



Chemometric authentication of Moroccan Picholine virgin olive oil by automatic classification based on the composition of fatty acids and sterols

M. Elbir¹, J. Oukrich², A. Moubarik³, A. Amhoud¹, M. Fakir², M. Berkani⁴, M. Naciri Bennani⁵, M. Mbarki^{1*}

¹ *Transdisciplinary Team of Analytical Sciences for Sustainable Development, Faculty of Science and Technology, Sultan Moulay Slimane University, PB 523, 23000 Beni Mellal, Morocco*

² *Team of Processing Information and Communications, Faculty of Science and Technology, Sultan Moulay Slimane University, PB 523, 23000 Beni Mellal, Morocco*

³ *Institut of Nanomaterials and Nanotechnology, ENSET, Madinat el Irfane, Rabat, Morocco*

⁴ *Laboratory of Applied Spectrochemistry and Environment, Faculty of Science and Technology, Sultan Moulay Slimane University, PB 523, 23000 Beni Mellal, Morocco*

⁵ *Laboratory of Chemical Biology Applied to the Environment, Faculty of Sciences, Meknes, Morocco*

Received 2 May 2013, Revised 23 July 2013, Accepted 23 July 2013

** Corresponding author. E mail: mbarkimeister@gmail.com; Tel: +212 6 66 72 20 38*

Abstract

The traceability of the samples is a recommended procedure to ensure the authenticity of virgin olive oil (VOO). The automatic classification is gaining more interest among growers and researchers in the field of olive cultivation, to manage the VOO quality control. The first objective of this work is to characterize, in terms of fatty acids and sterols, virgin olive oil from the Meknes Tafilalt area in Morocco since there is a paucity of research on such a land product as Mediterranean food presenting nutritive and therapeutic values. The second objective is to detect minimal gas chromatographic differences between samples coming from the same cultivar (Moroccan Picholine) and the same geographical area (Meknes Tafilalt). The third objective is to approach the effectiveness and efficiency of the coupling between the gas chromatography on the one hand and on the other chemometric treatment. So, using the Principal Components Analysis (PCA), K-means and Hierarchical Ascendant Clustering (HAC) methods, the present work is devoted to the characterisation and the authentication based on the composition of the Moroccan Picholine VOO variety in terms of fatty acids and sterols. Fatty acids are determined by gas chromatography (GC) after having transformed them to their corresponding metylic esters. The fact of not observing enough aberrant points confirms that the samples have similar varietal and regional origins. The PCA has allowed us to reduce a large number of variables (17 variables) to 7 new ones, respecting the minimization of the information loss. The results of the application of K-Means and HAC on the VOO samples have allowed detecting correlations between them. The results also showed that there is little variability in the content of fatty acids and sterols in the VOO produced in the Moroccan Meknes Tafilalt area.

Keywords: Virgin olive oil; Automatic classification; Principal Components Analysis, K-Means, Agglomerative Hierarchical Clustering.

1- Introduction

The virgin olive is a natural biochemical complex system presenting a combination of different components that is not easy to be identified only after the difference of the varietal or the geographic origin. Now days, an important role in determining the virgin olive oil (VOO) quality is rising thanks to the automatic classification since it can easily and rapidly identify, characterize and discriminate similar virgin olive oils. Furthermore, such a fast classification is also essential to manage eventual alteration and falsification in VOO [1-3]. The quality of VOO is defined as the set of chemical, physical and sensory parameters, to classify olive oils into different categories according to the effective standards of the International Olive Oil Council [4]. This quality is influenced by a combination of factors such as variety, method of harvesting, extraction process and storage [5-6]. Various works in the world have been invested in the VOO rewarding through analytical methods coupled to chemometrics [7-16]. Moreover, there is a need for fast and simple routine analytical methods to control the quality of the Moroccan Picholine VOO cultivar. PCA, K-Means and HAC methods are coupled to gas chromatography (GC) to obtain a fast, robust and less expensive means for this quality control. In fact, gas

chromatography as a still not routine analysis technique should be developed in order that laboratories ensure easily access to such a high cost analysis technique presented as a reference method. Results of analyses of VOOs gas chromatographic characteristics are saved in a data base and one do not need repeat the same analyses for checking each VOO sample whose data are close to some of gas chromatographic characteristics of VOOs in the data base. Coupling a software tool to a sufficient data base of gas chromatographic characteristics leads to results better than those before coupling. So the price/quality ratio would be reduced since the results errors would be at about 2 to 3%. In addition, the analysis time is very small compared to the analysis without such a coupling. Thus, such a coupling would be a profitable investment for routine and industrial laboratories. The present work aims to characterize, in terms of fatty acids and sterols, virgin olive oil from the Meknes Tafilalt area in Morocco and to detect minimal gas chromatographic differences between samples coming from the same variety (Moroccan Picholine) and the same geographical area (Meknes Tafilalt). Other objective of this work is to approach the coupling between the gas chromatography chemometrics. So, we carried out this work on VOO extracted from a Moroccan Picholine olive variety. Moreover, we investigated of fatty acids and sterols contents of Moroccan Picholine VOO in order to develop (GC) method since it remains a relatively expensive analytical means. Chemometrics coupled to GC technique is a way contributing to the easy establishment of different foods mapping. In fact, thanks to the automatic classification anterior studies have discriminated milk, wine and VOO [17-20], according to the geographic origin. Such fast discrimination, only according to the geographic origin of VOO is very interesting especially for samples coming from the same olive tree variety as the case of the Moroccan VOO since 96% of the olive cultivars are from the same Moroccan Picholine cultivar.

2. Materials and methods

2.1. Sample collection

46 samples of the Moroccan Picholine VOO have been collected in the Meknes Moroccan area, between October 2009 and March 2010. VOO has been extracted by olives mechanical press method, and then stored at 14 °C. We used the International Olive Oil Council [21] and the European Union [22] definitions of the quality of VOO based on the parameters of the quality that are mainly the degree of acidity (expressed as percentage of oleic acid), peroxide value, specific extinction values in the UV absorbance at 232 nm and 270 nm and organoleptic rating. We were especially interested in fatty acids and sterols, two non-volatile chemical constituents classes that we have determined by GC after having transformed them to their corresponding methyl esters that are volatiles constituents.

2.2. Gas chromatography

Among all the 46 VOO samples only 20 presented a quality according the International Olive Oil Council (IOOC) standards [21]. So, we prepared methyl esters of fatty acids from these 20 samples of VOO, using the derivation method. The methyl esters were then analyzed by gas chromatograph equipped with a FID. However, the chromatograms coming from the 20 VOOs are so similar that it is very difficult to differentiate between them and so that we proceeded to their automatic classification by algorithms. The chromatograms are then converted into digital data that were processed statistically. We established a database containing the 20 VOO samples that respond to the IOOC standards.

2.3. Analysis of chromatographic data

We analyzed the data to evaluate their dispersion and detect the most aberrant results (Scatterplots) [23]. The principal components analysis (PCA) is a multivariate analysis technique that reduces a large number of original correlated variables to a limited number of new uncorrelated variables with a minimal loss of information. Three steps of PCA were performed, identifying the most aberrant results, calculating a correlation matrix between the original variables and determining factors as new variables.

The main statistical indicators in the extraction of the main factors of inertia, called principal components in PCA, are:

- the eigenvalue λ_i : represents the inertia of the main axis of rank i ;
- the total inertia is equal to the sum of the eigenvalues $I = \sum_{i=1}^n \lambda_i$
- The contribution of each to the total inertia main axis is given by the percentage of inertia $CSV = \frac{\lambda_i}{I} * 100$

The choice of the number of principal components, to be used to represent data in the new basis of eigen vectors or inertia principal axes, can be done in ways:

Three rules apply:

- Rule Kaiser that retains only factors with eigen values greater than 1
- The number of axes in accordance with the minimum refund information (80%) is chosen
- The "Scree-test" or test of the Coude: It retains only the values that are to the left of the inflection point of the graph eigen values . Graphically, we join the points on a straight line and the components number to retain corresponds to points outside this line.

The linear correlation coefficient R is related to the average by the following relationship:

$$R = \frac{\sigma_{xy}}{\sigma_x * \sigma_y} = \frac{\sum_{i=1}^n (x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} * \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

If R = 0, the variables are not correlated and they are much better correlated if R is far from 0 (close to -1 or 1).

2.4. Classification methods

The main purpose of the clustering methods is to assign to each observation a "tag" class. However, in the case of supervised classification, the available observations would be already a class label, and the objective is to assign a new observation to a class.

2.4.1. K-means [24]

The K-Means algorithm, that has been proposed by Mac-Queen in 1967, is used the present work to classify or group objects based on attributes. The user sets the number of classes to a or k, in principle.

- K classes are constructed with the first k points that are the gravity centers of the classes.
- Among the nk remaining points, we take the first that is assigned to the class whose center of gravity is the closest and then we determine the new gravity center of the class that receives the point.
- The previous step is repeated until exhaustion of all points (n-(ku) = 0).
- To obtain the desired partition in a single pass.

2.4.2. Hierarchical ascending classification [25]

The algorithm of hierarchical ascendant clustering (HAC) has been implemented in four stages. The first is an initialization step which consists of the construction of the table of distances regardless of the formula used to construct the algorithm for HAC is independent of the metric. Thus, between each pair of points (x, y) of M, we have a value of d(x, y). The initial partition is finer M. The second step is the grouping that concerns the route table of distances to determine the torque (x*, y*) of the most nearby.

$$d(x^*, y^*) \leq \min_{x,y \in M} \{d(x, y)\}$$

It combines the two elements in the same class $A = x^* \cup y^*$ but the other classes remain unchanged. We get a new partition coarser than the previous one. At the third step we draw the table of distances. Class A will be seen as a single point, and then we calculate the distances to the A point in order to have a set cardinality greater than one, and so that all other items, that are not in A, can be singletons. In order ensuring a generality, we note B such as:

$$d(A, B); B \not\subset A$$

So, we have a new table of distances with one row and one column less than the previous one so that the difference is only by the row and column corresponding to point A.

The fourth and final step is the stop condition that it is achieved when we reached the desired partition level. The partition is generally coarse and has only one class together for all the points. Otherwise, we leave the stage "Combination" from the table of distances that are calculated according the previous grouping.

3. Results and discussion

3.1. Gas chromatography

The chromatograms of the 20 VOOs are similar in shape but not identical, and the discrimination between them is very difficult. Figure 1 shows a typical example of these chromatograms. The retention times of the fatty acids from these VOOs are illustrated in table 1.

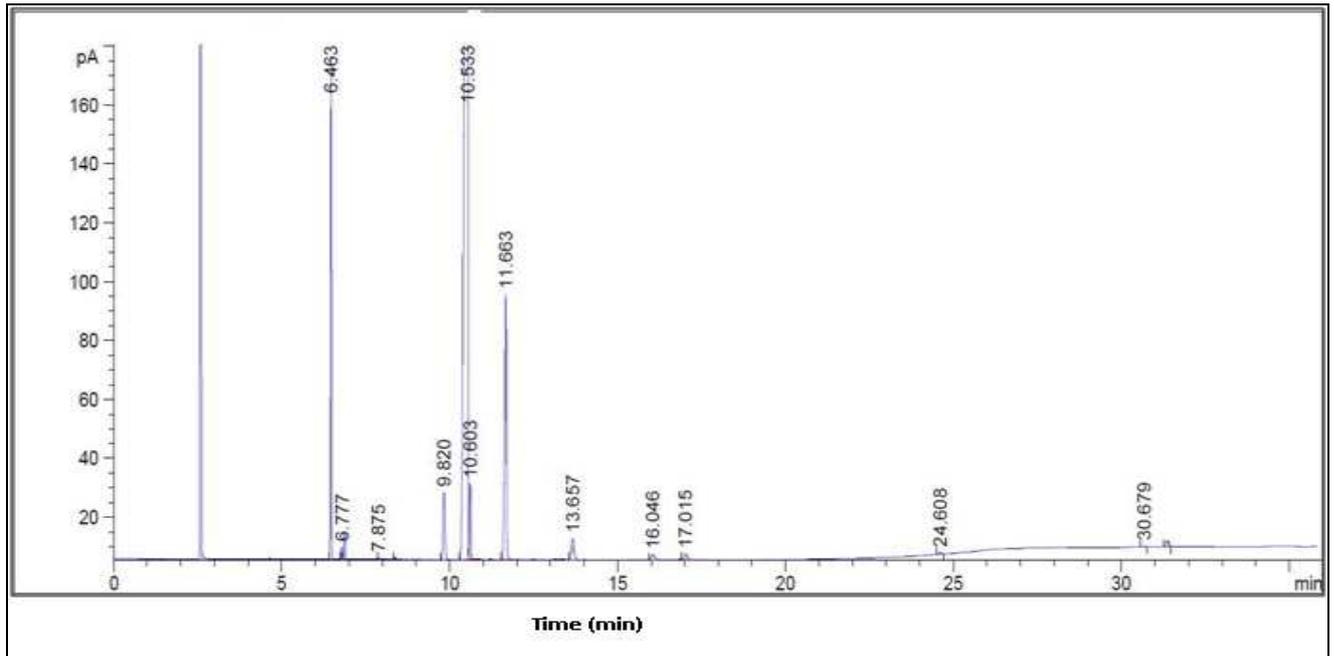


Figure 1: GC chromatogram of VOO from Meknes Moroccan area

Table 1: Main fatty acids in VOO from Meknes Moroccan Regions

Common name	Palmiti	Stearic	Oleic	Vaccenic	Linoleic	Linoleni.	Arachidi.	Gondoic
Shorthand	16:0	18:0	18:1w9	18:1w7	18:2w6	18:3w3	20:0	20:1w9
Retention time (min)	6.463	9.820	10.533	10.603	11.663	13.657	16.046	17.015

3.2. Principal Components Analysis

The aim of PCA is to reorganise the initial data (Table 2) with a large number of variables to a small number of variables obtained by combining the original variables with a relatively high correlation coefficient (table 3). The correlation matrix is shown in the table 3 while the cumulative sum of the variance is presented in the table 4.

Table 2: VOO date base:

MGi and Khj (in lines: samples codes and in columns: fatty acids (Ci) ans sterols (from Chol to Daveno)

	C160	C161	C170	C171	C180	C181	C182	C183	C200	C201	Chol	Brass	Camp	Stigm	CBAS	Dstegm	Daveno
MG04	15	2	0	0.1000	1.5000	62.6000	17.5000	0.3000	0.4000	0.3000	0.2000	0.2000	3.3000	3.3000	94.7000	0.4000	0.6000
MG07	10	0.8000	0	0.1000	2.8000	74.6000	10.7000	0.8000	0.4000	0.4000	0.1000	0	3.3000	0.7000	94.8000	0.4000	0.5000
MG13	8.4000	0.8000	0.1000	0.1000	1.9000	80.2000	7.2000	0.8000	0.3000	0.4000	0.1000	0	3.4000	0.6000	94.6000	0.3000	0.4000
MG23	8.9000	0.7000	0	0.1000	2.2000	76.9000	9.4000	0.8000	0.4000	0.3000	0.3000	0	3.5000	0.6000	94.8000	0.2000	0.2000
MG24	8.7000	0.7000	0	0.1000	2.3000	80.1000	6.1000	0.9000	0.3000	0.4000	0.2000	0	3.2000	1.3000	94.1000	0.4000	0.3000
MG25	8.2000	0.7000	0.1000	0.1000	1.8000	80.7000	7	0.8000	0.3000	0.5000	0.2000	0	3.4000	1.4000	93.5000	0.3000	0.3000
MG34	15.30...	1.8000	0	0.1000	2.9000	63.2000	15.2000	0.8000	0.3000	0.1000	0.2000	0	3.9000	1.4000	93.5000	0.2000	0.3000
MG37	11.30...	0.7000	0	0.1000	2.3000	70.3000	12.3000	1	0.4000	0.1000	0.4000	0	3.5000	1	93.9000	0.5000	0.4000
MG38	11	0.7000	0.1000	0.1000	2.3000	71.2000	12.1000	1	0.4000	0.1000	0.6000	0	3.4000	1.1000	93.9000	0.5000	0.4000
MG59	9.8000	0.8000	0	0.1000	2.5000	75	9.4000	1	0.3000	0.3000	0.1000	0	2.9000	2.8000	93	0.2000	0.3000
Kh05	9.0800	0.5800	0.0400	0.0600	2.6900	77.8200	8.3200	0.7900	0.3200	0.3000	0.1300	0.2100	2.7600	2.0200	90.0800	0.0900	0.2100
Kh14	8.6500	0.5400	0.0400	0.0700	2.2000	81.0200	6.1400	0.7600	0.2900	0.2900	0.2300	0.1700	3.0700	2.0800	92.0100	0.0900	0.1600
Kh21	8.6500	0.5400	0.0400	0.0700	2.2000	81.0200	6.1400	0.7600	0.2900	0.1600	0.2200	0.1500	1.7200	0.3400	87.9800	0.0500	0.1900
Kh23	9.7600	0.6900	0.0400	0.0700	2.1500	76.2600	9.4900	0.9300	0.3000	0.3100	0.2900	0.2900	3.0100	1.9100	91.5000	0.1300	0.1900
Kh24	8.5900	0.6100	0.0300	0.0600	2.5200	77.6100	9.2200	0.7400	0.2800	0.3400	0.0600	0.1000	3.2900	2.6600	92.5400	0.0500	0.2000
Kh25	9.3100	0.5900	0.0400	0.0600	2.2500	78.4300	8.1400	0.7600	0.3000	0.3000	0.1500	0.2300	3.3000	1.2400	92.0400	0.0400	0.3200
Kh27	9.6100	0.5800	0.0300	0.0600	2.5300	77.2800	8.5800	0.7500	0.3000	0.2700	0.0500	0.1300	3.5700	2.4000	88	0.0400	0.0700
Kh38	9.0500	0.6200	0.0300	0.0600	2.4100	76.5700	9.7800	0.8400	0.3100	0.3200	0.1100	0.1100	3.3600	1	92.9300	0.0900	0.2000
Kh43	8.8400	0.6200	0.0400	0.0600	2.7100	75.6000	10.7600	0.7600	0.2800	0.3400	0.1600	0.1400	3.3800	1.2300	93.6300	0.0500	0.1000
Kh46	9.0700	0.6300	0.0300	0.0600	2.5700	76	10.2600	0.7600	0.2800	0.3300	0.7000	0.0800	3.7600	3.8900	90.7700	0.0300	0.2100

The PCA does not give interesting results on tables large enough data: The number of statistical units should be greater than 15 and the number of variables to 4. In this study we have 20 samples and 17 variables.

Table 3: Correlation matrix (for the notations see the table 9 in the glossary)

	C160	C161	C170	C171	C180	C181	C182	C183	C200	C201	Chol	Brass	Camp	Stigm	CBAS	Dstegm	Daveno
C160	1.0000	0.9127	-0.3855	0.4066	-0.0277	-0.9613	0.8974	-0.3501	0.4265	-0.5329	0.1149	-0.0373	0.3203	0.2047	0.2723	0.3882	0.5406
C161	0.9127	1	-0.3308	0.4762	-0.2072	-0.8508	0.8048	-0.5509	0.3059	-0.2169	-0.0488	-0.0685	0.2993	0.2342	0.3880	0.3484	0.5845
C170	-0.3855	-0.3308	1.0000	-0.0805	-0.3345	0.3983	-0.3566	0.1431	-0.2190	0.1834	0.1507	0.0077	-0.0583	-0.2064	-0.0825	0.0031	-0.0704
C171	0.4066	0.4762	-0.0805	1	-0.3121	-0.3870	0.3055	0.1376	0.5898	-0.0615	0.1071	-0.7089	0.1956	-0.2769	0.6928	0.8582	0.7028
C180	-0.0277	-0.2072	-0.3345	-0.3121	1.0000	-0.0472	0.0211	0.3929	-0.2079	-0.2353	-0.0526	-0.1415	-0.0150	-0.2096	-0.3188	-0.3945	
C181	-0.9613	-0.8508	0.3983	-0.3870	-0.0472	1	-0.9766	0.2882	-0.5042	0.5215	-0.2009	0.1013	-0.4135	-0.2366	-0.3634	-0.4141	-0.5331
C182	0.8974	0.8048	-0.3566	0.3055	0.0211	-0.9766	1.0000	-0.3602	0.5264	-0.4294	0.2198	-0.0479	0.4499	0.2738	0.3896	0.3669	0.5216
C183	-0.3501	-0.5509	0.1431	0.1376	0.3929	0.2882	-0.3602	1.0000	-0.0558	-0.2195	0.2213	-0.4028	-0.0156	-0.3656	0.0158	0.1617	-0.2373
C200	0.4265	0.3059	-0.2190	0.5898	-0.2079	-0.5042	0.5264	-0.0558	1.0000	-0.2891	0.2618	-0.2938	0.1738	-0.2347	0.5041	0.7365	0.6580
C201	-0.5329	-0.2169	0.1834	-0.0615	-0.2353	0.5215	-0.4294	-0.2195	-0.2891	1	-0.3748	0.0164	0.0702	0.1245	0.1977	-0.1293	-0.0271
Chol	0.1149	-0.0488	0.1507	0.1071	-0.0526	-0.2009	0.2198	0.2213	0.2618	-0.3748	1	-0.1502	0.2149	0.1861	0.0363	0.2419	0.0955
Brass	-0.0373	-0.0685	0.0077	-0.7089	-0.1415	0.1013	-0.0479	-0.4028	-0.2938	0.0164	-0.1502	1.0000	-0.3640	0.2949	-0.5195	-0.5697	-0.2988
Camp	0.3203	0.2993	-0.0583	0.1956	0.1518	-0.4135	0.4499	-0.0156	0.1738	0.0702	0.2149	-0.3640	1.0000	0.2130	0.4844	0.2129	0.1419
Stigm	0.2047	0.2342	-0.2064	-0.2769	-0.0150	-0.2366	0.2738	-0.3656	-0.2347	0.1245	0.1861	0.2949	0.2130	1	-0.2270	-0.2613	-0.0529
CBAS	0.2723	0.3880	-0.0825	0.6928	-0.2096	-0.3634	0.3896	0.0158	0.5041	0.1977	0.0363	-0.5195	0.4844	-0.2270	1.0000	0.6704	0.6234
Dstegm	0.3882	0.3484	0.0031	0.8582	-0.3188	-0.4141	0.3669	0.1617	0.7365	-0.1293	0.2419	-0.5697	0.2129	-0.2613	0.6704	1.0000	0.7913
Daveno	0.5406	0.5845	-0.0704	0.7028	-0.3945	-0.5331	0.5216	-0.2373	0.6580	-0.0271	0.0955	-0.2988	0.1419	-0.0529	0.6234	0.7913	1.0000

After the correlation matrix, we observe that several variables are correlated (> 0.5). So, this allows to conclude that the factorization is possible.

Table 4: Cumulative Sum of the variance

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
CSV	36.69	54.94	67.04	75.65	83.39	87.87	91.61	93.84	95.79	97.39	98.33	99.09	99.53	99.92	99.98	99.99	100

The table 4 shows the extraction of the first five principal components that allow taking into account approximately 83.39% of the total inertia.

- Using Kaiser Rule that retain only factors with eigenvalues greater than 1, we can extract five principal components (Figure 2).
- Using the Scree-Test criterion from the eigen values results (Figure 2), the following component numbers 6,7,8,9,10,11,12,13,14,15,16 and 17 can be considered on the same straight line. Retained components are the components 1, 2, 3, 4 and 5. They correspond to 83.39% of the initial values of variance (figures 3). The examination of the graph shows a limitation to the extraction of the first five principal components that allow taking into account approximately 83.39% of the total inertia.

The extraction of the principal components from the graphs of the eigen values (Scree test criterion) do not provide the best result, in space. Such situation can be avoided by examining the correlation matrix on the one hand and the other hand by the principal components graph (Figure 4).

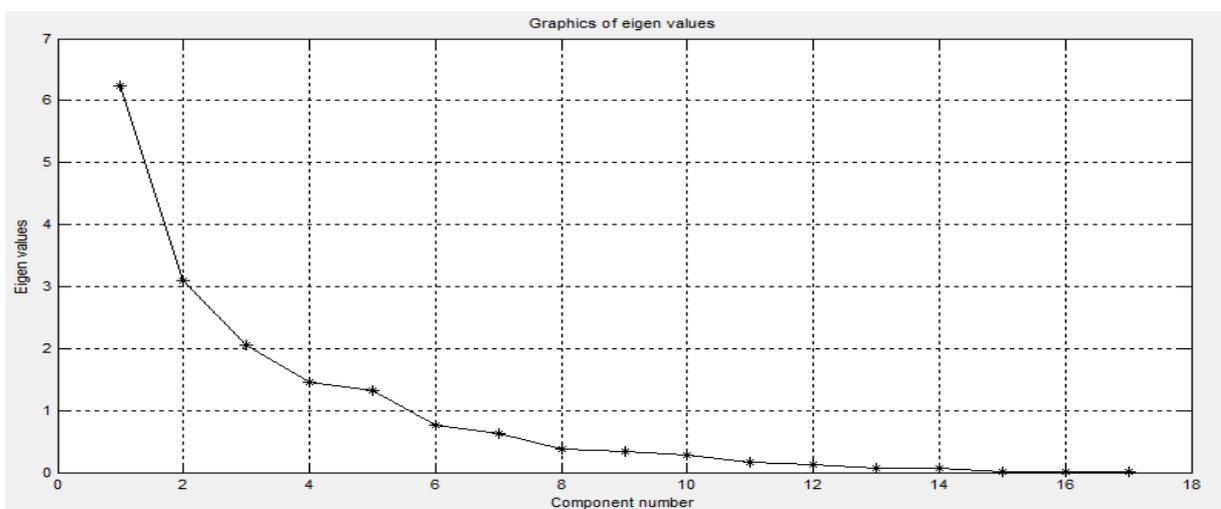


Figure2: Graph of Eigen values

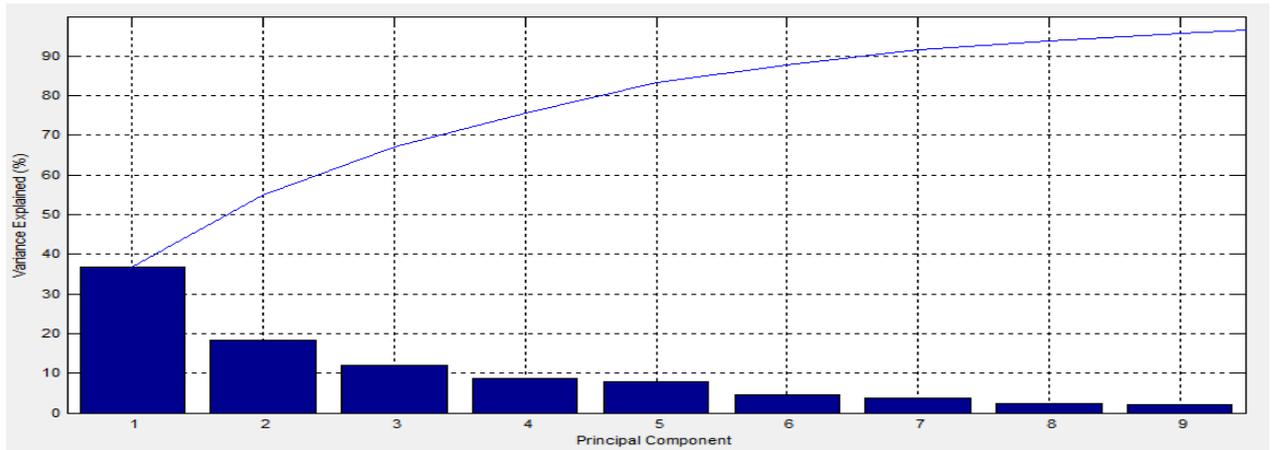


Figure 3: Total variance according to the retained principal components

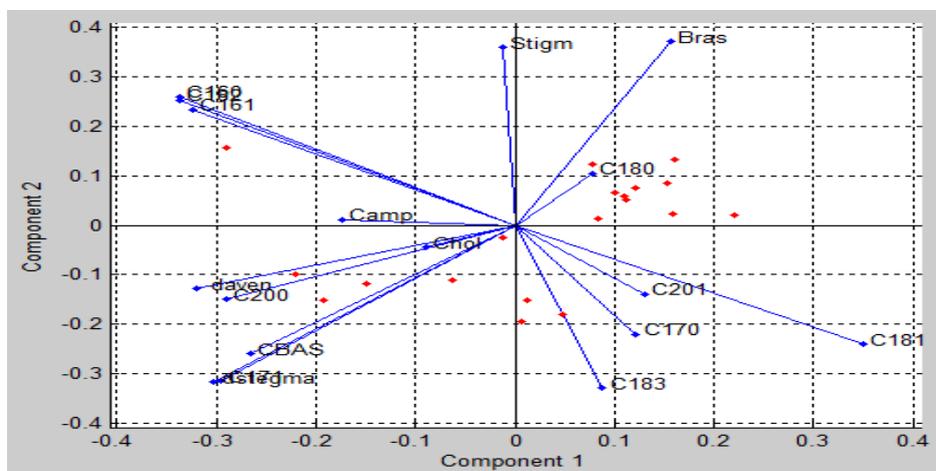


Figure 4: 2 dimensions principal components: PC1 and PC2.

The eligible elements for the 7 retained principal components are given in the table 5.

Table 5: Elements of retained principal components

Component	1	2	3	4	5	6	7
Elements	C160,C161,C181,C182,C183	C170, C201	stigm	C180	CHOL	C171,C200,Brass,Cbas, Dstegm,Daveno,	CAMP

3.3. Classification via query (Database)

The execution of the query from database name "oil" (Table 6) gives the results shown in Table 7.

Table 6: Application Selection

Querying a database "olive oil"	SELECT ALL Echantillons,C161,C200,Cholesterol FROM "oil" WHERE C161 BETWEEN 0.3 AND 3.5 AND C170 <= 0.5 AND C171 <= 0.6 AND C180 BETWEEN 0.5 AND 5 AND C181 BETWEEN 55 AND 83 AND C182 BETWEEN 3.5 AND 21 AND C183 <= 1.5 AND C200 <= 0.8 AND Cholesterol <= 0.5
---------------------------------	--

Table 7: Query Results

Samples	MG 04	MG 07	MG 13	MG 23	MG 24	MG 25	MG3 4	MG 37	MG 59	Kh 05	Kh 14	Kh 21	Kh 23	Kh 24	Kh 25	Kh 27	Kh3 8	Kh43
C161	2	0.8	0.8	0.7	0.7	0.7	1.8	0.7	0.8	0.58	0.54	0.54	0.69	0.61	0.59	0.58	0.62	0.62
C200	0.4	0.4	0.3	0.4	0.3	0.3	0.3	0.4	0.3	0.32	0.29	0.29	0.3	0.28	0.3	0.3	0.31	0.28
Cholesterol	0	0.1	0.1	0.3	0.2	0.3	0.2	0.4	0.1	0.13	0.23	0.22	0.29	0.06	0.15	0.05	0.11	0.16

The database allows us to conclude that the two samples KH46 and MG38 are not in concordance of the standards of the olive oil quality [4, 21].

3.4. Data dispersion Diagram

The data dispersion diagram is presented in the figure 5 allows us to deduce the most outliers.

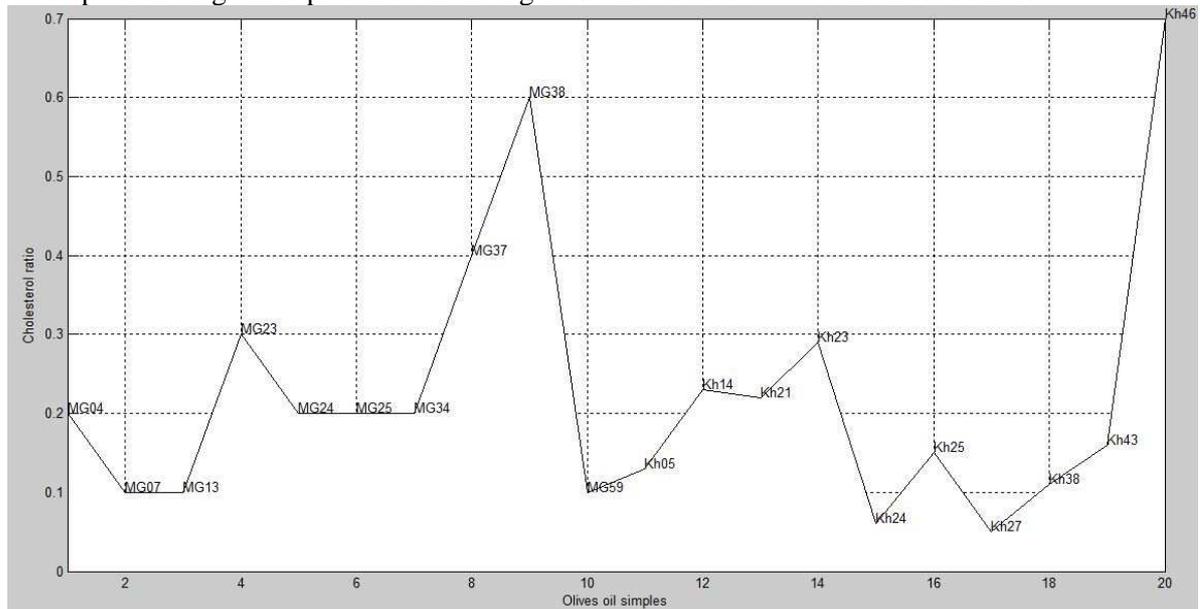


Figure 5: Samples dispersion diagram according to the cholesterol ratio

This diagram, in figure 5, confirms to us to conclude that the two samples KH46 and MG38 are not in concordance of the standards of the olive oil quality.

3.5. Classification by K-Means

The table 8 illustrates the classification results after using a K-means method.

Table 8: Classification by the K-means method (K=2, K=5, K=10 and K=14), K is number of classes.

	2		5		10		14
	N° Class		N° Class		N° Class		N° Class
MG04	2	MG04	4	MG04	2	MG04	4
MG07	1	MG07	5	MG07	3	MG07	1
MG13	1	MG13	1	MG13	6	MG13	10
MG23	1	MG23	5	MG23	3	MG23	6
MG24	1	MG24	1	MG24	6	MG24	9
MG25	1	MG25	1	MG25	6	MG25	9
MG34	2	MG34	4	MG34	1	MG34	5
MG37	2	MG37	3	MG37	8	MG37	7
MG38	2	MG38	3	MG38	7	MG38	7
MG59	1	MG59	5	MG59	5	MG59	11
Kh05	1	Kh05	2	Kh05	9	Kh05	3
Kh14	1	Kh14	1	Kh14	6	Kh14	13
Kh21	1	Kh21	2	Kh21	10	Kh21	14
Kh23	1	Kh23	5	Kh23	5	Kh23	11
Kh24	1	Kh24	5	Kh24	4	Kh24	2
Kh25	1	Kh25	1	Kh25	4	Kh25	2
Kh27	1	Kh27	2	Kh27	9	Kh27	8
Kh38	1	Kh38	5	Kh38	3	Kh38	6
Kh43	1	Kh43	5	Kh43	3	Kh43	6
Kh46	1	Kh46	5	Kh46	5	Kh46	12

After the increasing number of the classes and using the Euclidean gap, we observe that elements of classes change from one class to another according to the closest distance. The problem is the virtue of the choice of the number of classes in principle. To remedy this problem, we applied the HAC method.

3.6. Classification by HAC

HAC method overcomes the problem of initialization of the classes number. The classification result is shown in the figure 6.

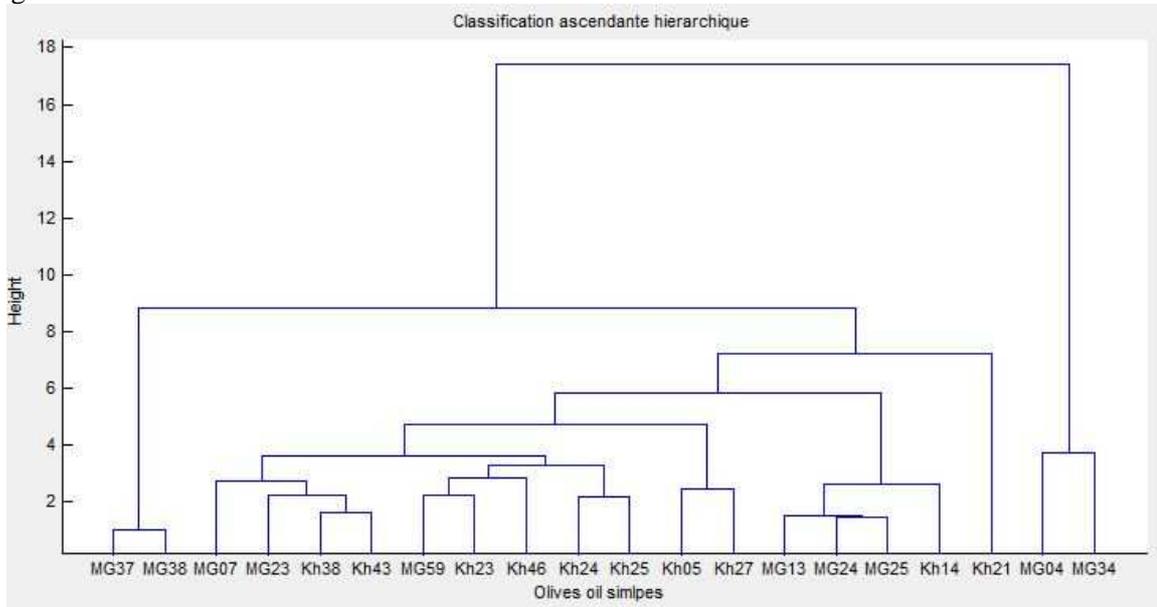


Figure 6: Classification dendrogram the by the HAC method

This dendrogram allows us to deduce that the two samples MG04 and MG34 have similar properties and are somewhat different from other samples such as MG37, KH43 and MG13. This method aims to remedy the problem of initialization of the classes number.

Several factors affects the quality of the VOO that is known as a food which is very exposed to the degradation, especially because of its oxidation and increasing free fatty acid contents in the vegetal oil. Among the parameters determining the VOO quality there is the variety of olive, maintenance of soil, climate, irrigation, processing plant, olives harvesting method and their conservation before crushing them, mode of extraction and storage of the oil. In the present work, the quality of the VOO is defined according to the IOOC standards and then determined after the procedures of such organism. Since the initial chromatogram (figure 1) of the samples cannot easily allow discriminating VOOs coming from the same Moroccan Picholine tree cultivar, it is very important that we used a fast robust means like the classification by algorithms. In the present work, the classification tools allowed us to conclude that almost all the 20 samples of the database are according to IOOC standards in terms of the olive oil quality, exception to the Kh46 and MG38 samples. The PCA has, also, allowed us to reduce a large number of variables (17 variables) to 7 new ones respecting the minimization of the information loss (figures 2-4). The diagram in the figure 5 concerning the samples scatterplots based on cholesterol confirm the results we obtained using the tool database. From this diagram, we conclude that the Kh46 and MG38 samples have cholesterol ratio that is greater than 0.5. Effectively, these two samples do not respond to the criteria of the VOO quality according to IOOC standards. Actually, scatterplots of data plays a very important role for the identification of the most aberrant observations and it is an essential step in the principal components analysis (PCA) and other analysis. The fact that we did not detect any aberrant observation confirms that the varietal origin is the same for all the samples. Moreover, the geographic origin is similar for all VOOs since collecting of olives were in a same geographic area, in Morocco.

After the K-Means application (table 8), if we make the number of class increasing successively we see that, using the Euclidean distance, the elements classes change from one class to another according to the closest distance. However, the problem is the virtue in the choice of the class number. To remedy this problem, we applied the CAH method. The results of K-Means application and HAC (figure 6) show that the MG04 sample

has proprieties that is similar to those of the MG34. In fact, the HAC dendrogram in the figure 6 allows us to deduce that the two samples MG04 and MG34 have similar properties and that are slightly different compared to other samples such as MG13, MG37, and Kh43. In addition, several previous studies have shown that chemometrics can be coupled to various analytical chemical methods to ensure authentication of foods like the VOOs especially when they present very similar macroscopic and microscopic characteristics. In fact, the Raman spectroscopy was performed, using linear discriminating analysis (LDA), to identify adulteration in VOOs by soybean oil [8]. To discriminate French VOOs that are registered as designations of origin (RDOS) and thanks to chemometrics, Ollivier D. and al. have in 2006 [12], also, used sensory and chromatography characteristics like fatty acid and triacylglycerol compositions. In fact, a linear discriminating analysis on samples that are described by 37 parameters allows them the differentiation of these RDOS VOOs. In 2011, using to the hierarchical ascendant clustering (HAC) analysis, derivative FTIR spectroscopy have allowed De Luca M. and al. [14] the classification of Moroccan VOOs. However, the results shown in the five figures, figure 2 to figure 6, would be original and interesting for studying the Moroccan Picholine olive cultivar. Actually, the present work allows a tool of coupling gas chromatography and chemometric methods (PCA, K-Means and HAC) for an automatic classification of the Moroccan virgin olive based on its fatty acids and sterols contents. In fact, these constituents are very important since the fatty acids are known for their nutritional and therapeutic values and the sterols present a cholesterol-lowering effect [26-28]. Such a tool is useful to overcome the problematique of finding fast and robust method in all the fields of quality control like in customs service that checks food quality and in routine food analysis laboratory. The coupling between methods, as it is presented in the present work, contributes to the development of the gas chromatography, as initially expensive analysis technique, in order that routine laboratories have easier access to it. Actually, such coupling ensures gain in terms of time and analysis cost. In addition, we are interested in extension of the present work by using other techniques of automatic classification, such as: methods of neural network, SVM, decision tree and Bayesian network method to calculate the probabilities of the elements influence on the quality of VOO.

Glossary

Table 9: Notations of the fatty acids and the sterols in the VOO.

	Characteristics	Notations	Common name
Fatty Acids	C16 :1	C161	Palmitic acid
	C17 :0	C170	Margaric acid
	C17 :1	C171	Heptadecenoic acid
	C18 :0	C180	Stearic acid
	C18 :1	C181	Oleic acid
	C18 :2	C182	Linoleic acid
	C18 :3	C183	linolenic acid
	C20 :0	C200	Arachidic acid
Sterols	Cholesterol	Chol	Cholesterol
	Campesterol	Camp	Campesterol
	Brassicasterol	Brass	Brassicasterol
	Delta7avenosterol	DAVEN	$\Delta 7$ Avenosterol
	Cholesterol + Betasitosterol + $\Delta 5$ Avenosterol + $\Delta 24$ stigmasterol	CBAS	Cholesterol + Betasitosterol + $\Delta 5$ Avenosterol + $\Delta 24$ stigmasterol

Conclusion

In the present work, the integral results of PCA, K-Means and HAC show that there are concordances and complementarities in the way of minimizing the gas chromatographic information loss. Now days, it is important to discriminate between virgin olive oils having a same varietal origin because such discrimination is not easy as an analysis means eventhough the olive geographic origin is different. In the present work the VOOs have the same varietal origin that is called the Moroccan Picholine and the area in were we collected olives is almost similar since it is the Moroccan Meknes Tafilalt area. Thanks to an automatic classification based on gas chromatographic characteristics in terms of fatty acids and sterols, the present work has allowed to autentificate very similar Moroccan Picholine virgin olives. The coupling between physical or chemical method and chemometrics optimizes the analysis cost in terms of time, human resources, equipment and chemical reagents.

Chemometrics opens new horizons to green chemistry and sustainable development as a means of developing analytical methods, in various fields applying the quality control of products like in the fraud prevention.

Acknowledgements

Our thanks go to the "Huilerie Céréales Industries" company in Meknes in Morocco, for ensuring the collection of the virgin olive oils.

References

1. Aguilera M.P., Beltràn G., Ortega D., Fernández A., Jiménez A., Uceda M. *Food Chemistry*. 89 (2005) 387.
2. Lai Y. W., Kemsley E. K., Wilson R. H. *J Agri Food Chem*. 42 (1994) 1154.
3. Luna G., Morales M. T., Aparicio R. *Food Chemistry*. 98 (2006) 243.
4. Conseil oléicole international, COI/T15/NC, n° 2/Rev., (1997)
5. Pinatel C., Petit C., Ollivier D., Artaud J. *Oléagineux, corps gras, lipides*. 11(2004) 217.
6. Tsimidou M. Z., Georgiou A., Koidis A., Boskou D. *Food Chemistry*. 93 (2005) 377.
7. Guillén M. D., Uriate P. S. *Food Chemistry*. 134 (2012) 162.
8. Mooeung K., Sanguk L., Kyeol C., Hoeil C., Young M. *J. Analy Chimica Acta*. 748 (2012) 58.
9. Maria L. C., Debora D., Tommaso G., Giancarlo C. *Food Chemistry*. 102 (2007) 571.
10. Pristouri G., Badeka A., Kontominas M.G. *Food Control*. 21 (2010) 412.
11. Elbir M., Moubarik A., Rakib E. M., Grimi N., Amhoud A., Miguel G., Hanine H., Artaud J., Vanloot P., Mbarki M. *Maderas ciencia y tecnologia*. 14 (2012) 361.
12. Ollivier D., Artaud J., Pinatel C., Durbec J.P., Guerer M. *Food Chemistry*. 97 (2006) 382-393.
13. Veillet S., Tomao V., Bornard I., Ruiz K., Chemat F. *C. R. Chimie*. 12 (2009) 895.
14. De Luca M., Terouzi W., Ioele G., Kzaiber F., Oussama A., Oliverio F., Tauler R., Ragno G. *Food chemistry* . 124 (2011) 1113.
15. Rubén M. M., Lorenzo C., Emma C., Teodoro S. K., Alessandra B. *Food Control*. 21 (2010) 890.
16. Galtier O., Dupuy N., Le Dréau Y., Ollivier D., Pinatel C., Kister J., Artaud J. *Analy Chimica Acta* 595 (2007) 136.
17. Collomb M., Butikofer U., Sieber R., Jeangros B., Bosset J.O. *Inter Dairy J*. 12 (2002) 649.
18. Etièvant P., Schlich P., Cantagrel R., Bertrand M., Bouvier J.C. *J Scie Food Agri*. 46 (2006) 421.
19. Ollivier D., Artaud J., Pinatel C., Durbec J. P., Guérière M. *J Agri Food Chem*. 51 (2003) 5723.
20. Shimoda M., Shibamoto T., Noble A. C. *J Agri Food Chem*. 41 (1993) 1664.
21. Conseil Oléicole International, COI/T.15/NC, n° 2/Rév., (2001)
22. EC. Commission regulation (EEC), *Official J L*, 248 (1991) 0001.
23. Selim S. Z., Ismail M. A. *Pattern Analy Mach Intelli*. 6(1984) 81.
24. Ritschard G. *Extraction et gestion des connaissances*. 1(4) (2001)185.
25. Greenacre M.G. *Academic Press, London*, (1984).
26. Turpeinen A.M., Kumpu M., Rönnback M., Seppo L., Kautiainen H., Jauhiainen T., Vapaatalo H., Korpela R. *Journal of Functional Foods*. 1(3) (2009) 260
27. Anne Doggrel S. *Complementary Therapies in Medicine*. 19 (1) (2011) 37
28. Musa-Veloso K., Poon T. H., Ann Elliot J., Chung C. *Prostaglandins, Leukotrienes and Essential Fatty Acids (PLEFA)*. 85(1) (2011) 9

(2014) ; <http://www.jmaterenvironsci.com>