Development of a food class-discrimination system by non-targeted NMR analyses using different magnetic field strengths

Rosa Ragone,[a] Stefano Todisco,[a] Maurizio Triggiani,[b,c] Stefania Pontrelli,[c] Mario Latronico,[a,c] Piero Mastrorilli,[a,c] Nicola Intini,[c,d] Claudia Ferroni,[e] Biagia Musio,[a,*] Vito Gallo[a,c,**]

[a]Dipartimento di Ingegneria Civile, Ambientale, del Territorio, Edile e di Chimica (DICATECh), Politecnico di Bari, via Orabona 4, I-70125, Bari, Italy

[b]Dipartimento di Ingegneria Elettrica e dell'Informazione, Politecnico di Bari, via Orabona 4, I-70125, Bari, Italy

[c]Innovative Solutions S.r.l. – Spin Off del Politecnico di Bari, zona H 150/B, I-70015, Noci (BA), Italy

[d]Agenzia Regionale per la Prevenzione e la Protezione dell'Ambiente, ARPA Puglia, corso Trieste 127, I-70126, Bari, Italy

[e]Istituto per la Sintesi Organica e la Fotoreattività Consiglio Nazionale delle Ricerche (ISOF-CNR), Via P. Gobetti 101, 40129 Bologna, Italy

*Corresponding author

**Corresponding author

E-mail addresses: biagia.musio@poliba.it (B. Musio), vito.gallo@poliba.it (V. Gallo)

Abstract

Non-targeted NMR-based approach has received great attention as a rapid method for food product authenticity assessment. The availability of a database containing many comparable NMR spectra produced by different spectrometers is crucial to develop functional classifiers able to discriminate rapidly the commodity class of a given food product. Nevertheless, variability in spectrometer features may hamper the production of comparable spectra due to inherent variations in signal resolution. In this paper, we report on the development of a class-discrimination model for grape juice authentication by application of non-targeted NMR spectroscopy. Different approaches for the pre-treatment of data will be described along with details about the model validation. The developed model performed excellently (95.4 to 100% correct predictions) even when it was tested against 650 spectra produced by 65 spectrometers with different configurations (magnetic field strength, manufacturer, age). This study may boost the use of non-targeted NMR methods for food control.

Keywords:

Non-targeted NMR-based metabolomics approach

Interlaboratory variability

Fingerprinting

Metabolite profiling

Method validation

Food authenticity

Grape juice

Chemometric analysis

1. Introduction

Food control has been historically achieved through a direct approach, namely by identification and quantification of a primary marker indicated as responsible for a food authenticity issue according to specific legal limits (*targeted approach*). Nevertheless, the possibility to obtain a larger amount of data more rapidly made the use of *non-targeted* approaches progressively more common for food control thanks also to the many advances in the analytical techniques and in the chemometric applications.(Granato et al., 2018; Medina, Pereira, Silva, Perestrelo, & Câmara, 2019; Oliveri, 2017) Non-targeted methods offer the possibility to extract rapidly and in non-destructive way information which can be advantageously used to unveil the compounds that may affect the authenticity of the food sample under investigation. Such analytical methods can be performed according to two alternative approaches, namely the *profiling* and the *fingerprinting*. In the first case (profiling) the identity of the compounds of interest is well known and established before the statistical data elaboration. Conversely, in the second case (fingerprinting) the analysis is performed with no *a priori* identification of the compounds contained in the sample mixture.(Ballin & Laursen, 2019) Both the aforementioned approaches can produce a large amount of data which can be exploited to assess the authenticity of a big variety of food products. Nuclear Magnetic Resonance (NMR) spectroscopy is gaining growing attention in this field, as demonstrated by an increasing number of applications reported in the recent literature (Consonni & Cagliani, 2019; Sundekilde, Eggers, & Bertram, 2019) The interest in non-targeted NMR methods is mainly due to its ability to generate highly reliable instrumental responses.(Emwas et al., 2019) Indeed, when a single sample is analyzed by different NMR spectrometers, statistically equivalent NMR spectra are generated. This aspect opens the way to the creation of a community-built system containing NMR spectra which can be safely compared and can be exploited to solve many analytical issues. For instance, for a given food fraud problem, as schematically represented in figure 1, NMR spectra of several samples, suitably selected to represent a class of a food product, may be provided either by a single spectrometer or by different instruments according to an agreed and validated procedure (including sampling, sample preparation, spectra acquisition, and processing details). The repeatability and the reproducibility of the produced spectra should be verified upon the application of opportunely defined criteria (figure 1, step 1). Then, only the laboratories producing comparable NMR spectra should be eligible for feeding the database containing NMR spectra of food samples (figure 1, step 2). The stored NMR spectra would be exploited to develop a classifier properly designed to unveil the fraud (figure 1, step 3). Finally, the same laboratories which resulted eligible to feed the database (admitted to step 2) could test the classifier by submission of the NMR spectra of an unknown sample. As a result, the commodity class, and, ultimately, the authenticity of the unknow sample should be established (figure 1, step 4).

Figure 1 here

Figure 1. Flowchart of the development of a classifier for food fraud detection by application of non-targeted NMR methods.

Despite the great interest in the described non-targeted NMR method to date no standardized procedures (protocols and materials) have been introduced to apply routinely this analytical strategy for the detection of food counterfeits and determining the authenticity of food products. In the context of an ongoing project, we gave a contribution to the harmonization of the experimental procedures of the NMR methods in food control. Based on the large amount of data produced by interlaboratory comparisons (ILCs),(Gallo, Intini, Mastrorilli, Latronico, Scapicchio, Cremonini, et al., 2015; Gallo et al., 2016, 2017) we demonstrated that targeted and non-targeted NMR methods can provide comparable results when the same sample is analyzed by spectrometers that are different in terms of magnetic field strength, manufacturer, hardware configurations and age. In particular, two selection criteria were adopted to assess the statistical equivalence of the spectra produced by different spectrometers during an interlaboratory comparison: a quality parameter, Qp-score, and the interlaboratory coefficient of variation, CV%.(Gallo, Intini, Mastrorilli, Latronico, Scapicchio, Triggiani, et al., 2015; Gallo et al., 2020) Besides, exploiting the unique capability of NMR spectroscopy compared to other analytical techniques to generate equivalent signal intensity regardless of the spectrometer configuration,(Bharti & Roy, 2012) we developed an NMR-based community-built calibration system which was able to assess the performance of the laboratories and to perform quantitative analysis (qNMR).(Musio et al., 2020) Nevertheless, one inherent issue observed when the same sample is analyzed by different spectrometers is that, while the intensity of the NMR signal is usually independent on the spectrometer configuration, the shape and the resolution of the signal is subjected to small variations which, not surprisingly, can affect the reliability of the non-target analysis. Indeed, the magnetic field strengths and the procedures adopted for the pre-treatment of data (normalization, peak alignment, scaling) play a crucial role to obtain high levels of repeatability and reproducibility of statistical results. (Craig, Cloarec, Holmes, Nicholson, & Lindon, 2006; Euceda, Giskeodegård, & Bathen, 2015; van den Berg, Hoefsloot, Westerhuis, Smilde, & van der Werf, 2006) In the present paper, we explored the effect of data-pre-treatment (buckets size and data scaling) on the performance of a class-discrimination system upon the statistical elaboration of the large number of data produced during an interlaboratory comparison. As proof of concept, samples of grape juice extracted from two different cultivars (cv.), Primitivo and Negroamaro, were analyzed by 65 different spectrometers applying the same protocol. Both the profiling and the fingerprinting approaches were explored and the chemometric analysis was based on *i*) a training set constituted of 100 NMR spectra recorded by a single spectrometer for 50 grape juice cv. Primitivo and 50 grape juice cv. Negroamaro and *ii*) a test set constituted of 650 NMR spectra produced by 65 different NMR spectrometers for one grape juice sample cv. Primitivo and one grape juice sample cv. Negroamaro (5 repetitions per sample per spectrometer). This study should demonstrate that the judicious pre-treatment of data is crucial to make the spectra produced by different spectrometers statistically equivalent. Only in this case, they may be used for the development of classifiers able to predict the commodity class of a food sample and, thus, allow to assess its authenticity. Considering the high throughput of non-targeted NMR methods, this potentiality is of great interest to the scientific community involved in food control.

Experimental section

2.1. Materials

3-(Trimethylsilyl)-2,2,3,3-tetradeutero-propionic acid sodium salt (TSP, CAS N. 24493-21-8, 99 %D, Armar Chemicals, Döttingen, Switzerland), sodium azide (NaN_3, CAS N. 26628-22-8; ≥99.5%, Sigma-Aldrich, Milan, Italy), deuterium oxide (D_2O, CAS. N. 7789-20-0, 99.86 %D, Eurisotop, Saclay, France) and methanol-d_4 (CD_3OD, CAS. N. 811-98-3, 99.80 %D, Eurisotop, Saclay, France) were used for sample preparation. NMR tubes (Norell 509-UP 7) were provided by Norell, Landisville NJ, US. The NMR samples were prepared using the automated system for liquid handling (SamplePro Tube, Bruker BioSpin). Grape samples (cv. Primitivo and cv. Negroamaro; Centro di Ricerca, Sperimentazione e Formazione in Agricoltura "Basile-Caramia (CRSFA), Locorotondo, Bari, Italy) were collected according to official recommendations (Regulations (CE) n. 834/2007, n. 889/2008, n. 1235/2008 and following modifications). 100 grape samples (50 of cv. Primitivo and 50 of cv. Negroamaro) were collected as follows: 30 berries were harvested randomly from different parts of the same plant for each sample. The samples were labeled according to the plant of origin, which was marked with a number and a letter, indicating respectively the vine-row and the sector of the vine-row to which the plant belonged. Two bigger samples (1 Kg of cv. Primitivo and 1 Kg of cv. Negroamaro) were collected randomly from 3 plants and labeled according to the same procedure. The samples were refrigerated at 4°C and transferred from the field to the laboratory, where they were stored at –20°C until analysis.

2.2. Experimental procedure

The interlaboratory comparison was organized according to EN ISO/IEC 17043:2010 and reference normative therein (Conformity assessment - General requirements for proficiency testing) with 52 registered participants, 76 available spectrometers of which 65 producing results spectrometers [300 MHz (2), 400 MHz (22), 500 MHz (18), 600 MHz (18) and 700 MHz (5); Bruker (52), Agilent (9) and Jeol (4) manufacturers]. The ILC participants were furnished with three test NMR tubes, labeled as T, X, and Y. Tube T contained pure methanol-d_4 (CD_3OD, 99.80 %D) and was used as an NMR thermometer to calibrate the temperature of each spectrometer at 298.1 ± 0.1 K. Tubes X and Y, containing aqueous solutions of grape juice (cv. Primitivo and cv. Negroamaro, respectively), were prepared as follows: 10 berries were defrosted at room temperature for 60 minutes. They were mechanically pressed and the resulting grape juice (~ 5 mL) was centrifuged (Ettich Rotofix 32A, 2500 g, 15 min). The supernatant (1.08 mL) was combined with a solution of NaN_3 (84.6 mg / 50 mL) in buffer [$(HC_2O_4)^-/(C_2O_4)^{2-}$ 0.11 M, pH 4.2]. 318 µL of this solution was combined stepwise with a volume of the buffer solution (222 µL) and a volume of a TSP/D_2O solution (60 µL, 0.10 g of TSP in 50 g of D_2O).

2.3. Data acquisition and processing

For each sample the participants were asked to perform five repetitions of a 1D ^{1}H NOESY NMR experiment,(Mckay, 2011) preceded by a selective pre-saturation step to remove the residual water signal. The 5-fold replication was needed to comply with conditions for intermediate precision, i.e. same NMR tube, same spectrometer, same user, at least 24 h delay between runs, removal of the NMR tube from the magnet from run to run. The participants received experimental guidelines for setting the acquisition parameters according to the spectrometer manufacturer requirements.

For Varian/Agilent spectrometers, guidelines included: pulse program (NOESY); size of fid (np, 128 K); spectral width (sw, 20 ppm); transmitter offset (tof): ca. 4.70 ppm (chemical shift value was set on the residual water signal); 90° hard pulse (pw, optimized by manual or automatic procedures keeping the pulse length as short as possible (< 10 µs)); steady state (ss, 8); number of transients (nt, 64); mixing time (mixN, 0.01 s); recycle delay (d1, 5 s); no sspul (sspul = 'n'); no ZQ filter (Gzqfilt = 'n'); no homo spoil during mixing time (gt1 = 0, gzlvl1 = 0 and gstab = 0); presaturation during the whole length of d1, centered at the HDO residual signal with a nutation frequency of about 25 Hz [satmode = 'yn', satdly = d1, satfrq = tof; satpwr was set to yield r1 of about 25 after running the command getpower(satpwr,tn):r1]; receiver gain optimization (once optimized for tube P, the obtained receiver gain value was also used for the tube N).

For Bruker spectrometers, guidelines included: pulse program: noesypr1d; size of FID (TD, 128 K); spectral width (SW, 20 ppm); transmitter offset, ca. 4.70 ppm (chemical shift value was set on the residual water signal); 90° hard pulse (p1, optimized by manual or automatic procedures keeping the pulse length as short as possible (< 10 µs); power level for presaturation (pl9, calculated by command "pulse 25Hz" after optimization of p1); dummy scans (ds, 8); number of scans (ns, 64); mixing time (d8, 0.01 s); recycle delay (d1, 5 s); receiver gain optimization (once optimized for tube P, the obtained receiver gain value was also used for the tube N).

For Jeol spectrometers, guidelines included: pulse program: noesy_abs; y_points = 1; size of fid (x_point = 131072); spectral width (x_sweep = 20); transmitter offset (x_offset = 4.7); 90º hard pulse (x_pulse = x90; x_atn = xatn) to be optimized by manual or automatic procedures, keeping pulse length as short as possible (< 10 µs); steady state (x_prescans = 8); number of transients (scans = 64); mixing time (mix_time = 0.01); recycle delay (relaxation_delay = 5); presaturation during the whole length of recycle delay, centered at the HDO residual signal with a γB2 power of about 25 Hz (irr_mode = presaturation; irr-offset = x_offset; presat_time_flag = y); the following formula was used to calculate the value of irr attenuator corresponding to 25 Hz: irr attenuation = x_atn + 20log(10.000/x90); receiver gain optimization (once optimized for tube P, the obtained receiver gain value was also used for the tube N).

The NMR raw data sets were uploaded by each laboratory on the website http://nmr.mxcs.it/index.php developed according to internationally agreed procedures.(International Organization for Standardization (ISO), 2005; ISO, 2012)

2.4. Data analysis

The 650 FIDs relative to the 1D ^{1}H NOESY NMR experiments produced for tubes X and tubes Y (5 replicates x 2 tubes x 65 spectrometers) and uploaded by the ILC participants onto the dedicated platform, were re-processed by a single operator and segmented into regular (0.04 ppm or 0.01 ppm as indicated) and variable-sized (as indicated) intervals (buckets) in the range

of [9.50, 0.50] ppm. For the sake of clarity, in the following tubes X and Y are indicated as P (cv. Primitivo) and N (cv. Negroamaro), respectively. The underlying area of each bucket was calculated and normalized to the total intensity. The areas of the buckets in the region [5.10, 4.15] ppm, corresponding to the residual water signal, were set to 0. The data matrices were imported into SIMCA 13.0.3 (Umetrics, Umea, Sweden), and buckets were mean-centered (Ctr) or subjected to Unit Variance (UV) scaling after mean-centering as indicated. Analogously, the 1D ^{1}H NOESY NMR spectra of 50 grape samples cv. Primitivo and 50 grape samples cv. Negroamaro were obtained using the spectrometer at the provider's disposal working at 400 MHz (Bruker Avance 400) and processed by the same operator. As an unsupervised approach, Principal Component Analysis (PCA) was used to have an overview of data. Partial Least Square – Discriminant Analysis (PLS-DA) was used as a supervised technique to build predictive statistical models for the *a priori* defined class of observations (in the present case, the class was the variety). Spectra constituted the observations and buckets were the x-variables.(Szymańska, Saccenti, Smilde, & Westerhuis, 2012) Predictions were based on the highest values predicted for the y-variable (YPredPS).

3. Results and discussion

The discrimination of samples having very similar compositions is a highly demanding task in analytical chemistry. To evaluate the power of non-targeted NMR methods in such task, the grape cultivars Primitivo (P) and Negroamaro (N) were chosen as a proof of concept because of the similarity of their metabolic compositions, especially if sugars and organic acids are considered. Moreover, the fact that test samples P and N gave NMR spectra which were almost superimposable (Figure S1), made this study even more challenging.

3.1. NMR data pre-treatment for class-discrimination

The PCA was applied to the mean-centered (Ctr) regular buckets (0.04 ppm) obtained from the spectra of tubes P (cv. Primitivo) and N (cv. Negroamaro). The first two principal components (PC1 *vs* PC2) explained 84% of the variance ($R^2X[1] = 0.732$ and $R^2X[2] = 0.105$, where R^2 indicated the goodness-of-fit), with a noticeable clustering of the observations in the score plot (figure 2a). The distinction between the two classes (P and N) could be visualized at higher principal components, PC4 *vs* PC5 (figure 2b), explaining together less than 6% of the variance of the x-variables ($R^2X[4] = 0.035$ and $R^2X[5] = 0.021$). The observed clustering at the first two principal components was strongly correlated to the magnetic field strength of the spectrometers used to generate the spectra (figure 2c), with the observations scattering primarily along the PC1 (at higher PC1 values corresponded lower magnetic field strengths).The loading plot p[1] *vs* p[2] revealed that the buckets included in the spectral region [3.15 – 4.15 ppm] referring to sugars (fructose and glucose) gave the highest contribution to the distribution of the observations according to the magnetic field strength (figure 2d).

Figures 2a-d here

Figure 2. PCA applied to the 325 spectra P and the 325 spectra N by using Ctr-scaled 0.04 ppm-sized bucketing. a, b) Score plots t[1] *vs* t[2] (a) and t[4] *vs* t[5] (b) where the observations are colored according to the belonging tube class: P (blue square), N (green circle); c) score plot t[1] *vs* t[2] where the observations are colored according to the magnetic field strength of the spectrometers: 300 MHz (red circle), 400 MHz (blue square), 500 MHz (green triangle), 600 MHz (light blue rhombus), 700 MHz (yellow inverted triangle); d) loading plot p[1] *vs* p[2].

Such behavior was observed also when smaller regular buckets (0.01 ppm) were considered, in which case the observations clustered remarkably along with the first principal component (PC1), explaining alone 72.1% ($R^2X[1] = 0.721$) of the x-variables variance (Figure S2, supporting materials). Still, the distinction between the two classes P and N could be appreciated only at higher PCs (PC6 and PC7, with $R^2X[6] = 0.00673$ and $R^2X[7] = 0.00489$).
In figure 3, the effect exerted by the magnetic field strength on the bucket areas is illustrated. Signals are distributed in a single bucket or adjacent buckets depending on the magnetic field strength. As a result, the area of the same bucket is also dependent on the operating magnetic field strength, giving rise to a magnetic field dependent variance contribution which may lead to misleading observations clustering. Considering the relatively high concentration of glucose and fructose compared to the other metabolites contained in grape juice, the high observed contribution of these buckets to scores distributions was justified.

Figure 3 here

Figure 3. Average spectra (3.15 – 4.15 ppm region) obtained for the tube P by spectrometers with different magnetic field strengths (300, 400, 500, 600, and 700 MHz). The dashed lines delimited the regular rectangular buckets (0.04 ppm).

With this information in our hand, the application of the variable-sized bucketing (VSB) was evaluated in this study as a useful alternative approach to overcome the intrinsic effect of the different spectrometers features on the sample grouping. For this purpose, 26 variable sized buckets (B1 – B26, figure 4) were selected upon evaluating the overlaid 650 spectra (325 samples P and 325 samples N). The metabolites associated with the selected buckets were characterized by referring to Chenomx library (table S1). The size of each bucket was defined in such a way that a signal or a group of signals (when more signals were overlapped) was always included in the bucket independently of the applied magnetic field strength.

Figure 4 here

Figure 4. Variable size bucketing scheme generating 26 buckets (B1 – B26). 1D ^{1}H NOESY spectra normalized to the total intensity of tube P (blue spectrum) and tube N (green spectrum), respectively.

When the variable size bucketing was applied, the observations clustered predominantly according to the class P *vs* N along the PC1 ($R^2X[1] = 0.706$) as shown in the score plot t[1] *vs* t[2] relative to PC1 *vs* PC2 (Figure S4a), while no relation with the

strength of the operative magnetic field could be extrapolated (figure S4b). According to the loading plot p[1] *vs* p[2] (figure S4c) the highest contribution to the class discrimination was due to the buckets related to ethanol (B26 at 1.16 ppm), fructose (B16 at 4.055 ppm), and malic, citric and succinic acids (B18 at 2.800 ppm and B19 at 2.625) and this result can be safely ascribed to the differences between the two cultivars.Since the application of mean-centering, without any scaling, could cause an over-estimation of the more intense buckets at the expense of the weaker ones, the effect of a different scaling on the distribution of the observations was evaluated upon application of the Unit Variance (UV) scaling ($1/SD_j$, where SD_j is the standard deviation of variable *j* computed around the mean).

Figures 5a-f here

Figure 5. PCA applied to the 325 spectra P and the 325 spectra N by using UV-scaled 0.04 ppm buckets (a, b, and c) and the 26 UV-scaled variable size buckets (d, e, and f). a,d) Score plots t[1] *vs* t[2]; the observations are coloured according to the belonging tube class: P (blue square), N (green circle); b, e) score plots t[1] *vs* t[2]; the observations are coloured according to the magnetic strength of the spectrometer: 300 MHz (red circle), 400 MHz (blue square), 500 MHz (green triangle), 600 MHz (light blue rhombus), 700 MHz (yellow inverted triangle); c, f) loading plots p[1] *vs* p[2].

Interestingly, when the PCA was performed on the UV-scaled regular buckets (0.04 ppm) the observations did not distribute according to the magnetic strength (figure 5b) as observed previously for the mean-centered 0.04 ppm regular buckets (figure 2). Still, a marked separation according to the different class P *vs* N could be observed only at higher PCs, namely along the PC4 ($R^2X[4] = 0.102$). On the contrary, good class discrimination was obtained already at the first two PCs when the 26 variable sized buckets were subjected to UV-scaling before PCA analysis, with observations grouping predominantly on the plane defined by PC1 ($R^2X[1] = 0.395$) and PC2 ($R^2X[2] = 0.272$) and explaining together 66.7% of x-variance (Figure 5d-e). Importantly, when the spectroscopic data were UV-scaled, the less abundant metabolites started to contribute markedly to the class discrimination, as demonstrated in the loading plots (compare figures 5c, 5f with figures 2c, 2f). Among them, chlorogenate (7.64 – 7.70 ppm), arginine (7.21 – 7.29 ppm), and phenylalanine (7.43 – 7.35) exerted the highest contribution. The observed results were confirmed by analyzing the mean spectra obtained in the four different studied cases. As shown in figure 7a, when the regular bucketed data (0.04 ppm) were mean-centered the variability was exclusively observed in the region related to the sugars (3.120 – 5.32 ppm) with a predominant influence of the operative magnetic strength (figure S5a). Such behavior was still observed in the case of UV-scaled regular bucketed (0.04 ppm) data, although the variability could be detected also in spectral regions related to less abundant metabolites (figure S5c). A drastic different variability was observed in the mean spectrum obtained from the variable size bucketed data both in the case of the mean-centering approach (figure S5b) and in the UV-scaling one (figure S5d). No correlation between the spectral variability and the magnetic field strength was observed, since the variable-sized bucketing approach allowed to overcome any signal shape variations, naturally existing in spectra produced by spectrometers operating at different magnetic field strength. Also, the contribution of almost all the selected buckets (B1 – B26) became crucial towards class discrimination.

3.2. Class-discrimination system development

A class-discrimination system was designed to identify the cultivar of two different groups of grape juice samples by comparison with a representative grape juice population (Figure 8). In particular, a limit case was evaluated with a single NMR spectrometer producing the training spectra and many spectrometers producing the test spectra (1 training spectrometer *vs* n testing spectrometers). The PLS-DA was applied to exploit the statistically significant difference between the two classes of grape juice (cv. Primitivo and cv. Negroamaro), as observed during the explorative PCA analysis. The training set was opportunely composed of 100 spectra related to 100 grape juice samples (50 samples cv. Primitivo and 50 samples cv. Negroamaro) deriving from 100 different plants. The test set consisted of 650 spectra related to 2 grape juice samples (5 repetitions of one sample cv. Primitivo, tube P; 5 repetitions of one sample cv. Negroamaro, tube N) produced by 65 different NMR spectrometers. The grape samples used to produce the test set were collected from plants that were different from those involved in the production of the training set. The test set was produced by a single spectrometer (Bruker Avance 400 MHz) and processed by a single operator (the same operator who processed the 650 spectra P and N) using Topspin – Amix application. The previously considered approaches for data processing (mean-centering and UV-scaling on both regular and variable-sized buckets) were explored also for the latter training dataset. The most extended grouping was observed when the PCA was conducted on the 26 UV-scaled variable-sized buckets (figure S3), which were ultimately selected as a training dataset for the PLS-DA. The best PLS-DA model was chosen based on the highest values of cumulative R^2Y and Q^2Y (where Q^2 was the goodness of prediction in 7-fold cross-validation) and upon passing the permutation tests (figure S4 and table S1). The optimal PLS-DA model was obtained upon removal of 6 samples (figures S5a and S5b), which resulted outliers, according to the Hotelling's T2 plot (confidence interval 95%) relative to the PCA-class performed singularly on 50 samples cv. Primitivo and 50 samples cv. Negroamaro (figure S6). The resulting values of $R^2Y(cum) = 0.851$ and $Q^2Y(cum) = 0.797$ were in agreement with the threshold values identifying good classification model capability ($Q^2Y(cum) > 0.5$ and $(R^2 - Q^2) < 0.2$).

Figure 6 here

Figure 6. Class discrimination system designed in the present study

The model was subjected to two validation steps, which were characterized by an increasing index of risk (Figure 6 and table S3). The first test set (TS1) was composed by 44 spectra related to 44 tubes (22 tubes P and 22 tubes N) randomly chosen from the 130 tubes (65 tubes P and 65 tubes N) delivered to the ILC participants but analyzed by the same spectrometer (400

MHz) used for producing the training spectra. The second test set (TS2) consisted of the 650 spectra generated by the 65 different spectrometers involved in the ILC. While the first validation step could be considered of medium risk (1 spectrometer *vs*:1 spectrometer approach), as both the training data set and the test data set (TS1) were generated by the same spectrometer/operator, the second validation step should be considered of high risk (1 spectrometer *vs* n spectrometers approach), because the test set was originated from different sources and was produced under high variability (different operators and/or instrumental features) compared to the training set. The prediction capability of the developed model resulted excellent when it was validated by querying TS1 with 100% of corrected prediction. Furthermore, 95.38% of spectra (620/650) were classified correctly when TS2 was queried. The prediction capability improved (98.21%) when 90/650 spectra, which resulted outliers according to the Hotelling's T2 test (95% confidence interval), were removed from TS2 (figure S6). No correlation could be observed between the nature of the outliers and the magnetic field strength of the spectrometers they derived from, thus confirming that variable sized bucketing approach (profiling) allows minimizing the effect of the signal shape variations on the outcomes of the statistical analysis.

4. Conclusion

We reported on the development of a grape juice class-discrimination system based on non-targeted NMR analysis. To the best of our knowledge, this is the first example of a food class-discrimination model validated by the approach "1 training spectrometer *vs* n testing spectrometers". The NMR spectra included in the training set were recorded by a single spectrometer and the NMR spectra used for the model validation were generated by a conspicuous number of differently configured NMR spectrometers. It was ascertained that, among the two data pre-treatment strategies, namely fingerprinting and profiling, the latter allows us to reach the best discrimination performance and to minimize the influence of the magnetic field strength on the results. The combination of fingerprinting and profiling strategies helped to identify the variables characterizing the two cultivars of the grape juice and demonstrated exemplarily the power of this analytical tool to get a huge amount of information in a short time and a non-destructive manner compared to other analytical spectroscopic techniques. The developed models for class prediction performed excellently (95.4% to 100%), despite the highly variable operative conditions (different spectrometers and many operators). Further efforts should be addressed to set up an "n training spectrometer *vs* n testing spectrometers" system with the aim to exploit to the largest extent the advantages of the non-targeted NMR analysis in food control. Indeed, the creation of a database containing NMR spectra of different food products suitably categorized on the bases of their commodity classes and the development of proper food classifiers based on the stored NMR data may give a boost to food fraud-fighting. Finally, the proposed discrimination strategy may help overcome, to some extent, the lack of official guidelines regulating the use of non-targeted NMR analyses and, promote the possible development of community-built discrimination/classification systems based on NMR spectroscopy.

ACKNOWLEDGMENT

This work was partially supported by Regione Puglia in framework of *Programma Sviluppo Rurale FEASR 2007 -2013 Mis. 214 Az. 4 sub az. a) "Progetti integrati per la Biodiversità" Reg. (CE) 1698/2005* and *Programma Sviluppo Rurale FEASR 2014-2020 Mis.10.2.1 "Progetti per la conservazione e valorizzazione delle risorse genetiche in agricoltura" Reg. (CE) 1305/2013*.

The authors acknowledge the following researchers to have voluntarily participated to the interlaboratory comparison: Antonino Rizzuti, Pasquale Scapicchio, Tommaso Dinoia, Domenico Acquotti, Cristina Airoldi, Michael Assfalg, Alessandro Barge, Francesca Benevelli, Davide Bertelli, Fabio Bertocchi, Aurimas Bieliauskas, Anna Borioni, Augusta Caligiani, Emanuela Callone, Ales Čamra, Flaminia Cesare Marincola, Dinesh Chalasani, Roberto Consonni, Paolo Dambruoso, Silvia Davalli, Taylor David, Bernd Diehl, James Donarski, Ana Gil, Roberto Gobetto, Luca Goldoni, Erwann Hamon, Andrea Kobrlová, Francesco Longobardi, Renzo Luisi, Denis Lynch, Domenico Mallamace, Stefano Mammi, Magali Martin-Biran, Pierluigi Mazzei, Andrea Mele, Salvatore Milone, Dolores Molero Vilchez, Roger Mulder, Claudia Napoli, Antonella Pepe, Daniele Ragno, Antonio Randazzo, Maria Cecilia Rossi, Archimede Rotondo, Algirdas Sackus, Elena Sáez Barajas, Elisabetta Schievano, Bhavaraju Sitaram, Livio Stevanato, Panteleimon G.Takis, Jan Teipel, Freddy Thomas, Elisabetta Torregiani, Daniela Valensin, Marina Veronesi, John Warren, Julien Wist, Elina Zailer, Cristiano Zuccaccia.

CONFLICT OF INTEREST STATEMENT

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.

REFERENCES

Ballin, N. Z., & Laursen, K. H. (2019). To target or not to target? Definitions and nomenclature for targeted versus non-targeted analytical food authentication. *Trends in Food Science and Technology*. https://doi.org/10.1016/j.tifs.2018.09.025

Consonni, R., & Cagliani, L. R. (2019). The potentiality of NMR-based metabolomics in food science and food authentication

assessment. *Magnetic Resonance in Chemistry*, *57*(9), 558–578. https://doi.org/10.1002/mrc.4807
Craig, A., Cloarec, O., Holmes, E., Nicholson, J. K., & Lindon, J. C. (2006). Scaling and normalization effects in NMR spectroscopic metabonomic data sets. *Analytical Chemistry*, *78*(7), 2262–2267. https://doi.org/10.1021/ac0519312
Emwas, A. H., Roy, R., McKay, R. T., Tenori, L., Saccenti, E., Nagana Gowda, G. A., … Wishart, D. S. (2019). Nmr spectroscopy for metabolomics research. *Metabolites*, *9*(7). https://doi.org/10.3390/metabo9070123
Euceda, L. R., Giskeodegård, G. F., & Bathen, T. F. (2015). Preprocessing of NMR metabolomics data. *Scandinavian Journal of Clinical and Laboratory Investigation*, *75*(3), 193–203. https://doi.org/10.3109/00365513.2014.1003593
Gallo, V., Intini, N., Mastrorilli, P., Latronico, M., Scapicchio, P., Cremonini, M. A., … Cafagna, I. (2015). *NMR Inter-laboratory Comparisons: Validation of a 1D 1H-NOESY experiment for fingerprinting of wheat and flour*. (NePEdizioni, Ed.). Rome. Retrieved from https://www.nepedizioni.com/index.php/collane/libri-collane/nmr-inter-laboratory-comparisons/266/validation-of-a-1d-h-noesy-experiment-for-fingerprinting-of-wheat-and-flour-detail.html
Gallo, V., Intini, N., Mastrorilli, P., Latronico, M., Scapicchio, P., Milella, A., … Mari, S. (2016). *NMR Inter-laboratory Comparisons: Validation of NMR fingerprinting methods: effects of processing on measure reproducibility and laboratory performance assessment*. (NePEdizioni, Ed.). Rome. Retrieved from https://www.nepedizioni.com/index.php/collane/libri-collane/nmr-inter-laboratory-comparisons/341/validation-of-nmr-fingerprinting-methods-effects-of-processing-on-measure-reproducibility-and-laboratory-performance-assessment-detail.html
Gallo, V., Intini, N., Mastrorilli, P., Latronico, M., Scapicchio, P., Triggiani, M., … Valerio, M. (2015). Performance Assessment in Fingerprinting and Multi Component Quantitative NMR Analyses. *Analytical Chemistry*, *87*(13), 6709–6717. https://doi.org/10.1021/acs.analchem.5b00919
Gallo, V., Intini, N., Mastrorilli, P., Latronico, M., Todisco, S., Rizzuti, A., … Ghelli, S. (2017). *NMR Inter-laboratory Comparisons: Validation of a 1D 1H-NOESY experiment for fingerprinting of grape juices*. (NeP edizioni, Ed.). Rome.
Gallo, V., Ragone, R., Musio, B., Todisco, S., Rizzuti, A., Mastrorilli, P., … Latronico, M. (2020). A Contribution to the Harmonization of Non-targeted NMR Methods for Data-Driven Food Authenticity Assessment. *Food Analytical Methods*, *13*(2), 530–541. https://doi.org/10.1007/s12161-019-01664-8
Granato, D., Putnik, P., Kovačević, D. B., Santos, J. S., Calado, V., Rocha, R. S., … Pomerantsev, A. (2018). Trends in Chemometrics: Food Authentication, Microbiology, and Effects of Processing. *Comprehensive Reviews in Food Science and Food Safety*, *17*(3), 663–677. https://doi.org/10.1111/1541-4337.12341
International Organization for Standardization (ISO). (2005). ISO 13528: Statistical methods for use in proficiency testing, *2005*, 76. Retrieved from https://www.iso.org/standard/35664.html
ISO. (2012). ISO 5725-1:1994 - Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions. Retrieved from https://www.iso.org/standard/11833.html
Mckay, R. T. (2011). How the 1D-NOESY suppresses solvent signal in metabonomics NMR spectroscopy: An examination of the pulse sequence components and evolution. *Concepts in Magnetic Resonance Part A*, *38A*(5), 197–220. https://doi.org/10.1002/cmr.a.20223
Medina, S., Pereira, J. A., Silva, P., Perestrelo, R., & Câmara, J. S. (2019, April 25). Food fingerprints – A valuable tool to monitor food authenticity and safety. *Food Chemistry*. Elsevier Ltd. https://doi.org/10.1016/j.foodchem.2018.11.046
Musio, B., Ragone, R., Todisco, S., Rizzuti, A., Latronico, M., Mastrorilli, P., … Gallo, V. (2020). A community-built calibration system: The case study of quantification of metabolites in grape juice by qNMR spectroscopy. *Talanta*, *214*, 120855. https://doi.org/10.1016/j.talanta.2020.120855
Oliveri, P. (2017). Class-modelling in food analytical chemistry: Development, sampling, optimisation and validation issues e A tutorial. https://doi.org/10.1016/j.aca.2017.05.013
Sundekilde, U. K., Eggers, N., & Bertram, H. C. (2019). NMR-Based Metabolomics of Food. In *Methods in Molecular Biology* (Vol. 2037, pp. 335–344). Humana Press Inc. https://doi.org/10.1007/978-1-4939-9690-2_18
Szymańska, E., Saccenti, E., Smilde, A. K., & Westerhuis, J. A. (2012). Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics*, *8*(S1), 3–16. https://doi.org/10.1007/s11306-011-0330-3
van den Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K., & van der Werf, M. J. (2006). Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics*, *7*, 1–15. https://doi.org/10.1186/1471-2164-7-142