Scientific paper

Characterization of Slovenian Wines Using Multidimensional Data Analysis from Simple Enological Descriptors

Adriána Bednárová,[1,*] Roman Kranvogl,[2] Darinka Brodnjak Vončina,[2] Tjaša Jug[3] and Ernest Beinrohr[1]

[1] *Department of Chemistry, Faculty of Natural Sciences, University of Ss. Cyril and Methodius in Trnava, Nám. J. Herdu 2, SK-917 01, Trnava, Slovakia,*

[2] *Faculty of Chemistry and Chemical Engineering, University of Maribor, Smetanova 17, 2000 Maribor, Slovenia*

[3] *Chamber of Agriculture and Forestry of Slovenia, Institute for Agriculture and Forestry, Pri hrastu 18, 5000 Nova Gorica, Slovenia*

* *Corresponding author: E-mail: adriana.bednarova@ucm.sk;*
Tel. +421 33 5921 403

This paper is dedicated to the memory of the late Professor Jan Mocak

Received: 19-09-2012

Abstract

Determination of the product's origin is one of the primary requirements when certifying a wine's authenticity. Significant research has described the possibilities of predicting a wine's origin using efficient methods of wine components' analyses connected with multivariate data analysis. The main goal of this study was to examine the discrimination ability of simple enological descriptors for the classification of Slovenian red and white wine samples according to their varieties and geographical origins. Another task was to investigate the inter-relations available among descriptors such as relative density, content of total acids, non-volatile acids and volatile acids, ash, reducing sugars, sugar-free extract, SO_2, ethanol, pH, and an important additional variable – the sensorial quality of the wine, using correlation analysis, principal component analysis (PCA), and cluster analysis (CLU). 739 red and white wine samples were scanned on a Wine Scan FT 120, from wave numbers 926 cm^{-1} to 5012 cm^{-1}. The applied methods of linear discriminant analysis (LDA), general discriminant analysis (GDA), and artificial neural networks (ANN), demonstrated their power for authentication purposes.

Keywords: Wine authentication, enological descriptors, classification techniques, ANN

1. Introduction

The authenticities of wines have been extensively investigated since wine, due to its chemical composition and availability world-wide, can be an easily adulterated product. The wine industry needs analytical tools for verifying the authenticities of high-value products, in order to protect their brands. Ideally, these tools should facilitate rapid and inexpensive analysis at any point along the distribution chain. Since the areas of production are expanding, and consequently the visible markings regarding the originalities and quality characteristics of the products are subsequently reflected in the final price, the determination of geographical origin is therefore one of the primary requirements when certifying a wine's authenticity. The question of geographical identification regarding wines has thus become crucial, especially when it relates to smaller production areas.

Currently, concentrations of minerals and trace elements,[1–7] polyphenolic compounds,[8–9] volatile compounds,[10–11] a combination of several types of analytes,[12–16] and isotope ratios,[17] are mostly employed as the discriminating variables in multidimensional data analysis for classification and authentication purposes. However, most of them are sophisticated and time-consuming and require tedious and complex processing for wines. Those institutions controlling wine quality, and especially the wineries, do not commonly possess sophisticated and modern analytical equipment. Therefore, it would be beneficial for all of them to evaluate the discriminating power of traditional and simple enological descriptors such as alcoholic grade, density, pH, content of extract or SO_2, that are already analysed at the mentioned institutions or wineries. When taking into account the economic reasons (costs in terms of money and time), the usage of these descriptors is usually more convenient than other applied analytical data.

The goal of this work was to investigate the discrimination ability of eleven simple enological descriptors obtained for authentication purposes, and to evaluate any differences between Slovenian varietal wines originating from different production areas and vintages. For this purpose, both the unsupervised and supervised techniques of multidimensional data analysis were applied in order to explain any inter-relations amongst the enological descriptors, and then to classify the wine samples into categories according to the target factors – the variety and geographical origin of wine.

2. Experimental

2. 1. Wine Samples

Analyses were conducted for 739 wine samples originating from four production regions of the Primorska area (Figure 1) in south-west Slovenia, namely Koper, Kras, Vipavska dolina, and Goriška brda. Four groups of red wines – the varieties Refosk (32 samples), Cabernet Sauvignon (110), Merlot (86), and mixtures of several red varietal wines (59 samples) were investigated, plus the variety Teran (37 samples) as a special type of Refosk wine. Seven sorts of white wine were additionally explored – Chardonnay (110), Sauvignon (59), Malvasia (43), Pinot Gris (55), Yellow Muscatel (24), Rebula (41), and samples and mixtures of several varietal white wines (83 samples). All the wine samples originated from four vintages (2003–2006) and were collected by KGZS-GO, Chamber of Agriculture and Forestry of Slovenia, Institute for Agriculture and Forestry.

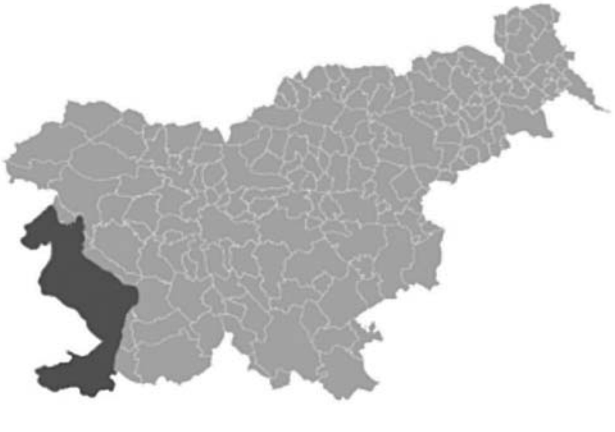

Figure 1: Primorska area in south-west Slovenia.[18]

2. 2. Analytical Methods

The following enological descriptors were determined: relative density (in g/mL at 20 °C), content of total extract (g/L), total acidity (g/L), non-volatile acidity (g/L), volatile acidity (g/L), ash (g/L), free SO_2 (mg/L), reducing sugars (g/L), sugar-free extract (g/L), ethanol (in %), and pH.

The wine analyses were performed at the KGZS – GO: analysis of ash (g/L) was performed gravimetrically, free SO_2 (mg/L) by iodometrical titration, relative density (in g/mL at 20 °C), content of total extract (g/L), total acidity (g/L), volatile acidity (g/L), reducing sugars (g/L), ethanol (in %), and pH by the Wine Scan FT 120 instrument and non-volatile acidity (g/L) and sugar-free extract (g/L) were calculated:

sugar-free extract = total dry extract
– (reducing sugars – 1)
non-volatile acidity = total acidity (g/L)
– volatile acidity (g/L)

An instrument utilising FTIR was employed for simultaneous determination of the mentioned wine descriptors. The samples of wine were filtered through filter paper to expel CO_2, and catch any sediment. The instrument was zeroed before any set of analyses with a zeroing solution. The samples were scanned from 926 to 5012 cm^{-1}.

For the calibration by Wine Scan FT 120, the official analytical methods were used for the real samples. Iodometric titration (Rebelein method) was applied for determining reducing sugars' contents. Potentiometric methods using glass electrode was employed for determining total acidity and pH. A hydrostatic balance was used for measuring density and alcohol, and total extract has been read from the official tables (OIV).

In addition, the sensorial variable *Mark* describing the sensory qualities of the wine samples was obtained by a group of experts evaluating the wine properties (colour, aroma, taste, harmony) using a twenty-point scale in total. All the analytical methods were accomplished according to the Official Gazette of the Republic of Slovenia No. 43/01, and sensorial analysis was performed according to the Official Gazette RS No. 32/00 determining methods for the analyses of wines.

2. 3. Statistical Analysis

Chemometrical data analysis was carried-out in order to discover any statistically significant differences between the samples grouped according to three categorical variables – *Variety*, *Area* (geographic origin), and *Vintage* as the target factors. Another task was to investigate any significant correlations between individual enological descriptors and the sensorial quality's variable *Mark*. The ratios of the volatile acidity to the total acidity, the non-volatile acidity to the total acidity, as well as the ratio of the sugar-free extract to the total extract were calculated as supplemental descriptors. All the employed descriptors are summarised in Table 1. Abbreviated designations were necessary for their shorter notations in graphical form. Statistical data treatment was performed using the program packages STATGRAPHICS Centurion v. 15, SPSS v. 15 and STATISTICA v. 7; Microsoft EXCEL was used for the data preparation and result outputs. The exploratory data analysis enabled the discovery of outliers within the data, and the departures from the normal distribution were evaluated by the Kolmogorov-Smirnov test. The logarithmic and reciprocal transformations of the descriptor data were used, if the descriptor data differed significantly from the normal distribution. The effect of the target categorical variables – *Variety*, *Area*, and *Vintage* upon the selected enological descriptors was investigated in four ways: (1) analysis of variance (ANOVA), (2) the Kruskal-Wallis non-parametric test, (3) the least significant difference post-hoc test (LSD) for pair-wise comparisons of the mean values, and (4) the Mann-Whitney test for non-parametric comparisons. Correlation analysis (CA), cluster analysis (CLU), and the principal component analysis (PCA) were applied in order to discover the inter-relations amongst the used descriptors. PCA and CLU were applied for displaying a natural grouping of the objects, i.e. the wine samples, in the multidimensional variable space; and for a better understanding of the investigated problems the sample categories (regarding the selected target variables) were sometimes displayed. Several supervised MDA techniques were employed for sample classification by the target factor, particularly linear discriminant analysis (LDA), general discriminant analysis (GDA), and artificial neural networks (ANN), specifically three-layer perceptrons.

3. Results and Discussion

3. 1. Exploratory Data Analysis

Exploratory data analysis was performed using the SPSS v. 15 package. Ten clear outliers (three red wine samples and seven white wine samples) were found and excluded from the data table. Departures from the normal distribution were demonstrated by the *Q–Q* plots and tested by the Kolmogorov-Smirnov test; 6 original descriptors were found as not normal (*RedSug, Extract, NVolAc, TotAc, VolAc, Mark*), and reciprocally or logarithmically transformed data forms were created in order to achieve correct results in the deviating cases.

3. 2. Correlation Analysis

A direct examination of any inter-relation between two continual variables is mostly realized by correlation analysis determining the extent to which the values of the two variables are mutually dependent. The more common Pearson correlation analysis is a parametric method. The Pearson (pair) correlation coefficient values of +1 or -1 indicate a perfect linear relationship between the two considered variables. If there are violations of the data's normality and constant variability assumptions, the Spearman correlation coefficient is an optimal equivalent, because it is the rank based robust statistical characteristic, and also works well for nonlinear correlations.[20]

Due to departure from the normal distribution exhibited by several descriptors, the non-parametric Spearman correlation analysis was performed and compared to the standard Pearson correlation analysis, where disagreement in the correlation results was mostly observed for the not normally distributed data, as expected. Statistically significant correlations were found for numerous pairs of descriptors (Table 2) when used the red and white wine samples together (n = 729). The highest correlation coefficients were observed between *TotAc* and *NVolAc* (0.98) and for all descriptor couples of the group *SFE*, *Extract*, and *Dens*. In addition, the descriptors of this group were highly significant and positively correlated to *Ash*, *TotAc,* and *NVolAc*. The high significant correlations between *TotAc* and *NVolAc* as well as between *SFE* and *Extract* were expected, as *NVolAc* and *SFE* are calculated from

Table 1: Abbreviated notations for the employed descriptors.

Descriptor	Denotation	Descriptor	Denotation	Descriptor	Denotation
sensorial quality	*Mark*	reducing sugars	*RedSug*	non-volatile acidity	*NVolAc*
relative density	*Dens*	extract	*Extract*	volatile acidity	*VolAc*
ethanol content	*Ethanol*	sugar-free extract	*SFE*	non-volatile acidity / total acidity	*NVA_TA*
pH	*pH*	free SO_2 conc.	*SO2Free*	volatile acidity / total acidity	*VA_TA*
ash content	*Ash*	total acidity	*TotAc*	sugar free extract/ total extract	*SFE_E*

Note: when using the logarithmic or reciprocal forms of a descriptor the abbreviation *log* or *rec* was used together with the abbreviation of the descriptor.

TotAc and *Extract*, respectively. Important correlations of the sensorial variable *Mark* with *Ethanol* (0.37), *NVolAc* (–0.22), *TotAc* (–0.22), *Dens* (–0.20) and *RedSug* (0.18) were found.

Due to large number of samples (n = 729), the critical value of correlation coefficient 0.07 corresponded to the statistical significance p = 0.05. Hence, only coefficients with $p < 10^{-6}$ were considered important as shown in Table 2. For the sake of differentiation, the important correlation coefficients with $p < 10^{-6}$ are bold and the strongest correlations are denoted by red colour ($p < 10^{-15}$) in Table 2.

The results of the Spearman correlation analysis were also obtained separately for the white wine and red wine samples. Compared to Table 2 that reflected all the wine samples, numerous differences were found for white wines (n = 408) when evaluated in more detail, for example, *Dens* in the white wines was highly significant correlated to the descriptors *RedSug* (0.63), *SO2Free* (0.33), and *VolAc* (–0.31), but the correlation to *SFE* (0.59), and mainly to *Ash* (0.18), were less pronounced. *Ethanol* in the case of white wines was significantly correlated to *VolAc* (0.33) and *RedSug* (–0.22) but the correlation to *SFE* (–0.02) was weak.

Differences in correlations in case of the red wines (n = 321) compared to the all wine samples of Table 2 were also observed for *Dens* – a considerable increase in the correlation coefficient with respect to *NVolAc* (0.60), and also to *RedSug* (0.30), *VolAc*, (–0.27), and *pH* (0.28); a decrease was observed in correlation with respect to *Ash* (0.15) and *Mark* (–0.09) – this meant that density plays an unimportant role when assessing the sensorial quality of red wine.

3. 3. Principal Component Analysis

Principal component analysis (PCA) is an unsupervised multidimensional method used for reducing the number of variables along with preserving the information contained in the data table. The most important prin-

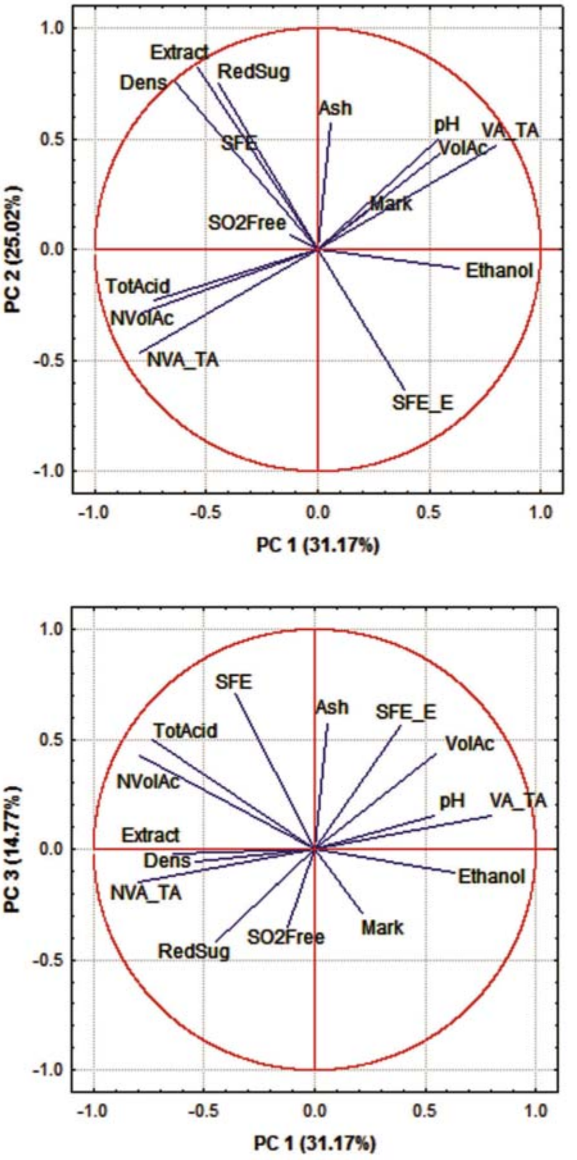

Figure 2: Loading plots as graphical outputs of PCA – the position of the original variables in space of *PC 1* to *PC 2* (top) and *PC 1* to *PC 3* (bottom). Software STATISTICA v. 7.

Table 2: Reduced correlation table for all wine samples (n = 729).

	Dens	*Ethanol*	*Extract*	*TotAc*	*NVolAc*	*VolAc*	*Ash*	*SO2Free*	*RedSug*	*SFE*	*pH*
Ethanol	**–0.62**										
Extract	**0.88**	**–0.24**									
TotAcid	**0.46**	**–0.40**	**0.35**								
NVolAc	**0.44**	**–0.42**	**0.31**	**0.98**							
VolAc	0.02	**0.19**	0.15	**–0.21**	**–0.35**						
Ash	**0.48**	–0.07	**0.58**	–0.06	–0.11	**0.27**					
SO2Free	0.10	–0.07	0.07	–0.01	0.02	**–0.20**	–0.03				
RedSug	0.14	0.03	**0.19**	0.02	0.02	–0.10	–0.14	**0.30**			
SFE	**0.80**	**–0.21**	**0.91**	**0.36**	**0.32**	**0.19**	**0.66**	–0.04	–0.11		
pH	0.07	**0.19**	**0.18**	**–0.59**	**–0.62**	**0.28**	**0.61**	0.05	–0.17	**0.25**	
Mark	**–0.20**	**0.37**	–0.03	**–0.22**	**–0.22**	0.11	0.00	0.04	**0.18**	–0.09	0.11

The significant Spearman correlations are bold ($p < 10^{-6}$) and the highly significant coefficients are denoted by red colour ($p < 10^{-15}$).

cipal components (PCs), calculated by a linear combination of the original variables, sufficiently represent the total variability of the original data. Moreover, the positions of the original variables in the space of the PCs (the loadings plot) represent their inter-relations. If the variables are in the opposite position, the given variables are negatively correlated; if the variables are closely located, their inter-relation is positive. The graphical representation of the investigated objects (e.g. wine samples) in the score plot is very useful for detecting their possible association. In addition, in the PCA bi-plot, demonstrating simultaneously the objects and the variables, it is possible to detect those variables that are associated with the formed group of closely located objects, and the mutual relations among the objects and variables can be discovered.[19]

The first three PCs calculated from all descriptors (variables) accounted for 70.96% of the total data variability, as shown in Figure 2. The mutual position of the descriptors is in accordance with the correlation analysis. The strongly correlated variables *Dens*, *Extract*, *RedSug,* and *SFE* are adjacent and are opposite to the variable *SFE_E*. Further highly correlated descriptors *NVA_TA*, *NVolAc,* and *TotAc* are in a negative relation with the variables *VA_TA*, *Mark*, *VolAc* and *pH* (Figure 2).

The most recognisable natural grouping of the samples according to the wine colour was observed in the plane *PC 3* vs. *PC 1* (Figure 3). Positive values for *PC 3* were observed for the red wine samples, the white wine samples being characterised by the negative *PC 3* values. At the highest *PC 3* values the Teran wine samples were situated (Figure 3), accordingly to that Teran is a special type of Refosk wine. It can be seen in Figure 3 that Refosk (1) and Teran (4) were not clearly separated. The *PC 3* axis clearly differentiated the red and white wines; *PC 3* represents the wines' 'redness'. The descriptors *Ash* (most positive) and *SO2Free* (most negative) have the highest *PC 3* loadings – they seemed to be mainly connected to the colour of the wines. In addition, the variety Yellow Muscatel was separated from the others and the samples recorded very negative *PC 3* and *PC 1* values (Figure 3), which were connected to its sweet taste and high level of reducing sugars compared to the other varieties; as follows from the loadings plot in Figure 2, on the bottom.

The first principal component *PC 1* was mostly related to the ethanol content; the representing descriptors for *PC 2* were *Extract* and *Dens* with the highest loadings.

3. 4. Cluster Analysis

Cluster analysis (CLU) belongs to the unsupervised multidimensional procedures that involve measuring the distances or similarities between the objects (or variables) to be clustered. Not only the objects but also the variables are grouped into the clusters in terms of their proximity in the multidimensional space, which is determined by various clustering algorithms; Ward's method being the more used nowadays. The results of hierarchical clustering are usually displayed by a dendrogram showing all the clustering steps in detail.[20]

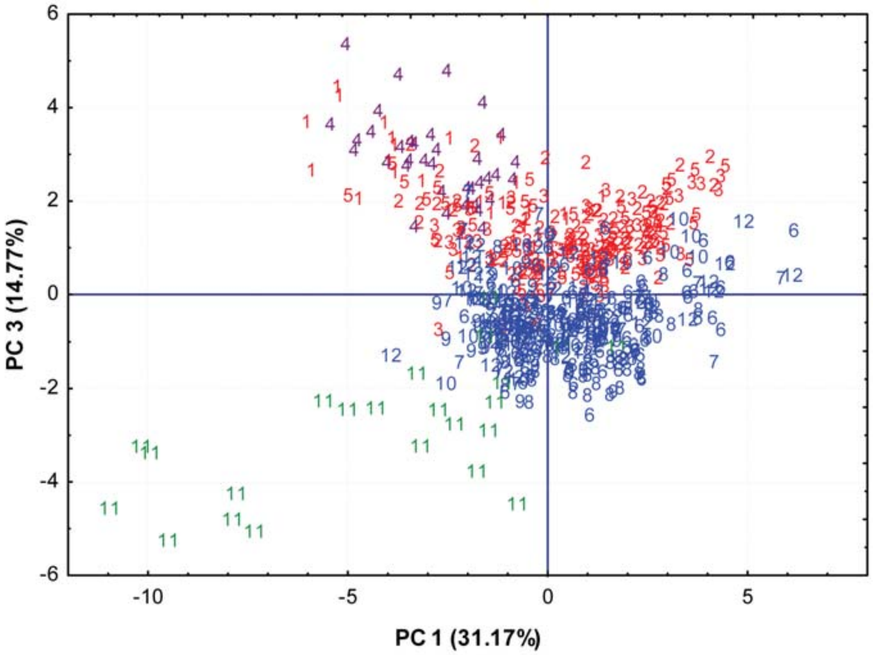

Figure 3: PCA score plot in the plane *PC 3* vs. *PC 1*. The objects are labelled by the wine variety – only two varieties are clearly separated from other varieties – Teran (4) and Yellow Muscatel (11). Teran (4) is a special type of Refosk wine (1) and these varieties are closely located in the figure. The remaining samples are denoted in blue (white wine) and red (red wine), resp. The *PC 3* values separate the samples according to the colour of the wines. Software STATISTICA v. 7.

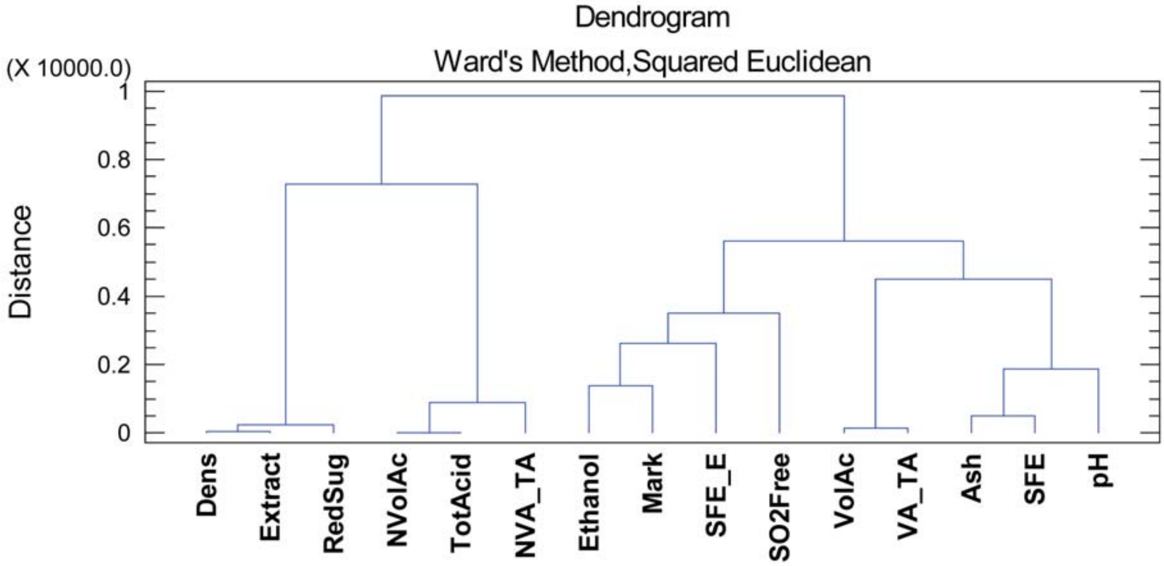

Figure 4: Dendrogram – the graphical output of the variable clustering; all wine samples being used. The sensory descriptor *Mark* is associated with *Ethanol*. Software STATGRAPHICS Centurion v. 15.

The most effective agglomerative clustering algorithm is used in Ward's method, which in this case was applied using the squared Euclidean distance as the similarity measure (Figure 4). The descriptor *Mark*, representing the sensorial quality of the wine samples, was most closely related to *Ethanol* regardless of whether all the wine samples were considered or either the red or white wines alone. In addition, *Mark* was clustered to *VolAc* for red wines and to *SO2Free* for white wines, which reflected the importance of the mentioned descriptors for the quality of the red and white wines, respectively. Further observed clusters were expected: (1) *TotAcid* and *NVolAc*, (2) *Dens* and *Extract*, (3) *Ash* and *pH*; which were observed in cluster analysis of all the samples, as well as for the red and white wine samples separately. In general, the mentioned results of cluster analysis were in good agreement with the PCA results. It is obvious that the descriptors composed by the descriptor ratio were closely clustered to the descriptor existing in the fraction numerator, therefore it was unnecessary to analyse them.

In the case of clustering the wine samples, no well-separated clusters were observed regardless of considering the target factors *Variety*, *Area,* and *Vintage.*

3. 5. ANOVA and Non-parametric Tests

The best way to examine the effect of a categorical variable (factor) on the magnitude of a continual variable (descriptor) is the application of analysis of variance (ANOVA). It is a statistical method for making simultaneous comparisons between two or more descriptors' means. A significant p-value (i.e. sufficiently small) resulting from any one-way ANOVA test would indicate that the continual variable is different in at least one of the groups analysed. If there are more than two groups being analysed, the one-way ANOVA approach does not specifically indicate which pairs of groups exhibit statistical differences. For this purpose, post hoc tests can determine which specific groups differ from each other. However, the necessary presumptions for ANOVA are the validity of normal distribution and the equality of variance of each population from which the sample is taken. If these assumptions are violated, the Kruskal-Wallis test – a non-parametric alternative has to be used. It is performed on the ranked data and, because of the loss of information involved in substituting the ranks for the original values it is a less powerful test than the one-way ANOVA approach. The Kruskal-Wallis test is an extension of the Mann-Whitney test that allows the comparison between just two independent groups, and is a non-parametric analogue to the Student t-test.[19]

With respect to differentiation between the red and white wines (a large data table with n = 729), all the descriptors were found to be statistically significant at the p-level $p < 0.001$ in the ANOVA outputs, as well as when using its non-parametric alternative – the Mann-Whitney test. Separate data for red and white wines were used for further statistical analysis.

The wine *Variety* was the first investigated target factor. Altogether 262 red wines belonging to four categories were analysed – Refosk (29 samples), Cabernet Sauvignon (110), Merlot (86), and Teran (37). The samples containing the mixture of several varietal red wines were not considered in these analyses. The results of ANOVA (Table 3) were concordant with the results of Kruskal-Wallis test. Similarly, in the case of white wines, the mixtures of several varietal white wines were also excluded from further analyses, and six white wines *Variety* categories were explored – Chardonnay (110 samples), Sauvignon (58), Malvasia (39), Pinot Gris (54), Yellow Muscatel

(24), and Rebula (40). The results of ANOVA agreed again with the Kruskal-Wallis test.

The second investigated factor *Area* expressed the geographical origin of the wine samples. The geographical origin of some samples (both – white and red wines) was not declared, so they were therefore excluded. Four categories of *Area* were considered for both sorts of wine – Goriška Brda, Koper, Kras, and Vipavska dolina. In the case of red wines, all descriptors were found statistically significant except *RegSug* and *Mark* (Table 3). In Kruskal-Wallis test, the descriptor *SFE_E* was additionally found as not significant ($p > 0.12$). In regard to the white wines, almost all descriptors were significant except *VolAc*, *SFE_E* and, *Mark*, which was confirmed by the Kruskal-Wallis test.

Mark was observed mainly for *Variety* and partially for *Vintage*.

In general, the results of ANOVA were in good agreement with the non-parametric Kruskal-Wallis test. The reason for some discordance between them was due to violation of the homogeneity of the variances (found by the Levene test) in the case of some descriptors, being one of the ANOVA assumptions. An attempt to improve the distribution normality by utilising the reciprocal or logarithmic forms of some descriptors led to *p*-values very similar to the case when using the original variable.

Additional LSD (least significant difference) post-hoc test and the alternative Mann-Whitney test were also applied for obtaining detailed information about the differences in the particular categories of the target factors.

Table 3: Summarization of results of ANOVA for the target factors – *Variety*, *Area,* and *Vintage*. The statistically significant *p*-values are in bold, and the highly-significant denoted by the blue colour.

	Target factors for red wines			**Target factors for white wines**		
Descriptor	**Variety**	**Area**	**Vintage**	**Variety**	**Area**	**Vintage**
	(*n* = 262)	(*n* = 243)	(*n* = 243)	(*n* = 325)	(*n* = 277)	(*n* = 277)
Ethanol	**6.12E–13**	**2.39E–09**	**8.39E–08**	**6.32E–17**	**3.33E–06**	**0.01**
Dens	**8.26E–23**	**4.72E–15**	**2.94E–12**	**6.87E–54**	**0.003**	**0.002**
Extract	**1.1E–08**	**3.65E–08**	**6.83E–16**	**1.09E–53**	**0.007**	**0.0005**
RedSug	0.64	0.06	0.07	**5.12E–53**	**0.009**	**0.0002**
SFE	**2.37E–12**	**3.22E–12**	**5.87E–20**	**5.19E–16**	**0.009**	0.08
SFE_E	0.32	**0.009**	0.52	**1.63E–85**	0.06	**0.001**
TotAc	**1.27E–56**	**2.16E–35**	**8.53E–10**	**4.86E–08**	**1.56E–07**	**0.001**
NVolAc	**7.64E–51**	**5.48E–33**	**2.63E–10**	**1.9E–06**	**3.13E–07**	**0.0002**
VolAc	0.24	**1.09E–05**	**0.0004**	0.59	0.11	**7.32E–05**
NVA_TA	**2.84E–08**	**5.66E–12**	**1.02E–08**	0.58	**0.005**	**9.33E–07**
VA_TA	**2.55E–08**	**5.37E–12**	**9.16E–09**	0.56	**0.004**	**7.96E–07**
pH	**9.12E–32**	**1.36E–20**	0.11	**2.74E–06**	**5.92E–05**	**0.004**
Ash	**0.0002**	**0.0003**	**0.02**	**6.54E–05**	**0.0001**	**0.004**
SO2Free	**0.001**	**0.006**	0.36	**0.003**	**0.002**	0.13
Mark	0.07	0.41	**9.05E–08**	**6.37E–11**	0.91	**0.04**

The third studied target factor was *Vintage*. Four categories of the year of wine production (2003, 2004, 2005, and 2006) were used for both the red and white wine samples. For the red wines, the results of ANOVA (Table 3) differed from the results of Kruskal-Wallis test in case of *RedSug* found as significantly affected by *Vintage* ($p < 0.006$). In addition, for white wines, higher significance for *SFE* ($p < 0.03$) and additional insignificance for the descriptors *Dens* ($p > 0.16$) and *Ethanol* ($p > 0.19$) resulted from the Kruskal-Wallis test.

In conclusion, the influences of the target factors varied upon the investigated descriptors when separately summarised for the red and white wine samples.

The sensorial quality descriptor *Mark* is used in Table 3 in the form of a continuous variable, and it is worth noting that it was affected differently by the three target factors. A strong effect of *Vintage* was visible for the red wines; but other target factors did not have any influence. On the other hand, for white wines the strong effect upon

3. 6. Linear and General Discriminant Analysis

Linear discriminant analysis (LDA) is a classification procedure that renders a number of orthogonal linear discriminant functions equal to the number of categories, minus one. The method maximises the variance between categories and minimises the variance within the categories.[20–21] General discriminant analysis (GDA) allows using also categorical variables in computation of the classification model, in contrast to LDA. These classification techniques belong to supervised multidimensional techniques, in which the classification model is constructed by means of the data containing the objects pre-categorised into the known categories (the training data set). The developed model is then employed for classification of those samples that were not used in designing the model (the validation data set). The validation of the classification model is necessary in order to verify its prediction ability. An important task in the optimisation of the discrimination model is an appropriate selection of input variables (descriptors).

The LDA calculations were performed using software SPSS v. 15. The stepwise selection method was used for optimising the set of input variables. The validation of the LDA model was accomplished by the 'leave-one-out' (LOO) method and compared to the results of 4-fold-cross validation. The GDA models were calculated by means of the STATISTICA v. 7 package. In all classification models, the sensory quality descriptor *Mark* was not utilised as an input variable.

The first task was to classify the wine samples according to *Variety*. For this purpose data were employed containing samples originating from all production areas and all vintages.

In the case of red wines, four categories of *Variety* were classified: (1) Refosk (29 samples), (2) Cabernet Sauvignon (110), (3) Merlot (86), and (4) Teran (37). The results of the achieved LDA classification model were significant, but not successful in classification of category 1 (Refosk) since only 55% of samples were correctly classified, the most part of incorrectly classified samples was assigned into category 4 (Teran). In LOO validation, the situation was similar. The classification rates for the three remaining categories were acceptably good and the overall correct classification ratio was 71.8% for the training set and 69.5% for the LOO validation (Table 4). The most discriminating variables were optimised by the stepwise selection procedure: *TotAc*, *NVolAc*, *Ethanol*, *pH*, *Ash*, *SO2Free*, *Dens*, *NVA_TA*, *Extract,* and *RedSug*.

Nevertheless, the discrimination of Refosk and Teran categories was problematic because Teran is a special type of Refosk wine therefore the classification was performed using joined samples of both categories, resulting in a more homogeneous distribution of the objects. Using the following optimised descriptors *SFE*, *Ethanol*, *pH*, *NVolAc*, *TotAc*, *NVA_TA*, *SO2Free, Dens, recRedSug,* the classification results were better (Table 4); the rates for correctly classified objects in each category exceeded 70%, and the average percentages of the correctly classified objects were 77.5% for the training set and 76.3% for LOO validation. When 4-fold cross-validation was applied, the results were very similar to the LOO method – 77.5% and 76.3% of samples were correctly classified for the training and validation samples, respectively. Since only 33% classification success expected by a random classification into three categories, we can conclude that the selected enological descriptors have considerable discrimination ability for classification of selected red wine's varieties.

On the other hand, the results of classification of the white wine samples by *Variety* were not so successful, when using all six categories: (1) Chardonnay (110 samples), (2) Sauvignon (58), (3) Malvasia (39), (4) Pinot Gris (54), (5) Rebula (40), and (6) Yellow Muscatel (24).

Despite very good classification rates for Chardonnay (78.2% in training and 72.7% in validation), and Yellow Muscatel (91.7% and 87.5%, resp.), the remaining categories were not well distinguished; especially in case of Rebula (only 22.5% correctly classified). This resulted in an overall classification rate of 54.8% for the training data and 50.2% for validation employing the optimally selected variables *SFE_E*, *Dens*, *Ethanol*, *TotAc*, *NVolAc, pH*, *recRedSug*, *SO2Free*, *logVolAc*, *SFE,* and *Ash*. When using independent LDA calculation covering four categories, without the non-problematic varieties Chardonnay and Yellow Muscatel, the classification rates were improved only slightly (Table 4). According to this, we can conclude, that the varieties Sauvignon, Malvasia, Pinot Gris and Rebula are not clearly distinguishable when using simple enological descriptors for their classification. Consequently, the analysed descriptors exhibited satisfactory discrimination efficiency in case of the varieties Chardonnay and especially Yellow Muscatel when all six categories were classified, since the results were far above the 16.7% limit – the classification success corresponding to a random classification into six categories.

Table 4: Summarization of the results of LDA for the classification of red and white wine samples according to the *Variety*.

	Number of categories	Classification model	Leave-one-out validation	Input variables
Red wines	4	71.8%	69.8%	*TotAc, NVolAc, Ethanol, pH, Ash, SO2Free, Dens, NVA_TA, Extract, RedSug*
	3	77.5%	76.3%	*SFE, Ethanol, pH, NVolAc, TotAc, NVA_TA, SO2Free, Dens, recRedSug*
White wines	6	54.8%	50.2%	*SFE_E, Dens, Ethanol, TotAc, NVolAc, pH, recRedSug, SO2Free, logVolAc, SFE, Ash*
	4	60.7%	52.4%	*Ethanol, VolAc, pH, Extract, recRedSug, Ash, SO2Free, NVA_TA*

Considering other research employing enological descriptors for classifications according wine's variety[15] similarly, the total content of acids and ash along with tartaric acid content have been selected as optimal input variables in classification of three Slovakian white varieties resulting in excellent classification rates in stepwise LDA.

Further, the target factor *Area* was used for categorisation of wine samples according to the geographical origin. When all varieties of red wine were used, the classification was significant but not too successful. Four categories (1) Goriška Brda (134 samples), (2) Vipavska dolina (83), (3) Koper (35), and (4) Kras (39) were considered, and the results were 67.7% of correctly classified training samples and 64.6% success in LOO validation; the descriptors *recNVolAc*, *pH*, *SFE*, *VolAc*, *VA_TA*, *SFE_E*, *Ethanol*, *SO2Free* and *recRedSug* were found as optimal by stepwise selection. The most problematic was the classification of wines from the Koper region (only 22.9% correct). Taking into account that the Koper and Kras regions are closely located, a new classification was performed with the samples from Koper and Kras joined together, which resulted in a more homogeneous distribution of the samples into created categories. The rate of correctly classified samples into three categories improved evidently – 73.2% for the training data and 70.1% for LOO validation; in this case the descriptors *TotAc*, *pH*, *SFE*, *Ash*, *SO2Free*, *VolAc*, *VA_TA* and *SFE_E* were designated as being optimal (Table 5). Regarding only a 33% success expected by a random classification into three classes, the results of the proposed classification model showed, that the simple enological descriptors exhibited a considerable discrimination power for classification of red wines according to the region of their production, even when different varieties of red wines were included in all classified categories of *Area*.

Table 5: Summarization of results of LDA, GDA and ANN for the classification of red wine samples according to the *Area*.

	Number of categories	Training set	Validation set	Input variables
LDA	4	67.7%	64.6%	*recNVolAc, pH, SFE, VolAc, VA_TA, SFE_E, Ethanol, SO2Free, recRedSug*
	3	73.2%	70.1%	*TotAc, pH, SFE, Ash, SO2Free, VolAc, VA_TA,SFE_E*
GDA	4	74.2%	68.9%	*Variety, recNVolAc, pH, SFE, VolAc, VA_TA, SFE_E, Ethanol, SO2Free, recRedSug*
	3	75.3%	70.5%	*Variety, TotAc, pH, SFE, Ash, SO2Free, VolAc, VA_TA, SFE_E*
ANN	3	93.6%	75.0%	*pH*, *VolAc*, *Dens*, *SFE*, *Ethanol*, *NVA_TA*, *TotAc*, *logNVolAc*, *VA_TA*, *Ash, SO2Free*

Additionally, the variable *Variety* – as an additional categorical input – was implemented in the classification of the red wine samples by *Area* using GDA and hence, the samples originated from mixtures of wine varieties had not been employed in classification. For the GDA model validation, a special validation data set was created which consisted of randomly selected samples comprising 25% of their total number. The addition of categorical input *Variety* did not improve the classification results significantly. Using four categories of *Area*, the percentage of correctly classified samples was 74.2% for training data and 68.9% for the validation data set. When three classes of *Area* were classified the result of classification was 75.3% and 70.5% for the training and validation data, respectively (Table 5).

In case of white wines, the distribution of samples into four categories of *Area* was considerably non-homogenous – 144 samples were from the region Goriška brda, 83 from Vipavska dolina, 40 from Koper, and only 10 samples from Kras. Therefore the samples originating from the adjacent Koper and Kras regions were joined but the classification model was still not successful. The same was observed even if the category Goriška brda had been randomly diminished to only 74 samples to obtain more homogeneous distribution. Therefore an additional approach was attempted – to classify the samples by *Area* separately for particular varieties of white as well as red wine samples. The corresponding LDA results were very good and are summarised in Table 6.

All selected input descriptors have been chosen by stepwise selection method. It is interesting, that in case of red wines, the content of non-volatile acids, and particularly also ethanol level, were very important for differentiation of regions of production (Table 6). In addition, the ability of a single descriptor to discriminate two *Area* categories was evident in few cases, especially in the case of Malvasia data, when only the content of reducing sugars was efficient to distinguish the wine samples according to their region of production. Furthermore, the content of free SO_2 was important feature for discrimination between two regions for Pinot Gris data. As it is known, the content of free SO_2 is typically connected to technology of wine's production. However, according to geographical distinctness of the regions analysed, the climatic and edaphical conditions are different. Thus, the maturation of grape (acid consumption, sugar accumulation, phenolic synthesis, etc.) is developed in conditions associated with the production region. Therefore, the geographical proximity is reflected in similar traditional wine making practices which contribute to the factors determining geographical authenticity of products. Further important properties for distinguishing the regions of production of white wines were the density, content of ash and pH, specifically for Sauvignon and Chardonnay wines. The graphical output of classification of the samples belonging to Chardonnay variety according to their geographical origin is shown in Figure 5.

Concerning other research utilising the enological properties for classification according to the geographical origin,[14] the contents of ethanol and total SO_2 have been

Table 6: Summarization of the results of LDA for the classification of white and red wine samples according to *Area,* separately by *Variety*.

	Variety	Area	Number of samples	Classification model	Leave-one-out validation	Input variables
White wine	**Sauvignon**	Goriška brda Vipavska dolina	21 23	90.9%	90.9%	*Dens, Ash, pH*
	Malvasia	Koper Vipavska dolina	21 9	90.0%	90.0%	*RedSug*
	Pinot Gris	Goriška brda Vipavska dolina	32 7	84.6%	84.6%	*SO2Free*
	Yellow Muscatel	Goriška brda Koper Vipavska dolina	6 7 10	78.3%	60.9%	*TotAc, Dens, Ethanol*
	Chardonnay	Goriška brda Koper Kras	59 9 6	93.2%	89.2%	*Dens, Extract, NVolAc, Ash, SO2Free, RedSug, pH*
Red wine	**Cabernet Sauvignon**	Vipavska dolina Goriška brda	40 55	77.9%	76.8%	*Ethanol, NVolAc*
	Merlot	Vipavska dolina Goriška brda	23 48	66.2%	66.2%	*NVolAc*
	Refosk and Teran	Goriška brda Koper Kras	18 8 37	84.1%	71.4%	*Ethanol, recTotAc, NVolAc, SO2Free, SFE*

also considered among other wine's properties as the most important variables for the differentiation of Spanish rose wines from three production areas. Together with several elements and phenolic compounds, these variables provided successful classification rates using stepwise LDA.

In addition, LDA calculations have been utilised in classification of different wines originated from Slovenia with employing more sophisticated analytical techniques. For example, a combination of SNIF-NMR (site-specific natural isotopic fractionation nuclear magnetic resonance) and IRMS (isotope-ratio mass spectrometry) analyses achieved efficient classification of three Slovenian production areas, when the coastal production area differed evidently from the continental areas.[17, 22] Perfect classification rates have been accomplished in classification of ten samples of red Slovenian wines into three varieties' categories with use of anthocyanins as input variables in regularised discriminant analysis.[23] Similarly, the regularised discriminant analysis has been applied efficiently to classify the Slovenian and Apulian wines by utilising ^{1}H NMR (nuclear magnetic resonance) signals in comparison with data comprising contents of organic acids and trace elements.[24] In all mentioned researches, the discrimination power of data used for classification was evidently higher in comparison with the enological properties applied in this paper. However, the analytical procedures necessary to obtain these significant descriptors are time and cost demanding.

Consequently, the simple enological descriptors provided considerable information for the discrimination of wines by their variety and geographic origin, and therefore could contribute to authentication process of wine, especially in connection with other types of analytes characterised by stronger differentiation ability e.g. elements or isotopic ratios for geographic origin determination and different organic compounds or spectral signals in case of designation of wine's variety.

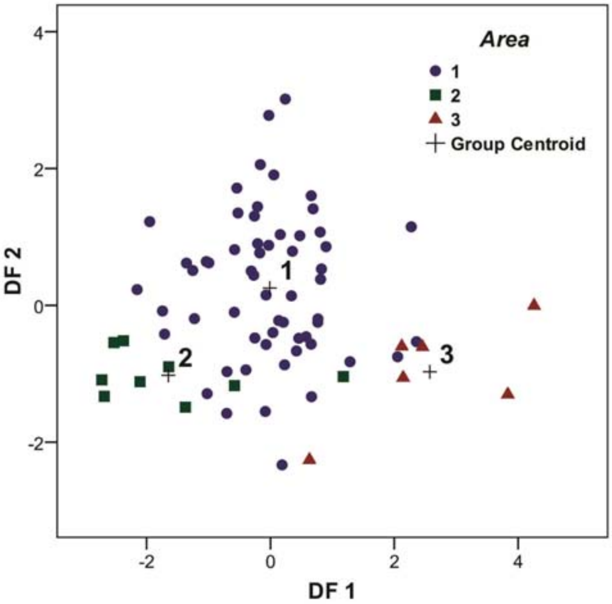

Figure 5: Graphical output of LDA in plane of first two discriminant functions – classification of Chardonnay wine samples according to the region of production (1 – Goriška brda, 2 – Koper, 3 – Kras). The percentage of correct classification: for training set 93.2% and for validation set 89.2% (leave-one-out). Software SPSS v. 15.

3. 7. Classification by Artificial Neural Networks

The use of ANN for data processing can be characterised by analogy using biological neurons. The artificial neural network itself consists of interconnected neurons situated in an input layer, one or more hidden layer(s), and an output layer. The input neurons receive the input data characteristics for each observation; the output neurons provide the predicted value or pattern of the studied objects. In most cases, the ANN architecture consists of two active layers – one hidden and one output layer. The neurons of two adjacent layers are mutually connected and the importance of each connection is expressed by weights.[10]

The major advantage of the artificial neural networks (ANNs) over traditional multivariate techniques is their efficiency in handling more complex and non-linear problems. Besides, the construction of ANNs is not significantly affected by the imbalance in the number of samples in the chosen categories, and does not impose any requirements with respect to the structure of the data.[25–26]

In general, the implementation of ANNs resulted in improving the classification rates. The key decision on the network type, and the number and structure of the hidden layers was made by the Automated Neural Networks option in STATISTICA v. 7, by which a variety of algorithms for different network types were automatically tested and the best alternatives determined. A three-layer perceptron was indicated as the best network type, i.e. with a single hidden layer. The validation of the ANN classification models was performed with the help of an independent validation sample set, which consisted of 20% of the total randomly selected number of samples.

The classification of the red wine samples by *Variety* into all four categories was improved to 80.7% for the training subset and was significant but not too good for the validation set – 61.5%. The following descriptors were used as the ANN inputs: *recTotAc*, *pH*, *logNVolAc*, *VolAc*, *NVA_TA*, *Ethanol*, *SFE*, *Ash*, *Dens,* and *SO2Free*. Six neurons were selected as optimal in the hidden layer and the BFGS (Broyden-Fletcher-Goldfarb-Shanno) algorithm with 31 epochs was employed. The hyperbolic tangent function in the hidden layer and the softmax function in the output layer were used as the activation functions. Similarly as in LDA classification, the varieties Teran and Refosk were joined in one category and the percentage of correctly classified samples improved to 86.0% for training and 72.3% for the validation samples. A three-layer perceptron was employed with 8 hidden neurons and the back-propagation within a 100 epoch algorithm. The optimised inputs were: *pH*, *VolAc*, *NVolAc*, *TotAc*, *SFE*, *VA_TA*, *RedSug*, *Ethanol*, *Extract*, *Ash*, and *SO2Free*.

The classification of the red wine samples into three categories according to *Area* provided better results when compared to the LDA: 93.6% of the samples were correctly classified in the training subset and 75.0% in the validation subset (Table 5). The employed training algorithm was the BFGS with 55 epochs, and six neurons used in the hidden layer. The logistic activation function in the hidden layer and the softmax activation function in the output layer were used. The following input descriptors were selected: *pH*, *VolAc*, *Dens*, *SFE*, *Ethanol*, *NVA_TA*, *TotAc*, *logNVolAc*, *VA_TA*, *Ash,* and *SO2Free*. When the categorical descriptor *Variety* was added to the continuous descriptors at the input, the classification rate was not improved.

Several architectures of three-layer perceptron and different activation functions were explored when optimising the classification of all investigated varieties of white wines into three *Area* categories, but the results were not encouraging – similarly to the LDA classification.

Considering other research involving enological descriptors to classify the wine samples according to the production area by three-layer perceptron,[14] the contents of total acids and free SO_2 have been similarly chosen as the input variables together with different five analytes resulting in perfect classification.

Since two different supervised pattern recognition techniques were applied together with diverse procedures of selection of input variables in classification models without significant improvement of classification rates, we conclude, that only addition of different analytical descriptors with higher discrimination power would improve the classification of white wines according to the target factors substantially, as it was reported in numerous already published results.

4. Conclusions

All the studied descriptors selected for wine samples' characterization were found to be statistically significant in relation to the differentiation between red and white wines. Numerous differences among them were found when compared their importance with regard to the categories of the target variables *Variety*, *Area*, and *Vintage*, as detected by ANOVA as well as by the non-parametric Kruskal-Wallis test. *Mark*, the sensory quality descriptor (evaluated by the group of experts), was significantly positive correlated with the ethanol content for both red and white wine samples and was negatively correlated to the content of non-volatile acids, total acids, and the density. However, the correlation of this sensorial quality descriptor to density was significant, particularly for the white wine samples. Further considerable inter-relations were observed by the correlation analysis; noteworthy among them was significant positive correlation between the ash content and the wine' pH, as well as negative high correlation between the ethanol content and relative density.

The three most important principal components (*PC 1*, *PC 2,* and *PC 3*), calculated by PCA from all descriptors, comprised 70.96% of the total data variability. *PC 1* was mainly influenced by the ethanol content, *PC 3* was the most important for the red and white wines' separation – red wine samples were characterised by significantly higher *PC 3* values. The highest *PC 3* values appertained to the Teran wine samples (special type of Refosk wine). The samples of the variety Yellow Muscatel clearly differed from the others, they exhibited the lowest *PC 3* values and their distinctness was also revealed in the wine classification by variety. The PCA loading plots were concordant with the outputs of correlation analysis and in several aspects similar to the results of cluster analysis.

In classification by variety of the red wine samples, the most problematic classification was between varieties Teran and Refosk, because Teran is a special type of Refosk wine. When joining the varieties (Refosk plus Teran together), utilisation of the optimally selected descriptors and linear discrimination model resulted in a satisfactory classification rate.

Application the three-layer perceptron ANN improved the classification performance compared to LDA even if both, Refosk and Teran varieties were included in the calculations. The classification of white wines was considerably less successful. Despite very good classification rates for the varieties Yellow Muscatel and Chardonnay, the discrimination of other white wine varieties was insufficient.

Classification of the red wine samples into the three categories according to the wine production region was successful. An important circumstance should be repeated, that the production regions are located very close together and belong to one wine's production area in Slovenia. For the red wine classifications all the varieties and vintages were included, thus showing that the employed traditional enological descriptors proved very valuable tools for discriminating the red wines according to their production areas. The situation concerning the white wines was not analogous; in this case the classification according to geographical origin was successful, but only when the samples of a single variety were used.

5. Acknowledgements

The authors thank the Ministry of Higher Education, Science and Technology of the Republic of Slovenia, contract number P2-006 for financial support. The support of Slovak grant agency VEGA 1/0233/12 is highly acknowledged. The cooperation with the Chamber of Agriculture and Forestry of Slovenia, Institute for Agriculture and Forestry, is highly acknowledged.

6. References

1. S. M. Rodrigues, M. Otero, A. A. Alves, J. Coimbra, M. A. Coimbra, E. Pereira, A. C. Duarte, *J. Food Compos. Anal.* **2011**, *24*, 548–562.
2. M. Álvarez, I. M. Moreno, Á. Jos, A. M. Cameán, G. González, *Microchem. J.* **2007**, *87*, 72–76.
3. P. Paneque, M. T. Álvarez-Sotomayor, A. Clavijo, I. A. Gómez, *Microchem. J.* **2010**, *94*, 175–179.
4. M. J. Baxter, H. M. Crews, M. J. Dennis, I. Goodal, D. Anderson, *Food Chem.* **1997**, *60*, 443–450.
5. P. Paneque, M. T. Álvarez-Sotomayor, I. A. Gómez, *Food Chem.* **2009**, *117*, 302–305.
6. M. P. Fabani, R. C. Arrua, F. Vazquez, M. P. Diaz, M. V. Baroni, D. A. Wunderlin, *Food Chem.* **2010**, *119*, 372–379.
7. J. Šperková, M. Suchanek, *Food Chem.* **2005**, *93*, 659–663.
8. N. H. Beltrán, M. A. Duarte-Mermoud, M. A. Bustos, S. A. Salah, E. A. Loyola, A. I. Peña-Neira, J. W. Jalocha, *J. Food Eng.* **2006**, *75*, 1–10.
9. D. P. Makris, S. Kallithraka, A. Mamalos, *Talanta*. **2006**, *70*, 1143–1152.
10. D. Kruzlicova, J. Mocak, B. Balla, J. Petka, M. Farkova, J. Havel, *Food Chem.* **2009**, *112*, 1046–1052.
11. J. S. Câmara, M. A. Alves, J. C. Marques, *Talanta*. **2006**, *68*, 1512–1521.
12. S. Rebolo, R. M. Peña, M. J. Latorre, S. García, A. M. Botana, C. Herrero, *Anal. Chim. Acta*. **2001**, *417*, 211–220.
13. A. Sass-Kis, J. Kiss, B. Havadi, N. Adányi, *Food Chem.* **2008**, *110*, 742–750.
14. S. Pérez-Margariño, M. Ortega-Heras, M. L. González-San José, Z. Boger, *Talanta*. **2004**, *62*, 983–990.
15. K. Šnuderl, J. Mocak, D. Brodnjak-Vončina, B. Sedláčková, *Acta Chim. Slov*. **2009**, *56*, 765–772.
16. M. A. Brescia, V. Caldarola, A. De Giglio, D. Benedetti, F. P. Fanizzi, A. Sacco, *Anal. Chim. Acta*. **2002**, *458*, 177–186.
17. I. J. Košir, M. Kocjančič, N. Ogrinc, J. Kidrič, *Anal. Chim. Acta*. **2001**, *429*, 195–206.
18. http://sl.wikipedia.org/wiki/Slika:Primorska_vino_si.pngD.L.
19. D. L. Massart, B. G. M. Vandeginste, S. N. Deming, Y. Michotte, L. Kaufman, *Handbook of chemometrics and qualimetrics: Part A*. Elsevier, Amsterdam, **1997**.
20. K. Varmuza, P. Filzmoser, *Introduction to Multivariate Statistical Analysis in Chemometrics*. CRC Press, Boca Raton, **2009**.
21. R. G. Brereton, *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*. Wiley, Chichester, **2003**.
22. N. Ogrinc, I. J. Košir, M. Kocjančič, J. Kidrič, *J. Agric. Food Chem*. **2001**, *49*, 1432–1440.
23. I. J. Košir, B. Lapornik, S. Andrenšek, A. Golc Wondra, U. Vrhonšek, M. Kocjančič, J. Kidrič, *Anal. Chim. Acta*. **2004**, *513*, 277–282.
24. M. A. Brescia, I. J. Košir, V. Caldarola, J. Kidrič, A. Sacco, *J. Agric. Food Chem*. **2003**, *51*, 21–26.
25. S. Haykin, *Neural networks. A comprehensive Foundation*. Pearson Education, Delhi **2001**.
26. B. G. M. Vandeginste, D. L. Massart, L. M. Buydens, S. De Jong, P. J. Lewi, J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics: Part B*. Elsevier, Amsterdam, **1998**.

Povzetek

Določevanje izvora je ena od osnovnih zahtev, ko želimo certificirati pristnost vin. Raziskava opisuje možnosti napovedovanja izvora vin z uporabo učinkovitih metod analize parametrov vin in multivariantno analizo. Glavni namen študije je proučevanje možnosti razlikovanja enostavnih enoloških deskriptorjev za klasifikacijo vzorcev slovenskih rdečih in belih vin glede na vrsto in geografski izvor. Drugi cilj je bil proučevanje razmerij med deskriptorji, kot so: relativna gostota, vsebnost skupnih kislin, nehlapne kisline, hlapne kisline, pepel, reducirajoči sladkor, prosti sladkor, SO_2, etanol, pH in med pomembnimi dodatnimi spremenljivkami, kot je senzorična kakovost vina z uporabo korelacijske analize, metode glavnih osi (PCA) in analizo grupiranja podatkov (CLU). 739 vzorcev rdečih in belih vin je bilo posnetih na aparatu Wine Scan FT 120, od valovnega števila 926 cm^{-1} do 5012 cm^{-1}. Uporabljene metode linearne diskriminantne analize (LDA), splošne diskriminantne analize (GDA) in umetnih nevronskih mrež (ANN) potrjujejo sposobnost določanja pristnosti vin.