DE GRUYTER OPEN

ASN, Vol 5, No 1, Pages 24-34, 2018

Acta Scientifica Naturalis

Former Annual of Konstantin Preslavsky University – Chemistry, Physics, Biology, Geography

Journal homepage: http://www.spisanie.shu.bg

Received: 30.11.2017

Accepted: 11.01.2018

Descriptor Fingerprints and Their Application to WhiteWine Clustering and Discrimination.

I. P. Bangov, M. Moskovkina, B. P. Stojanov

"Konstantin Preslavsky" University of Shumen, Faculty of Natural Sciences, Department of Chemistry, 115 Universitetska Str., 2, 9712, Shumen, Bulgaria
e-mail: ivan.bangov@gmail.com, mmarinan@abv.bg

Abstract: *This study continues the attempt to use the statistical process for a large-scale analytical data. A group of 3898 white wines, each with 11 analytical laboratory benchmarks was analyzed by a fingerprint similarity search in order to be grouped into separate clusters. A characterization of the wine's quality in each individual cluster was carried out according to individual laboratory parameters.*

Key words: *similarity search, descriptor fingerprints, white wine clustering*

Introduction

Wine as a valuable natural product historically and deservedly occupies an important place on our table. The interest in winemaking grows, the market for wine-making products expands its range, ever-increasing offers on the market make it difficult for the consumer to choose the desired quality for an adequate price.
Wine certification is generally estimated by physicochemical and sensory tests [1]. Routine laboratory tests used to evaluate wine quality include analytical determination of some physicochemical wine descriptors such as sugar content, density, alcohol, pH values, while sensory tests rely mainly by experts. The sensory tests are usually carried out by human senses such as flavor and taste and they require extremely experienced persons. The relationships between the physicochemical and sensory analysis are too much complex and still poorly understood [2], thus wine classification becomes a serious problem. The wine quality is characterized by a 10-point system (point 0 means very bad wine quality; point 10 is equal to excellent quality), is perceived subjectively by a flavor and taste sense perception and basically determines the price of wine production on the market.
Our investigation aimed to check and expose the possibility of the descriptor fingerprints procedure to rank and distinguish different classes of white wines based on laboratory test data. This work aims at the prediction of wine preferences from objective analytical tests that are available at the certification step.
The investigation was performed to test the potentials of a fingerprint clustering algorithm for a set of 3898 white wines in relation to some wine properties, comprised in the notion '*wine quality*".

Corresponding author: ivan.bangov@gmail.com
DOI 10.2478/asn-2018-0004

DE GRUYTER OPEN

ASN, Vol 5, No 1, Pages 24-34, 2018

Materials and Methods

A set of physicochemical laboratory data routinely used for wine characterization (fixed and volatile acidity; residual sugar, total and free sulfur dioxide, citric acid, chlorides, sulfates, density, pH and alcohol content) for a group of 3899 white wines (vinho verde samples from Northen regions of Portugal) was taken from literature [3] and applied for descriptor fingerprints creation for each object of the mentioned group. Numeric data were used to create fingerprints for each individual object by the methodology described earlier [4]. As a next step similarity search was carried out and followed with clustering procedure of Butina [5].

Results and Discussion

We have obtained a distribution of 3898 wine objects into 31 different clusters as a result of similarity search based on descriptor fingerprints. The quality of the wine samples in our case varied between 4-8 points. In the end of similarity search each cluster was composed of objects with close numerical values of laboratory parameters and a wine quality certificate. Some wines of the same quality has fallen into different clusters with different values of the laboratory descriptor parameters and vice verse the wines falling in the same cluster clearly show either the same or closely similar laboratory test parameters.

We have calculated *mean* values for each individual laboratory benchmark used to create fingerprints. These values are presented in Table 1.

In order to see what indicators distinguish wines grouped in a common cluster, we have performed a comparative analysis of each individual laboratory indicator. To facilitate visualization, the results are graphically represented in Figures 1-11, where the "*mean*" values of laboratory descriptors for each cluster being created are compared. The starting point in each figure shows the "*mean*" value for the descriptor characteristic of the entire series of wines.

Corresponding author: ivan.bangov@gmail.com
DOI 10.2478/asn-2018-0004

DE GRUYTER OPEN

ASN, Vol 5, No 1, Pages 24-34, 2018

Table 1. Laboratory test values (Mean) for 31clusters created under similarity search

descriptor / number	Fixed acidity g(tartaric acid)/dm^3)	Volatile acidity g(acetic acid)/dm^3)	Free sulfur dioxide mg/dm^3	Alcohol % vol	Sulphates g(potassium sulphat)/ dm^3)	pH	Total sulfur dioxide mg/dm^3	Chlorides g(sodium chloride/ dm^3)	Density g/ cm^3	Citric acid g/dm^3)	Residual sugar g/dm^3
a whole group	**6.90**	**0.300**	**35.00**	**10.4**	**0.50**	**3.10**	**138.0**	**0.050**	**0.9940**	**0.300**	**6.40**
Cluster 1 (776 objects)	6.65	0.312	19.15	10.9	0.48	3.21	89.1	0.037	0.9919	0.283	2.18
Cluster 2 (1249 objects)	7.01	0.255	44.48	9.8	0.49	3.19	171.9	0.049	0.9958	0.308	8.29
Cluster 3 (281 objects)	6.77	0.332	40.04	9.6	0.52	3.22	169.4	0.053	0.9968	0.396	10.06
Cluster 4 (371 bjects)	7.11	0.328	22.27	10.8	0.45	3.17	117.9	0.036	0.9938	0.293	6.66
Cluster 5 (449 objects)	7.28	0.231	24.10	10.9	0.49	3.18	119.6	0.046	0.9925	0.397	2.28
Cluster 6 (329 objects)	6.55	0.233	43.77	11.1	0.51	3.23	140.7	0.044	0.9922	0.345	3.79
Cluster 7 (34 objects)	8.15	0.351	40.20	9.4	0.62	3.06	194.9	0.049	0.9996	0.619	16.18
Cluster 8 (102 objects)	6.47	0.323	46.38	11.0	0.49	3.20	136.3	0.047	0.9930	0.257	6.25
Cluster 9 (38 objects)	7.17	0.273	52.37	9.3	0.47	3.10	154.2	0.044	0.9976	0.441	13.09
Cluster 10 (11 objects)	6.46	0.261	65.64	9.7	0.44	3.10	171.5	0.060	0.9953	0.479	8.46
Cluster 11 (40 objects)	6.78	0.355	44.20	9.3	0.47	3.13	154.0	0.048	0.9971	0.464	12.50
Cluster 12 (37 objects)	7.24	0.316	38.06	11.0	0.50	3.13	138.6	0.033	0.9945	0.379	10.19
Cluster 13 (50 objects)	5.65	0.330	38.52	12.0	0.48	3.24	108.9	0.031	0.9896	0.246	2.56
Cluster 14 (35 objects)	6.80	0.306	23.88	12.0	0.47	3.18	98.4	0.030	0.9911	0.371	5.06
Cluster 15 (15 objects)	7.2	0.28	44	9.3	0.48	3.06	156.	0.052	0.9966	0.47	9.8
Cluster 16 (14 objects)	5.65	0.32	47.92	10.78	0.49	3.28	170.0	0.041	0.9932	0.33	6.7
Cluster 17 (11 objects)	7.3	0.27	46.0	10.6	0.44	3.10	141	0.059	0.9937	0.516	5.5
Cluster 18 (7 objects)	5.8	0.33	46	10.18	0.43	3.18	153	0.0825	0.99365	0.33	7.7
Cluster 19 (3 objects)	8	0.25	49	9.7	0.46	2.96	219	0.036	0.9996	0.13	17.2
Cluster 20 (6 objects)	7.4	0.23	21.2	11.22	0.41	3.16	100.8	0.038	0.9933	0.442	6.2
Cluster 21 (3 objects)	8.7	0.26	71	9.15	0.57	3.08	242	0.049	0.9997	0.425	14.6
Cluster 22 (2 objects)	7.2	0.58	40	13	0.53	3.17	118	0.032	0.9909	0.27	5.8
Cluster 23 (12 objects)	7.5	0.23	43	11.2	0.41	3.05	118.8	0.029	0.9920	0.28	4.19
Cluster 24 (1 object)	7.1	0.49	146.5	11	0.37	3.24	307.5	0.047	0.9924	0.22	2.0
Cluster 25 (5 objects)	6.5	0.38	41.5	11.5	0.53	3.17	119	0.068	0.9926	0.40	6.5
Cluster 26 (4 objects)	6.6	0.29	64.3	9.4	0.40	3.22	194.7	0.152	0.9956	0.32	6.17
Cluster 27 6 objects)	6.66	0.188	30.4	11.82	0.486	3.164	72.8	0.0242	0.9901	0.29	2.24

Corresponding author: ivan.bangov@gmail.com
DOI 10.2478/asn-2018-0004

ASN, Vol 5, No 1, Pages 24-34, 2018

Cluster 28 (3 objects)	5.8	0.24	23	9.3	0.4	3.25	130	0.038	0.9956	0.21	7.25
Cluster 29 (1 object)	6.9	0.34	13	9.8	0.52	3.07	145	0.032	0.9921	0.36	1.4
Cluster 30 (2 objects)	5.8	0.28	26	10.8	0.55	3.66	159	0.039	0.9965	0.66	9.1
Cluster 31 (1 object)	6.7	0.61	40	9.3	0.57	3.11	240	0.117	0.9938	0.21	1.65

Corresponding author: ivan.bangov@gmail.com
DOI 10.2478/asn-2018-0004

DE GRUYTER OPEN

ASN, Vol 5, No 1, Pages 24-34, 2018

The data exposed in figure 1 shows the distribution of alcohol content (mean value % vol.) typical of each of the created clusters. An increase in the alcohol tends to result in a higher quality of the wine. Formation of alcohols occurs at the biochemical fermentation processes: in processing of grapes under *anaerobic conditions* to produce dry wines and champagne. The alcohol content in all tested samples ranges from 9 to 14%, which defines them as *table wines*.

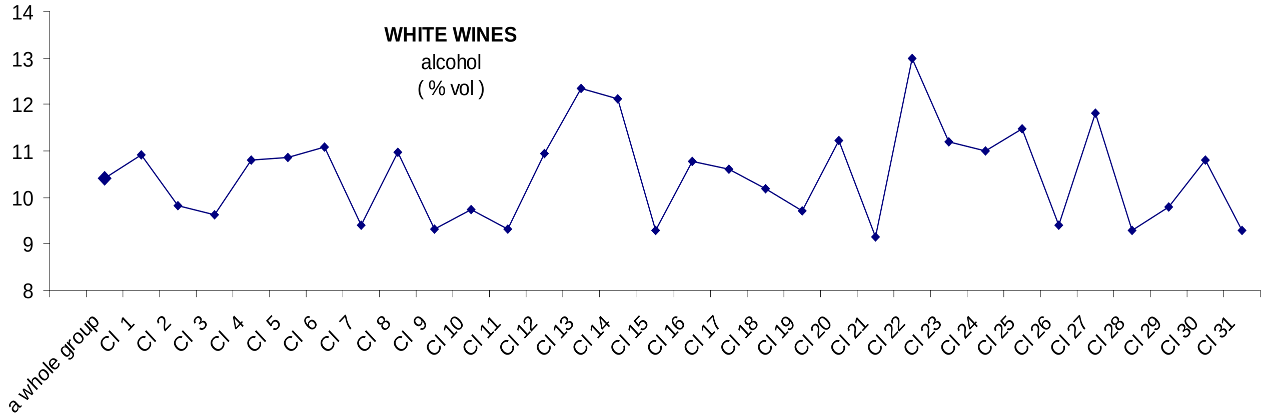

Figure1. Comparing of alcohol content data (*mean* value % vol) in white wines distributed into separate clusters (1-31).

It can be seen on figure 1 that wines with the highest alcohol content are in the 13, 14, 22, 25 and 27 clusters. Considering that the average alcohol content for all white wines is 10.4%, it can be assumed that in the 1, 4-6, 8 , 12,16-18, 23, 24 and 30 clusters are white wines with high alcohol content. Poor alcoholic beverages are represented in 2, 3, 7, 9 – 11, 15, 19, 21, 26, 28, 29 and 31 clusters.

Sulfate ions are introduced in wine production to increase fermenting nutrition, which is very important to improve the wine aroma. For white wines, this indicator is of particular importance. What values are characteristic of the descriptor "concentration of sulfate ions" can be seen in fig.2.

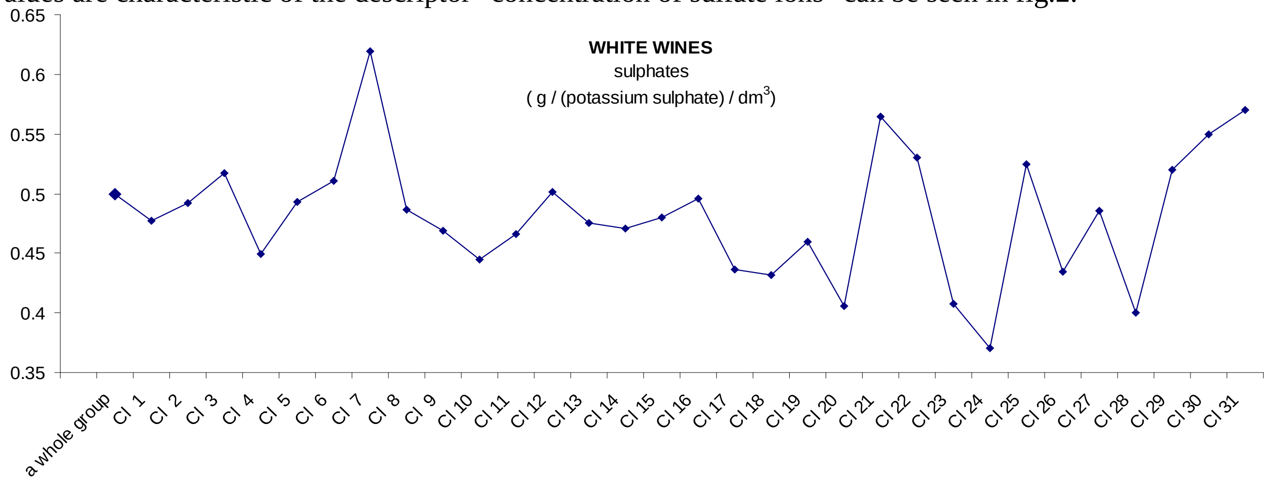

Figure2. Comparing of sulphate ions concentration data (g (potassium sulphate)/dm^3) in white wines distributed into separate clusters (1-31).

The concentration of sulphate ions in the final product is highest for wines in the 7 cluster as well as in 21, 22, 25, 29-31, and the lowest is in the 4, 10, 17, 18, 20, 23-24, 26 and 28 clusters; for the remaining clusters 1-3, 5-6, 8-9, 11-16, the sulphate ions content is about the average.

Sulphurous dioxide is a traditional additive in wine making. Sulphurous acid is is widely used in winemaking thanks to its *antiseptic and antioxidant* property. Sulphurous acid is introduced in wine

Corresponding author: ivan.bangov@gmail.com
DOI 10.2478/asn-2018-0004

DE GRUYTER OPEN
ASN, Vol 5, No 1, Pages 24-34, 2018

making to direct fermentation and oxidation processes to alcohol synthesis, but not to acetic acid. It is used for the purpose of transforming taste and aroma. It acts as inhibitor and suppresses the oxidative activity of enzymes and prevents the formation of the oxidized products, spoiling the taste of wine bouquet. It blocks some of the oxidative enzymes and also paralyzes the function of the yeast cell.

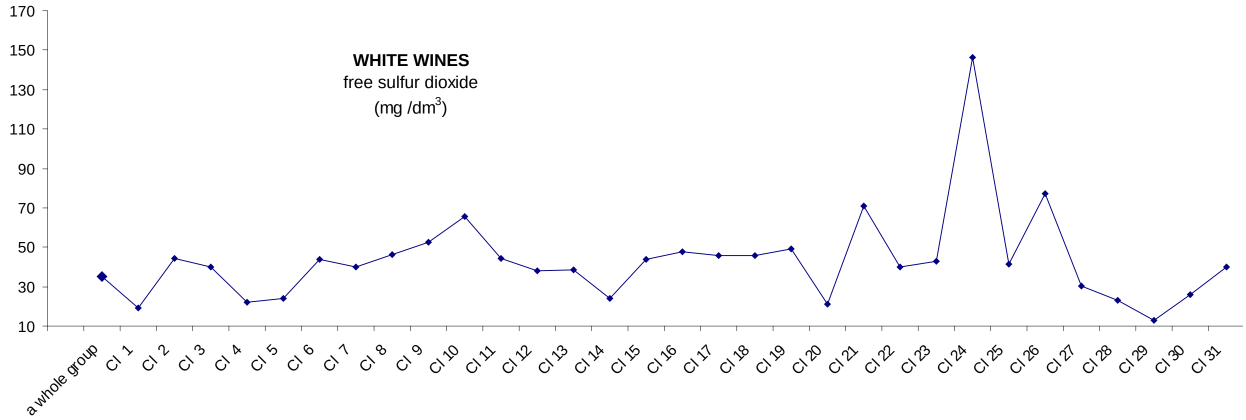

Figure 3. Comparing of free sulphur dioxide content data (mg /dm^3) in white wines distributed into separate clusters (1-31).

Cultural yeast (Sacch.vini) also proved deficiency of oxygen. But they switch *to anaerobic metabolism* of the matter, i.e. to fermentation. The sulphurous acid may exist in wine wort either as an undissociated form, as a gas SO_2, or in the form of $(SO_3)^{-2}$ or $(HSO_3)^-$ anions. Just the non dissociated form of sulphurous acid has antiseptic properties. The concentration of the non dissociated form in the wine solution is very low (Fig.3) and depends on pH and on the temperature of the solution.

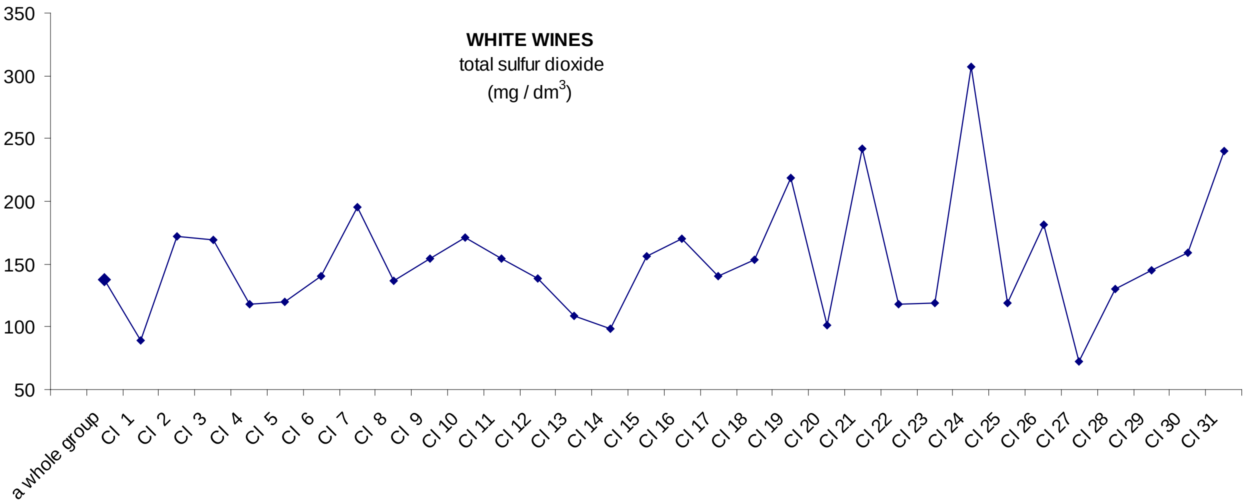

Figure 4. Comparing of total sulphur dioxide content data (mg /dm^3) in white wines distributed into separate clusters (1-31).

Corresponding author: ivan.bangov@gmail.com
DOI 10.2478/asn-2018-0004

DE GRUYTER OPEN

ASN, Vol 5, No 1, Pages 24-34, 2018

The total SO_2 content in the wine is related to the alcohol content. From Figures 3 and 4, the quantification of this laboratory benchmark can be traced. For the clusters with low alcohol content wines (eg, Cluster 2, 3, 7, 9-11, 16, 19, 21, 26, 28, 31), the amount of sulfur dioxide (summarized in both forms) is medium or medium-high. For clusters where the SO_2 amount is high (eg. Cluster 24), - the alcoholic degree is medium-high. And vice versa - for clusters with wines with a high alcoholic degree (Cl 13, 14, 22, 27) - the concentration of SO_2 is extremely low. An analogous correlation is also observed for the other clusters.

A very important indicator of the wine's taste qualities is the acidity of the wine. It is expressed as *fixed acidity* and *volatile acidity* (Figures 5 and 6). Volatile acidity is related to the amount of acetic acid.The increasing in acidity content is due to the development of bacteria in contact with air. However, small amounts of it increase the wine’s flavor.

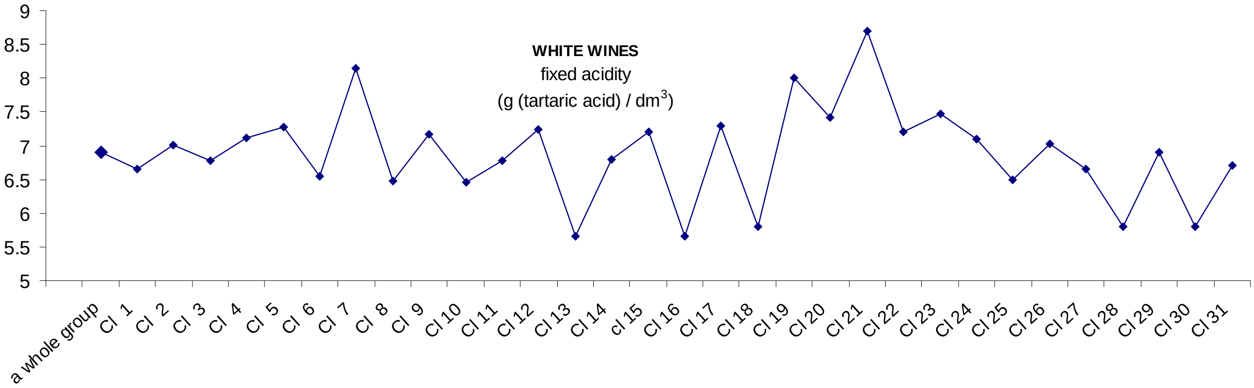

Figure 5. Comparing of fixed acidity content data (g (tartaric acid)/dm^3) in white wines distributed into separate clusters (1-31).

Volatile acids in large doses result in the aroma of acetone or acetic acid, but at lower concentrations they can produce a complex and attractive aroma in prestigious wines. These two indicators change synergistically within each individual cluster, but give a nuanced flavor to the wines.

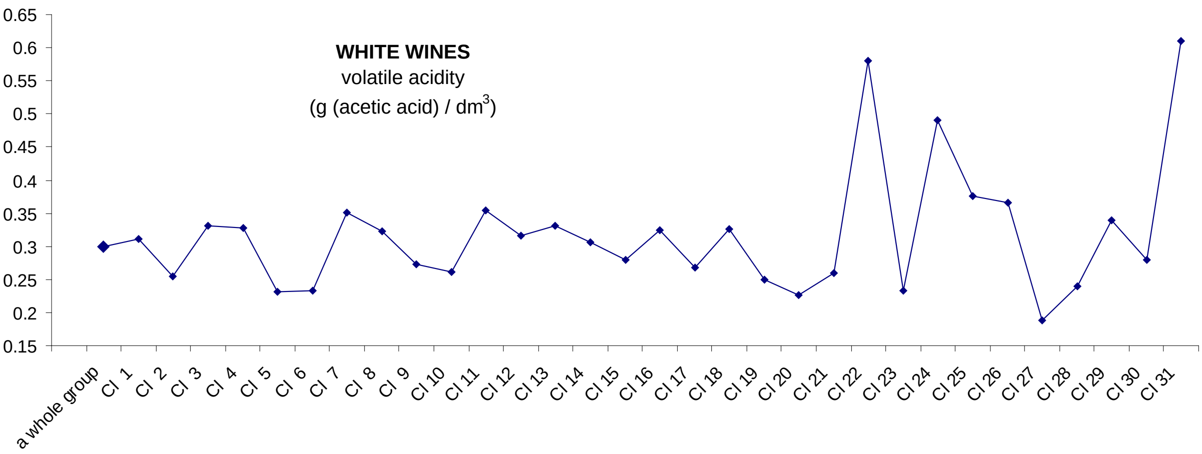

Figure 6. Comparing of volatile acidity content data (g (acetic acid)/dm^3) in white wines distributed into separate clusters (1-31).

Commonly the laboratory measure for a solute acidity is pH. Acidity in the case is related to the natural freshness of the wine.The main acids in grapes are tartaric and malic acids.

Corresponding author: ivan.bangov@gmail.com
DOI 10.2478/asn-2018-0004

DE GRUYTER OPEN

ASN, Vol 5, No 1, Pages 24-34, 2018

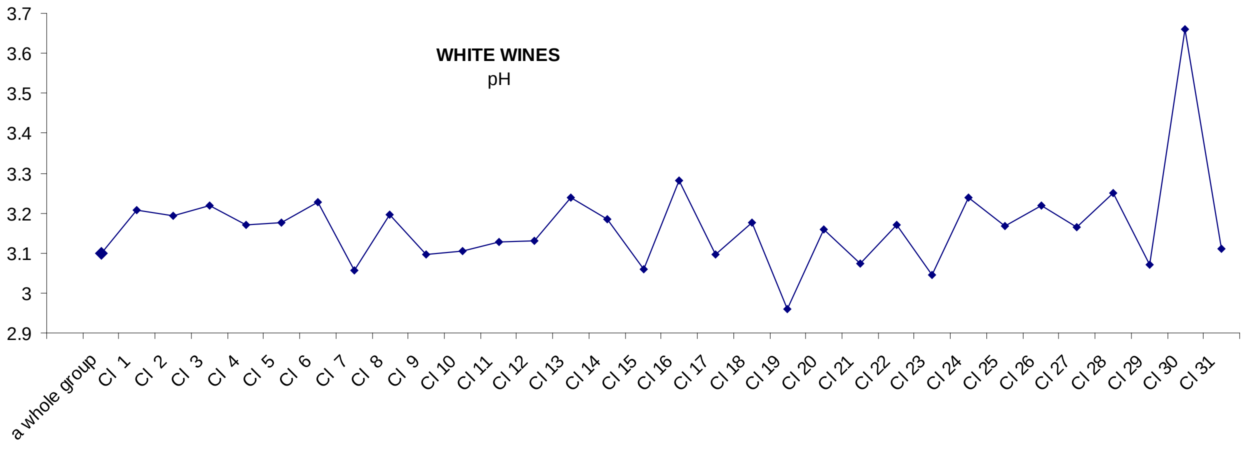

Figure 7. Comparing of pH data in white wines distributed into separate clusters (1-31).

Citric, lactic and succinic acids are also present in small amounts. The average pH =3.1 is a mean value for the whole study group (Figure 7).

Lowering the pH value (such as in the case of clusters 7, 15, 19, 21, 23, 29, 31)) means increasing the acidity of the wine and this is reflected in higher levels of *fixed acidity* and *volatile acidity*.

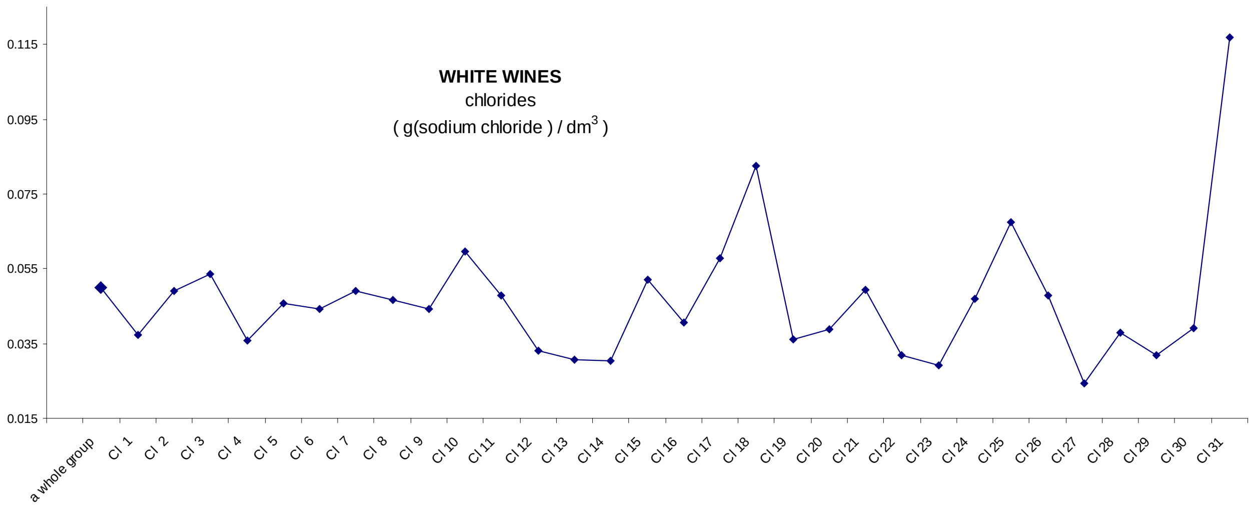

Figure 8. Comparing of chloride ions concentration data (g (sodium chloride)/dm^3) in white wines distributed into separate clusters (1-31).

For the other groups of wines in clusters, the described dependence is observed and can be seen in Fig.5-7. In the case of white wines, the richness of the flavor depends on the presence of *citric acid* and of chloride ions as well (Figures 8, 9). According to oenological theory the citric acid and the residual sugar levels are more important in white wines, where the equilibrium between the freshness and sweet taste is more appreciated.

Corresponding author: ivan.bangov@gmail.com
DOI 10.2478/asn-2018-0004

DE GRUYTER OPEN

ASN, Vol 5, No 1, Pages 24-34, 2018

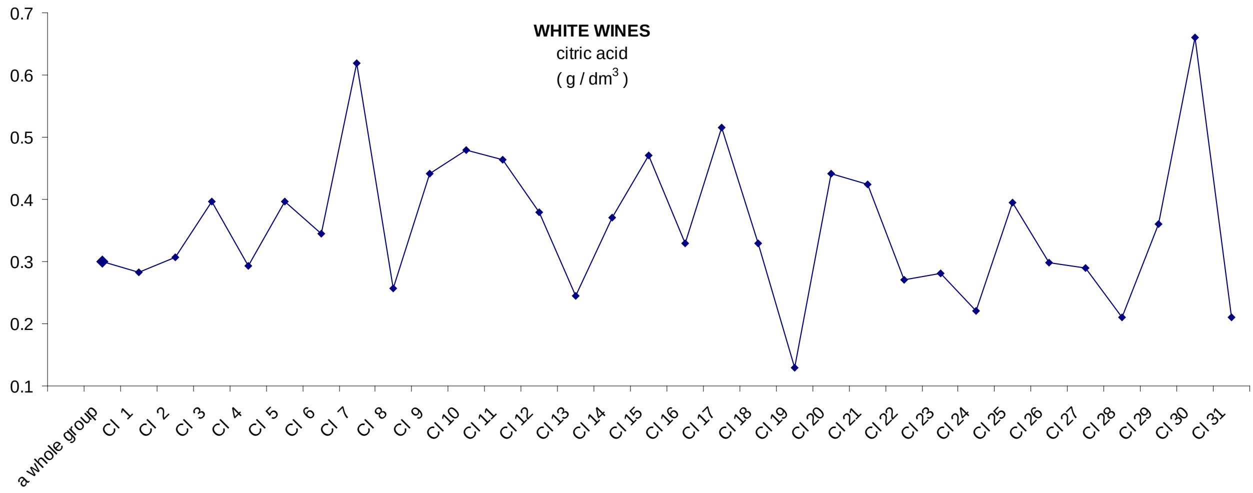

Figure 9. Comparing of citric acid content data (g (citric acid)/dm^3) in white wines distributed into separate clusters (1-31).

The wines collected in the 10, 15, 17 and 25 clusters are characterized by relatively high contents of these components and will be distinguished by strong fruit and citrus flavors. The wines in the 7, 20, 21 and 30 clusters are characterized by citrus aroma; fruit flavors will be present in wines from the 18 and 25 clusters. A specific "bouquet" will distinguish the wine in the 31 cluster. Tender and soft aroma characteristics will be present in wines of 19, 22-23, 28 clusters.

White wines differ in *residual sugar* content. Wines in which detectable residual sugar is less than 0.3% are *dry* wines. With increased sugar content(0.5-3)%, wines are distinguished as *semi-dry*, *semi-sweet* (3-8)% or *sweet*. Sometimes, however, wines with a residual sugar of 5 g / l can be "*dry*" to taste.

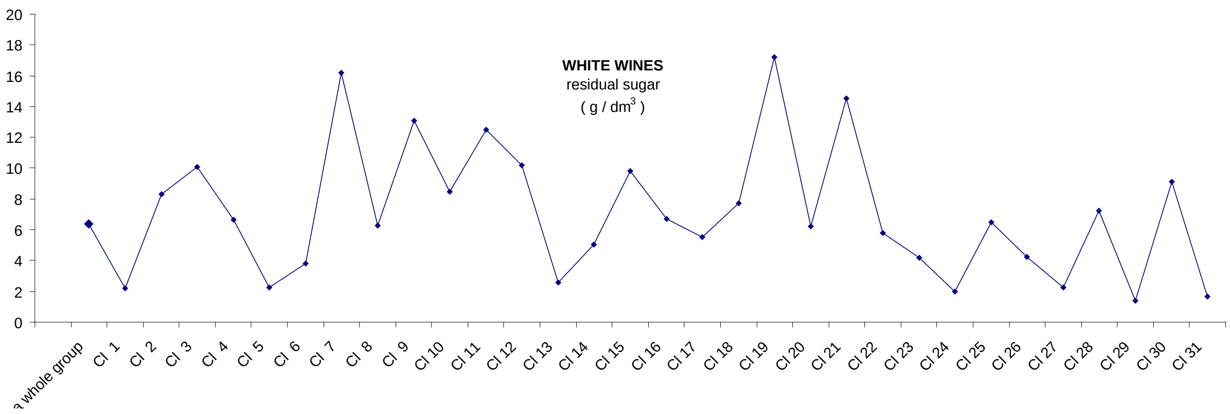

Figure 10. Comparing of residual sugar content data (g /dm^3) in white wines distributed into separate clusters (1-31).

Wines with higher sugar contents refer to *dessert wines* (semi-sweet, sweet and liqueur) where the sugar content is (5-12)%; (14-20)% and (21-35)% respectively. Each of the clusters created combines wines with a certain sugar content. From Figure 10 it can be seen that wines in clusters 1, 5, 24, 27, 29 and 31 has a low residual sugar content and are known as *semi-dry table wines*; in 2, 4, 6, 8, 10, 14, 16-18, 20, 22, 23, 25, 26 and 28 are grouped *semi-sweet* wines and in 3, 7, 9, 11, 12, 15, 19 and 21 clusters are the *dessert wines*.

Corresponding author: ivan.bangov@gmail.com
DOI 10.2478/asn-2018-0004

DE GRUYTER OPEN

ASN, Vol 5, No 1, Pages 24-34, 2018

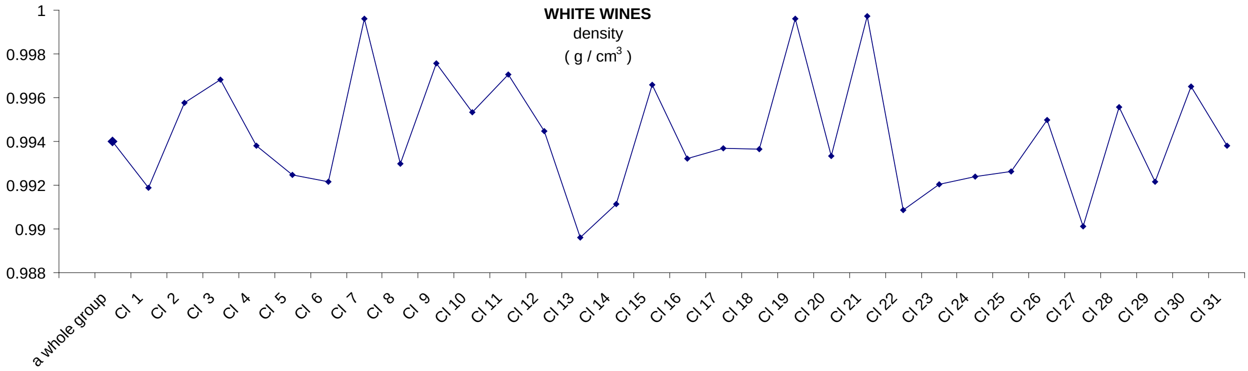

Figure 11. Comparing of density data (g /cm^3) in white wines distributed into separate clusters (1-31).

Wines may differ in whether they are dense, heavy or light. The benchmark for this quality is *density*. Wines gathered in clusters differ in their density. The average density for the whole group of white wines is 0.994 g / cm^3.

From Fig. 11 it can be seen that the *most dense* wines are collected in 7, 19 and 21 clusters. The lightest wines are in 13 , 14, 22 and 27 clusters. A relatively higher density is characteristic of the wines in the second, third, ninth, 11, 15, 26, 28 and 30 clusters.

The comparative analyses of the laboratory test values for the created wine clusters make it possible to distinguish the some features of the cluster's quality.

The main part of the wine samples is collected in the first two clusters. The first cluster groups light, semi-dry white wines with a medium-high alcohol content. The lowered sulphite content is the distinctive quality for the cluster. These are wines with near-average laboratory test values, estimated for the whole studied group. As a distinctive quality for the wines in the first cluster is the lowered sulphite content. They belong to the so-called "*elegant table wines*", in which the quality of wine is expressed in a fine and delicate way; they have a delicate fruity aroma and a harmonious taste balance.

In a second cluster are grouped the wines with a medium-high residual sugar content - semi-sweet table wines with a delicate fruity aroma and a medium-high alcoholic degree. These are dense soft wines with a slim and balanced taste.

Even softer, sweet and dense wines with a bright citrus flavor and depth are harvested in a third cluster. The residual sugar content of the samples in these cluster make it possible to treat them as a *soft dessert wines.*

The fourth cluster contains light wines with an average residual sugar amount, belonging to semi-sweet wines with a very soft fruity aroma and average alcohol concentration.

In the fifth cluster are light wines with a low residual sugar content. These are semi-dry, aromatic wines with over-average alcohol content.

Wines in the sixth cluster are of similar taste, but have a little more residual sugar and are of a higher alcoholic degree.

Desert sweet "*ladies*" wines can be found in 3, 7, 9, 11, 12, 19 and 21 clusters. These are very thick wines with a relatively lower alcoholic degree and relatively high residual sugar content. In all of them, the tartaric acid content is over-average. They are distinguished by the aromatic qualities - the least aromatized are the samples in 19 cluster; the richest of citrus and flower fragrance are in the 7 cluster wines; a smoother citrus and fruity presence has odors of the samples in 9 and 21 clusters.

Other clusters with semisweet wines are the 15 and 30. These are the table wines. These objects are not included in the 2 cluster, because they differ in their aromatic qualities. The wines are very aromatic in 15 cluster, and in 30 cluster are examples of wines with high alcohol content and remarkably bright taste bouquets.

Corresponding author: ivan.bangov@gmail.com
DOI 10.2478/asn-2018-0004

DE GRUYTER OPEN

ASN, Vol 5, No 1, Pages 24-34, 2018

Other clusters in which semi-sweet table wines are harvested are 10, 14, 16, 17, 20, 22, 23, 25 and 26. They are separated in different clusters as they have some distinctive distinctions.
For example, wines in 14 and 22 clusters have a higher alcohol content; in 10, 17 and 20 clusters are very aromatic wines with citrus smell. The 16 cluster are wines with an increased content of SO_2. In the 23 cluster are samples with increased acidity and low aromatic properties.
The 25 cluster contents the semi-sweet table wines with a high alcohol persent and a rich bouquet of color-fruit flavors. The 26 cluster's wines are strongly inflated by SO_2 presence.
Applying the fingerprint method to analyze white wines allows to group them into separate clusters with specific flavors. The comparative analyses of the laboratory test values for the created wine clusters make it possible to distinguish some features of the cluster's quality. It is obvious that laboratory descriptors of some chemicals present in a wine lead to a different clustering than the sensory tests of test men. Nevertheless, they can throw some light on the qualities of studied wines. This expands the taste information of each object (white wine brand) and makes it easier for the consumer to make the desired shopping choices.

Acknowledgments:

This work has been supported by project №RD-08-98/06.02.2017 of "Scientific Research" section of University of Shumen "Konstantin Preslavsky".

References

[1]. Ebeler, S., Flavor Chemistry - Therty Years of Progress, Kluwer Academic Publishers, **1999**, 409-422.
[2]. Legin, A. A.; Rudnitskaya, L.; Luvova, Y.; Vlasov C., Natale and D'Amico, A., Evaluation of Italian wine by the electronic tongue: recognition, quantitative analysis and correlation with human sensory perception, *Analytica Chimica Acta*, **2003**, 484(1), 33-34.
[3]. Cortez, P.; Cerdeira, A.; Almeida, F.; Matos, T.; Reis, J., Modeling wine preferences by data mining from hysicochemical properties, http://repositorium.sdum.uminho.pt/bitstream/1822/10029/1/wine5.pdf
[4]. Bangov, I.; Moskovkina, M.; Stojanov, B.; Descriptor Fingerprints and Their Application to Red Wine Clustering and Discrimination, *Acta Scientifica Naturalis*, **2017**, 1, 29-34.
[5]. Butina, D., Unsupervised data base clustering based on Daylight's fingerprint and Tanimoto similarity: A fast and automated way to cluster small and large data sets, *J. Chem. Inf. Comput. Scie.*, **1999.** 39, 747 -750.

Corresponding author: ivan.bangov@gmail.com
DOI 10.2478/asn-2018-0004