LWT - Food Science and Technology 42 (2009) 1455–1460

Contents lists available at ScienceDirect

LWT - Food Science and Technology

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

Classification of Brazilian vinegars according to their ^{1}H NMR spectra by pattern recognition analysis

Elisangela F. Boffo [a,*], Leila A. Tavares [a], Márcia M.C. Ferreira [b], Antonio G. Ferreira [a]

[a] Departamento de Química, Universidade Federal de São Carlos, P.O. Box 676, CEP 13565-905 São Carlos, SP, Brazil
[b] Instituto de Química, Universidade Estadual de Campinas, P.O. Box 6154, CEP 13084-971 Campinas, SP, Brazil

ARTICLE INFO

Article history:
Received 23 January 2009
Received in revised form
4 May 2009
Accepted 5 May 2009

Keywords:
Vinegar
^{1}H NMR
SIMCA
KNN
PLS-DA

ABSTRACT

This work describes using ^{1}H NMR data and pattern recognition analysis to classify vinegars. Vinegar authenticity is linked to raw ingredient source and manufacturing conditions. Application of PCA and HCA methods resulted in the natural clustering of the samples according to the raw material used. Wine vinegars were characterized by a high concentration of ethyl acetate, glycerol, methanol and tartaric acid, while glycerol and ethyl acetate signals were not visible in alcohol/*agrin* vinegars. Apple vinegars showed to be richer in alanine. The KNN, SIMCA and PLS-DA methods were used to build predictive models for classification of vinegar type wine, apple and alcohol/*agrin* (27 samples – 22 as training set). The models were tested using an independent set (5 samples), no samples were wrongly classified. Validated models were used to predict the class of 21 commercial samples, which, as expected, were correctly classified. Eight commercial vinegars (honey, orange, pineapple and rice) were discriminated from these samples using PCA method. Honey vinegars did not present ethanol signals and pineapple vinegars presented the largest amount of tartaric acid. Rice and orange vinegars are richer in lactic acid and did not present the methanol signal. Alanine signals were not visible in orange vinegars.

1. Introduction

The vinegar is a product extensively consumed in Brazil and is mainly used as condiment in salads. It is produced by an alcoholic fermentation followed by an acetic fermentation (García-Parrilla, González, Heredia, & Troncoso, 1997), from an enormous variety of raw materials like grape, apple, honey, rice, orange and pineapple (Natera, Castro, García-Moreno, Hernández, & García-Barroso, 2003).

Many processed foods can be adulterated in function of its high aggregate value (Ogrinc, Kosir, Spangenberg, & Kidric, 2003) or climatic variations and that affects its productivity (Pupin et al., 1998). The vinegar also can pass by an adulteration process. One mode is to add a vinegar derived from a C_4 plant (Hatch–Slack pathway for carbon fixation), such as the sugar-cane (named alcohol vinegar), of easy obtainment and low cost in our country. This increases the production of vinegar from a C_3 plant (Calvin cycle for carbon fixation), as grape (wine vinegar) and to reduce the final cost (Hermann, 2001). Therefore, the determination of vinegars authenticity is of great importance in the food industry, since raw material costs are directly related to their origin.

SNIF-NMR method (Site Specific Natural Isotopic Fractionation studied by Nuclear Magnetic Resonance) is an efficient technique to control the food authenticity and products' adulteration that has been introduced recently to verify the addition of an exogenous product by comparing their isotopic content with that from authentic material (Lindner, Bermann, & Gamarnik, 1996; Martin et al., 1988; Martin, Zhang, Naulet, & Martin, 1986). For vinegars, this method is based on the ratio of deuterium and hydrogen content at the methyl site of acetic acid, which can be obtained from plant sugars with different biosynthetic mechanisms of CO_2 fixation as C_3, C_4 or CAM (Crassulacean Acid Metabolism pathway for carbon fixation) and results in different isotopic ratios (Boffo & Ferreira, 2006; Remaud, Guillou, Vallet, & Martin, 1992).

^{1}H NMR spectroscopy joint to chemometric methods is another technique that can be successfully used for the vinegars' analysis (Caligiani, Acquotti, Palla, & Bocchi, 2007). Spectroscopic NMR methods provide information on a wide range of compounds present in the food matrix in a single experiment, offering advantages in terms of simplicity of sample preparation and short time of analysis (Le Gall, Puaud, & Colquhoun, 2001). Because the richness of information often results in high spectral complexity, it requires the use of multivariate data analysis to study a large number of

* Corresponding author. Tel.: +55 16 3351 8066; fax: +55 16 3351 8350.
E-mail address: efboffo@yahoo.com.br (E.F. Boffo).

0023-6438/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.lwt.2009.05.008

spectra and extract the meaningful information. The chemometric methods (Beebe, Pell, & Seasholtz, 1998) that group samples with common characteristics and distinguish them from the samples with different characteristics have been applied successfully in NMR spectral data (Alcantara, Honda, Ferreira, & Ferreira, 2007; Consonni et al., 2008; Defernez & Cololquhoun, 2003; Tavares, Ferreira, Ferreira, Correa, & Mattoso, 2005).

Principal Component Analysis (PCA) and Hierarchical Clusters Analysis (HCA) are unsupervised pattern recognition methods and were used in this work for exploratory data analysis. PCA is a multivariate procedure, which rotates the data such that maximum variability is projected onto the axes. The main use of PCA is to reduce the dimensionality of a data set while retaining as much information as is possible. It computes a compact and optimal description of the data set. The first principal component is the combination of variables that explains the greatest amount of variation. The second principal component defines the next largest amount of variation and is independent to the first principal component (Chen et al., 2008; Consonni & Cagliani, 2008; Jalali-Heravi, Masoum, & Shahbazikhah, 2004).

In PCA, a J-dimensional matrix $\mathbf{X}$ ($I \times J$) is projected into a lower, R-dimension space by decomposing $\mathbf{X}$ into a score matrix $\mathbf{T}$ ($I \times R$) and a loading matrix $\mathbf{P}$ ($J \times R$) whose product models the systematic variation in the data (in equation (1)).

$$\mathbf{X} = \sum_{r=1}^{R} \mathbf{t}_r \mathbf{p}_r^{\mathrm{T}} + \mathbf{E} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \quad (1)$$

In equation (1), $\mathbf{E}$ ($I \times J$) is the residual matrix, $\mathbf{t}_r$ is the score vector for the rth component and $\mathbf{p}_r$ is the loading vector for the rth component. A plot of two columns of the score matrix *versus* each other gives a two-dimensional projection of original data set, whereas a plot from columns of the loading matrix $\mathbf{P}$ displays the correlation among the variables (Beebe et al., 1998).

In HCA, the main goal is to display the data in a manner which emphasizes its natural grouping. The results are usually presented in the form of a dendrogram, allowing the visualization of clusters and correlations among either samples or variables. The Euclidean distances among samples or variables are calculated and transformed into similarity indices ranging from 0 to 1. A small distance corresponds to a large index and indicates a large similarity (Beckonert et al., 2003; Brereton, 2002).

Moreover, the methods K-Nearest Neighbor (KNN), Soft Independent Modeling of Class Analogies (SIMCA) and Partial Least Squares – Discriminant Analysis (PLS-DA) were used for classificatory analysis. These techniques are termed supervised pattern recognition methods, whereby a training set with known classes are used to build a mathematical model which is then evaluated by an independent validation data set (Beebe et al., 1998). After this external validation procedure, the models are ready for classification of unknown samples.

KNN attempts to categorize an unknown based on its proximity to samples from the training set. Specifically, the predicted class of an unknown depends on the class of its K-Nearest Neighbors. In a fashion analogous to polling, each of the K closest training set samples votes once for its class; the unknown is then assigned to the class mostly voted. An important part of the process is determining an appropriate value for K, the number of neighbors polled (Lindon, Holmes, & Nicholson, 2001). In contrast to KNN, the supervised method SIMCA is based on modeling each class by a separate PCA model. As PCA is applied to each class separately, SIMCA provides additional information about each separate class, such as the relevance of the different variables and outlier detection. Information about relations between samples and classes in SIMCA is given through probability measures. Unknown samples are projected into each PCA model and the degree of the fit is determined (Alan & Alan, 2005). Thus, it is possible to use SIMCA to determine whether a sample is uniquely assigned to a class, fits to several classes or whether it does not fit in any of the classes (Massart, Vandeginste, Deming, Michotte, & Kaufman, 2001). PLS-DA is used as a discriminant analysis technique by constructing a $\mathbf{Y}$ matrix consisting of dummy variables used to indicate class membership. For each binary class, a column of $\mathbf{Y}$ is constructed by assigning a value of 0 or 1 to each sample, according to class category. For example, a value of 0 is used for untreated samples and 1 for bleached samples. A PLS model is then built between the set of I spectra ($\mathbf{X}$ matrix) and $\mathbf{Y}$. The set of responses predicted by the model is rounded to either 0 or 1, and the true and predicted class memberships are then compared to evaluate how successful the model is at classifying the given samples. Given new $\mathbf{X}$ data, the PLS-DA model can be used to make class membership predictions for new samples (Masoum, Bouveresse, Vercauteren, Jalali-Heravi, & Rutledge, 2006; Nouwen et al., 1997).

In this study, the principal purpose is to distinguish vinegars produced from different types of raw materials such as wine, apple, alcohol, grape, honey, orange and pineapple. Each product contains chemical compounds which are characteristic and originated from the type of raw material used. Unsupervised pattern recognition approaches (PCA and HCA), that allow the grouping of samples with common characteristics based on their ^{1}H NMR spectra and supervised classification methods (KNN, SIMCA and PLS-DA) were employed in the classification of vinegar samples according to raw material used, and compared.

2. Materials and methods

2.1. Vinegar samples

Fifty-six vinegar samples obtained from different raw materials [white and red wine, apple, honey, rice, orange, pineapple, alcohol and *agrins*, which are blends of C_3 (wine) and C_4 vinegars (alcohol)] were analyzed. All of them were industrial or handmade vinegars produced in Brazil.

The samples were prepared, in triplicate, using 600 μL of vinegar and 100 μL of a 0.16 g/100 g solution of sodium-3-trimethylsilyl-propionate (TMSP-2,2,3,3-d_4) prepared in D_2O. TMSP was used as a chemical shift reference (δ 0.0). D_2O (99.9%) and TMSP (98%) were from Cambridge Isotope Laboratories, Inc. (USA).

2.2. NMR spectra

The NMR spectra were recorded at room temperature using a Bruker DRX400 spectrometer operating at 9.4 T, equipped with a 5-mm inverse-detection probe with z-gradient and observing ^{1}H at 400.2 MHz.

All ^{1}H NMR spectra have been acquired with low power water signal irradiation and using the same parameter conditions of acquiring and processing data. Typically, 16 scans (FIDs) were collected into 64k data points using a 8.5 μs pulse width (90° pulse angle), spectral width of 4401 Hz, acquisition time of 7.4 s and relaxation delay of 1.0 s. Each ^{1}H NMR spectrum was acquired in 2 min and 15 s. In the processing were used 32k points and an exponential weighing factor corresponding to a line broadening of 0.3 Hz was applied. After this, the NMR spectra were phased and baseline corrected using a simple polynomial curve fit included in xwinnmr software, version 3.1 and referenced to the TMSP peak (δ 0.0). Phase correction was performed manually for each spectrum, and the baseline correction was applied over the intire spectral range. The resulting spectra were converted into JCAMP format to build the data matrix.

Two-dimensional (2D) experiments were acquired using the standard spectrometer library pulse sequences. ^{1}H–^{1}H COSY experiment was obtained with spectral width of 8278 Hz in both dimensions; 4K × 256 data matrix; 50 scans per t_1 increment and relaxation delay of 1.8 s. COSY experiment was acquired in 7 h and 30 min. One-bond ^{1}H–^{13}C HSQC experiment was acquired with an evolution delay of 1.7 ms for an average ^{1}J(C,H) of 145 Hz; 4K × 256 data matrix; 64 scans per t_1 increment; spectral widths of 8278 Hz in f_2 and 21,633 Hz in f_1 and relaxation delay of 1.2 s. HSQC experiment was acquired in 6 h and 30 min. Long-range ^{1}H–^{13}C HMBC experiment was recorded with an evolution delay of 62.5 ms for ^{LR}J(C,H) of 8 Hz; 4K × 256 data matrix; 100 scans per t_1 increment; spectral width 8278 Hz in f_2 and 22,640 Hz in f_1 and relaxation delay of 1.2 s. HMBC experiment was acquired in 10 h.

2.3. Multivariate data analysis

Pirouette® version 3.11 (InfoMetrix, Woodinville, Washington, USA) was the software used for data analysis. The data matrix was built with 4650 variables (columns) and 168 spectra (lines – 56 samples in triplicate). PCA was applied to explore the data and for feature selection. Each NMR spectrum was normalized to norm one (the area under the sample profile was set equal to one) and first derivative was taken (to correct minor variations in the spectra baseline), supplying the best conditions for discrimination of the samples. No smoothing was necessary before taking the first derivative due to the good spectral S/N ratio. The data were also autoscaled, i.e., mean centered and scaled to unit variance, in order to give equal weight to each variable and so, large and small peaks gained the same importance. In HCA, the Euclidean distances among samples were calculated by using the incremental linkage method and transformed into similarity indices ranging from 0 to 1.

The supervised pattern recognition models (KNN, SIMCA and PLS-DA) were built with all replicates of wine, apple, alcohol and *agrin* authentic vinegars as shown in Table 1 (twenty-two samples in triplicate, $\mathbf{X} = (66 \times 4650)$). Five authentic vinegar samples (in triplicate) were used to validate the models and twenty-one commercial samples (in triplicate) were used in the prediction of their class identities.

KNN, SIMCA and PLS-DA methods were used in order to attain classification rules for predicting the raw material mainly used for the vinegars' production. Using KNN, the Euclidean distance was used as the criterion for calculating the distance between samples from the training set, and the optimum number of nearest neighbors (K) was selected after studying the success in classification with different K values. For all neighbors tested (1–10) none of the samples were misclassified, therefore K = 3 was selected since the number of samples in each class varied from 4 to 11. For SIMCA modeling, the number of principal components (PCs) used in each class model was determined as 3, 3 and 4 PCs for wine, apple and alcohol/*agrin* categories, respectively. For PLS-DA model, the number of PCs was chosen based on the standard error of validation, as estimated by the leave-one-out (LOO) procedure. The total prediction error [root mean square error of prediction (RMSEP)] and the residual error for each model sample were evaluated.

Table 1
Class and number of samples used for the building the models and for external validation.

	Class	Modeling	Validation
Wine	1	7	2
Apple	2	4	1
Alcohol/*Agrin*	3	11	2

Some vinegar types like rice, orange, pineapple and honey are commercialized by few or even just one brand. For this reason, only two samples of each type (8 samples) were analyzed and they were not included in the classification study. Nevertheless, a second PCA analysis was applied to all vinegar types, to verify their characteristics with respect to the other types studied previously.

3. Results and discussion

3.1. Spectral data for multivariate analysis

Fig. 1A shows all ^{1}H NMR spectra, with water suppression, analyzed. The signals observed were identified without recording NMR spectra of pure compounds or the use of standards for the vinegars. The chemical shift assignments obtained from 2D NMR (*g*COSY, *g*HSQC and *g*HMBC) spectra for apple vinegar are summarized in Table 2 and Fig. 1C–F which includes the NMR spectral regions enlarged, showing the different quantities of some constituents present in the vinegars.

Different quantities of organic acids, acetic, lactic and tartaric acids were observed in the samples. In some vinegars, lactic and tartaric acids were not detectable in NMR spectra. The only sugar peaks in vinegars were α- and β-glucose. However, the signals were observed in very small intensity and they did not show any influence in the statistical analyses. Among amino acids, only alanine was included in the chemometric analysis. The signals of phenylalanine and tyrosine were very small.

Consequently, comparing the spectra from all vinegars some similarities can be observed. However, the amount of combined data, with no obvious relation between the intensities of certain lines and the authenticity of the vinegars, makes the visual analysis of these sets virtually impossible. The need for multivariate analysis tools is obvious. An exploratory data analysis applied to the vinegar spectra using the PCA method had shown the main signals which could be involved in the identification and discrimination of the vinegar types, and they can be observed in Fig. 1B (uncolored). Thus, the regions appearing in gray were excluded from the study.

3.2. Vinegar discrimination with PCA and HCA methods

The graphical representation of the PCA scores (Fig. 2) for component 1 (describing 25.2% of the total variance) plotted against component 2 (22.5% of total variance) shows the similarities/differences among the samples, wherein similar samples tend to form clusters and dissimilar samples are found at larger distances. In this plot, the vinegars are grouped into three distinct clusters, one for wine vinegars (seven samples); one for apple (four samples) and another for alcohol/*agrin* (eleven samples). *Agrin* samples, which are blends of C_3 (wine) and C_4 vinegars (alcohol), showed to be more similar to alcohol than to wine vinegars. This similarity occurred probably, because the *agrin* samples presents higher content of alcohol vinegars originated from the sugar-cane that is extensively cultivated in our country. In addition, the alcohol/*agrin* vinegars are closely clustered, indicating that these samples are the most homogeneous in terms of chemical composition. By contrast, in the wine vinegar main group, the samples are scattered and it can be observed the discrimination between handmade and industrial vinegars. Handmade vinegars are found to the left side because they presented higher concentration of ethyl acetate and glycerol than industrial ones. Analyzing the loading plot (not shown), it was observed that the wine vinegars presented higher concentration of ethyl acetate, glycerol, methanol and tartaric acid than the others. Apple vinegars showed to have higher quantities of alanine while the glycerol and ethyl acetate signals were not visible in alcohol and *agrin* vinegars' spectra.

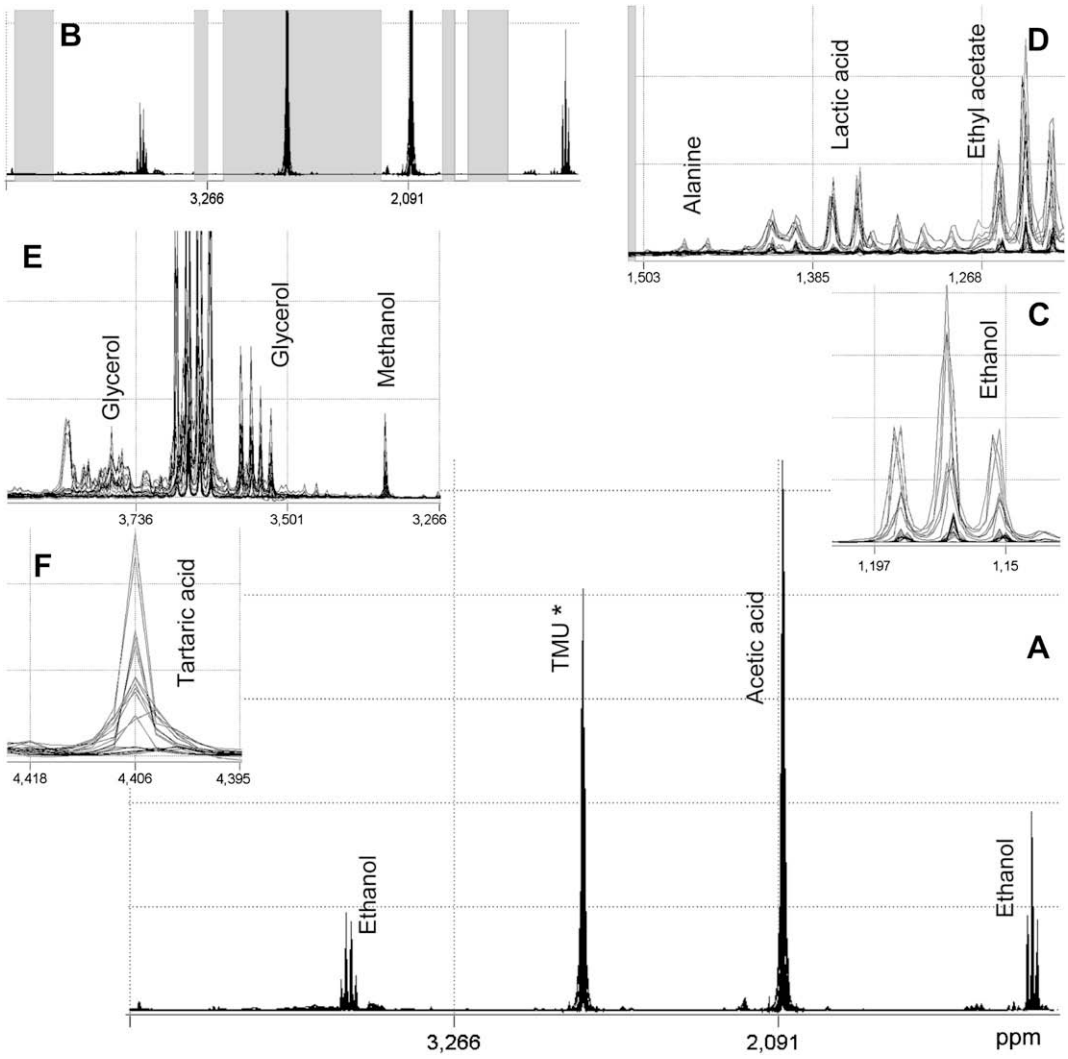

Fig. 1. (A) ^{1}H NMR spectra, with water suppression, of all vinegar analyzed; (B) selected regions used in statistical analyses (in white); (C–F) spectral expansions showing different quantities of vinegar constituents – *TMU was used in the determination of vinegar authenticity study (Boffo & Ferreira, 2006).

Table 2
Characteristics of ^{1}H NMR signals δ (chemical shift) and J (coupling constant) observable in vinegars.

Compound	Group identified	δ (ppm)	Multiplicity	J (Hz)
Acetic acid	Methylic group	2.08	s	–
Alanine	Methylic group	1.47	d	6.5
Ethanol	Methylic group	1.17	t	7.1
	Methylenic group	3.65	q	7.1
Ethyl acetate	Methylic group	1.24	t	7.1
	Methylenic group	4.13	q	7.1
Lactic acid	Methylic group	1.37	d	7.2
	Methylenic group	4.42	q	7.2
Methanol	Methylic group	3.35	s	–
α-Glucose	Anomeric hydrogen	5.18	d	3.7
β-Glucose	Anomeric hydrogen	4.63	d	7.9
	Methynic in the position 2	3.23	d	7.9; 9.3
Glycerol	Methylenic group	3.55	m	–
	Methylenic group	3.65	m	–
	Methynic group	3.77	m	–
Phenylalanine	Aromatic (positions 2 and 6)	7.31	m	–
	Aromatic (positions 3, 4 and 5)	7.38	m	–
Tartaric acid	Methynic group	4.41	s	–
Tyrosine	Aromatic (positions 2 and 6)	7.17	d	8.5
	Aromatic (positions 3 and 5)	6.84	d	8.5

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

Moreover, the score plot (Fig. 2) confirms the good reproducibility expected for this method since the replicates of each sample are well grouped and discriminated from the other vinegars.

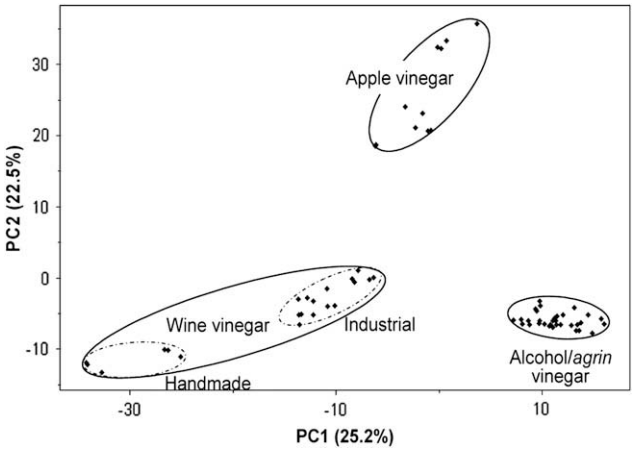

Fig. 2. PCA score plot (PC1 × PC2) showing the discrimination between vinegar samples studied.

The dendrogram obtained from HCA is shown in Fig. 3. It is important to underline that the branches' length in the dendrogram is related to the distances between the various clusters and hence, is a measure of their similarity. Therefore, two similar clusters are represented by two connected small branches, and so, have a high similarity index. With a very low similarity index, of approximately 0.06, two main clusters can be visualized in Fig. 3, one of them consists of samples produced from C_3 plants (wine and apple), and the second cluster contains vinegars originated from C_4 plants (alcohol). As in PCA analysis, *agrin* vinegars are grouped with alcohol vinegars due to their similar compositions. With a higher similarity index, of 0.37, three subgroups are identified. It can be observed that the C_3 cluster was divided into two subgroups. One of them includes apple vinegars and the other, the industrial and handmade wine vinegars. This separation agreed well with the Principal Component Analysis results. Besides, with about 90% similarity, even the replicates can be nicely identified.

The results of PCA and HCA show that the application of the chemometric methods to the ^{1}H NMR data is a powerful tool for the discrimination of vinegar types.

3.3. Vinegar classification with KNN, SIMCA and PLS-DA methods

After the success obtained by the above exploratory analysis, classification methods were applied to the samples. An important step in the development of a classification strategy is the splitting of the data set into a training and a validation or test set (used to evaluate the performance of the prediction ability of classification model for new samples). In this work, five authentic samples were selected to form the external validation set, as indicated in Table 1. Excellent results were obtained by all models since none of the validation samples were misclassified.

Finally, new commercial samples were predicted and excellent results were obtained by KNN, SIMCA and PLS-DA models, since the prediction abilities were 100% for all categories. These results have shown that all of the classification methods tested are efficient to resolve the problem proposed.

3.4. Discrimination of all vinegars types

The PCA score plot, presented in Fig. 4, shows that applying the chemometric methods to ^{1}H NMR spectra it is still possible to discriminate a more complex data set including vinegars produced by honey, orange, pineapple and rice to that analyzed previously (wine, apple and alcohol/*agrin*). The first principal component (PC1) describes 18.5% of total variance, while the second component (PC2) describes 17.4%; the two PCs together express 35.9% of the total variability. In this figure it can be observed that the replicates of each vinegar type are well grouped, showing once again, the good reproducibility of ^{1}H NMR spectra. Moreover, each vinegar type formed clusters well defined and discriminated from other groups. The honey vinegars were located more distant with respect to others, because in contrast to all of others, they did not present the ethanol signals. Moreover, pineapple vinegars presented the highest amount of tartaric acid, rice vinegars are richer in lactic acid compared to the others and did not present the methanol signal. The orange vinegars also contain more lactic acid than others, but did not present the methanol and alanine signals.

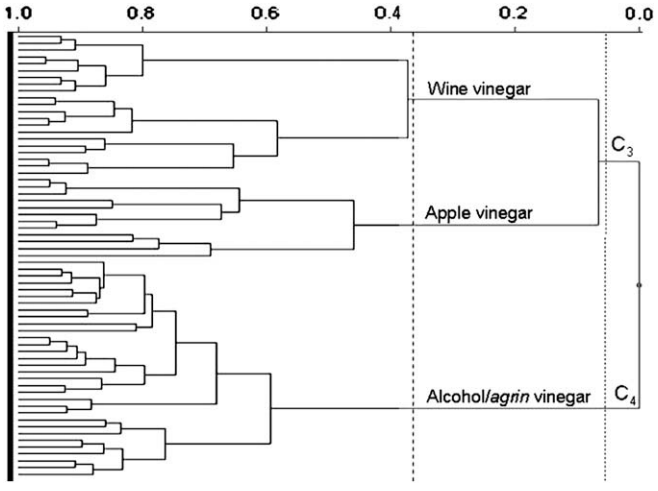

Fig. 3. HCA dendrogram obtained from ^{1}H NMR spectra from different vinegar types (dotted line: similarity index of 0.06 and dashed line: similarity index of 0.37).

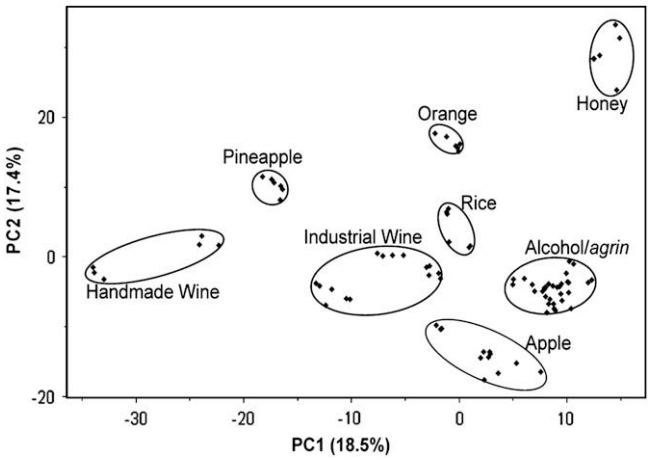

Fig. 4. PCA score plot (PC1 × PC2) showing the discrimination between all vinegar types studied.

In conclusion, the present study demonstrated the potential of a ^{1}H NMR spectroscopic approach in discriminating vinegars according to the raw materials as wine, apple and alcohol/*agrin* vinegars. Moreover, the pattern recognition methods had shown that it is possible to classify the commercial vinegar samples according to the raw material they are generated from. The results obtained in the differentiation of commercial vinegars showed that it is possible to discriminate between honey, orange, pineapple, and rice with respect to the others vinegars. In addition, the characteristic components which have the most influence on the separation among different groups of samples were found out. They could be used as chemical markers in NMR quality control of vinegars in the future. The generated data provided valuable insight into the application of NMR fingerprint coupled with powerful chemometrics analysis in the analysis and quality control of vinegars.

Acknowledgments

This work has been financially supported by CAPES, CNPq, FAPESP and FINEP.

References

Alan, T. M., & Alan, M. K. (2005). Chemometric analysis of NMR spectroscopy data: a review. *Annual Reports on NMR Spectroscopy, 54*, 41–80.

Alcantara, G. B., Honda, N. K., Ferreira, M. M. C., & Ferreira, A. G. (2007). Chemometric analysis applied in ^{1}H HR-MAS NMR and FT-IR data for chemotaxonomic distinction of intact lichen samples. *Analytica Chimica Acta, 595*, 3–8.

Beckonert, O., Bollard, M. E., Ebbels, T. M. D., Keun, H. C., Antti, H., Holmes, E., et al. (2003). NMR-based metabonomic toxicity classification: hierarchical cluster analysis and k-nearest-neighbor approaches. *Analytica Chimica Acta, 490*, 3–15.

Beebe, K. R., Pell, R. J., & Seasholtz, M. B. (1998). *Chemometrics: A practical guide*. New York: John Wiley & Sons.

Boffo, E. F., & Ferreira, A. G. (2006). Determinação da origem biossintética de ácido acético através da técnica SNIF-NMR. *Química Nova, 29*, 456–458.

Brereton, R. G. (2002). *Chemometrics: Data analysis for the laboratory and chemical plant*. Chichester: John Wiley & Sons.

Caligiani, A., Acquotti, D., Palla, G., & Bocchi, V. (2007). Identification and quantification of the main organic components of vinegars by high resolution ^{1}H NMR spectroscopy. *Analytica Chimica Acta, 585*, 110–119.

Chen, Y., Zhu, S. B., Xie, M. Y., Nie, S. P., Liu, W., Li, C., et al. (2008). Quality control and original discrimination of *Ganoderma lucidum* based on high-performance liquid chromatographic fingerprints and combined chemometrics methods. *Analytica Chimica Acta, 623*, 146–156.

Consonni, R., & Cagliani, L. R. (2008). Geographical characterization of polyfloral and acacia honeys by nuclear magnetic resonance and chemometrics. *Journal of Agricultural and Food Chemistry, 56*, 6873–6880.

Consonni, R., Cagliani, L. R., Benevelli, F., Spraul, M., Humpfer, E., & Stocchero, M. (2008). NMR and chemometric methods: a powerful combination for characterization of balsamic and traditional balsamic vinegar of Modena. *Analytica Chimica Acta, 611*, 31–40.

Defernez, M., & Cololquhoun, I. J. (2003). Factors affecting the robustness of metabolite fingerprinting using ^{1}H NMR spectra. *Phytochemistry, 62*, 1009–1017.

García-Parrilla, M. A., González, G. A., Heredia, F. J., & Troncoso, A. M. (1997). Differentiation of wine vinegars based on phenolic composition. *Journal of Agricultural and Food Chemistry, 45*, 3487–3492.

Hermann, A. (2001). Determination of D/H isotope ratio in acetic acid from vinegars and pickled products by ^{2}H-NMR-spectroscopy. *European Food Research and Technology, 212*, 683–686.

Jalali-Heravi, M., Masoum, S., & Shahbazikhah, P. (2004). Simulation of ^{13}C nuclear magnetic resonance spectra of lignin compounds using principal component analysis and artificial neural networks. *Journal of Magnetic Resonance, 171*, 176–185.

Le Gall, G., Puaud, M., & Colquhoun, I. J. (2001). Discrimination between orange juice and pulp wash by ^{1}H nuclear magnetic resonance spectroscopy: identification of marker compounds. *Journal of Agricultural and Food Chemistry, 49*, 580–588.

Lindner, P., Bermann, E., & Gamarnik, B. (1996). Characterization of citrus honey by deuterium NMR. *Journal of Agricultural and Food Chemistry, 44*, 139–140.

Lindon, J. C., Holmes, E., & Nicholson, J. K. (2001). Pattern recognition methods and applications in biomedical magnetic resonance. *Progress in Nuclear Magnetic Resonance Spectroscopy, 39*, 1–4.

Martin, G. J., Guillou, C., Martin, M. L., Cabanis, M. T., Tep, Y., & Aerny, J. (1988). Natural factors of isotope fractionation and the characterization of wines. *Journal of Agricultural and Food Chemistry, 36*, 316–322.

Martin, G. J., Zhang, B. L., Naulet, N., & Martin, M. L. (1986). Deuterium transfer in the bioconversion of glucose to ethanol studied by specific isotope labeling at the natural abundance level. *Journal of the American Chemical Society, 108*, 5116–5122.

Masoum, S., Bouveresse, D. J.-R., Vercauteren, J., Jalali-Heravi, M., & Rutledge, D. N. (2006). Discrimination of wines based on 2D NMR spectra using learning vector quantization neural networks and partial least squares discriminant analysis. *Analytica Chimica Acta, 558*, 144–149.

Massart, D. L., Vandeginste, B. G. M., Deming, S. M., Michotte, Y., & Kaufman, L. (2001). *Data handling in science and technology*. Amsterdam: Elsevier.

Natera, R., Castro, R., García-Moreno, M. V., Hernández, M. J., & García-Barroso, C. (2003). Chemometric studies of vinegars from different raw materials and processes of production. *Journal of Agricultural and Food Chemistry, 51*, 3345–3351.

Nouwen, J., Lindgren, F., Hansen, B., Karcher, W., Verhaar, H. J. M., & Hermens, J. L. M. (1997). Classification of environmentally occurring chemicals using structural fragments and PLS discriminant analysis. *Environmental Science & Technology, 31*, 2313–2318.

Ogrinc, N., Kosir, I. J., Spangenberg, J. E., & Kidric, J. (2003). The application of NMR methods for detection of adulteration of wines, fruit juices, and olive oil. A review. *Analytical and Bioanalytical Chemistry, 76*, 424–430.

Pupin, A. M., Dennis, M. J., Parker, I., Kelly, S., Bigwood, T., & Toledo, M. C. F. (1998). Use of isotopic analyses to determine the authenticity of Brazilian orange juice (*Citrus sinensis*). *Journal of Agricultural and Food Chemistry, 46*, 1369–1373.

Remaud, G., Guillou, C., Vallet, C., & Martin, G. J. (1992). A coupled NMR and MS isotopic method for the authentication of natural vinegars. *Fresenius' Journal of Analytical Chemistry, 342*, 457–461.

Tavares, L. A., Ferreira, A. G., Ferreira, M. M. C., Correa, A., & Mattoso, L. H. (2005). Analysis of blends of raw coffees of Arabica and Robusta varieties by ^{1}H NMR and chemometric methods. *Magnetic Resonance in Food Science,* 80–88.