

## Vineyard zonation based on natural terroir factors using multivariate statistics – Case study Burgenland (Austria)

Lukáš Karlík<sup>1\*</sup>, Marián Gábor<sup>2</sup>, Vladimír Falt'an<sup>2</sup> and Marek Havlíček<sup>3</sup>

<sup>1</sup>Comenius University in Bratislava, Faculty of Natural Sciences, Department of Regional Geography, Planning and Environment, Mlynská dolina, Ilkovičova 6, 842 15, Bratislava, Slovakia

<sup>2</sup>Comenius University in Bratislava, Faculty of Natural Sciences, Department of Physical Geography and Geoecology, Mlynská dolina, Ilkovičova 6, 842 15, Bratislava, Slovakia

<sup>3</sup>Silva Tarouca Research Institute for Landscape and Ornamental Gardening, v. v. i. Lidická 25/27, 602 00 Brno, Czechia

### Abstract

**Aim:** The aim of the study was to explicitly develop a methodology of delineation of natural Protected Designation of Origins (PDO) terroir regions for Austria, where PDO viticulture regions reflect natural conditions only in a formal manner.

**Methods and results:** There is increasing competition in the wine market from globalized trends, where the European Union (EU) and non-EU wine producers have adopted different market strategies to promote their wines and gain larger market share. The EU has therefore established protective agricultural product categories such as PDO and Protected Geographical Indication (PGI) based on the terroir concept which contrasts with USA strategies. Here, we first collected and derived as many as possible relevant physical geographical data for a total of 66,673 officially registered vineyard areas in Burgenland (Austria). Next, we applied factor analysis to these data with the aim to shrink their size and reduce their dimensionality. For each vineyard plot was derived a factor score which was used for performing k-means clustering. The best count of clusters, k-parameter, was estimated using five internal validity indices. Five homogenous management zones were created as a result of clustering. Correctness and accuracy of the clustering was evaluated by multidimensional discriminant analysis. The final zones were compared to current Districtus Austriae Controllatus (DAC) of Burgenland.

**Conclusion:** It was found by the comparison of DAC regions of Burgenland and our drafted zones that some of the DAC regions do not respect natural terroir zones, while these regions were created as PDO regions which should respect their natural terroir.

**Significance and impact of the study:** The presented methodology can be applied all over Austria and, with some modifications caused by different input data, to each EU member country where it is necessary to revise PDO regions' borders.

**Key words:** PDO regions, zonation, vineyard areas, terroir, factor analysis

Received : 1 August 2017; Accepted : 30 May 2018; Published : 25 June 2018

doi: 10.20870/oeno-one.2018.52.2.1907

## Introduction

Discussion between the European Union (EU) and the United States on food product labeling has become a central topic in ongoing trade negotiations. The EU asserts the concept of protection based on geographical indication (GI) and this contrasts sharply with the USA business strategy used in their geographical-origin labeling (Gervais, 2015; Van Caenegem, 2015).

The main GI concept was developed in France at the beginning of the 20<sup>th</sup> century, aiming to promote local wine products amid escalating competition in the international wine market. This system also protected the producer's reputation by applying the appellation system. Other European countries, including Spain, Italy and Portugal, rapidly followed France in adopting the appellation system (Barham, 2003). In 2008, the EU has initiated registration of GIs under "Protected Designation of Origins" (PDO) for certain wine sector products (Council Regulation (EC) No 510/2006; European Commission, 2006). Although member states employ some differences in applying GI protection rules, the main GI protection concept is always based on the terroir concept.

The resolution of the International Organization of Vine and Wine (OIV) 333/2010 defines the terroir concept as "applying to the region where the interaction between the organic and inorganic parts of the land and the applied agro-practices constitute the combined characteristics recognizing the uniqueness of the product from these areas" (OIV, 2010). France initiated the terroir concept in connection with wine. These attributes can have direct or indirect effect on the typicality and quality of the region's wine.

Additional interpretations of terroir include:

(A) In addition to environmental factors, human geographical factors such as tradition connected with the transfer of information and the cultural environment can also influence wine character (Vaudour, 2002).

(B) Carey (2001) defines terroir as a complex of natural factors not easily changeable or modifiable by anthropogenic activity.

(C) Gladstones and Smart (1997) consider terroir to be more complex, so that influences on the wine cannot be measured simply as a combination of the landscape's natural components.

GIs are protected not only by the terroir but also by the implementation of specific standards and procedures. In practice, this means the wine must not only come from a specific region, but must also be made by approved procedures and meet specific attributes. The EU's third major approach to GI is that the GI product has total protection from unauthorized use so that new products cannot be labeled in any way that can be mistaken for a GI product (Gangjee, 2012).

United States' protection of all agricultural products, including wine, is based on application for trademark; the producers gain exclusive rights and protection of their product and buyers gain greater guarantee of high-quality products. The U.S. Patent and Trademark Office generally excludes registration of products with trademark names based on geographical locations (United States Patent and Trademark Office, 2016). Further, geographical features can be protected only where these features

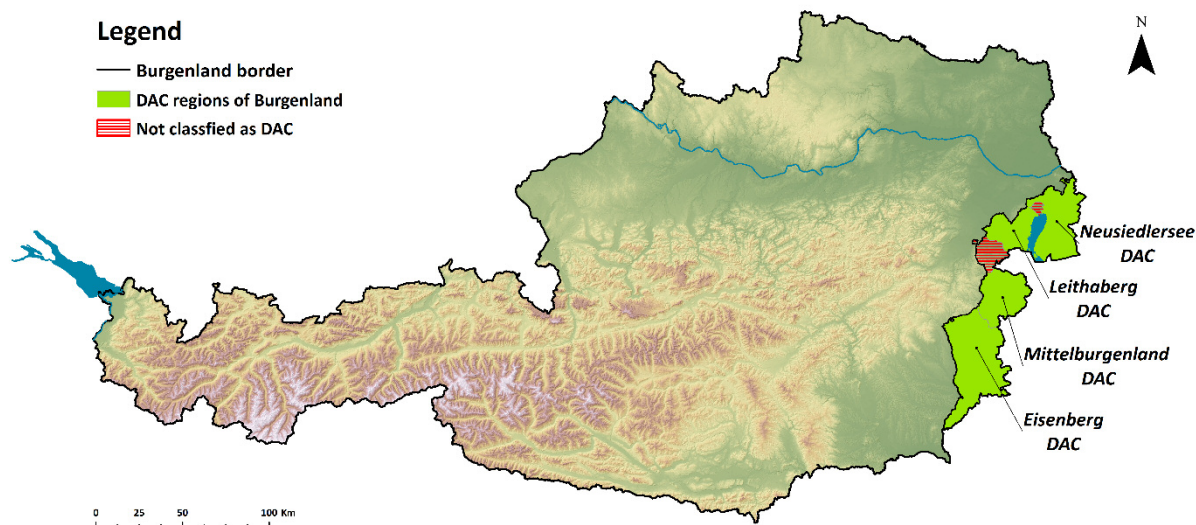


Figure 1. Location of Burgenland DAC regions in Austria.

refer to properties other than the product's geographical origin. America opposes the European GI approach mainly because it considers the European model as a form of protectionism, protecting traditional EU producers and ensuring them advantages over other global market producers. In rebuttal, the EU argues that its approach mainly protects consumers. Under the EU system, consumers are ensured explicit information on products they purchase, and it is harder to confuse similar products with each other (Farrand, 2016).

Discussions on food labeling remain an ongoing topic between the EU and USA at World Trade Organization (WTO) and Trans-Pacific Partnership (TPP) and also in acceptable formulation of the Transatlantic Trade and Investment Partnership (TTIP). An effective argument against implementation of the European GI model is to challenge "region" as the foundation of the PDO. Barham (2003) and Vaudour *et al.* (2015) support that the PDO borders are not frequently representative of the terroir borders. For this reason, countries which have adopted the European GI model should concentrate on eliminating this conflict.

This study proposes zonation for the Austrian Burgenland vineyard areas. This EU member state introduced the European GI model and created PDO regions - so called Districtus Austriae Controllatus (DAC) - for the most expensive and best quality wines. Hypothesis was formed to test if the present PDO borders correspond to the geographical attributes of the natural terroir units. Partial goals were created to validate the hypothesis:

- Creation of the Burgenland wine-growing zonation map based on the physical geographical characteristics which form the natural substance of the terroir determined by multivariate statistics.
- Validation of the actual extent of the Burgenland DAC region borders with the newly formed zonation.
- Suggestions for possible changes to the DAC borders to initiate discussion on this topic and instigate negotiation with authorities and involved organizations.

## Materials and methods

### 1. Study area

In 2009 Austria divided wine into two categories based on viticultural law: wine with and without geographical labels (Weingesetz, 2009). This study concentrates on wines with geographical labels and

specifically on wines which are produced within DAC protected regions. The Austrian wine-producing area is subdivided into 16 wine-growing regions with wine geographical indication: seven with "profiling variety" and nine so called DAC regions. In the former regions are included and protected only wines made from selected, regionally typical varieties, while in the DAC regions other factors like respect of prescribed processing techniques defined by regional commission are considered. Four of these DAC regions are situated in Burgenland (Figure 1), with territory covering over 3,961.8 km<sup>2</sup> and 138.4 km<sup>2</sup> total vineyard area. Most Burgenland territory is in one of the four DAC regions: Neusiedlersee, Leithaberg, Mittelburgenland and Eisenberg. There are only two areas not in a DAC region and these surround Leithaberg DAC in the northern part of the federal state. The most common Burgenland wine varieties are Zweigeltrebe, Welschriesling, White Burgundy, Chardonnay and Lemberger.

### 2. Data

Our homogenous zones could be created by as many relevant physical geographical factors as possible. The input data was collected from all 66,673 officially registered vineyard areas in Burgenland federal state (Amt der Burgenländischen Landesregierung, 2016). Climate indicator data is from the accessible public database WorldCLIM (Hijmans *et al.*, 2005), WIND ATLAS of Austria (Krenn *et al.*, 2012) and the only soil information source is the E-BOD database which has data gathered from one-kilometer-gapped soil-probing areas (BFW, 2016). The digital terrain model enabled calculation of topographic characteristics and Table 1 shows the input data used for zonation.

### 3. Methodology

Research methodology is divided into two main steps: (1) dimensionality reduction and finding of common relationships in the multidimensional dataset by factor analysis and (2) creation of zones by cluster analysis. Validation of accuracy of proposed vineyard zonation is represented in a separate step.

#### 3.1 Factor analysis

Factor analysis was applied to reduce the multidimensional dataset in order to shrink the size of the original data with as little information loss as possible. It comes from the assumption that the relationships between the parameters are the result of a smaller number of unmeasurable dimensions that are labeled as factors. The first important step is to

**Table 1. Selected descriptive statistical indicators of vineyards (n=66,673)**

Variable	Median	Minimum	Maximum	Unit	Source/calculation
ELEV	162.504	114.570	465.618	m a.s.l.	DTM
TWI	8.112	3.740	13.134	dimensionless number	DTM / Böchner <i>et al.</i> (2001)
SLOPE	2.346	0.039	34.874	degree (°)	DTM
TPI	0.003	-8.588	12.247	dimensionless number	DTM / Weiss (2001)
EVAP	710.706	652.071	729.932	mm	WorldCLIM / Allen <i>et al.</i> (1998)
GST	15.410	13.590	15.840	°C	WorldCLIM
HI	1,785.735	1,395.398	1,867.740	dimensionless number	WorldCLIM / Huglin (1978)
GDD	1,094.850	819.650	1,151.247	°C	WorldCLIM / Amerine and Winkler (1944)
DTR	10.060	9.600	10.160	°C	WorldCLIM/ Tomasi <i>et al.</i> (2011)
CI	9.800	8.400	10.100	°C	WorldCLIM / Tonietto and Carbonneau (2004)
PERCIP	425.000	400.000	602.000	mm	WorldCLIM
GLOBIR	1,272.148	971.357	1,428.905	kWh/m <sup>2</sup>	DTM / Böchner and Antoni <sup>ć</sup> (2009)
WIND	5.000	3.500	6.000	m/s	WIND ATLAS
SOM	2.200	0.000	53.200	%	E-BOD
PH	7.127	4.000	8.700	%	E-BOD
CLAY	27.000	0.000	70.000	%	E-BOD
SILT	18.000	0.000	61.000	%	E-BOD
CA	3.300	0.000	72.600	%	E-BOD
SAND	36.000	0.000	84.000	%	E-BOD
DEPTH*	-	1	5	category	E-BOD
WATER**	-	1	14	category	E-BOD

Topographic data: ELEV – height above sea level; TWI – topographic wetness index; TPI – topographic position index. Climate data (analysis conducted on surfaces derived from the 1950-2000 WorldCLIM database (1km resolution). Data source Hijmans *et al.*, 2005); EVAP – potential evapotranspiration; GST – average temperature of atmosphere during the vegetation season; HI – Huglin index; GDD – growing degree days; DTR – diurnal temperature range (difference of temperatures between day and night); CI – cool night index; PERCIP – total rainfall during the vegetation season; GLOBIR – potential global irradiance; WIND – average speed of wind (data obtained from the 2009-2011 WIND ATLAS of Austria. Data source Krenn *et al.*, 2012). Soil and substrate data obtained from BFW (2016): SOM – percentage of organic matter; PH – soil reaction; CLAY – percentage content of clay; SILT – percentage content of silt; CA – percentage content of calcium; SAND – percentage content of sand; DEPTH\* – categorized data of depth of soils (1 most shallow – 5 most deep); WATER\*\* – categorized data of soil water content (1 most dry – 14 most wet).

standardize the data by z-score and evaluate the usability of the dataset for the factor analysis based on the Kaiser-Meyer-Olkin test (KMO) (Kaiser and Dickman, 1959). In addition, data with value larger than 0.6 was appropriate for factor analysis.

In the next step we performed factor analysis on the rest of the data. As one of the goals of the factor analysis is to reduce the input data, we performed a reduction of factors based on the Kaiser criterion. The criterion recommends to keep factors with eigenvalue larger than 1 (Kaiser and Dickman, 1959).

The last part of the factor analysis was the rotation of the factors. Most of the rotation methods are based on a simplicity function which is a function of all factor loadings. It is constructed to reach maximal or minimal values of factor loading. Rotation method selection is researcher-specific and herein we choose orthogonal rotation Varimax.

### 3.2 Cluster analysis

Regions were created by non-hierarchical k-means cluster analysis, because statistical literature advises its use for larger datasets and it is simple to apply (Kurasova *et al.*, 2014). The input dataset was divided into k-clusters so that intra-cluster similarity was as high as possible and inter-cluster similarity as low as possible. Cluster similarity is calculated on the average value of selected characteristics in each cluster. K-means clustering involves first selecting the number of objects or clusters to give parameter k, and each of these objects provides the cluster center of gravity. The remaining objects are assigned to the most similar clusters. Similarity is then calculated by the distance of the object from the cluster center of gravity. A new center of gravity of the newly formed clusters is then calculated and this process is repeated until there are no other changes to improve the relationship of inter- and intra-cluster distances. The



**Table 2. KMO test of sampling adequacy applied to input data**

Variable	KMO
ELEV	0.952982
TWI	0.773093
SLOPE	0.8865
TPI	0.396292
EVAP	0.959631
GST	0.834244
HI	0.835869
GDD	0.898026
DTR	0.904139
CI	0.903007
PERCIP	0.943935
GLOBIR	0.91991
WIND	0.911219
SOM	0.630198
PH	0.89363
CLAY	0.691821
SILT	0.728273
CA	0.799709
SAND	0.760614
WATER	0.675705
DEPTH	0.655018
Overall KMO	0.893747

KMO values lower than 0.6 indicate the variable is not adequate for factor analysis and should be removed. For this reason, TPI index was removed from further analysis. Abbreviations are explained in Table 1.

sum of the squared error is the criterion used for distance.

This criterion was set to create the most compact and separated clusters which suit the conditions of the created vineyard production zones. The k-parameter cannot be estimated from the actual status, but must be set by other methods. Internal validity indices are commonly used to specify k-parameter. No one has yet been able to determine which index is the most precise, and therefore we tested the estimation of the parameter k using five different internal validity indices. We calculated: Ball/Hall (Ball and Hall, 1965), Hartigan (Hartigan, 1975), WBI (Zhao *et al.*, 2009), Calinski-Harabasz (Calinski and Harabasz, 1974) and Xu index (Xu, 1997). We tested clustering from k=2 to k=25, with initial points of gravity in each process set to maximize initial intra-cluster distances. Internal indices were visualized on linear graphs and the optimal parameter value was determined at local maxima or minima (Zhao, 2012).

In addition to application and consequent comparison of index validation we used the input data to estimate the number of clusters by Generalized Cluster Analysis tool in Statistica 10. Automated estimation of the k parameter by this method was chosen to test the validity of k parameter selection by indices. Cluster analysis was then performed on the factor analysis results. Each of the 66,673 analyzed vineyards was subsequently matched to the identified zones.

The final step set the proximity of links between the members of each cluster. The strength of the force, the quality of cluster analysis and the statistical confidence of each object in the clusters were evaluated by multidimensional discriminant analysis (Browne and McNicholas, 2012). This analysis gives the percentage of correctly included members of each cluster using the classification matrix. The inclusion accuracy is then calculated by comparing predefined facilitation of the object to a cluster. Distances in

**Table 3. Results of factor analysis based on the KMO test**

Factor number	Eigenvalues	Percent of variance	Cumulative percentage
1	9.02806	45.140	45.140
2	1.89353	9.468	54.608
3	1.23863	6.193	60.801
4	1.15853	5.793	66.594
5	1.05963	5.298	71.892
6	0.954819	4.774	76.666
7	0.893446	4.467	81.133
8	0.849936	4.250	85.383
9	0.674782	3.374	88.757
10	0.60904	3.045	91.802
11	0.530933	2.655	94.457
12	0.334668	1.673	96.130
13	0.317736	1.589	97.719
14	0.174036	0.870	98.589
15	0.148171	0.741	99.330
16	0.0634455	0.317	99.647
17	0.0471023	0.236	99.883
18	0.0181766	0.091	99.973
19	0.00496228	0.025	99.998
20	0.000360405	0.002	100.000

According to Kaiser and Dickman (1959), factors with eigenvalues greater than 1 were selected for cluster analysis.

discrimination analysis are calculated on Mahalanobis distances in parameter correlation. The analysis is independent of parameter value range and each member is assigned to the cluster with the shortest distance to the center of gravity. This creates a classification matrix with percentage representation of accuracy in both the created cluster and the entire cluster analysis.

Finally, DAC regions and their borders were visually compared to the created homogenous zones and proposals were drafted for discussion with local/regional authorities.

## Results

The delineation of Burgenland vineyard zones based on physical geographical characteristics describing the natural terroir was achieved by detailed analysis of the relevant scientific sources in Table 1. The topographic position index (TPI) was regarded unsuitable for factor analysis and was not used after evaluation of parameter suitability (Table 2). Factor analysis was then repeated without TPI. The overall KMO test for all parameters was 0.89, thus providing appropriate initial conditions for factor analysis without TPI.

Factor analysis results enabled extraction of five factors with eigenvalue over 1 (Table 3) and factors were also extracted outside Kaiser criteria based on scientific studies delimiting natural terroir. These studies predetermined the conditions of occurrence of the common factors influencing wine production. We then applied Varimax rotation on the results of the initial factor analysis (Table 4).

The first extracted factor is characterized according to the high factor load of climate parameters with elevation as TOPOCLIMATE. The correlation coefficient for the analyzed region is above 0.96, so it is obvious that climate conditions strongly correlate with elevation. The second factor is connected with soil reaction and the silt and calcium content. Although silt is a physical characteristic of soil, we refer to this factor as SOIL CHEMISTRY because soil reaction is allied with soil calcium content and this is strongly linked to deep silty soils. The third factor is called SOIL WATERLOGGING because it reflects the change in soil water absorption and the clay particle content. The fourth factor depicts sand particle changes. While other parameters influencing the fourth factor were not above 0.5, most affected the soil's physical properties; therefore PHYSICAL PROPERTIES OF SOIL was used for this factor. The

**Table 4. Factor loadings of retained variables after Varimax rotation**

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
	Topoclimate	Soil chemistry	Soil water-logging	Physical properties of soil	Topography
ELEV	-0.957643	-0.126356	-0.0227931	-0.0617702	-0.107573
TWI	0.311146	0.0286387	-0.0517444	-0.00989252	0.820675
SLOPE	-0.612572	-0.0164482	0.0609348	-0.181047	-0.53443
EVAP	0.909105	0.0731152	-0.0306384	0.0568519	0.0734469
GST	0.974275	0.102026	0.0129797	0.0797623	0.0401743
HI	0.975718	0.109578	0.0141117	0.0720309	0.0476076
GDD	0.977622	0.110673	0.0189262	0.071202	0.0537263
DTR	0.961554	0.140967	0.0226473	0.0493315	0.0848718
CI	0.933426	0.0472726	0.00764502	0.112055	0.0201029
PERCIP	-0.896718	-0.12265	-0.0799133	-0.00698279	-0.197829
GLOBIR	-0.3916	0.151515	0.0699982	-0.475182	0.261027
SOM	0.093325	0.264664	-0.0449342	-0.424565	-0.308935
PH	0.493718	0.634691	-0.162522	0.0444839	0.0721408
CLAY	0.0495993	0.0391226	0.678125	0.395764	0.0798881
SILT	0.00643576	0.710337	0.166929	0.045122	0.0447121
CA	0.229908	0.692381	-0.116364	-0.265293	-0.0698598
SAND	0.13498	0.0744505	0.144605	0.707568	0.0214897
WATER	0.0269432	0.0255475	0.656637	0.025093	-0.0659048
WIND	0.596934	0.083237	0.152006	-0.0856553	0.297428
DEPTH	-0.0576788	0.501648	-0.58587	0.367333	0.00549108

Relationships of each variable to the retained factors expressed by the factor loadings. Bold values indicate significant association of the variable to the extracted factors. Abbreviations are explained in Table 1.

fifth factor was named TOPOGRAPHY because it was strongly linked to both the topographic wetness index and slope.

Parameter k reflects the optimal number of factors and this was set on the internal validity indices values in figure 2. The result was confirmed by V-fold cross validation in Statistica 10 general cluster analysis.

The second most common homogenous vineyard zone is zone 1 (Table 5, Figure 3) with 27.98% of the total Burgenland vineyard area. This is situated almost entirely in the Neusiedlersee and Leithaberg DACs and has very deep silty soils with neutral to slightly alkaline PH denoting the SOIL CHEMISTRY factor.

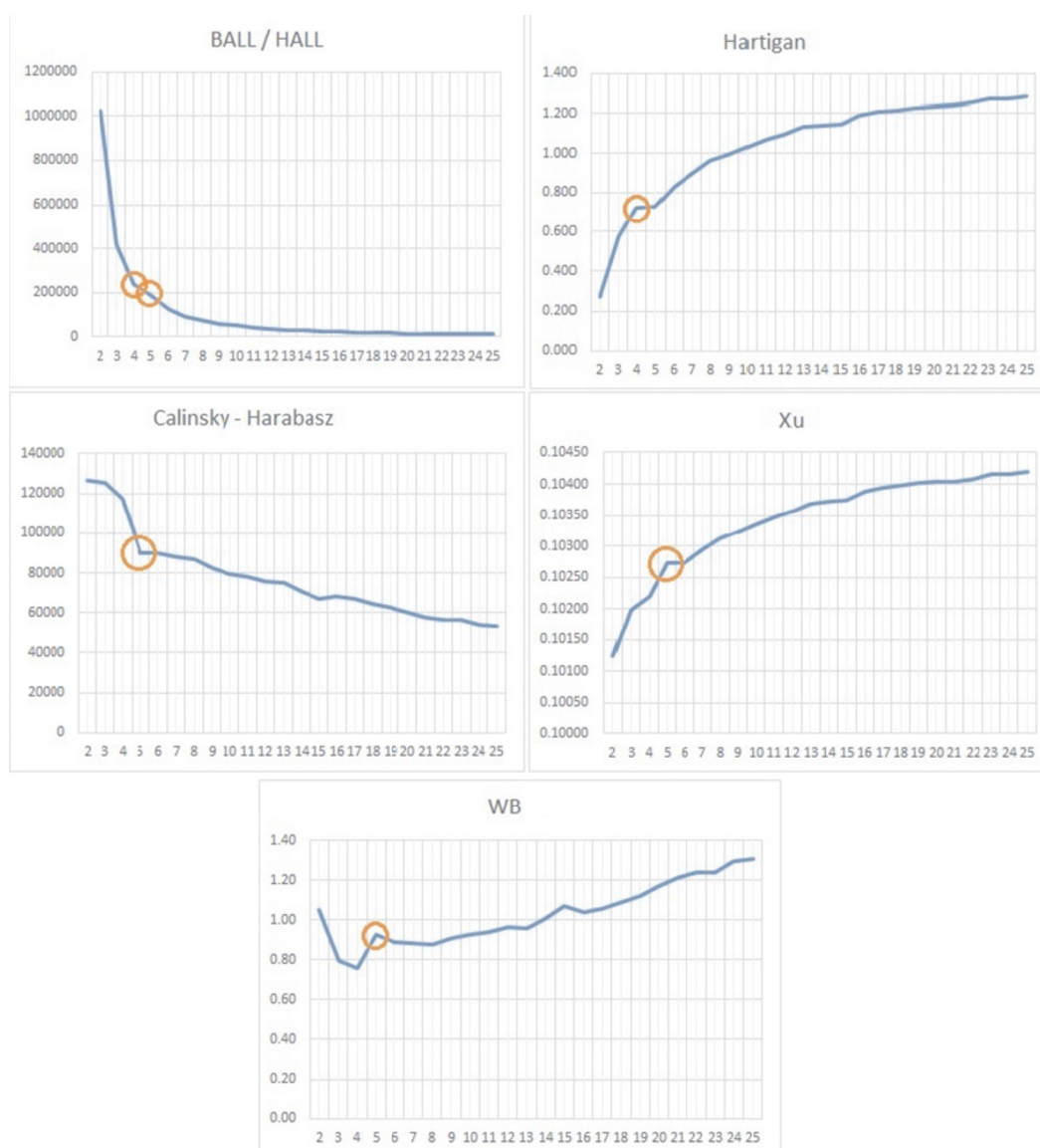
Zone 2 has the third most common type of vineyard site and covers 19.06% of the total Burgenland vineyard area. It is mostly situated in Mittelburgenland DAC in central Burgenland and comprises 77% of vineyard area. This zone differs from zones 1 and 4 through topographic and climatic factors and the vineyards are at higher altitude with higher rainfall during the growing season and also

lower atmospheric temperature; as designated under TOPOCLIMATE factor. The vineyards in this zone are predominantly on slopes; therefore also included in the TOPOGRAPHY factor. The soils in zone 2 are relatively shallower than in zones 1 and 4, thus imposing different hydrogeological conditions on this zone.

Zone 3 in Eisenberg DAC in southern Burgenland is the smallest zone. Its soils are typically shallower, acidic and with less calcium; as noted under SOIL CHEMISTRY factor. This soil also dries faster because it contains less of the clay particles prevalent in the SOIL WATERLOGGING factor.

Zone 5 comprises 41% of Leithaberg DAC vineyards and is at lower altitude, with the strong Pannonian climate influence noted in zone 4. The water level in this SOIL WATERLOGGING and the higher clay fraction differentiate it from other zones.

The 95.95% cluster analysis quality confirmed that discriminant analysis validated the correctness and accuracy of the chosen methodology, and accuracy of the classification of each zone also attained a



**Figure 2. Internal validity indices values used to estimate optimal value of the parameter k for k=2 to k=25. The red circle highlights optimal number of clusters (parameter k).**

satisfactory level (Zone 1 – 92.10%; Zone 2 – 92.31%; Zone 3 – 96.62%; Zone 4 – 89.50%; Zone 5 – 99.95%).

The homogenous zones were created for the purpose of the PDOs to approach the actual terroir borders. We propose changes to the DAC regional borders in the analyzed zones such as division of Neusiedlersee DAC into two equal zones, where zone 4 equals Neusiedlersee DAC and the second part is joined to Leithaberg DAC because it is also part of zone 1. We also propose creation of a DAC in the northern part of Burgenland. This can be considered by the regional and national committee which reviews inclusion of DAC wines and a certificate for DAC

trademark use can be issued if plans include areas currently not part of any DAC (Figure 3). In addition, while Neusiedler Lake comprises the remainder of the Leithaberg DAC which should become a separate DAC region, a different situation exists for Mittelburgenland and Eisenberg DACs where zones 2 and 3 do not cross the borders of the DAC regions currently defined in Austrian law. Therefore we recommend that these regions' borders retain their present form.

### Discussion

This research presents a specific step towards delimiting homogenous vineyard production zones



**Table 5. Mean values of used grouping indicators for each created zone**

	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5
ELEV	146.97	236.62	313.134	130.55	168.94
TWI	8.33	7.85	6.82	8.42	7.98
SLOPE	2.38	4.40	8.84	1.161	3.53
EVAP	715.39	698.80	682.74	716.45	710.52
GST	15.52	15.06	14.65	15.64	15.411
HI	1,805.72	1,708.14	1,617.46	1,828.82	1,781.81
GDD	1,110.74	1,033.70	963.94	1,127.85	1,093.07
DTR	10.084	9.96	9.83	10.11	10.05
CI	9.83	9.48	9.19	9.92	9.70
PERCIP	420.30	448.65	501.66	414.533	426.67
GLOBIR	1,277.97	1,275.52	1,308.41	1,270.14	1,270.68
SOM	2.93	2.08	2.81	2.47	1.84
PH	7.40	6.24	5.50	7.23	6.04
CLAY	16.91	26.90	18.68	29.48	43.69
SILT	23.28	19.67	15.56	21.93	18.59
CA	12.93	4.55	1.23	6.94	1.41
SAND	30.81	32.93	27.95	45.29	39.94
WATER	3.09	4.26	5.81	7.17	7.91
WIND	5.57	5.81	4.14	5.76	5.52
DEPTH	4.92	4.82	4.43	4.65	3.04

Abbreviations are explained in table 1.

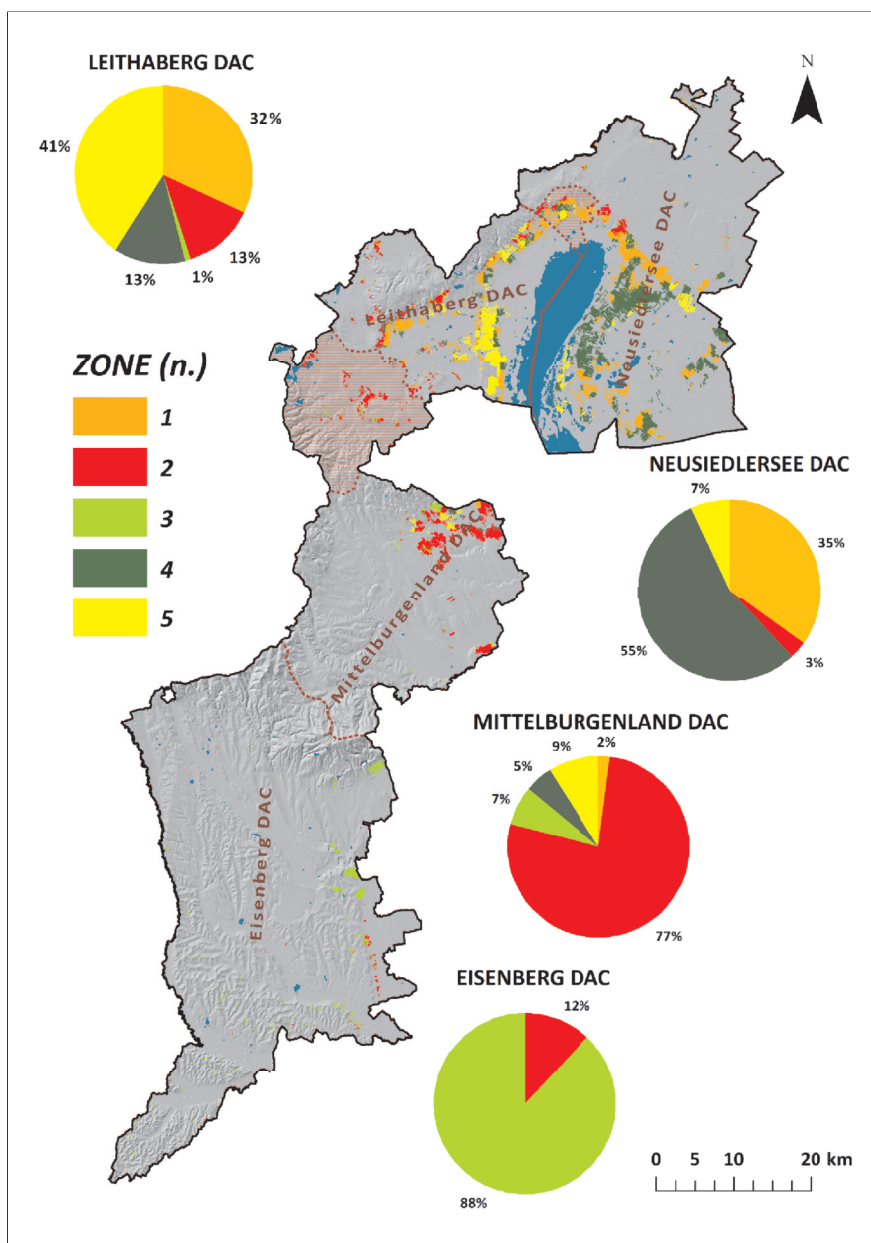
whose borders correspond with those of natural terroir. This should be implemented in each DAC region that labels products based on GIs. The cluster analysis method of identifying vineyard zones has been used in research (Herrera Nuñez *et al.*, 2011; Hugues *et al.*, 2012; Priori *et al.*, 2014), and these authors delimit homogenous zones of natural terroir units by data reduction in principal component analysis (PCA). They then selected the variables for cluster analysis from their input dataset, based on resultant factor scores.

We used factor analysis for data reduction, because factor analysis takes hidden structure of multidimensional data into account. This does not occur in PCA analysis. Factor analysis is a better tool than PCA analysis when there is assumption of knowledge of the factors which describe the input data and it is essential to capture their structure (Beavers *et al.*, 2013). When extracting our factors, we used Kaiser criterion (Kaiser and Dickman, 1959) and also the assumption that the physical and chemical factors of climate, topography and soil property all influence natural terroir. Moreover, PCA

does not search for inter-data relationships in datasets, it only explains the maximum amount of variance with the fewest number of principal components and leads to less accurate, if sometimes similar, results. We therefore consider factor analysis more efficient than PCA.

Cluster analysis of k-means based on the extracted and rotated factors employs two methods of estimating parameter k. This is a strong aspect of our research when compared with Priori *et al.* (2014). The number of our clusters was set by both the Generalized Cluster Analysis using V-fold cross validation and five different validity indices. The resultant estimation of parameter k for k-means clustering was stable, with the number of clusters close to the actual number of DAC regions in Burgenland. This justifies the assumption outlined in our proposed changes to these regions' borders.

It is important to note that our classified natural terroir zones cross not only intra-regional borders but also administrative borders; therefore DAC zone changes should also be made in neighboring Burgenland regions. We did not incorporate this



**Figure 3. Vineyard homogenous zones based on the natural terroir concept in the federal state of Burgenland.**

Short description of created zones: Zone 1 – deep silty soils with neutral to slightly alkaline pH; Zone 2 – the highest situated with shallow soils on slopes; Zone 3 – shallow, relative acid and dry soils; Zone 4 – the lowest elevation with high sand content; Zone 5 – the highest clay content. Mean values of indicators used in cluster analysis are given in Table 5.

factor in our analysis because of the following: (1) it was important to create classification independent of other model regions; (2) we cannot identify Austria and Hungary as a single unit because this would contravene the legal EU zonation purpose as a protective mechanism for products from a specific region; and (3) changes cannot be effectively applied to existing zonation.

Herrera Nuñez *et al.* (2011) used similar methodology to Priori *et al.* (2014), where cluster analysis uses the very important topoclimate indicator of potential solar radiation, but the final clustering did not incorporate soil-substrate data which is considered essential in terroir-oriented research (White *et al.*, 2007). In addition, Carey *et al.* (2008) employed overlain maps with climate, topographic,

geologic and soil input data classified according to their production potential. These authors thus achieved a detailed description of the natural terroir units which described qualitative characteristics, but this can lead to problems in the field because qualitative research alone is less effective than when combined with appropriate research based on quantitatively interpreted data. The final synthesis excludes solar radiation which has a strong influence on wine production in all global vineyard regions (Weiss *et al.*, 2003), and cluster analysis benefits include available descriptive characteristics of specific vineyard areas in the created homogenous zones. This method enables zone comparison and determination of quantitative differences which are difficult to achieve in overlain maps. The unit borders set by this method are also harder to apply when DAC regional borders change as required in our hypothesis and also to achieve the goals of our research.

In contrast, Hugues *et al.* (2012) employed hierarchical cluster analysis where vineyard zones are created on a larger scale with similar data capturing only soil conditions. This creates problems when applying hierarchical cluster methodology to the clustered data size. Calculation problems also arise when there is a large number of records and parameters because hierarchical clustering is more difficult than k-means clustering. We therefore found that applying this method to large areas creates unwanted problems.

Widely popular research concentrates on evaluating the potential of terroir vineyard sites where vineyard zonation is based on combined physical geographical attributes (Boyer and Wolf, 2000; Jones and Duff, 2011). These authors do not evaluate zone homogeneity although this actually results from vineyard site zonation and potential areas. Much of this research cannot extract information to use in GI protection; it can only evaluate the total suitability of a specific site. This type of research is more characteristic of agro-economics than market value even though it separates the most valuable zones from those less valuable or unsuitable for quality wine production.

### Conclusion

One of the major disputes concerning the GIs of products based on the terroir concept is discrepancy between regional borders where products are made and natural terroir. The borders of the physical geographical environment forming the vineyard region's terroir are often ignored when delimiting the

borders of the regions from the "table", or mainly on the basis of cultural and socio-economic characteristics. The methodological approach based on combining factor and cluster analyses proved suitable in identifying homogenous vineyard production zones. The borders of these zones can be used to delimit PDO regions and our research results suggest creation of five production zones instead of four and inclusion of areas omitted from zonation. The process of identification of PDO regions which considers terroir factors can be applied throughout Austria. It can also be used with some modifications caused by input data in other EU member states, which have accepted the European model of protection of agricultural products and especially in those where the PDO regional system is not yet developed. Austria is a perfect example because, similar to other EU member states, it has only formally adopted the European rules of protected labeling for food products. The process of identifying homogenous vineyard zones is certainly applicable to other EU member states which produce wine and have available datasets of soil characteristics. Based on the results and proposed changes, we emphasize that the strongest asset of our research is the applicability of this approach to other regions.

### Acknowledgements

This research was conducted with support from the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and Slovak Academy of Science under VEGA grants 1/0540/16 and 1/0421/16 and APVV-15-0597, Comenius University Grant No. 398/2016 and the institutional support of VUKOZ-IP-00027073.

### References

- Allen R.G., Pereira L.S., Raes D. and Smith M., 1998. Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56. FAO, Rome.
- Amerine M.A. and Winkler A.J., 1944. Composition and quality of musts and wines of California grapes. *Hilgardia*, 15, 493-675. doi:10.3733/hilg.v15n06p493
- Amt der Burgenländischen Landesregierung, 2016. Weinbaukataster, Eisenstadt: Amt der Burgenländischen Landesregierung.
- Ball G.H. and Hall D.J., 1965. Isodata, a novel method of data analysis and pattern classification. Menlo Park: Stanford Research Institute.
- Barham E., 2003. Translating terroir: the global challenge of French AOC labeling. *Journal of Rural Studies*, 19, 127-138. doi:10.1016/S0743-0167(02)00052-9

- Beavers A.S., Lounsbury J.W., Richards J.K., Huck S.W., Skolits G.J. and Esquivel S.L., 2013. Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research & Evaluation*, 18, 1-13. <http://pareonline.net/pdf/v18n6.pdf>
- BFW (Bundesforschungs und Ausbildungszentrum für Wald, Naturgefahren und Landschaft), 2016. Digitale Bodenkarte von Österreich, 1km-Raster, Wien: Institut für Waldökologie und Boden.
- Böhner J. and Antonić O., 2009. Land-surface parameters specific to topo-climatology. *Developments in Soil Science*, 33, 195-226. doi:10.1016/S0166-2481(08)00008-1
- Böhner J., Köthe R., Conrad O., Gross J., Ringeler A. and Selige T., 2001. Soil regionalisation by means of terrain analysis and process parameterisation, pp. 213-222. In: Micheli, E., Nachtergaele, F. O., Jones, R. J. A., Montanarella, L. (eds.) *Soil Classification 2001*. Luxembourg: European Soil Bureau.
- Boyer J. and Wolf T., 2000. GIS and GPS aid the exploration of viticultural potential in Virginia. *Vineyard and Winery Management*, Nov/Dec 2000, 48-54.
- Browne R. and McNicholas P., 2012. Model-based clustering, classification, and discriminant analysis of data with mixed type. *Journal of Statistical Planning and Inference*, 142, 2976-2984. doi:10.1016/j.jspi.2012.05.001
- Calinski T. and Harabasz J., 1974. A dendrite method for cluster analysis. *Communications in Statistics*, 3, 1-27.
- Carey A., 2001. Spatial characterisation of natural terroir units for viticulture in the Bottelaryberg-Simonsberg-Helderberg winegrowing area (Doctoral dissertation, Stellenbosch: Stellenbosch University).
- Carey V.A., Saayman D., Archer E., Barbeau G. and Wallace M., 2008. Viticultural terroirs in Stellenbosch South Africa. I. The identification of natural terroir units. *Journal International des Sciences de la Vigne et du Vin*, 42, 169-183. doi:10.20870/oeno-one.2008.42.4.809
- European Commission, 2006. Council Regulation (EC) No 510/2006 of 20 March 2006 on the protection of geographical indications and designations of origin for agricultural products and foodstuffs.
- Farrand B.M., 2016. Two continents, divided by deep philosophical waters? Why geographical indications pose a challenge to the completion of the TTIP. *European Journal of Risk Regulation*, 7, 269-273. doi:10.1017/S1867299X00005663
- Gangjee D., 2012. *Relocating the Law of Geographical Indications*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139030939
- Gervais D., 2015. Irreconcilable differences? The Geneva Act of the Lisbon Agreement and the Common Law. *Houston Law Review*, 53, 339-371.
- Gladstones J. and Smart R., 1997. Terroir. In: Robinson, J. (ed.) *The Oxford Companion to Wine*. Oxford: Oxford University Press.
- Hartigan J.A., 1975. *Clustering Algorithms*. New York: Wiley.
- Herrera Nuñez J., Ramazzotti S., Stagnari F. and Pisante M., 2011. A multivariate clustering approach for characterization of the Montepulciano d'Abruzzo Colline Teramane area. *American Journal of Enology and Viticulture*, 62, 239-244. doi:10.5344/ajev.2010.10008
- Hijmans R.J., Cameron S.E., Parra J.L., Jones P.G. and Jarvis A., 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965-1978. doi:10.1002/joc.1276
- Huglin P., 1978. Nouveau mode d'évaluation des possibilités héliothermiques d'un milieu viticole. *Comptes-rendus de l'Académie d'Agriculture de France*, 64, 1117-1126.
- Hugues P., McBratney A.B., Malone B.P. and Minasny B., 2012. Development of terrons for the Lower Hunter Valley wine-growing region, pp. 31-36. In: Minasny, B., Malone, B. P., McBratney, A. B. (eds.) *Digital Soil Assessments and Beyond*. London: Taylor & Francis.
- Jones G.V. and Duff A.A., 2011. The climate and landscape potential for quality wine production in the Snake River Valley AVA. Open report to the Idaho Wine Commission.
- Kaiser H.F. and Dickman K.W., 1959. Analytic determination of common factors. *American Psychologist*, 14, 425.
- Krenn A., Winkelmeier H., Cattin R., Müller S., Truhetz H., Biberacher M. and Eder T., 2012. Austrian wind atlas and wind potential analysis. 12. *Symposium Energieinnovation*, 10.
- Kurasova O., Marcinkevicius V., Medvedev V., Rapečka A. and Stefanovič P., 2014. Strategies for big data clustering, pp. 740-747. In: *Proceedings of the IEEE 26th International Conference on Tools with Artificial Intelligence*. Washington: IEEE Computer Society. doi:10.1109/ICTAI.2014.115
- OIV, 2010. Résolution VITI 333/2010. Definition of vitivinicultural "Terroir".
- Priori S., Barbetti R., L'Abate G., Bucelli P., Storchi P. and Costantini E., 2014. Natural terroir units, Siena province, Tuscany. *Journal of Maps*, 10, 466-477. doi:10.1080/17445647.2014.885853
- Tomasi D., Jones G. V., Giust M., Lovat L. and Gaiotti F., 2011. Grapevine phenology and climate change: relationships and trends in the Veneto region of Italy



- for 1964-2009. *American Journal of Enology and Viticulture*, 62, 329-339. doi:10.5344/ajev.2011.10108
- Tonietto J. and Carbonneau A., 2004. A multicriteria climatic classification system for grape growing regions worldwide. *Agricultural and Forest Meteorology*, 124, 81-97. doi:10.1016/j.agrformet.2003.06.001
- United States Patent and Trademark Office (USPTO), 2016. Trademark Manual of Examining Procedure, §1210. Available at: <http://tmep.uspto.gov> [Cit. 12.10.2016].
- Vaudour E., 2002. The quality of grapes and wine in relation to geography: notions of terroir at various scales. *Journal of Wine Research*, 13, 117-141. doi:10.1080/0957126022000017981
- Vaudour E., Costantini E., Jones G.V. and Mocali S., 2015. An overview of the recent approaches to terroir functional modelling, footprinting and zoning. *Soil*, 1, 287-312. doi:10.5194/soil-1-287-2015
- Van Caenegem W., 2015. Quality local food products – some aspects of trademarks law and GIs. Corporate Governance eJournal, Paper 35. <https://epublications.bond.edu.au/cgej/35/>
- Weingesetz, 2009. BGBl. I Nr. 111/2009: Österreichisches Bundesgesetz über den Verkehr mit Wein und Obstwein.
- Weiss A., 2001. Topographic position index and landform analysis. San Diego CA, Esri User Conference.
- Weiss S., Luth D. and Guerra B., 2003. Potential solar radiation in a Vertical Shoot Positioned (VSP) trellis at 38° N latitude. *Practical Winery & Vineyard*, 3, 16-23.
- White R., Balachandra L., Edis R. and Chen D., 2007. The soil component of terroir. *Journal International des Sciences de la Vigne et du Vin*, 41, 9-18. doi:10.20870/oeno-one.2007.41.1.860
- Xu L., 1997. Bayesian Ying-Yang machine, clustering and number of clusters. *Pattern Recognition Letters*, 18, 1167-1178. doi:10.1016/S0167-8655(97)00121-9
- Zhao Q., 2012. Cluster validity in clustering methods (Dissertation in Forestry and Natural Sciences, Joensuu: University of Eastern Finland).
- Zhao Q., Xu M. and Fränti P., 2009. Sum-of-squares based cluster validity index and significance analysis, pp. 313-322. In: *Proceedings of the 17th International Conference on Adaptive and Natural Computing Algorithms*. doi:10.1007/978-3-642-04921-7\_32.