Comparative analysis of red wines on the base of optical and chemical characteristics by statistical methods

St Minkov1, M Milev2, Kr Nikolova1*, Ir Ivanova3, N Hristova-Avakumuva4, V Hadjimitova4, M Kakalova1, V Vladev2

1 Medical University, 55 Marin Drinov Blvd., 9000 Varna, Bulgaria
2 University of Food Technologies, 26 Maritsa Blvd., 4000 Plovdiv, Bulgaria
3 Agricultural University, 12 Mendeleev Blvd., 4000 Plovdiv, Bulgaria
4 Medical University, Str. Zdrave 5., Sofia, Bulgaria

*E-mail: kr.nikolova@abv.bg

Abstract. Thirty – five samples of red wines from different geographic regions have been characterised according to 14 basic optical and chemical characteristics - colour parameters, content of anthocyanins and antioxidant activity. The goals of this study is to compare the tested wines using different statistical methods. Conducting hierarchical cluster analysis the samples have been divided into 4 clusters according to their difference and similarity. The result is presented graphically by a dendrogram. It could be concluded that the optical parameters and related chemical indicators are more decisive for the division of wines based on the region of production than on the basis of the predominant grape variety. The total number of studied parameters was reduced to 4 factors by factor analysis using Principal Component Analysis (PCA). These new factors explained 86.75% of the entire variation. The first factor contained chemical parameters and color coefficients in SIELab colorimetric system. The second one was connected with the quality characteristics of wine, which were important for consumers such as brilliance of red, color intensity. The third and the fourth factors were related with the Hues angles, color parameter b and saturation for the red wines.

1. Introduction

The red wine is very rich in polyphenols (catechin and epicatechin), flavanols (quercetin, rutin, myristin), phenolic acids (gallium, caffeic and p-cumaric acid). Phenolic compounds play a significant role in the sensory properties of wines such as flavor, color and taste and have a positive effect on health [1]. These components are important because of their proven bactericidal effect [2]. The content of polyphenols in grapes is different for various types of grapes in different years. The quality of grape is different from year to year, because the rainfall, sunshine-exposure and average temperature of each harvest year are not the same. These factors affect the polyphenol content and color parameters of wine samples as reported in the literature [3-7].

The purpose of the present study is to group the examined red wines examined into clusters on the basis of the degree of similarity between them, to assess the influence of factors such as grape variety, geographic region or year of production on the modelling of the samples, as well as by the method of principle component analysis (PCA) to reduce the huge number of examined variables to several new artificial factors having a determining significance for the respective groups.

2. Materials and methods
2.1. Samples
The database includes 35 types of red wines from 4 countries - Bulgaria, Italy, Chile and Sought Africa. The samples are purchased from the Bulgarian and foreign supermarkets or obtained directly from winery. Details for the geographical region and grape variety for the investigated samples are presented in table 1.

Table 1. The wine collection.

| Country       | Code | Region          | Variety grapes | Year | Country Code | Region     | Variety grapes       | Year |
|---------------|------|-----------------|----------------|------|--------------|------------|-----------------------|------|
| Bulgaria      | 1    | Harmanly        | Merlot         | 2012 | Bulgaria     | 19         | Montana               | 2013 |
| Bulgaria      | 2    | Harmanly        | Merlot         | 2013 | Bulgaria     | 20         | Tracian Plain         | 2013 |
| Bulgaria      | 3    | Harmanly        | Merlot         | 2014 | Bulgaria     | 21         | Tracian Plain         | 2013 |
| Bulgaria      | 4    | Harmanly        | Merlot         | 2015 | Bulgaria     | 22         | Harmanly              | 2016 |
| Bulgaria      | 5    | Harmanly        | Merlot         | 2016 | Bulgaria     | 23         | Stara Zagora          | 2013 |
| Serbia        | 6    | Pleven          | Merlot         | 2013 | Bulgaria     | 24         | Montana               | 2013 |
| Bulgaria      | 7    | Tracian Plain   | Merlot         | 2013 | Bulgaria     | 25         | Vidin                 | 2013 |
| Bulgaria      | 8    | Yambol          | Merlot         | 2013 | Bulgaria     | 26         | Sakar Mountain        | 2013 |
| Holland       | 9    | Asenovgrad      | Merlot         | 2013 | Bulgaria     | 27         | Sakar Mountain        | 2013 |
| Bulgaria      | 10   | Stara Zagora    | Merlot         | 2013 | Bulgaria     | 28         | Stara Zagora          | 2013 |
| Bulgaria      | 11   | Veliki Preslav  | Merlot         | 2013 | Bulgaria     | 29         | Harmanly              | 2016 |
| Bulgaria      | 12   | Danube Plain    | Merlot         | 2013 | Bulgaria     | 30         | Svilengrad            | 2014 |
| South Africa  | 13   |                  | Merlot         | 2013 | Bulgaria     | 31         | Tracian Plain         | 2014 |
| Chile         | 14   | Harmanly        | Merlot         | 2014 | Bulgaria     | 32         | Harmanly              | 2016 |
| Bulgaria      | 15   | Suhindol        | Merlot         | 2014 | Bulgaria     | 33         | Malbec+                | 2016 |
| Bulgaria      | 16   | Stara Zagora    | Merlot         | 2014 | Bulgaria     | 34         | Trakian Plain         | 2015 |
| Bulgaria      | 17   | Tracian Plain   | Merlot         | 2013 | Italy         | 35         | Barbera d'asti        | 2013 |

2.2. Used methods
Color measuring: CIELab coordinates [8] have been measured by using spectrophotometer (Lovibond Trintometer PFX 195, UK). The content of carotene is determined by special software. The samples were used without any dilution, poured into a 10 mm cuvette.

Absorption spectra: The absorption spectra for investigated samples are obtained by spectrometer Brolight, using optical scheme Czerny-Turner. The parameters such as color intensity, tint, saturation, logarithmic color density, angle of Hue are used for estimation of quality and ages of wines. The
absorbance values (D) at 420 nm, 520 nm and 620 nm [9] are used to obtain numerical values of color intensity etc.

Antioxidant activity: Determination of the antioxidant capacity was performed by two complementary methods, evaluating different aspects of the reduction activity of the samples:

- ABTS method - The method is carried out according to Re et al. [10]
- DPPH method - The method is performed according to Goupy et al. [11].

2.3. Statistical analysis

To compare and group the types of wines on the base of all investigated physico-chemical parameters a hierarchical cluster analysis with a measure of similarity Euclidean distance [12, 13] has been used, applying the method of average linkage between groups. The data have been preliminary standardized with z-standardization in order to avoid the influence of different dimensions. The results have been represented graphically by a dendrogram.

A factor analysis was conducted. Thus the input set of correlating data was transformed into a new set with a smaller number of uncorrelated artificial variables, so-called „factors” or „principal components”, which explain the greatest part of the data variation. The reduction in the number of original variables was achieved by grouping the correlated variables in a common factor and the separation of uncorrelated ones in different factors [14-15]. A stronger distinction of the variables to one or another factor is performed by further factor rotation using Varimax method. All data analyses have been performed using the SPSS software program [16].

3. Results and discussion

The results from cluster analysis are presented on the figure 1. The samples from the different from geographical regions which have been investigated are grouped in fourth clusters. The distance between samples is from 1.030 to 100.943. From the dendrogram it can be seen that fourth clusters. Table 2 shows both - the percentage distribution of the samples into the 4 clusters and the samples participation in each of the 4 clusters. Samples 1, 19, 22, 25, 27, 29, 30, 32, and 35 do not belong to these clusters, because they have some different values for investigated parameters.

Most of the samples in the first cluster belong to the Eastern Thracian Lowland and only 3 of them in the northern Black Sea region. A sample from South Africa, it is the closest to red wine from the Yambol region. The second cluster is joined samples of Western Thracian Lowland and a wine from northwestern Bulgaria with predominant sort Cabernet Sauvignon. Heterogeneity is observed from varieties, regions and years in the third cluster. The fourth includes samples from the Harmanly region 2015 and 2016 year. This can be explained by the high antioxidant activity in that region over the period considered. There are data that SV$_{50}=1\, \mu l$ in the same region [17]. The last fact is associated with very high antioxidant activity. Although at a much distance to them, samples 22 and 32 are joined in the same cluster. They are from the same region and year of production but from another grape variety. Samples 1 and 25 are at the longest distance. This is understandable since they are from two climatically opposite regions – the Vidin and the Harmanly. They, as well as samples 30 and 35, could not belong to any of the four clusters. This is natural because sample 35 is Italian wine and the weather conditions there are not similar to ours weather conditions. Sample 30 is from the region of Svilengrad and according to distances from the dendrogram is closest to the case 1, which belongs to Harmanly region.

Therefore, not grape varieties, but the geographical area and the year of production are decisive factors for the classification of red wines according to the physico-chemical characteristics.

The investigated characteristics (total number 14) have been reduced by grouping in different factors. These, which have correlations one with other, have been combined in one factor. The others - uncorrelated have been divided in different factors.
Table 2. Distribution of types wines in clusters.

| Variants in clusters | Samples, number | Distribution of the samples of wines, % |
|----------------------|-----------------|----------------------------------------|
| 6, 7, 8, 11, 12, 13, 17, 20, 23, 26, 27, 28, 31, 34 | 14 | 40 |
| 9, 18, 21, 24 | 4 | 11.5 |
| 2, 3, 10, 14, 16, 29 | 6 | 17 |
| 4, 5, 15, 33 | 4 | 11.5 |

Figure 1. Dendrogram for investigated wines

Figure 2. Results from Factor analysis

The determinant of the correlation matrix is different from zero, which allows conducting the factor analysis. The data have to posses normal distribution for obtaining the adequate model. It can be determined with model of Kaizer-Maier-Olkin (KMO test), which value has to be greater than 0.5, in our case this is fulfilled. The Bartlett’s test of sphericity is meaningful (p<0.001). This gives us reason to perform a factor analysis and that the model will be adequate.

It is determined from the PCA that four factors have values of an own vectors greater than one, which determined the choice of 4 main components, which explained 86.75% of total variation. It is presented on the figure 3. The component plot in a rotated space is shown on the figure 2.

The first factor (principle component one - PC 1) explains 42.47% of total variation and strongly correlated with color parameters “a” in color system CIELab, lightness L, Chroma C, anthocyanins, antioxidant activity, determined by ABTS and DPPH methods and nuance. We could summarize that this factor is generalizing, as these parameters have the greatest weight in separating wines into clusters according to consumer sensory assessment. It combines parameters that have correlations between them with a level of significance of 0.1%. The second component summarizes indicators tied with an evaluation of red wine by quality and age–color intensity, brilliance and hue - T, which explains 20.92% of total variation. The third and fourth components are linked to one indicator and explain 13.92% and 9.44% of total variation respectively. So the greatest weight has the parameters that come in the first and second factors.

The variation of the factor weight and distribution between four main components are presented in table 3.
Table 3. Factor matrix, obtained by method PCA

| №  | Investigated Parameters          | Main components (factors) | 1   | 2   | 3   | 4   |
|----|----------------------------------|---------------------------|-----|-----|-----|-----|
| X1 | AOA by ABTS methods             | -0.582                    | -0.210 | 0.025 | -0.246 |
| X2 | AOA by DPPH methods             | -0.582                    | -0.332 | -0.022 | -0.031 |
| X3 | Anthocyanins 1                  | 0.858                     | 0.277 | -0.067 | 0.208 |
| X4 | Anthocyanins 2                  | 0.859                     | 0.276 | -0.069 | 0.208 |
| X5 | Lightness L                     | -0.917                    | -0.170 | -0.065 | 0.290 |
| X6 | “a” color parameter             | 0.959                     | 0.329 | 0.046  | 0.108 |
| X7 | “b” color parameter             | -0.266                    | 0.536 | 0.699  | 0.334 |
| X8 | “C” chroma                      | 0.899                     | 0.181 | 0.166  | -0.106 |
| X9 | H angle                         | -0.768                    | 0.252 | 0.515  | 0.238 |
| X10| Saturation                      | 0.242                     | 0.163 | 0.557  | -0.744 |
| X11| Nuance                          | 0.856                     | 0.181 | 0.166  | -0.106 |
| X12| Hue                             | -0.404                    | 0.781 | -0.436 | -0.172 |
| X13| Brilliance                      | -0.733                    | 0.650 | -0.178 | -0.064 |
| X14| Color Intensity                 | 0.761                     | -0.587 | 0.231 | 0.139 |

Total Variation, %
Cumulative present from total variation, %

Table 4. Extraction method: Principal Component analysis

| Component | Initial Elgenvalues | Rotation sum of squared loadings |
|-----------|--------------------|----------------------------------|
|           | % of Variance      | Cumulative % | % of Variance | Cumulative % |
| X1        | 53.226             | 53.226       | 5.946        | 42.468       | 42.468       |
| X2        | 16.095             | 69.322       | 2.929        | 20.923       | 63.392       |
| X3        | 10.075             | 79.397       | 1.948        | 13.917       | 77.309       |
| X4        | 7.355              | 86.751       | 1.322        | 9.443        | 86.751       |
| X5        | 5.883              | 92.634       |              |              |              |
| X6        | 3.407              | 96.042       |              |              |              |
| X7        | 2.343              | 98.384       |              |              |              |
| X8        | 1.223              | 99.607       |              |              |              |
| X9        | 0.187              | 99.794       |              |              |              |
| X10       | 0.151              | 99.945       |              |              |              |
| X11       | 0.035              | 99.980       |              |              |              |
| X12       | 0.018              | 99.998       |              |              |              |
| X13       | 0.002              | 100.00       |              |              |              |
| X14       | 0.000              | 100.00       |              |              |              |

4. Conclusion
The strongest influence indicates the anthocyanin content, the antioxidant activity and the luminosity that groups the wines into separate clusters. The fourth cluster brings together samples of red wines from southeastern Bulgaria (Harmanli and Svilengrad) with the highest antioxidant activity, anthocyanin content and color indicators. The first one combines the most samples of the most widely consumed red wines from the southern region of Bulgaria with relatively good antioxidant properties.

References
[1] Woraratphoka K, Intarapichet K. 2007 Food Chemistry 104 1485
[2] Singleton V, Esau P 1969 *Phenolic Substance in Grapes and Wine and Their Significance* (Academic Press, New York) p. 125

[3] Lee J E, Hwang G S, Van Den Berg, Lee C H & Hong Y S 2009 *Anal Chim Acta* **648** 71

[4] Lorrain B, Chira K, Teissedre P L 2011 *Food Chemistry* **126** 1991-1999

[5] Pereira G E, Gaudillere J P, Leeuwen C V, Hilbert G, Maucourt M, Deborde C, Rolin D 2006. *Anal Chim Acta* **563** 346

[6] Ferrandino A, Lovisolo C 2014 *Environmental and Experimental Botany* **103** 138

[7] Siegl K, Eder R., Rak G, Abranko L, Koellensperger G, Hann S 2010 *Food Chemistry* **122** 366372

[8] 2004 *C. I. E. Colorimetry* (edited by Central Bureau of the Commision Internationale de L’Eclairage ( Publication C. I. E., Vienna) p.35

[9] Yıldırım H, Elmacı Y, Ova G, Altuğ T and Yücel U 2007 *International Journal of Food Properties* **93** 102

[10] Re R, Pellegrini N, Protegente A, Pannala A, Yang M and Rice-Evans C 1999 *Radical Biology and Medicine* **26** 1231

[11] Goupy P, Dufour C, Loonis M and Dangles O 2003 *Journal of Agricultural & Food Chemistry* **51** 615

[12] Duran B and Odelle P 1977 *Cluster analysis* (Moscow, Russia: Statistics) p 115

[13] Ward JH 1963 *Journal of the American Statistical Association* **58** 236

[14] Jolliffe I 2002 *Principal Component Analysis* (USA, New York, Springer-Verlag New York) p 76

[15] Vundev D 2003 *Applied Statistic* (Sofia, Bulgaria: SU Kliment Ohridski) p 76

[16] Statistics Toolbox™ User’s Guide, R2014a, Copyright 1993–2014 by The MathWorks, Inc

[17] Hristova –Avakumova N G, Atanasova L A, Atanassova S L , Vangelova T V, Hadjimitova V A 2018 *Bulgarian Chemical Communications* **50** 321