 4.75 and 5 ppm was removed since it contained the residual water signal not completely removed by the instrumental presaturation step. To correct for the inhomogeneous pH-dependent chemical shifts, all spectra were aligned by means of icoshift [49] whereas weighted least squares (WLS) [50] was used for baseline correction.

Then, MCR was applied. The whole ^{1}H-NMR data was divided into 52 intervals of different size in order to avoid splitting the single NMR signals. This task was performed manually by making use of the previous knowledge of NMR chemical shifts of the main wine and vinegar compounds [33,51,52]. These intervals are shown in Fig. II. Supplementary Material. The MCR settings were the same for each interval: the number of components was determined by inspection of PCA explained variance and SIMPLISMA [53] was used to obtain the initial estimation of the pure spectral profile. The peak areas of the resolved concentration profiles (chemical components) within each interval were calculated by integration and used as features for the subsequent fused data set.

In order to achieve a tentative assignment of the ^{1}H-NMR resolved components, both Chenomx NMR Suite 7.0 (Chenomx, Edmonton, Canada), as well as assignments reported in literature [33,51,52,54,55] were used. Sixty-two components were resolved and integrated; thirty-five of these were tentatively assigned. Those components that were not possible to assign, are named as "X" plus a number. The fact that several regions of the NMR spectra could not be associated to a single signal is due to the many overlapped multiplets present, which impair certain identification. On the other hand, they could be attributed to overall contribution of a class of compounds, such as sugars (between 3-4 ppm). In our case, in this region, only glucose and fructose could be separately assessed.

2.3.5. Data fusion strategies

2.3.5.1. Low-level fusion of MIR and NIR data

In the low-level strategy, fusion occurs by concatenating the original data matrices, opportunely pretreated and then analyzing the resulting array as a single data block.

The MIR and NIR spectra were single preprocessed as described in Sections 2.3.4.1. Then, the matrices describing the individual blocks were concatenated to obtain a single one, having as many rows as samples analyzed and as many columns as spectral wavelengths selected for each data set. This new matrix was additionally normalized in order to compensate for the different measuring scales and variability of each technique in order to prevent one block from being dominant in the subsequent data analysis [2]. Thus, block-scaling, to equalize variance, and mean centering were applied. Doing so, each block presented variance equal to one, but the ratio of the variance between any two variables inside a single block was preserved.

After preprocessing, a PCA model based on 8 principal components, accordingly to Scree plot and explaining 99.87% of total variance, was selected and the extracted score vectors were used as MIR/NIR features to build the fused dataset.

A possible alternative approach consists of applying PCA to the separate MIR and NIR spectral data and then using the extracted features (distinct set of PCs) in mid-level DF; this approach was also considered and gave very similar results.

2.3.5.2. Mid-level data fusion

In the mid-level strategy, fusion occurs at the level of features extracted from the different data blocks. In this study, as Fig.1 shows, the final fused array was assembled using the 8 PCA scores from MIR and NIR, the 5 factors from the PARAFAC model of EEM data, and the peak areas of the 62 resolved components by MCR of ^{1}H-NMR data.

As in the case of low-level fusion, since the extracted features in mid-level data fusion can have different numerical characteristics, scaling of the fused matrix [2,15,17] was performed. Different preprocessing tools were assessed: autoscaling and block-autoscaling (each data set corresponding to an analytical technique was considered as a block). In block-autoscaling, each variable is first scaled to unit variance (autoscaling), and then each block is scaled to equal variance. As a result, each block presented unit variance and each variable inside a block had its variance equal to $1/n_{\text{block}}$, where n_{block} is the number of variables in a given block.

2.3.5.3. *P-ComDim*

The recently proposed P-ComDim (*i.e.*, Predictive ComDim) method [24], which is the extension of the multiblock method ComDim to the supervised context, has also been evaluated as a different data fusion strategy. For details on P-ComDim algorithm the reader is referred to literature [24,25]. Briefly we recall the main feature of the method. P-ComDim can be applied to any number of data blocks, of which the

dependent one is denoted by $\boldsymbol{Y}$ and the independent ones by $\boldsymbol{X}_k$. The first step in P-ComDim algorithm is calculating the kernel matrices:

$$\boldsymbol{S}_k=\boldsymbol{X}_k\boldsymbol{X}_k^T\boldsymbol{Y}\boldsymbol{Y}^T \qquad (1)$$

Then a "common singular value decomposition" is conducted, by minimizing the criterion:

$$\sum_{k=1}^{K}\|\mathbf{S}_k - \lambda_k\mathbf{t}\mathbf{u}^T\|^2 \qquad (2)$$

obtaining a first common component for the **X**-blocks, i.e. $\mathbf{t}_1$, as well as a component in **Y**-space, i.e. $\mathbf{u}_1$. Further components are then calculated sequentially after deflation of both **X**-blocks and **Y**-block. As for standard ComDim, each single **X**-block ($\boldsymbol{X}_k$) contributes to a common component according to its salience, λ_k [29]. It is also possible to associate to each block $\boldsymbol{X}_k$ a local component by calculating:

$$\mathbf{t}^{(k)} = \mathbf{X}_k\mathbf{X}_k^T\mathbf{t} \qquad (3)$$

i.e. eq. (3) maps t into a latent variable which lies in the space spanned by the variables in X_k. This latent variable $t^{(k)}$ is used to recover and interpret the specific contribution of the X_k-block variables to the global latent variable t.

To accomplish classification, the **Y**-block holds the class membership information, as described in section 2.3.3. and a classification model can be built by applying PLS-DA to the **u**-scores obtained by P-ComDim. Prediction is accomplished by first estimating, in prediction, the **u**-scores for the test samples (**u**-test) in P-ComDim, then using the **u**-test in PLS-DA as prediction set. In our case, the number of PLS-DA components was estimated according to minimum classification error in CV using the same splits and classification rule as described in Section 2.3.3. Also the subdivision in training and test sets was the same as described in Section 2.3.1.

Moreover, in P-ComDim methodology, two different strategies were performed and compared. In the first, ComDim was developed using the raw spectra of MIR, NIR, EEM and ^{1}H-NMR as X-blocks after applying the same spectral preprocessing as described in Section 2.3.4. MIR and NIR data were mean centered, the ^{1}H-NMR data was block-scaled by dividing the spectra into six regions (0.84-1.15, 1.15-1.5, 1.5-2.0, 2.0-2.25, 2.25-3.2, 3.2-9.8 ppm) to compensate for major differences in spectral region signal intensities and the EEM data array of dimensions *I*-samples × *J*-excitation × *L*-emission wavelengths, was unfolded to a matrix of dimensions $I \times JL$.

In the second, the extracted features of each data block (PCA scores from MIR/NIR, MCR peaks areas of resolved components, and PARAFAC factors) were used as ***X***-blocks.

Both in the first and second cases, each data table $\boldsymbol{X}_k$ was normalized in order to obtain the data tables having the same inertia as usually done in ComDim algorithm [56].

The interpretation of each model and comparison of two approaches (i.e. with raw spectra and with the features) was performed by studying the saliences, global and local scores/loadings [28,29], and the classification performance.

2.3.6. Software

Preprocessing, PARAFAC, PCA, PLS-DA and NPLS-DA models were calculated by using routines of PLS Toolbox 6.5 (Eigenvector Research Inc.,WA, USA) working under MATLAB environment v.2016a (Mathworks, MA, USA). LDA was calculated by using the Statistics and Machine Learning Toolbox v. 10.1. Multivariate curve resolution was carried out by using the MCR-ALS GUI (http://www.mcrals.info) and a MATLAB routine implemented to automatically work on spectral intervals, courtesy from Prof. R. Bro's group. ^{1}H-NMR data acquisition, Fourier transformation and spectral preprocessing were carried out using Bruker TopSpin 3.0 and Chenomx NMR Suite 7.0 (Chenomx, Edmonton, Canada) was used to obtain a tentative assignment of the ^{1}H-NMR resolved components.

P-ComDim models were obtained by using routines developed by Prof. D. Rutledge and the SAISIR package for MATLAB [57,58].

3. RESULTS AND DISCUSSION

This section is articulated in three main parts. In the first one, the description of exploratory analysis results for the individual data sets, as well as the feature extraction step (Sections 3.1), and the respective classification models (Section 3.2) are reported. In the second part (Section 3.3), the fused dataset is considered and the application of the mid-level approach is described in detail. The third part (Section 3.4) presents the results obtained by P-ComDim in order to study the complementarity of the techniques.

3.1. Exploratory analysis of individual data matrices

MIR and NIR data were preprocessed and fused as described in Section 2.3.5.1, the results of exploratory PCA analysis (8 PCs, accounting for 99.8% of the total variance) are reported in Figure III of the Supplementary Material. The three categories

strongly overlap and a partial trend of separation was only observed on the scores plot of the PC1, PC3 and PC8 (Fig. III.A), inspecting the corresponding loading plots (Fig. III.B) it can be observed that PC1 mainly distinguishes the sweet Pedro Ximenez sub-category which is present in both "Vinagre de Jerez" and "Vinagre de Montilla-Moriles" PDOs (the contributing spectral regions have been associated with the presence of grape sugars, furfural and Maillard compounds [6,12,46,59]). "Vinagre de Montilla-Moriles" PDO samples are partially separated from "Vinagre de Jerez" PDO along PC3 to which are contributing peaks (Fig. III.B) that have been assigned to chemical compounds that change during aging, e.g. some alcohol, aldehydes, esters, ethers and acids [6, 12,46,60].

The EEM data array was preprocessed and decomposed by PARAFAC as described in Section 2.3.4.2 obtaining a five factors model (explained variance 99%), which is in good agreement with the three individual PARAFAC models obtained in our previous work [7] for each one of the three PDOs. Fig. 2.A and B includes the PARAFAC loadings for mode 2 and 3 (excitation and emission spectra) of the extracted factors. The excitation and emission maxima of these extracted factors, as well as their possible matching fluorophores according to the literature and our previous knowledge [7,62], are listed in Table 2.

Figure 3 to be inserted about here

Fig. 3.C shows the average value of the scores (first mode loadings) for samples belonging to each PDO vs. the number of PARAFAC factors. The "Vinagre de Montilla-Moriles" PDO presents higher values on the first and the second factors, with respect to the other two PDOs. Hence, higher presence of components coming from raw materials, which is indicative of less aging, as well as more amount of caramel and 5-Hydroxymethylfurfural (Table 2). However, it is difficult to highlight a clear separation of samples belonging to each class in any of the scatter plots of PARAFAC scores (plots not shown for sake of brevity).

The NMR data set built with the integrated areas of the sixty-two resolved components (Table 3), obtained by MCR analysis of the ^{1}H-NMR spectra (as is detailed in Section 2.3.3.4) was preprocessed by autoscaling prior to PCA analysis (six components, explained variance 90.3%). The score and loading plots of the PCs that better highlighted the separation between the three PDOs are shown on Figure IV Supplementary Material. Also in this case a strong overlap is present and only a partial

separation trend of "Vinagre de Montilla-Moriles" PDO samples from "Vinagre de Condado de Huelva" can be observed. The loadings plot (Fig. IVB) highlight, similar to MIR-NIR PCA results, that: i) the first component distinguishes the Pedro Ximenez sweet samples from the rest (contribution from the sugar spectral region, compounds labeled from 34 to 43 in Table 3) and ii) samples from "Vinagre de Condado de Huelva" PDO seem to have higher amount of acetic acid (feature named 18 in Table 3) and ethanol (features 8 and 37, Table 3 with respect to the other two PDOs (separation on PC5).

3.2. Classification results of individual datasets.

In a first stage, separate classification models (PLS-DA for MIR+NIR and ^{1}H-NMR data sets and NPLS-DA as described in previous sections for EEM data) were built on the data coming from the different instrumental techniques. The distinct datasets were split in the same training and test sets of 47 and 19 samples as described in Section 2.3.

The classification results obtained by the application of PLS-DA and NPLS-DA on each separate data set, according to the classification criterion described in Section 2.3.3, are reported in Table 4, which reports for each spectroscopic technique the data preprocessing, the model dimensionality (assessed by cross-validation) and the classification performance. PLS-DA was built on the PCA scores (8 PCs) for the MIR-NIR data set, and on the sixty-two peak areas of MCR resolved components for the NMR data set, respectively. While for EEM data set, NPLS-DA was directly built on the spectral data array (samples x excitation wavelengths x emission wavelengths).

The classification results, in calibration, are promising for ^{1}H-NMR models (correct classification rates higher than 90% for all categories). The model dimensionality, i.e. 7 components, is lower with respect to MIR+NIR, i.e. 10, and EEM, i.e. 12, probably because in this case peak areas of resolved spectral components are used instead of the spectroscopic signal itself. In contrast, the models built on MIR+NIR data and EEM show quite good classification rates, only for one of the category, namely "Condado" and "Jerez" for MIR+NIR and EEM, respectively. These results agree with what already observed in our previous studies [6,7], in which it was shown that these techniques had a better ability to distinguish between categories (aging and sweet) than among the different PDOs.

It can also be observed, that NPLS-DA requires an higher number of latent variables, with respect to the number of PARAFAC factors obtained for EEM data (i.e. five), this

could be explained by the fact that NPLS-DA (as PLS usually does) modulates the main fluorophores present in the matrix as well as the environment effects ant the interferences.

On the other hand, the predictive capability (external validation) was almost similar for all the techniques. In general, the results could be considered fairly good, taking into account that, due to the limited number of test samples, for example, in the case of ^{1}H-NMR, 75% correct prediction rates for the classes "Vinagre de Jerez" and "Vinagre de Montilla-Moriles" PDO correspond to 2 and 1 misclassified samples, respectively. In all the prediction models, the same sample of "Vinagre de Jerez" PDO sample was misclassified; also one sample of "Vinagre de Montilla-Moriles" PDO was always misclassified.

Furthermore, it can be observed that prediction rates were higher for "Vinagre de Condado de Huelva" (MIR-NIR and ^{1}H-NMR models) and "Vinagre de Jerez" (EEM model) with respect to "Vinagre de Montilla-Moriles"**.** This fact could be mainly explained by the relative new recognition of this PDO (included in the European Register of Protected Geographical Indications and Protected Designation (PGI) in 2015), in comparison with the other two PDOs, "Vinagre de Condado de Huelva" and "Vinagre de Jerez" PDOs, and specially the last one that was the first wine vinegar PDO of Spain [3]. Furthermore, this is in agreement with our previous studies [7].

To summarize, even though the results are quite promising, the quality of each model was not enough good for the characterization and classification purpose and it varied significantly from one technique to another.

3.3. Mid-level data fusion

The results described in Sections 3.1.4 showed that classification models built on each of the individual data matrices are not accurate enough, indicating that a single instrumental fingerprint is not completely able to correctly predict the high-complex samples under study. For this reason, the possibility of combining the information from the different instruments by means of mid-level data fusion strategy was investigated.

The features obtained from the decomposition of the single data blocks (i.e. the eight MIR+NIR PCA scores, the five factors EEM PARAFAC scores and the peak areas of the sixty-two resolved ^{1}H-NMR MCR components) were merged in a unique block as described in Section 2.3.4 (Fig. 1). Since scaling is a critical issue both block-autoscaling and autoscaling (Section 2.3.4.) were compared.

Explorative PCA models were built with the fused data preprocessed by both scaling's methods and results shown in Fig. 3. The autoscaled data (Fig. 3.A) showed a similar clustering of the three PDOs as the one observed in the score plot of ^{1}H-NMR PCA reported in Fig. IV.A Supplementary Material. In particular, PC1 distinguish the samples belonging to the sweet category at positive values of PC1. "Vinagre de Montilla-Moriles" PDO showed positive scores values on PC5, whereas "Vinagre de Condado de Huelva" PDO samples showed negative scores values for this component and samples of "Vinagre de Jerez" PDO are placed again in the middle. Fig. 3.B shows the loading plot of the same principal components, in which the contribution of several of the features, both from ^{1}H-NMR and MIR-NIR was observed. PC5, PC2 and PC8 from MIR-NIR PCA, as well as several of the NMR features, seem to be the main responsible features for the improvement in the separation of "Vinagre de Condado de Huelva" and "Vinagre de Montilla-Moriles" samples. In fact, they have high negative loadings values on the fifth component of the PCA on fused data, while at positive loadings values there are PC4 and PC6 from MIR-NIR PCA and F1-F3 from PARAFAC. PC1 from MIR-NIR PCA seems of relevance in the Pedro Ximenez samples separation from the rest, since its high positive loadings on the first component of the PCA on fused data.

Figure 3 to be inserted about here

Even if few minor differences were noticed with respect to ^{1}H-NMR data analysis, some improvements in the separation of PDOs occurred. The similarity between the fused autoscaled data and the ^{1}H-NMR data block is explained by the fact that using autoscaling as merging strategy, a higher importance is given to the block of variables more numerous, hence, the ^{1}H-NMR data.

Regarding the block-autoscaling PCA results (Fig. 3B), the principal components that better shows a separation were PC1, PC2 and PC5. In this scores plot, the separation of PDOs seems to be worse than with autoscaling procedure. Thus, a higher overlapping between "Vinagre de Jerez" and "Vinagre de Condado de Huelva" samples was observed. In spite of this, "Vinagre de Jerez" PDO seems to be mainly placed in the negative side of PC1 while "Vinagre de Condado de Huelva" in the positive side of PC1 and PC5, and "Vinagre de Montilla-Moriles" PDO in the positive side of PC2. The loadings plot (Fig. 3.D) shows in this particular scaling procedure that ^{1}H-NMR components had lower relevance and the MIR-NIR and EEM variables became more influential. Thus, PC3 (MIR-NIR) and F5 (EEM) showed the most negative

contribution of PC1, while F4, F1 and PC1 the most positive, as well as PC2 and PC8 of MIR-NIR data had the most positive values of PC5, relevant for the separation of "Vinagre de Condado de Huelva" PDO.

Then, PLS-DA models were built using six and seven latent variables for autoscaling and block-autoscaling procedures, respectively (chosen accordingly to minimum cross validation classification errors). The results obtained are reported in Table 5. They confirmed the improvement with respect to the classification models obtained for the separate data blocks. In fact, 100% of correct classification was obtained for the predicted samples (test set) of all the PDOs, as well as 100% of "Vinagre de Condado de Huelva" PDO samples were correctly classified in both fit and the prediction. The two scaling procedures give very similar PLS-DA classification rates, only the number of latent variables were different.

In order to identify the most effective variables in discriminating the PDO samples, the values of the PLS-DA regression vectors and the variable importance in projection (VIP) index were studied; for interpretative purposes all the predictors having a VIP>1 are considered to be relevant [69]. Despite the different scaling procedure, the variables with VIP higher than one quite matched in both PLS-DA models and are reported in Table 6 together with the sign of the corresponding regression coefficients. Accordingly, the most relevant variables for the discrimination of the "Vinagre de Condado de Huelva" PDO were mainly MIR-NIR PC2, PC3, PC5 and PC8 previously described as the spectral regions related to the presence of acetic acid and ethanol (~1410 and ~1290 cm^{-1} and 1045 cm^{-1} in MIR spectra) as well as alcohol compounds, aldehydes, and some esters and ethers that matched with PC3 loadings. Other important variables were EFM F1 and F4 that matched with the presence of phenolic compounds and NMR7, NMR11, NMR16, NMR18 and NMR27 that were interpreted as isopropanol, acetic acid, acetoin and some other compounds such as 6-acetylglucose, beta-alanine and succinates.

Regarding "Vinagre de Jerez" PDO, this PDO was described mostly by the variables PC3 and PC8 of MIR-NIR PCA, related to alcohol compounds, aldehydes, esters, ethers and acids and commonly presented in grapes, wine and vinegar; EEM F5 related to grape sugars, furfural and Maillard compounds more presented the Pedro Ximenez category included in this PDO, together with F1 and F4 again; and NMR16, NMR26 and NMR59 identified as 6-acetylglucose, aminoacids as malate, glutarate or n-acetylglutamate and formic acid, respectively.

Finally, the variables that seems to give a relevant contribution for the classification of "Vinagre de Montilla-Moriles" PDO were mainly: MIR-NIR PC5 and PC8 whose loadings mainly showed a peak at 1045 cm^{-1} and PC1 again related to the Pedro Ximenez samples of this PDO; EFM F1 and F5, which brings mainly the information of the compounds commonly presented in grapes and wine such as cumarins, tannins, phenols, flavonols, and moreover, compounds related to the sweet category such as HMF and sugars also related to the NMR most relevant variables according to the VIPs (i.e. compounds from NMR32 to NMR52). These results agree with those obtained in the loadings plot of the PCA model previously described (Fig. 3).

3.4. P-ComDim.

P-ComDim was carried out with the raw spectral data (Fig. 4a and Fig. 5a) and the data of the extracted features (Fig. 4b and Fig. 5b) in order to study the best approach that show the complementarity of the techniques and therefore also their differences. Fig. 4 and Fig. 5 shows the saliences and the global loadings obtained [28,70] for each technique, respectively.

Figure 4 to be inserted about here

In Fig. 4 on the top is shown the percentage of variance extracted by each common component (graph on top left), the sum of saliences of all data tables for each common component (graph on top middle) and the sum of saliences for each data table over all the calculated common components (graph on top right). Taking into account the normalization of the single data table, the sum of saliences in the latter plot can be at maximum equal to 1, when no residual variance is left, for that data table after extracting the common components. In the bottom part of Fig. 4 are shown the saliences of each data table on each common component. The sum of the saliences reported on top of each graph corresponds to the values reported on the top middle graph. The first two components explain most of the data variance but taking into account eight components allows describing all data tables.

The analysis of salience for the raw spectral data (Fig. 4.A) show that MIR and NIR share mainly one common component, i.e. CC1, while EEM and ^{1}H-NMR data seem to capture most distinctive information, contributing to different components, namely CC2 for EEM and CC3, CC4, CC5 and CC6 for ^{1}H-NMR. Despite with lower weights, CC8 is common to MIR, NIR and ^{1}H-NMR data blocks and CC7 to all of them. Regarding the loadings vectors associated to each block (Fig. 5.A), CC1 seemed to be related to

the Pedro Ximenez category due to the intense band showed in MIR and NIR loadings plot (between 1000-1150 cm-1 and 5200 and 6500 cm^{-1}, respectively) and in ^{1}H-NMR data point to a higher intensity in the sugar region of the spectra (from 3.22-4.12 ppm); CC2, considering the excitation and emission wavelengths of the EEM reshaped landscapes, resemble the first PARAFAC factor (Fig. 2), while CC3 was related to the first region of the ^{1}H-NMR spectra were acids (e.g. acetic acid), alcohols (e.g. ethanol) and some esters (isobutyrate) appear. Finally, CC7 seemed to be associated to the presence of acetic acid and ethanol that could be observed by NIR, MIR and ^{1}H-NMR techniques and, as far as EEM loadings are concerned, the profile resembles those of the fourth PARAFAC factor which was associated to phenols compounds.

Figure 5 to be inserted about here

In the case of P-ComDim model, obtained with the extracted features of each data block (Fig. 4.B), EEM (data table numbered as 2 in the figure) has again little in common with the other data tables and mainly contribute to CC1 and CC4, which by inspection of loadings are related to the first four PARAFAC factors (CC1) and second, third and fifth factors (CC3), respectively. ^{1}H-NMR data contribute mainly to CC5 and CC6 together with MIR and NIR data, i.e. these global components are shared by these data tables and, hence, should reflect the samples trends common to ^{1}H-NMR and MIR-NIR. CC2 is mainly contributing the NMR data table and the respective loadings (Fig. 5B) show high influence of the first region of the ^{1}H-NMR spectra (alcohols and acids). CC3, CC7 and CC8 are mainly contributing the MIR-NIR data table, in particular, according to the loadings plot (Fig. 5B), the PC6 and PC8 scores of PCA decomposition of NIR-MIR spectra.

Fig. 6 illustrates the global scores scatter plot obtained by P-ComDim analysis (the bottom plot in Fig. 6A and the bottom right one in Fig. 6B). In comparison to PCA analysis of individual spectral data sets (Fig. II and Fig. IV Supplementary Material), ComDim clearly shows an increased separation trend according to the PDO, even though this separation was slightly worse than in the PCA obtained on the mid-level fused data (Fig. 3). These results are consistent with the fact that the global scores scatter plot of P-ComDim obtained on the extracted features data tables, i.e. corresponding to the data used for the mid-level data fusion, show a better separation among PDOs than the ComDim performed on the raw spectral data. These results could be better observed by the scores plot of the PLS-DA models obtained for each approach

(Fig. 6). Thus, this latter figure showed that more overlapping occurs when PLS-DA is carried out with raw data than by using the extracted features of each data set (i.e. six samples were not correctly predicted by the raw data model with respect to the two samples wrongly predicted by the model with extracted features). Nonetheless, one advantage of performing P-ComDim directly on the raw spectra is the interpretation of the spectral regions contribution by visualization of the corresponding local loadings.

Figure 6 to be inserted about here

The classification results expressed as percentage of corrected classified by means of PLS-DA model carried out with P-ComDim results are reported in Table 5 together with the classification results obtained by the mid-level data fusion models. Looking at the table it can be noticed once more that the results obtained by the PLSDA performed on the P-ComDim scores from the extracted features were better than the PLS-DA results obtained by each data set individually studied, only comparable to the ^{1}H-NMR results, as well as they were better than the P-ComDim classification model developed with raw data. However, in spite the promising classification rates obtained by the P-ComDim with the extracted features, the classification results were inferior to the results obtained by Mid-level data fusion.

4. CONCLUSIONS

This study demonstrates the potential of the combination of four spectroscopic analytical methods (MIR, NIR, EFM and ^{1}H-NMR) when they were combined. The application of data fusion methods improved the characterization and authentication of PDO wine vinegars, providing a more efficient differentiation than the models based on single methods. The obtained results support the approach of combining these methods to achieve synergies for an optimized differentiation of the PDO of wine vinegars. With regard to single analytical methods, especially the classification results of ^{1}H-NMR models were promising. On the other hand, the application of P-ComDim method was useful for describing, in a simple and synthetic manner, the overall spectral information collected and reveal the complementarity and differences of the spectroscopic techniques, assessing the importance of each technique to each of the common variables. However, for a PDO classification objective, the results of the present work showed that Mid-level data fusion can be the better option in comparison to the classification models obtained by P-ComDim. In spite of this fact, this study presents

promising results related to the development of efficient classification models by P-ComDim carried out with the extracted features of spectroscopic data.

5. ACKNOWLEDGEMENTS

The authors would like to thank the Spanish Regulatory Councils of each wine vinegar PDO— "Vinagre de Jerez" ("Sherry wine vinegar"), "Vinagre de Montilla-Moriles" and "Vinagre de Condado de Huelva"— for their invaluable help with the acquisition of the samples. Authors also acknowledge the Nuclear Magnetic Resonance services of the University of Seville (CITIUS) for the ^{1}H-NMR analyses. This work was supported by "Consejería de Economía, Innovación y Ciencia, Junta de Andalucía" [P12-AGR-1601]; and the FPU scholarship of "Ministerio de Educación, Cultura y Deporte" [FPU014/01247]. The authors are also very grateful to Prof. D. Rutledge for providing access to the P-ComDim code and for helping to apply it.

6. REFERENCES

[1] M. Cocchi, Chemometrics for Food Quality Control and Authentication, in: R.A. Meyers (Ed.), Encycl. Anal. Chem., John Wiley & Sons, Ltd, 2017. doi:10.1002/9780470027318.a9579.

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

[3] Council Regulation (EC) No 510/2006 of 20 March 2006 on the protection of geographical indications and designations of origin for agricultural products and foodstuffs., 2006.

[4] M. Cocchi, C. Durante, a Marchetti, C. Armanino, M. Casale, Characterization and discrimination of different aged "Aceto Balsamico Tradizionale di Modena" products by head space mass spectrometry and chemometrics., Anal. Chim. Acta. 589 (2007) 96–104. doi:10.1016/j.aca.2007.02.036.

[5] A. Marrufo-Curtido, M.J. Cejudo-Bastante, E. Durán-Guerrero, R. Castro-Mejías, R. Natera-Marín, F. Chinnici, C. García-Barroso, Characterization and differentiation of high quality vinegars by stir bar sorptive extraction coupled to gas chromatography-mass spectrometry (SBSE-GC-MS), LWT - Food Sci. Technol. 47 (2012) 332–341. doi:10.1016/j.lwt.2012.01.028.

[6] R. Ríos-Reina, R.M. Callejón, C. Oliver-Pozo, J.M. Amigo, D.L. García-González, ATR-FTIR as a potential tool for controlling high quality vinegar categories, Food Control. 78 (2017) 230–237. doi:10.1016/j.foodcont.2017.02.065.

[7] R. Ríos-Reina, S. Elcoroaristizabal, J.A. Ocaña-González, D.L. García-González, J.M. Amigo, R.M. Callejón, Characterization and authentication of Spanish PDO wine vinegars using multidimensional fluorescence and chemometrics, Food Chem. 230 (2017) 108–116. doi:10.1016/j.foodchem.2017.02.118.

[8] M.I. Guerrero, C. Herce-Pagliai, A.M. Cameán, A.M. Troncoso, a. G. González, Multivariate characterization of wine vinegars from the south of Spain according to their metallic content, Talanta. 45 (1997) 379–386. doi:10.1016/S0039-9140(97)00139-2.

[9] P. Paneque, M.L. Morales, P. Burgos, L. Ponce, R.M. Callejón, Elemental characterisation of Andalusian wine vinegars with protected designation of origin by ICP-OES and chemometric approach, Food Control. 75 (2017) 203–210. doi:http://dx.doi.org/10.1016/j.foodcont.2016.12.006.

[10] R.M. Callejón, M.L. Morales, A.C. Silva Ferreira, A.M. Troncoso, Defining the typical aroma of Sherry vinegar: Sensory and chemical approach, J. Agric. Food Chem. 56 (2008) 8086–8095. doi:10.1021/jf800903n.

[11] R. Ríos-Reina, M.. Morales, D.L. García-González, J.M. Amigo, R.M. Callejón, Sampling methods for the study of volatile profile of PDO wine vinegars. A comparison using multivariate data analysis., Food Res. Int. 105 (2018) 880–896. doi:10.1016/j.foodres.2017.12.001.

[12] R. Ríos-Reina, D.L. García-González, R.M. Callejón, J.M. Amigo, NIR spectroscopy and chemometrics for the typification of Spanish wine vinegars with a protected designation of origin, Food Control. 89 (2018) 108–116. doi:10.1016/j.foodcont.2018.01.031.

[13] M. Silvestri, L. Bertacchini, C. Durante, A. Marchetti, E. Salvatore, M. Cocchi, Application of data fusion techniques to direct geographical traceability indicators, Anal. Chim. Acta. 769 (2013) 1–9. doi:10.1016/j.aca.2013.01.024.

[14] D. Lahat, T. Adali, C. Jutten, Multimodal Data Fusion : An Overview of Methods ,Challenges, and Prospects, Proc. IEEE. 103 (2015) 1449–1477. doi:10.1109/jproc.2015.2460697.

[15] 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. doi:10.1016/j.aca.2014.02.024.

[16] C. Pizarro, S. Rodríguez-Tecedor, N. Pérez-Del-Notario, I. Esteban-Díez, J.M. González-Sáiz, Classification of Spanish extra virgin olive oils by data fusion of visible spectroscopic fingerprints and chemical descriptors, Food Chem. 138 (2013) 915–922. doi:10.1016/j.foodchem.2012.11.087.

[17] M. Silvestri, A. Elia, D. Bertelli, E. Salvatore, C. Durante, M. Li Vigni, A. Marchetti, M. Cocchi, A mid level data fusion strategy for the Varietal Classification of Lambrusco PDO wines, Chemom. Intell. Lab. Syst. 137 (2014)

181–189. doi:10.1016/j.chemolab.2014.06.012.

[18] B.P. Geurts, J. Engel, B. Ra, L. Blanchet, A. Suppers, E. Szyma, J.J. Jansen, L.M.C. Buydens, Chemometrics and Intelligent Laboratory Systems Improving high-dimensional data fusion by exploiting the multivariate advantage, 156 (2016) 231–240. doi:10.1016/j.chemolab.2016.05.010.

[19] I. Van Mechelen, A.K. Smilde, Chemometrics and Intelligent Laboratory Systems A generic linked-mode decomposition model for data fusion, Chemom. Intell. Lab. Syst. 104 (2010) 83–94. doi:10.1016/j.chemolab.2010.04.012.

[20] T.G. Doeswijk, A.K. Smilde, J.A. Hageman, J.A. Westerhuis, F.A. Van Eeuwijk, Analytica Chimica Acta On the increase of predictive performance with high-level data fusion, Anal. Chim. Acta. 705 (2011) 41–47. doi:10.1016/j.aca.2011.03.025.

[21] E.M. Qannari, I. Wakeling, P. Courcoux, H.J.H. MacFie, Defining the underlying sensory dimensions, Food Qual. Prefer. 11 (2000) 151–154. doi:http://dx.doi.org/10.1016/S0950-3293(99)00069-5.

[22] E.M. Hanafi, M. Qannari, Nouvelles propriétés de l'analyse en composantes communes et poids spécifiques, 149 (2008) 75–97.

[23] M. Hanafi, G. Mazerolles, E. Dufour, E.M. Qannari, Common components and specific weight analysis and multiple co-inertia analysis applied to the coupling of several measurement techniques, (2007) 172–183. doi:10.1002/cem.

[24] A. El Ghaziri, V. Cariou, D.N. Rutledge, E.M. Qannari, Analysis of multiblock datasets using ComDim : Overview and extension to the analysis of (K + 1) datasets, (2016) 420–429. doi:10.1002/cem.2810.

[25] V. Cariou, E.M. Qannari, D.N. Rutledge, E. Vigneau, ComDim: From multiblock data analysis to path modeling, Food Qual. Prefer. (2016) 1–8. doi:10.1016/j.foodqual.2017.02.012.

[26] M. Hohmann, Y. Monakhova, S. Erich, N. Christoph, H. Wachter, U. Holzgrabe, Differentiation of Organically and Conventionally Grown Tomatoes by Chemometric Analysis of Combined Data from Proton Nuclear Magnetic Resonance and Mid-infrared Spectroscopy and Stable Isotope Analysis, J. Agric.

Food Chem. 63 (2015) 9666–9675. doi:10.1021/acs.jafc.5b03853.

[27] S. Erich, S. Schill, E. Annweiler, H.-U. Waiblinger, T. Kuballa, D.W. Lachenmeier, Y.B. Monakhova, Combined chemometric analysis of 1H NMR, 13C NMR and stable isotope data to differentiate organic and conventional milk, Food Chem. 188 (2015) 1–7. doi:10.1016/j.foodchem.2015.04.118.

[28] G. Mazerolles, M.. Devaux, E. Dufour, E.. Qannari, P. Courcoux, Chemometric methods for the coupling of spectroscopic techniques and for the extraction of the relevant information contained in the spectral data tables, Chemom. Intell. Lab. Syst. 63 (2002) 57–68. doi:10.1016/S0169-7439(02)00036-9.

[29] Y.B. Monakhova, M. Hohmann, N. Christoph, H. Wachter, D.N. Rutledge, Improved classification of fused data: Synergetic effect of partial least squares discriminant analysis (PLS-DA) and common components and specific weights analysis (CCSWA) combination as applied to tomato profiles (NMR, IR and IRMS), Chemom. Intell. Lab. Syst. 156 (2016) 1–6. doi:10.1016/j.chemolab.2016.05.006.

[30] R. Karoui, C. Blecker, Fluorescence Spectroscopy Measurement for Quality Assessment of Food Systems-a Review, Food Bioprocess Technol. 4 (2011) 364–386. doi:10.1007/s11947-010-0370-0.

[31] C. Alamprese, M. Casale, N. Sinelli, S. Lanteri, E. Casiraghi, Detection of minced beef adulteration with turkey meat by UV-vis, NIR and MIR spectroscopy, LWT - Food Sci. Technol. 53 (2013) 225–232. doi:10.1016/j.lwt.2013.01.027.

[32] G. Downey, R. Briandet, R.H. Wilson, E.K. Kemsley, Near- and Mid-Infrared Spectroscopies in Food Authentication: Coffee Varietal Identification, J. Agric. Food Chem. 45 (1997) 4357–4361. doi:10.1021/jf970337t.

[33] R. Consonni, L.R. Cagliani, F. Benevelli, M. Spraul, E. Humpfer, M. Stocchero, NMR and Chemometric methods: A powerful combination for characterization of Balsamic and Traditional Balsamic Vinegar of Modena, Anal. Chim. Acta. 611 (2008) 31–40. doi:10.1016/j.aca.2008.01.065.

[34] M. Hohmann, N. Christoph, H. Wachter, U. Holzgrabe, 1H NMR profiling as an approach to differentiate conventionally and organically grown tomatoes, J.

Agric. Food Chem. 62 (2014) 8530–8540. doi:10.1021/jf502113r.

[35] R.D. Snee, Validation of Regression Models: Methods and Examples, Technometrics. 19 (1977) 415–428. doi:10.1080/00401706.1977.10489581.

[36] R. Bro, PARAFAC. Tutorial and applications, Chemom. Intell. Lab. Syst. 38 (1997) 149–171. doi:10.1016/S0169-7439(97)00032-4.

[37] R. Bro, Multi-way analysis in the food industry. Models, algorithms, and applications., (1998).

[38] A. de Juan, J. Jaumot, R. Tauler, Multivariate Curve Resolution (MCR). Solving the mixture analysis problem, Anal. Methods. 6 (2014) 4964. doi:10.1039/c4ay00571f.

[39] A. de Juan, R. Tauler, Multivariate Curve Resolution (MCR) from 2000: Progress in Concepts and Applications, Crit. Rev. Anal. Chem. 36 (2006) 163–176. doi:10.1080/10408340600970005.

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

[41] M. Salvatore, E., Bevilacqua, M., Bro, R. Marini, F. and Cocchi, Classification Methods of Multiway Arrays as a Basic Tool for Food PDO Authentication, in: A. De La Guardia, M, Gonzalvez (Ed.), Compr. Anal. Chem. Food Prot. Des. Orig. Vol. 60, Barcelo, D, Wiley, 2013: pp. 339–379.

[42] A.G.E.K. Smilde, Comments on multilinear pls, J. Chemom. 11 (1997) 367–377.

[43] U.G. Indahl, H. Martens, T. Næs, From dummy regression to prior probabilities in PLS-DA, J. Chemom. 21 (2007) 529–536. doi:10.1002/cem.1061.

[44] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra, Appl. Spectrosc. 43 (1989) 772–777.

[45] M.J. Sáiz-Abajo, J.M. González-Sáiz, C. Pizarro, Classification of wine and alcohol vinegar samples based on near-infrared spectroscopy. Feasibility study on the detection of adulterated vinegar samples, J. Agric. Food Chem. 52 (2004) 7711–7719. doi:10.1021/jf049098h.

[46] M. Casale, M.J. Sáiz Abajo, J.M. González Sáiz, C. Pizarro, M. Forina, Study of the aging and oxidation processes of vinegar samples from different origins during storage by near-infrared spectroscopy, Anal. Chim. Acta. 557 (2006) 360–366. doi:10.1016/j.aca.2005.10.063.

[47] S. Elcoroaristizabal, R. Bro, J.A. García, L. Alonso, PARAFAC models of fluorescence data with scattering: A comparative study, Chemom. Intell. Lab. Syst. 142 (2015) 124–130. doi:10.1016/j.chemolab.2015.01.017.

[48] R. Bro, H.A.L. Kiers, A new efficient method for determining the number of components in PARAFAC models, J. Chemom. 17 (2003) 274–286. doi:10.1002/cem.801.

[49] G. Tomasi, F. Savorani, S.B. Engelsen, Icoshift: An effective tool for the alignment of chromatographic data, J. Chromatogr. A. 1218 (2011) 7832–7840. doi:10.1016/j.chroma.2011.08.086.

[50] H.F.M. Boelens, R.J. Dijkstra, P.H.C. Eilers, F. Fitzpatrick, J.A. Westerhuis, New background correction method for liquid chromatography with diode array detection, infrared spectroscopic detection and Raman spectroscopic detection, J. Chromatogr. A. 1057 (2004) 21–30. doi:10.1016/j.chroma.2004.09.035.

[51] E.F. Boffo, L.A. Tavares, M.M.C. Ferreira, A.G. Ferreira, Classification of Brazilian vinegars according to their 1H NMR spectra by pattern recognition analysis, LWT - Food Sci. Technol. 42 (2009) 1455–1460. doi:10.1016/j.lwt.2009.05.008.

[52] C. Fotakis, K. Kokkotou, P. Zoumpoulakis, M. Zervou, NMR metabolite fingerprinting in grape derived products: An overview, Food Res. Int. 54 (2013) 1184–1194. doi:10.1016/j.foodres.2013.03.032.

[53] W. Winding, in: S. Brown, R. Tauler, R. Walckzak (Eds.), Comprehensive Chemom. Vol. 2, Elsevier, Oxford, 2009: pp. 275–307.

[54] G. Papotti, D. Bertelli, R. Graziosi, A. Maietti, P. Tedeschi, A. Marchetti, M. Plessi, Traditional balsamic vinegar and balsamic vinegar of Modena analyzed by nuclear magnetic resonance spectroscopy coupled with multivariate data analysis, LWT - Food Sci. Technol. 60 (2015) 1017–1024. doi:10.1016/j.lwt.2014.10.042.

[55] A. Caligiani, D. Acquotti, G. Palla, V. Bocchi, Identification and quantification of the main organic components of vinegars by high resolution 1H NMR spectroscopy, Anal. Chim. Acta. 585 (2007) 110–119. doi:10.1016/j.aca.2006.12.016.

[56] E. Dubin, M. Spiteri, A.S. Dumas, J. Ginet, M. Lees, D.N. Rutledge, Common components and specific weights analysis: A tool for metabolomic data pre-processing, Chemom. Intell. Lab. Syst. 150 (2016) 41–50. doi:10.1016/j.chemolab.2015.11.005.

[57] C.B.Y. Cordella, D. Bertrand, SAISIR: A new general chemometric toolbox, TrAC - Trends Anal. Chem. 54 (2014) 75–82. doi:10.1016/j.trac.2013.10.009.

[58] D. Bertand, C. Cordella, SAISIR package. free toolbox for chemometrics in the Matlab, Octave or Scilab environments Available at http:/www.chimiometrie.fr/saisir_ webpage.html2011, (n.d.).

[59] R. Giangiacomo, G.G. Dull, Near Infrared Spectrophotometric Determination of Individual Sugars in Aqueous Mixtures, J. Food Sci. 51 (1986) 679–683. doi:10.1111/j.1365-2621.1986.tb13910.x.

[60] D. Cozzolino, H.E. Smyth, M. Gishen, Feasibility Study on the Use of Visible and Near-Infrared Spectroscopy Together with Chemometrics to Discriminate between Commercial White Wines of Different Varietal Origins, J. Agric. Food Chem. 51 (2003) 7703–7708. doi:10.1021/jf034959s.

[61] J. Moros, F.A. Iñón, S. Garrigues, M. De la Guardia, Determination of vinegar acidity by attenuated total reflectance infrared measurements through the use of second-order absorbance-pH matrices and parallel factor analysis, Talanta. 74 (2008) 632–641. doi:10.1016/j.talanta.2007.06.046.

[62] D. Airado-Rodríguez, I. Durán-Merás, T. Galeano-Díaz, J.P. Wold, Front-face fluorescence spectroscopy: A new tool for control in the wine industry, J. Food Compos. Anal. 24 (2011) 257–264. doi:10.1016/j.jfca.2010.10.005.

[63] S.M. Azcarate, A. De Araújo, M.R. Alcaraz, M.C. Ugulino, D. Araújo, J.M. Camiña, H.C. Goicoechea, Modeling excitation- emission fluorescence matrices with pattern recognition algorithms for classification of Argentine white wines according, 184 (2015) 214–219.

[64] R.M. Callejón, J.M. Amigo, E. Pairo, S. Garmón, J.A. Ocaña, M.L. Morales, Classification of Sherry vinegars by combining multidimensional fluorescence, parafac and different classification approaches, Talanta. 88 (2012) 456–462. doi:10.1016/j.talanta.2011.11.014.

[65] É. Dufour, A. Letort, A. Laguet, A. Lebecque, J.N. Serra, Investigation of variety, typicality and vintage of French and German wines using front-face fluorescence spectroscopy, Anal. Chim. Acta. 563 (2006) 292–299. doi:10.1016/j.aca.2005.11.005.

[66] D. Zhu, B. Ji, H.L. Eum, M. Zude, Evaluation of the non-enzymatic browning in thermally processed apple juice by front-face fluorescence spectroscopy, Food Chem. 113 (2009) 272–279. doi:10.1016/j.foodchem.2008.07.009.

[67] M.C. García Parrilla, F.J. Heredia, A.M. Troncoso, Sherry wine vinegars: Phenolic composition changes during aging, Food Res. Int. 32 (1999) 433–440. doi:10.1016/S0963-9969(99)00105-2.

[68] J. Sádecká, J. Tóthová, P. Májek, Classification of brandies and wine distillates using front face fluorescence spectroscopy, Food Chem. 117 (2009) 491–498. doi:10.1016/j.foodchem.2009.04.053.

[69] T. Mehmood, K.H. Liland, L. Snipen, S. Sæbø, A review of variable selection methods in Partial Least Squares Regression, Chemom. Intell. Lab. Syst. 118 (2012) 62–69. doi:10.1016/j.chemolab.2012.07.010.

[70] G. Mazerolles, M. Hanafi, E. Dufour, D. Bertrand, E.M. Qannari, Common components and specific weights analysis: A chemometric method for dealing with complexity of food products, Chemom. Intell. Lab. Syst. 81 (2006) 41–49. doi:10.1016/j.chemolab.2005.09.004.

FIGURE CAPTIONS

Fig. 1. Graphical representation of the data sets, data analysis flow and data fusion process.

Fig. 2. Emission and Excitation spectra (PARAFAC loadings) of the main fluorophores present in the PDO wine vinegars (A and B). Mean PARAFAC scores of each PDO for the five resolved components (C). The acronyms for the different vinegar PDOs are defined in Table 1.

Fig. 3. 3-D plot of PCA scores and loadings obtained for both data fusion strategies (with autoscaling and block-autoscaling preprocessing). The acronyms for the different vinegar PDOs are defined in Table 1.

Fig. 4. Graph of saliences and sum of saliences obtained by the P-ComDim method developed with the raw data (A) and with the extracted features (B).

Fig 5. Global loadings for each data block and global scores plot obtained by P-ComDim method carried out by using the raw spectral data of MIR, NIR, ^{1}H-NMR and EFM scores (A) and the data of extracted features obtained by MIR-NIR PCA, EFM PARAFAC and ^{1}H-NMR MCR compounds (B).

Fig. 6. Scores for the first two latent variables of the PLS-DA classification model obtained by P-ComDim with the raw data (A) and extracted features (B). The acronyms for the different vinegar PDOs are defined in Table 1. Test samples are represented by filled symbols. The labels (letter indicate the category predicted by the model) highlight misclassified samples.

Table 1. Samples included in the study.

PDO	Category	Ageing	n
"Vinagre de Jerez" (J)	Crianza	≥6 months	11
	Reserva	≥2 years	13
	Pedro Ximenez	-	4
	Total		28
"Vinagre de Condado de Huelva" (C)	Without ageing	0 months	5
	Solera	≥6 months	5
	Reserva	≥2 years	8
	Añada	≥3 years (static system)	3
	Total		21
"Vinagre de Montilla-Moriles" (M)	Crianza	≥6 months	8
	Reserva	≥2 years	3
	Pedro Ximenez	-	5
	Total		16

Table 2. Emission and Excitation maxima of the 5 factor PARAFAC model and their possible matching fluorophores.

	F1	F2	F3	F4	F5
Ex/Em (nm)	**380/450**	**425/520**	**475/565**	**380/425**	**550/630**
Fluorophores	Cumarins, tannins, phenols, flavonols from wine	5-Hydroxmethylfurfural caramel	Vitamin B2 and its principal forms	Phenolic compounds, Maillard products, oxidation products	Unknown related to Pedro Ximenez vinegars

Table 3. MCR resolved, integrated and interpreted components for ^{1}H-NMR data.

RT	Type*	Code	Interpretation
0.86-0.9	t	NMR1	2-Hydorxy-3-methylvalerate
0.9-0.97	d + m	NMR2	X1
0.98-1.02	t+q	NMR3	X3
	-	NMR4	X4
1.03-1.06	d	NMR5	Isobutyrate
1.06-1.11	t	NMR6	Propionate
1.11-1.16	d	NMR7	Isopropanol
1.17-1.20	t	NMR8	Ethanol
1.22-1.29	q	NMR9	X5
1.30-1.34	d+q	NMR10	X6
1.35-1.38	d	NMR11	Acetoin
1.39-1.43	d	NMR12	Lactate/2-Phenylpropionate
1.48-1.53	s+t	NMR13	X7
	-	NMR14,NMR15	X8, X9
1.77-1.81	q	NMR16	6-Acetylglucose
1.97-2.00	s	NMR17	Acetamide
2.02-2.12	s	NMR18	Acetic Acid
2.12-2.14	s/d	NMR19	X10
2.13-2.16	s/d	NMR20	X11
2.16-2.19	s	NMR21	Acetoin
2.21-2.25	s+d	NMR22	Acetone
	dd	NMR23	Acetone
2.28-2.30	s	NMR24	Acetoacetate,Acetylsalicilate
2.32-2.34	d	NMR25	X12
2.37-2.40	s+t	NMR26	Malate, Glutarate, N-Acetylglutamate…
2.59-2.62	t	NMR27	Beta-Alanine, Succinate…
2.64-2.67	s	NMR28	Succinic Acid
2.81-2.85	d	NMR29	X13
2.96-3.01	d	NMR30	X14
3.18-3.21	s	NMR31	Acetylcholine
3.22-3.31	m	NMR32	Glucose
3.30-3.36	d	NMR33	Methanol
3.37-3.51	m+m	NMR34	Glucose
3.51-3.58	m	NMR35	Glucose

3.57-3.65	d	NMR36	Glucose+Fructose
3.63-3.67	q	NMR37	Ethanol
3.67-3.74	m	NMR38	Fructose+Glucose
3.74-3.78	dd	NMR39	Glucose
3.78-3.84	m	NMR40	Fructose
3.84-3.86	d	NMR41	X15
3.87-3.91	dd	NMR42	Fructose+Glucose
3.98-4.03	d+s	NMR43	Frcutose
4.09-4.12	t	NMR44	Frcutose
4.11-4.15	q	NMR45	X17
4.51-4.54	d+s	NMR46	X20
4.56-4.60	d	NMR47	X21
4.62-4.68	d	NMR48	Glucose
4.68-4.71	s	NMR49,NMR50,NMR51	5-HMF
5.21-5.26	d	NMR52	Glucose
5.35-5.39	d	NMR53	X22
	-	NMR54	X23
6.67-6.70	d	NMR55	X24
	-	NMR56	X25
7.52-7.55	d	NMR57	X26
8.25-8.28	s	NMR58, NMR59	Formic Acid
9.43-9.47	s	NMR60, NMR61	5-HMF
9.65-9.68	q	NMR62	X27

* Peak multiplicities: s, singlet; d, doublet; t, triplet; dd, doublet of doublets; q, quadruplet; m, multiplet.

Table 4. Classification results for each individual data block.

DATA	CLASSIFICATION METHOD	PRETREATMENT	LV [a]	% CORRECTED CLASIFIED					
				Train [b]			Test [b]		
				C	J	M	C	J	M
MIR+NIR	PLS-DA	Block Scaling + Mean Centering	10	90.0	85.0	79.2	100	87.5	62.5
EEM	NPLS-DA	Mean centering	12	66.7	95.0	75.0	50.0	100	83.3
^{1}H-NMR peak areas	PLS-DA	Autoscaling	7	100	97.5	91.7	100	75.0	75.0

[a] *LVs number determined on the basis of minimum classification error in CV (Venetian blind 7 splits, keeping replicates in the same set).* [b] *Independent train and test sets, average correct classification rate for 5 random training/ test splitting is reported.*

Table 5. PLS-DA RESULTS OBTAINED BY MID-LEVEL FUSED DATASET WITH TWO DIFFERENT SCALING PROCEDURES

DATASET	CLASSIFICATION METHOD	PRETREATMENT	LVs[a]	% CORRECTED CLASSIFIED					
				Train[b]			Test[b]		
				C	J	M	C	J	M
Mid-Level Data Fusion	PLSDA	Autoscaling	6	100	100	91.7	100	100	100
		Block-Autoscaling	7	100	97.5	91.7	100	100	100
P-Comdim Raw		Autoscaling	2	90.0	97.5	75.0	50.0	75.0	75.0
P-Comdim Extracted Features		Autoscaling	2	96.7	100	87.5	91.7	87.5	87.5

[a] *LVs number determined on the basis of minimum RMSECV with Venetian blind cross validation (7 splits, 2 samples per split).* [b] *Independent test set, average correct classification rate for 5 random training/ test splitting is reported.*

Table 6. Salient variables for discrimination for each PDO category according to PLS-DA VIP values, which were concordant in both DF PLS-DA models, i.e. autoscaling and block-autoscaling. In parenthesis, the sign of the corresponding regression coefficients is reported.

PDOs	NIR-MIR	^{1}H-NMR	EEM
"Vinagre de Condado de Huelva"	PC2(+), PC3(-), PC5(+), PC8 (+)	NMR7(-), NMR11(-), NMR16(+), NMR17(+), NMR18(+), NMR24(-), NMR26(-), NMR27(+), NMR29(+), NMR30(+), NMR31(-)	F1(-),F4(+), F5(+)
"Vinagre de Jerez"	PC1(-), PC2(-), PC3 (+), PC4(-), PC7(-), PC8(-)	NMR14(+), NMR16(-), NMR26(+), NMR27(-), NMR29(-), NMR31(+), NMR59(-)	F1(-), F4(-), F5(+)
"Vinagre de Montilla-Moriles"	PC1(+), PC5(-), PC8(-)	NMR16(-), NMR26(-), NMR27(-), NMR32(+), NMR35(+), NMR36(+), NMR39(+), NMR44(+), NMR48(+), NMR49(+), NMR51(+), NMR59(+), NMR61(+)	F1(+), F5(-)

Figure 1
Click here to download high resolution image

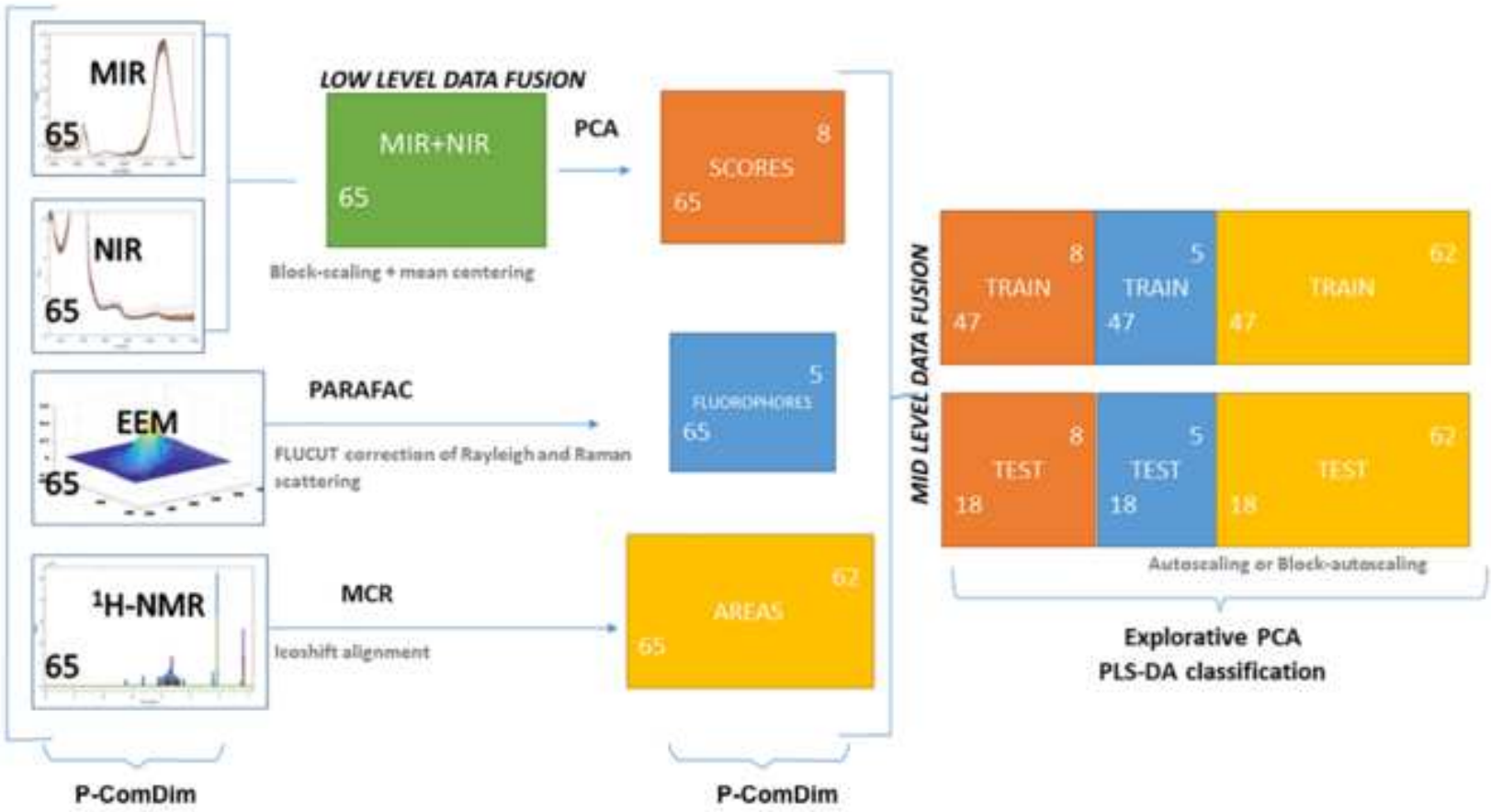


Figure 2
Click here to download high resolution image

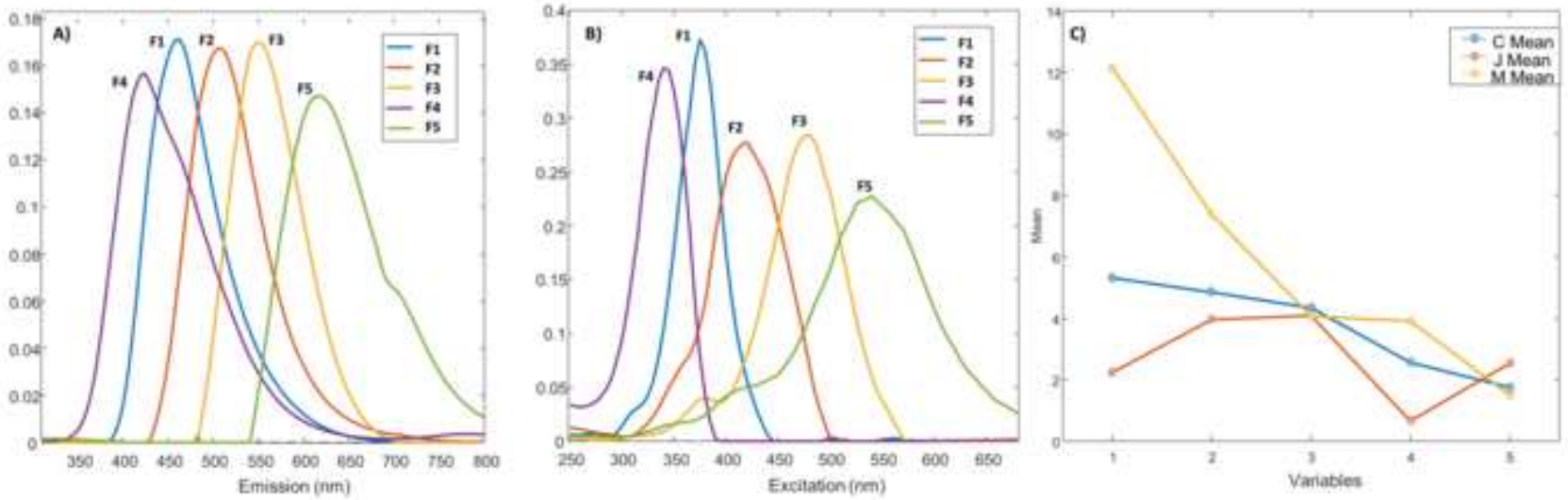


Figure 3
Click here to download high resolution image

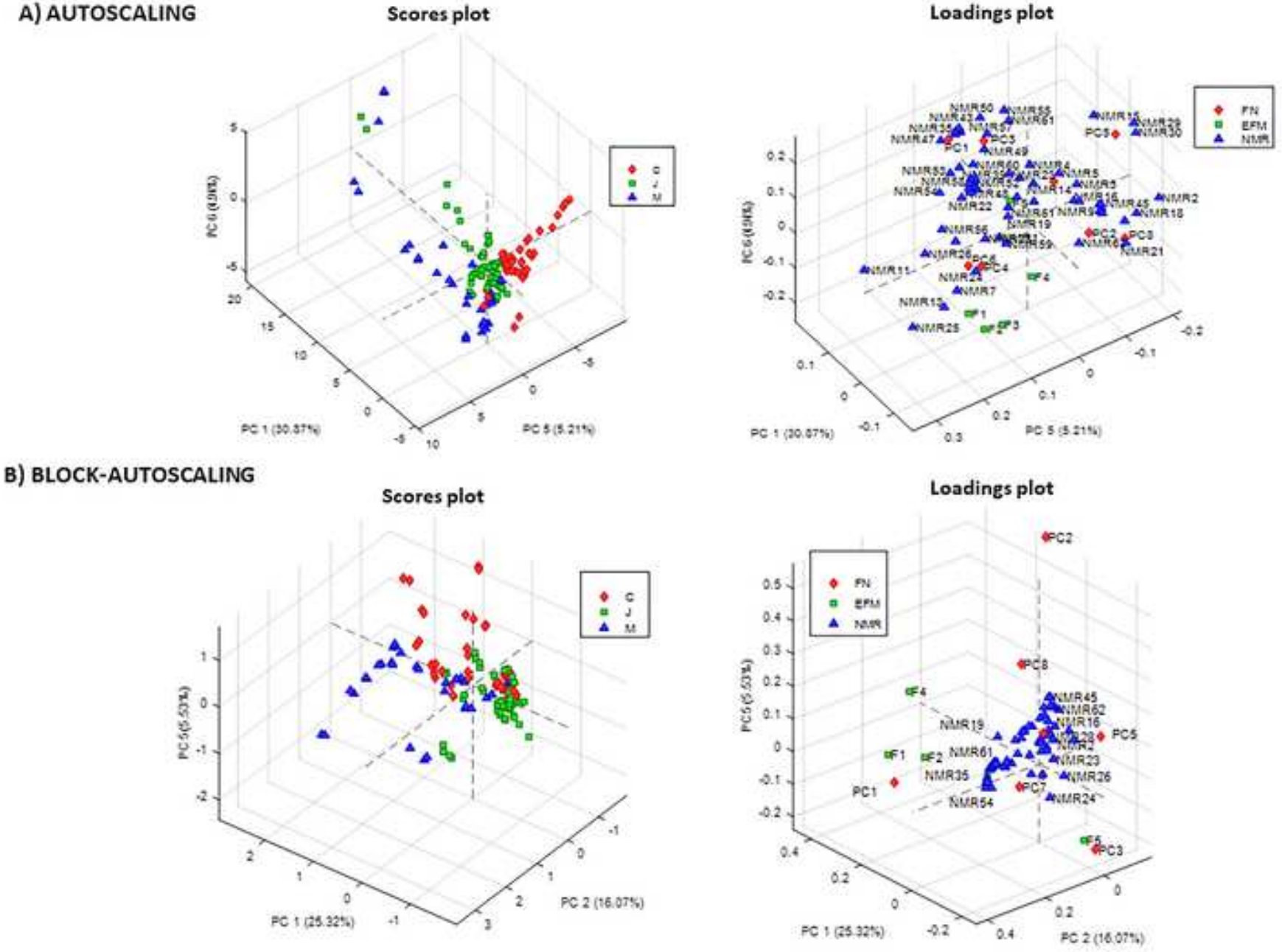


Figure 4

Click here to download high resolution image

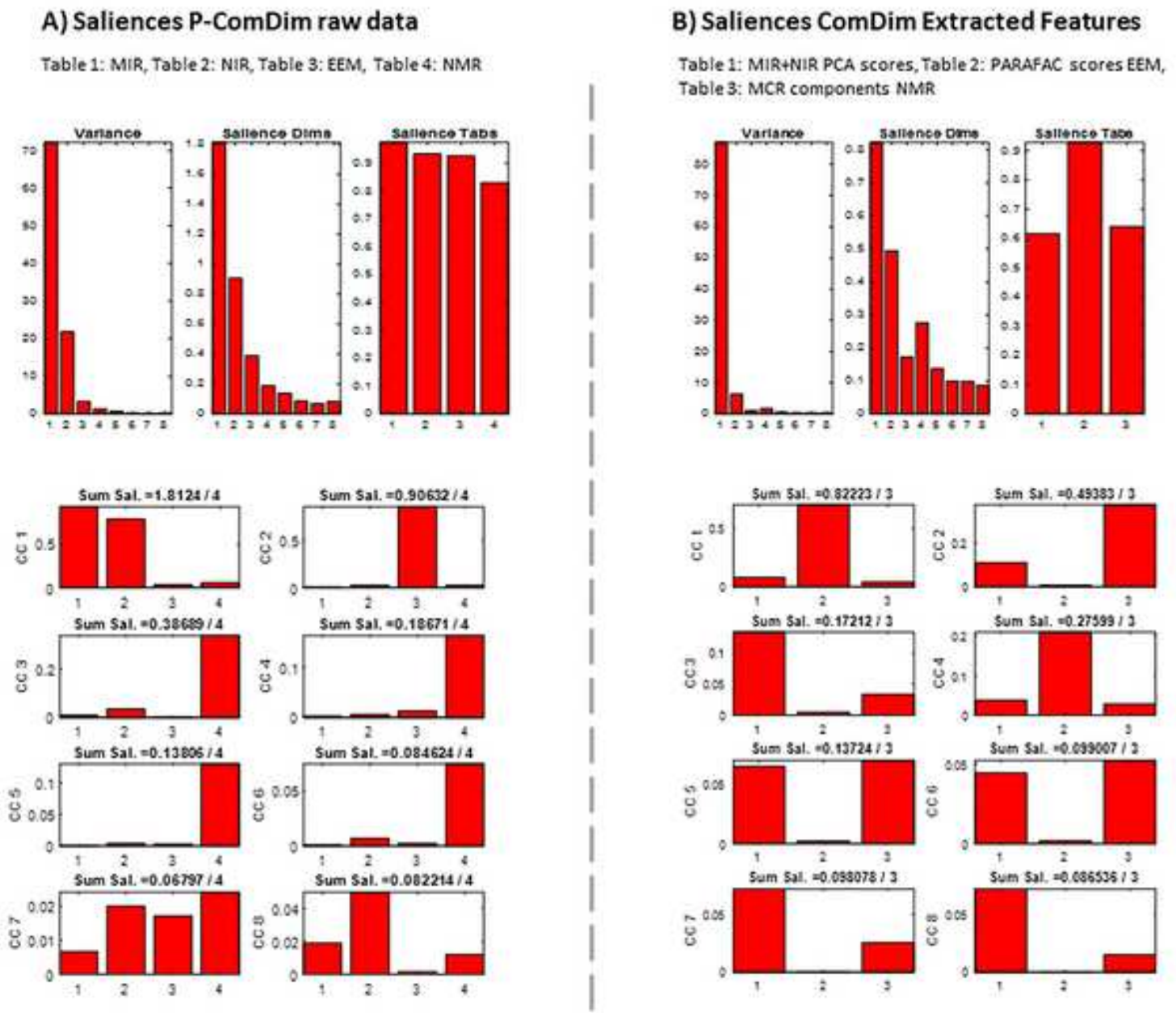


Figure 5
Click here to download high resolution image

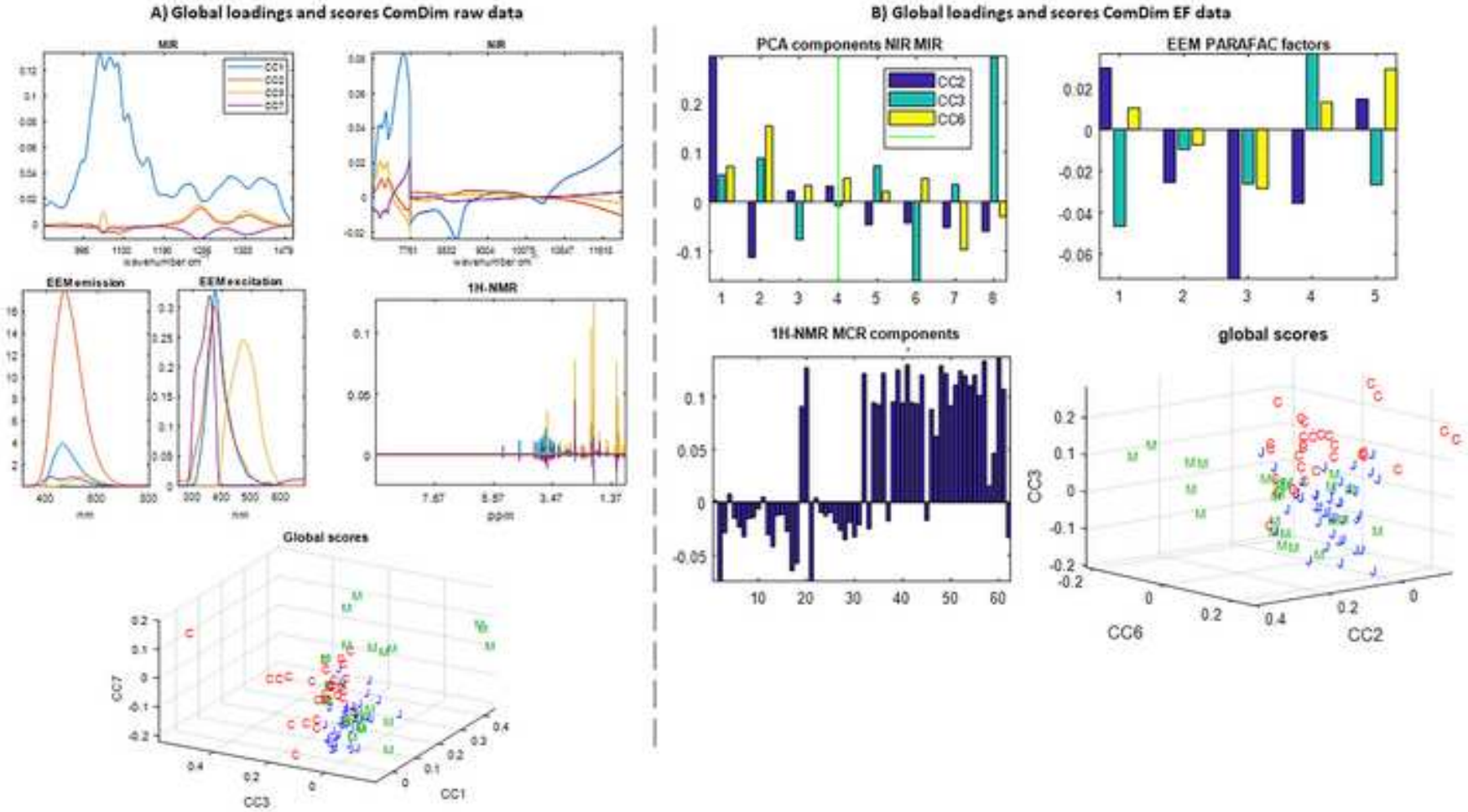
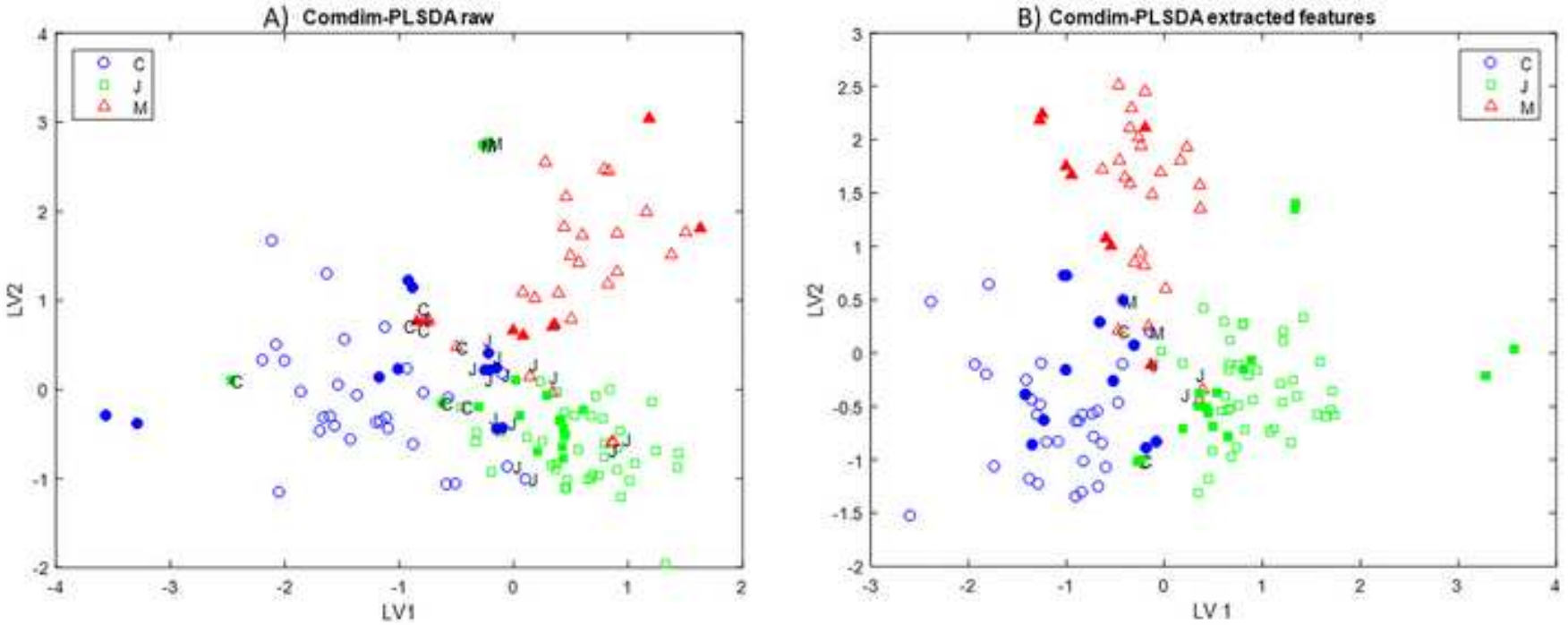


Figure 6
Click here to download high resolution image



Supplementary Material

Click here to download Supplementary Material: Supplementary material Talanta.pdf