Classification of Slovak juniper-flavoured spirit drinks

VERONIKA URÍČKOVÁ – JANA SÁDECKÁ – PAVEL MÁJEK

Summary
Synchronous fluorescence (SF) spectra measured at constant wavelength differences from 10 nm to 100 nm in the excitation wavelength range 250–350 nm and pattern-recognition methods were used for searching the natural grouping among Slovak juniper-flavoured spirit drinks. The best result was achieved using SF spectra recorded at a constant wavelength difference of 10 nm. Supervised methods were used to classify samples considering two types of classification criteria: distinguishing between drinks from different producers or distillates of different geographical indications and others. The supervised methods were based on two different data sets: the first principal components of the principal component analysis performed on the SF spectra and the SF spectra. Linear discriminant analysis was applied to the first principal components. General discriminant analysis (GDA), *k*-nearest neighbour (*k*NN) and support vector machine (SVM) were applied to SF spectra. Regarding different producers, both use of both GDA and SVM resulted in 100% correct classification. Regarding geographical indiciation, 100% correct classification was obtained using GDA.

Keywords
beverages; synchronous fluorescence; multivariate analysis

Regulation EC No 110/2008 [1] lays down rules on the definition and description of spirit drinks as well as on the protection of geographical indications. A geographical indication identifies a spirit drink as originating in the territory of a country, where a given quality or other characteristic of that spirit drink is essentially attributable to its geographical origin. According to this Regulation, juniper-flavoured spirit drinks are produced by flavouring ethyl alcohol of agricultural origin and/or grain spirit and/or grain distillate with juniper (*Juniperus communis* L. and/or *Juniperus oxicedrus* L.) berries. Some geographical indications (GI) include Spišská borovička, Slovenská borovička Juniperus, Slovenská borovička, Inovecká borovička and Liptovská borovička (Slovakia).

Identification of the geographical origin of beverages is an important issue in food chemistry. A powerful method for determining the geographical origin is multivariate analysis of the data provided by analytical instruments [2–5]. Chromatographic methods are relatively expensive, time-consuming and require highly skilled operators. Common approach is the use of a multi-elemental analysis followed by stable-isotope ratio-based methods. Recently, attention has been focused on the development of non-invasive and non-destructive instrumental techniques such as ultraviolet (UV), visible (VIS) and infrared (IR) spectroscopy [6–10].

The application of fluorescence spectroscopy to the analysis of beverages is particularly attractive due to its high sensitivity. Fluorescence spectra have allowed the classification of wines according to variety, typicality and manufacturer [11–13]. Classification of French and German wines was done using their excitation spectra [14]. The combination of absorption (UV/VIS, near IR) and fluorescence spectroscopic data demonstrated the possibility of grouping single-malt whiskies according to their geographic area of production [15]. In a previous work, we demonstrated the feasibility of synchronous fluorescence spectroscopy (SFS) to differentiate Slovak and foreign juniper-flavoured spirit drinks [16].

Large amounts of spectral data, containing useful analytical information, noise, variabilities, uncertainties and unrecognized features, are

Veronika Uríčková, Jana Sádecká, Pavel Májek, Institute of Analytical Chemistry, Faculty of Chemical and Food Technology, Slovak University of Technology, Radlinského 9, SK-812 37 Bratislava, Slovakia.

Correspondence author:
Veronika Uríčková, tel.: +421-2-59325722, fax: +421-2-52926043, e-mail: veronika.urickova@stuba.sk

usually obtained from spectroscopic instruments. Thus, pattern recognition methods are required to extract as much relevant information from spectral data as possible. Non-supervised pattern recognition methods do not require any a priori knowledge about the group structure in the data, but instead produce the grouping, i.e. clustering, itself. This type of analysis is often very useful at an early stage of an investigation and can be performed with simple visual techniques, such as hierarchical cluster analysis (HCA) or principal component analysis (PCA) [17]. When employing HCA, the original data are separated into a few general groups, each of which is further divided into still smaller groups until, finally, the individual objects themselves remain [17, 18]. PCA is usually the first step in spectroscopic data exploration. The aims of performing PCA are two. Firstly, PCA reduces the dimensions of the spectral dataset by explaining a large part of the variance using synthetic factors, principal components (PCs). Therefore, the whole range of wavelengths can be compressed into the first few PCs. Secondly, PCA performed on spectral data makes it possible to draw similarity maps of the samples and to get spectral patterns [19, 20]. With supervised pattern recognition methods, the number of groups is known in advance and representative samples of each group are available. This information is used to develop a suitable discrimination rule or discriminant function with which new, unknown samples can be assigned to one of the groups.

There is also a difference between parametric and non-parametric methods. In the parametric methods, such as linear discriminant analysis (LDA) and general discriminant analysis (GDA), a multivariate normal distribution of the data is assumed. Non-parametric methods, such as *k*-nearest neighbor (*k*NN) and support vector machine (SVM), are not based on distribution statistics [21]. LDA is concerned with determining the so-called discriminant functions as linear combinations of the descriptors, which best separate the classes according to minimization of the ratio of within-class and between-class sum of squares. LDA requires that the number of variables (wavelengths) must be smaller than the number of samples in each group. Consequently, large spectral datasets with few samples cannot be analysed using LDA. Combining LDA with PCA overcomes this problem [22]. GDA utilizes a general multivariate linear model to solve the discriminant function analysis problem. The discrimination is determined not only by the most significant wavelengths but also by all spectra [23]. In *k*NN method, the objects are classified according to the classes of the *k* closest objects. The best way of selecting *k* is by testing a set of *k* values (e.g. from 1 to 10), then, the *k* giving the lowest classification error can be selected as the optimum one [24]. SVM maps the sample data with specific kernel functions to a higher dimensional feature space to linearize the boundary and generate the optimal separating hyperplane. A number of kernels can be used in SVM models. These include linear, polynomial, radial basis function (RBF) and sigmoid [25]. The supervised methods always comprise selecting the cross-validation method. Cross-validation methods separate the calibration dataset to training and validation subsets. The former is used to build the model, the latter is used to test and validate the model. When there are not enough samples to have an independent training and validation sets, leave-one-out cross-validation (LOOCV) is the best alternative. In LOOCV, the training set itself is used to validate the model. The model is repeatedly re-fit leaving out a single sample and then used to derive a prediction for the left-out sample [26].

The aim of this study was to assess the potential of SFS and pattern-recognition methods for distinguishing between commercial samples of Slovak juniper-flavoured spirit drinks. Non-supervised pattern-recognition methods, HCA and PCA, were used for searching the natural grouping among drinks. Supervised methods were used to classify samples considering two types of classification criteria: distinguishing between drinks from different producers or distillates of different geographical indication and others. The supervised methods were based on two different data sets: (1) the first PC of the PCA performed on the SF spectra and (2) the SF spectra. LDA was applied to the first PC. GDA, *k*NN and SVM were applied to the SF spectra.

MATERIALS AND METHODS

Samples

A total of 52 commercially available samples from five Slovak producers (code S1–S5) were collected. Different products from the same producer and four (or five) bottles of the same product were sampled. Thus, sample coding included producer, product name, bottle (e.g. S1S1 meant producer S1, product S, bottle 1). All samples belonged to the "juniper-flavoured spirit drinks" category according to EEC Regulations. Distillates of different geographical indications included 33 samples (Borovička Slovenská, St. Nicolaus, Liptovský Mikuláš, Slovakia (S1S1–5); Borovička Inovecká,

Tab. 1. Samples used in the study.

Code of producer	Code of product	Number of bottles	Producer	Product name
Geographical indications				
S1	S	5	St. Nicolaus (Liptovský Mikuláš, Slovakia)	Borovička Slovenská
	I	5		Borovička Inovecká
	L	5		Liptovium borovička
S2	J	5	Old Herold (Trenčín, Slovakia)	Slovenská borovička Juniperus
	S	5		Borovička Slovenská
S4	P	4	Frucona Trade Košice (Košice, Slovakia)	Spišská borovička
S5	P	4	GAS Família (Stará Ľubovňa, Slovakia)	Original Spišská borovička*
Others				
S1	Z	5	St. Nicolaus (Liptovský Mikuláš, Slovakia)	Pravá Zbojnícka borovička Original**
S2	K	5	Old Herold (Trenčín, Slovakia)	Slovenská borovička Koniferum
S3	M	5	Imperator (Bratislava, Slovakia)	Original Slovak Juniper brandy
S5	G	4	GAS Família (Stará Ľubovňa, Slovakia)	Original Spiš borovička**

* – containing juniper berries, ** – containing juniper twig.

St. Nicolaus (S1I1–5), Liptovium borovička, St. Nicolaus (S1L1–5); Slovenská borovička Juniperus, Old Herold, Trenčín, Slovakia (S2J1–5); Borovička Slovenská, Old Herold (S2S1–5); Spišská borovička, Frucona Trade Košice, Košice, Slovakia (S4P1–4) and Original Spišská borovička containing juniper berries, GAS Família, Stará Ľubovňa, Slovakia (S5P1–4). The other 19 samples were Slovak commercial brands of "juniper-flavoured spirit drinks": Pravá Zbojnícka borovička Original containing juniper twig, St. Nicolaus (S1Z1–5); Slovenská borovička Koniferum, Old Herold (S2K1–5); Original Slovak Juniper brandy, Imperator, Bratislava, Slovakia (S3M1–5); Original Spiš borovička containing juniper twig, GAS Família (S5G1–4) (Tab. 1). The alcoholic degree ranged within 35–42% ethanol. The samples were stored at room temperature and analysed without any prior treatment.

Twelve samples (brands S4P, S5G and S5P) were detected as outlying samples and the remaining 40 samples were treated by the pattern recognition methods. The samples were divided into two groups by a method of He et al. [27, 28]. The method selects the first n-1 samples in a group of every n samples. Thus, four samples selected from each group (five samples of the same product) were assigned to the calibration set, and used to build as well as to validate the models using LOOCV. The remaining samples were assigned to the prediction set and used as 'unknown' samples in the external prediction procedure. Thus, the calibration and prediction set contained 32 samples (S1S1–4, S1I1–4, S1L1–4, S2J1–4, S2S1–4, S1Z1–4, S2K1–4 and S3M1–4) and 8 samples (S1S5, S1I5, S1L5, S2J5, S2S5, S1Z5, S2K5 and S3M5), respectively.

Fluorescence spectroscopy

Fluorescence spectra were recorded using a Perkin-Elmer LS 50 Luminescence spectrometer (Perkin-Elmer, Waltham, Massachusetts, USA) equipped with a Xenon lamp. Samples were placed in a 10 mm × 10 mm × 45 mm quartz cell. Excitation and emission slits were both set at 5 nm. SF spectra were collected by simultaneously scanning the excitation and emission monochromator in the excitation wavelength range of 200–700 nm (with an interval of 1 nm), with constant wavelength differences $\Delta\lambda$ between them. SF spectra were recorded for $\Delta\lambda$ interval from 10 nm to 100 nm, in steps of 5 nm. Fluorescence measurements were done in triplicate for each sample. The spectrometer was interfaced to a computer supplied with FL Data Manager Software (Perkin-Elmer) for spectral acquisition and data processing. Fluorescence intensities were plotted as a function of the excitation wavelength. Contour maps of SF spectra were plotted using Windows-based software OriginPro 7.5 (OriginLab, Northampton, Massachusetts, USA).

Pattern-recognition methods

HCA and PCA were used for searching the natural grouping among drinks. Agglomerative cluster analysis, where similarity extent was

measured by Manhattan (city-block) distances, and cluster aggregation based on Ward's method were used [17, 18]. PCA performed on fluorescence spectra made it possible to draw similarity maps of the samples and to get spectral patterns. Classification of objects was done by constructing similarity maps of the samples, using PCs chosen. The number of PCs was based on the eigenvalue criterion and the total variance explained [19, 20]. The spectral patterns corresponding to the PCs provided information about the characteristic peaks, which were the most discriminating for the samples observed on the similarity maps.

Supervised methods were used to classify samples considering two types of classification criteria: distinguishing between drinks from different producers (S1, S2 and S3) or distillates of different geographical indications (S1S, S1I, S1L, S2J and S2S) and other drinks (S1Z, S2K and S3M). The methods were based on two different data

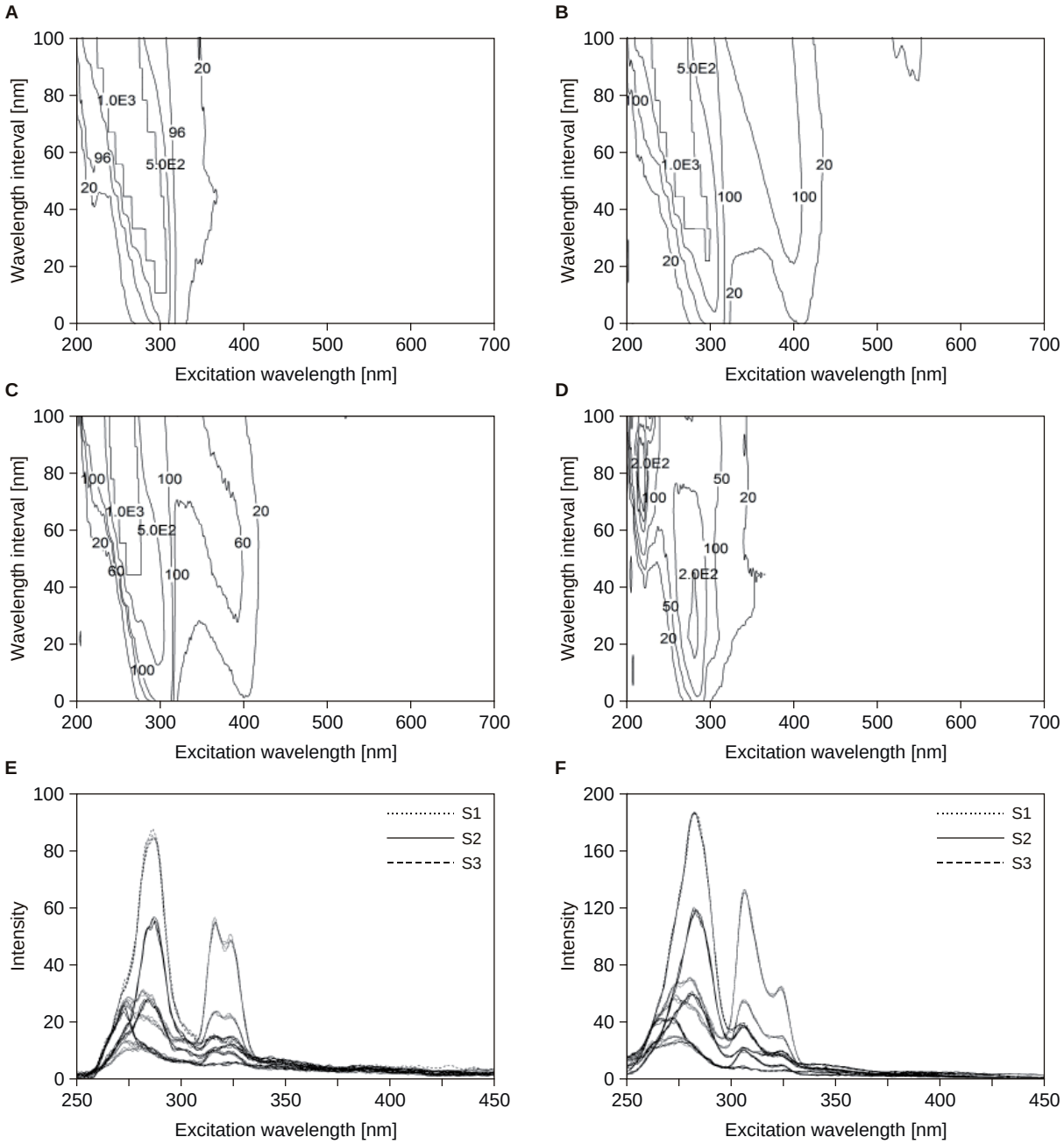

Fig. 1. Total and individual synchronous fluorescence spectra of juniper-flavoured spirit drinks.

Total fluorescence spectra: A – S4P, B – S5G, C – S5P, D – S3M, individual fluorescence spectra E – $\Delta\lambda$ = 10 nm, F – $\Delta\lambda$ = 20 nm.

sets: (1) the first PCs of the PCA performed on the SF spectra; PCA was used as a tool for dataset size and co-linearity reduction [22], and (2) the SF spectra in order to keep all the original information. LDA was applied to the first PCs, while GDA, *k*NN and SVM were applied to the SF spectra. The second approach was used when the first one did not produce a good classification. The performance of various supervised methods was evaluated and compared using a LOOCV method.

Data were converted to ASCII and processed with Microsoft Office Excel 2010 software (Microsoft, Redmond, Washington, USA), Statistica version 7.0 (StatSoft, Tulsa, Oklahoma, USA) and MATLAB Version 7.0 (MathWorks, Natick, Massachusetts, USA).

RESULTS AND DISCUSSION

Synchronous fluorescence spectra

The contour plots of total SF spectra were obtained by plotting the fluorescence intensity as a function of excitation wavelength (λ_{ex}) and wavelength interval $\Delta\lambda$. Fig. 1 (A–D) shows some of the total SF spectra. Brands S4P, S5G and S5P had an abnormally high fluorescence intensity below $\lambda_{ex} = 300$ nm (Fig. 1A–1C), which could be attributed to (bi)phenyl derivatives [29, 30]. Therefore, it was always easy to detect these products by visual inspection of the spectra. At higher excitation wavelengths, there were two less intense bands, the former with excitation in the wavelength region of 350–440 nm (S5G, S5P), and the latter with excitation at 520–600 nm (S5G), providing a much better way of differentiating between these three brands. A spectral region around 400 nm and around 600 nm were preliminarily attributed to coumarins and pigments of the chlorophyll group [31–35], respectively. It is noteworthy that the brands S4P, S5P and S5G are regarded as good quality brands coming from Eastern Slovakia. Moreover, S4P is a "pure" drink, in contrast to S5P and S5G, which contain berries and twig, respectively. All other brands (32 samples, S1, S2, S3) beyond those mentioned above showed less intense bands in the wavelength region below 350 nm. An example of this is the total SF spectrum of brand S3M (Fig. 1D).

The shape and intensity of the SF spectra depended on the difference between excitation and emission wavelengths ($\Delta\lambda$). Fig. 1 (E, F) presents the SF spectra of samples recorded at $\Delta\lambda$ of 10 nm and 20 nm.

For $\Delta\lambda$ = 10 nm (Fig. 1E), SF spectra of S1 brands showed a band at wavelengths of 260–305 nm with a maximum at about 272 nm or 282 nm, and two overlapping bands at wavelengths of 307–334 nm with maxima at 317 nm and 324 nm. SF spectra of S2 brands showed a maximum at about 272 nm or 287 nm, and a less intensive band at wavelengths of 310–330 nm with a maximum at about 324 nm. SF spectra of S3 brand showed a high intensive band with a maximum at about 287 nm, and two overlapping bands with maxima at 317 nm and 324 nm.

For $\Delta\lambda$ = 20 nm (Fig. 1F), the fluorescence intensity of all bands increased and changes in their relative intensities were noted. Fluorescence maxima were black-shifted to 270 nm, 280 nm, 306 nm and 323 nm for S1 brands; to 266 nm, 282 nm and 304 nm for S2 brands; and to 282 nm, 305 nm and

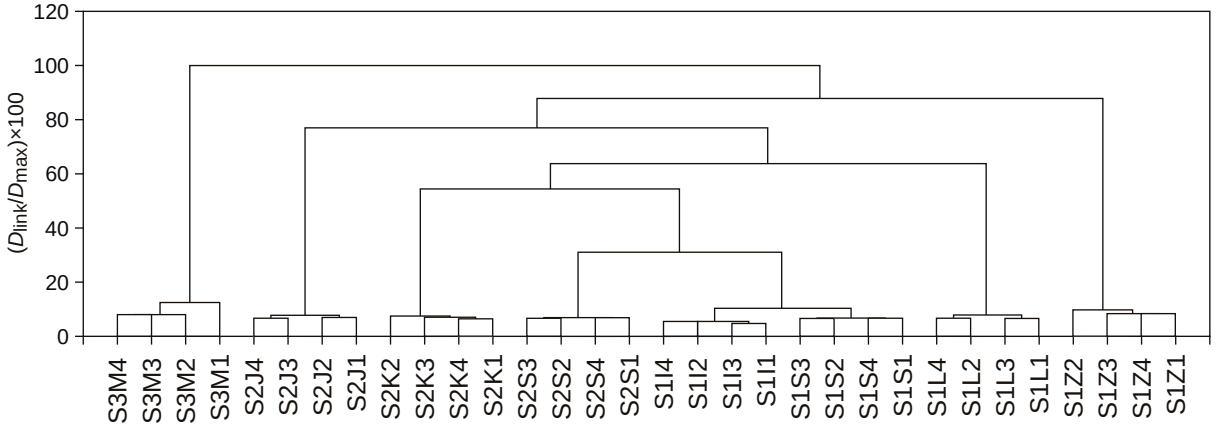

Fig. 2. Hierarchical cluster analysis dendrogram for synchronous fluorescence spectra recorded at $\Delta\lambda$ = 10 nm.

Tree diagram for 32 variables. (*Dlink*/*Dmax*) × *100* represents the percentage of the range from the maximum to the minimum Manhattan distance in the data.

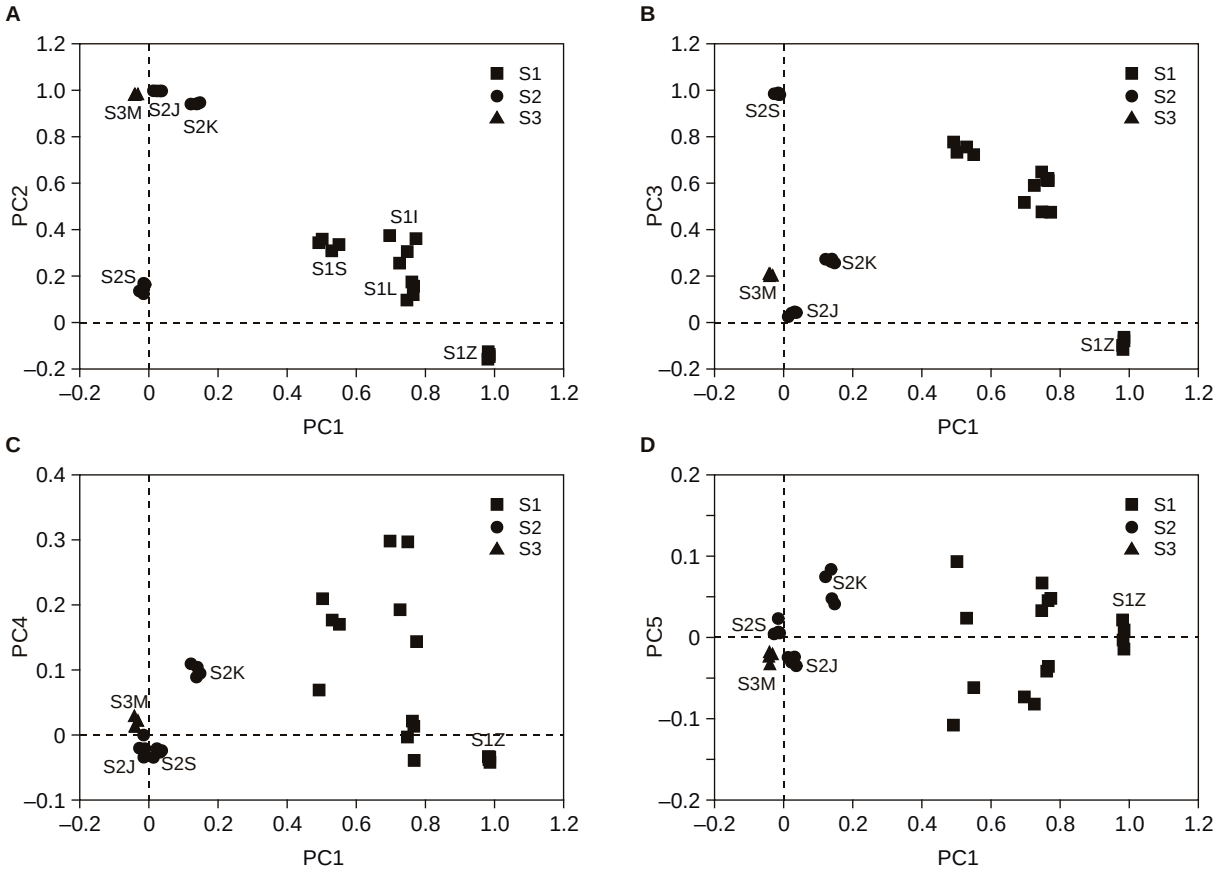

Fig. 3. Principal component analysis similarity map (score plot) for synchronous fluorescence spectra obtained at $\Delta\lambda$ = 10 nm.

A – PC1 vs PC2, B – PC1 vs PC3, C – PC1 vs PC4, D – PC1 vs PC5.

323 nm for S3 brand. A further increase of fluorescence intensity, together with band broadening, was apparent for the higher value of $\Delta\lambda$. Based on Fig. 1 (E, F), it can be suggested that SFS offers a promising approach for differentiation of drinks from different producers. Wavelength range selection was done by visual inspection of SF spectra. In general, the scatter bands are observed below 240 nm and the spectra do not present relevant information in the region above 350 nm. Therefore, the spectral range between 250 and 350 nm was used to build the pattern recognition models.

Non-supervised pattern-recognition methods

HCA performed separately on SF spectra recorded at $\Delta\lambda$ of 10–100 nm in the excitation wavelength range 250–350 nm was used for searching the natural grouping among drinks S1, S2 and S3. The best result was achieved using fluorescence spectra recorded at $\Delta\lambda$ = 10 nm (Fig. 2). At a similarity level of 48%, six clusters were found. Five of them consisted of four samples of individual brands (S3M, S2J, S2K, S1L and S1Z). One cluster included two subclusters; the first of them included four samples of S2S brand, and the second subcluster included eight samples of brands S1I and S1S.

PCA was used to examine the similarity among the S1, S2 and S3 samples. Applying PCA to SF spectra recorded at $\Delta\lambda$ = 10 nm in the excitation wavelength range of 250–350 nm, the first five PCs explained 99.0% of the total variance, where PC1, PC2, PC3, PC4 and PC5 accounted for 56.8%, 28.4%, 12.8%, 0.8% and 0.2% of the total variance, respectively. Eigen values accounted for by each principal component (PC1–PC5) were 18.2, 9.1, 4.1, 0.2 and 0.1, respectively. Fig. 3A shows the score plot of the first two PCs. PC1 promoted the separation of three groups: the group of S1 brands (high positive score), the group of S2J and S2K brands (low positive score) and the group of S2S and S3M brands (negative score). In addition, PC1 indicated the proximity between S1I and S1L brands, and PC2 indicated the proximity

between S3M, S2J and S2K. PC3, similar to PC2, roughly separated brand S1Z from other brands (Fig. 3A–3B), as S1Z had a negative score. The differentiation between samples was partially improved by including PC4 and PC5 (Fig. 3C–3D). The loading for PC1 showed the importance of the bands in the range of 280–295 nm, while PC2 corresponded to variations in the bands at 275 nm and 325 nm. PC3 related to the changes in the bands with maxima at 270 nm and 306 nm, and PC4 corresponded to the bands at 305 nm (data not shown).

Supervised pattern-recognition methods

Discrimination between the drinks from three producers

The ability of SF spectra to differentiate between the drinks from three producers was investigated by applying LDA to the first five PCs of PCA performed on the SF spectra ($\Delta\lambda$ = 10 nm, 250–350 nm). In both the calibration set and the prediction set, 100% correct classification was observed for S1 and S3 samples (Tab. 2), while only 67% of S2 samples were properly classified. S2J samples were classified as belonging to S3 group. Because this classification was unsatisfactory, we chose to discriminate the samples by GDA, *k*NN and SVM based on the SF spectra (at $\Delta\lambda$ = 10 nm) in the range of 250–350 nm with an interval of 1 nm. Using GDA, 100% correct classification was observed for S1, S2 and S3 samples (Tab. 2). Classification performed using the *k*NN algorithm gave a significantly lower rate of correct classification of S2 samples. The optimum number *k* of neighbours used to predict an unknown was determined from the highest number of total correctly classified samples obtained in cross-validation with *k* set at 1 through 10 (Tab. 3). The *k* values were chosen in the range from 1 to 10 due to the size of our sample set, which was too small for larger values of *k*. Samples S1 and S3 were correctly classified by *k*NN for *k* values in the range from 1 to 5, while only 58% of S2 samples were correctly classified in the cross-validation step. All S2S samples and one S2J sample were classified as belonging to S1 group. In the prediction step, S2S and S2J samples were again found in the S1 group. The total rate of correct classification decreased at higher *k* values. Finally, SVM with different kernel functions (linear, polynomial, RBF and sigmoid) was tested. The performance of various SVM-based algorithms was evaluated and compared using a LOOCV method. Because the results were

Tab. 2. Discrimination of juniper-flavoured spirit drinks from different Slovakian producers using various pattern recognition methods.

Producer		S1		S2		S3		Total	
		c	p	c	p	c	p	c	p
PCA-LDA	[%]	100	100	67	67	100	100	87	87
	S1	16	4	0	0	0	0	16	4
	S2	0	0	8	2	0	0	8	2
	S3	0	0	4 (S2J)	1 (S2J)	4	1	8	2
GDA	[%]	100	100	100	100	100	100	100	100
	S1	16	4	0	0	0	0	16	4
	S2	0	0	12	3	0	0	12	3
	S3	0	0	0	0	4	1	4	1
*k*NN	[%]	100	100	58	33	100	100	84	75
	S1	16	4	5 (S2S, S2J)	2 (S2S, S2J)	0	0	21	6
	S2	0	0	7	1	0	0	7	1
	S3	0	0	0	0	4	1	4	1
SVM	[%]	100	100	100	100	100	100	100	100
	S1	16	4	0	0	0	0	16	4
	S2	0	0	12	3	0	0	12	3
	S3	0	0	0	0	4	1	4	1

PCA-LDA – principal component analysis – linear discriminant analysis, GDA – general discriminant analysis, *k*NN – *k*-nearest neighbor method, SVM – support vector machine method, *c* – calibration set (results from cross-validation), *p* – prediction set, in brackets misclassified samples.

worse with non-linear kernels, it was selected to use linear kernels (Tab. 4). Thus, using SVM with linear kernels, 100% correct classification was observed for S1, S2 and S3 samples (Tab. 2). The comparison of the results showed that drinks from producer S1 and S3 were correctly classified regardless of the supervised method used. However, GDA and SVM performed better than other methods for discriminating the drinks from producer S2. It is worth mentioning that PCA-LDA and *k*NN, although extensively used in chemometrics, did not provide results as good as GDA and SVM, at least for this particular problem.

Discrimination between distillates of different geographical indications and other drinks

In order to discriminate between distillates of different geographical indications (S1S, S1I, S1L, S2J and S2S) and other drinks (S1Z, S2K and S3M), LDA was again applied to the first five PCs of PCA performed on SF spectra ($\Delta\lambda$ = 10 nm, 250–350 nm), or GDA to SF spectra ($\Delta\lambda$ = 10 nm) in the range of 250–350 nm with an interval of 1 nm. LDA based on PCs resulted in totally 87% correct classification (Tab. 5); samples S2J with geographical indication were classified incorrectly in both the calibration set and in the prediction set. Using SF spectra, GDA produced in total 100% correct classification (Tab. 5). Classification performed with the *k*NN algorithm gave a lower rate of correct classification. The optimum number *k* of neighbours, based on the highest number of

Tab. 3. Total correct classification in the cross-validation using the *k* nearest neighbour method.

k	Rate of correct classification [%]	
	Between producers	Between geographical indications and others
1	84	87
2	84	87
3	84	87
4	84	87
5	84	69
6	78	69
7	78	44
8	78	41
9	78	41
10	72	48

Tab. 4. Total correct classification in the cross-validation using the support vector machine.

Kernel function	Rate of correct classification [%]	
	Between producers S1, S2 and S3	Between geographical indications and others
Linear	100	84
Polynomial	75	87
RBF	96	79
Sigmoid	96	84

RBF – radial basis function.

Tab. 5. Discrimination between distillates of different geographical indications and other drinks using various pattern recognition methods.

Sample		GI		Other		Total	
		c	p	c	p	c	p
PCA-LDA	[%]	80	80	100	100	87	87
	GI	16	4	0	0	16	4
	Other	4 (S2J)	1 (S2J)	12	3	16	4
GDA	[%]	100	100	100	100	100	100
	GI	20	5	0	0	20	5
	Other	0	0	12	3	12	3
*k*NN	[%]	90	80	83	67	87	75
	GI	18	4	2 (S2K)	1 (S2K)	20	5
	Other	2 (S2J)	1 (S2J)	10	2	12	3
SVM	[%]	100	100	67	67	87	87
	GI	20	5	4 (S2K)	1 (S2K)	24	6
	Other	0	0	8	2	8	2

PCA-LDA – principal component analysis – linear discriminant analysis, GDA – general discriminant analysis, *k*NN – *k*-nearest neighbor method, SVM – support vector machine method, GI – geographical indications, c – calibration set (results from cross-validation), p – prediction set, in brackets misclassified samples.

total correctly classified samples obtained in cross-validation, was in the range from 1 to 4 (Tab. 3). Two S2J and two S2K samples were classified incorrectly in cross-validation, leading to totally 87% correct classification. In the prediction step, S2J and S2K samples were again misclassified (Tab. 5). The same S2J samples misclassified by *k*NN were also misclassified by PCA-LDA. The total rate of correct classification decreased at higher *k* values. The results showed that, in SVM classification, the maximum correct classification in cross-validation was achieved using polynomial kernels (Tab. 4). In this case, 100% correct classification was observed for geographical indication group and 67% for group of others, because S2K samples were again misclassified (Tab. 5). The discrimination between samples with and without geographical indication seems to be more challenging than that between producers. Only GDA resulted in totally 100% correct classification.

CONCLUSIONS

The results suggest that SFS is a promising approach for classification of Slovak juniper-flavoured spirit drinks. The SF spectra recorded at constant wavelength difference of 10 nm in the excitation wavelength range 250–350 nm were found to provide the best results, with 100% correct classification of juniper-flavoured spirit drinks of three producers using GDA or SVM. In addition, 100% correct classification was observed for distillates of different geographical indications and other drinks using GDA. The advantages of the suggested procedures are that neither sample preparation nor specifically qualified personnel are required, data acquisition is relatively simple and total time of the analysis is three minutes.

Acknowledgements

This research was supported by the Scientific Grant Agency of the Ministry of Education of Slovak Republic and the Slovak Academy of Sciences VEGA No 1/0051/13. This publication was supported by the Competence Center for SMART Technologies for Electronics and Informatics Systems and Services, ITMS 26240220072, funded by the Research and Development Operational Programme from the ERDF.

REFERENCES

1. Regulation (EC) No. 110/2008 of the European Parliament and of the Council of 15 January 2008 on the definition, description, presentation, labelling and the protection of geographical indications of spirit drinks and repealing Council Regulation (EEC) No. 1576/89. Official Journal of the European Union, *L39*, 2008, p. 51.
2. Luykx, D. M. A. M. – van Ruth, S. M.: An overview of analytical methods for determining the geographical origin of food products. Food Chemistry, *107*, 2008, pp. 897–911. DOI: 10.1016/j.foodchem.2007.09.038.
3. Saurina, J.: Characterization of wines using compositional profiles and chemometrics. Trends in Analytical Chemistry, *29*, 2010, pp. 234–245. DOI: 10.1016/j.trac.2009.11.008.
4. Drivelos, S. A. – Georgiou, C. A.: Multi-element and multi-isotope ratio analysis to determine the geographical origin of foods in the European Union. Trends in Analytical Chemistry, *40*, 2012, pp. 38–51. DOI: 10.1016/j.trac.2012.08.003.
5. Savchuk, S. A. – Vlasov, V. N. – Appolonova, S. A. – Arbuzov, V. N. – Vedenin, A. N. – Mezinov, A. B. – Grigor'yan, B. R.: Application of chromatography and spectrometry to the authentication of alcoholic beverages. Journal of Analytical Chemistry, *56,* 2001, pp. 214–231. DOI: 10.1023/A:1009446221123.
6. Azcarate, S. M. – Cantarelli, M. A. – Pellerano, R. G. – Marchevsky, E. J. – Camiña, J. M.: Classification of Argentinean Sauvignon blanc wines by UV spectroscopy and chemometric methods. Journal of Food Science, *78*, 2013, pp. 432–436. DOI: 10.1111/1750-3841.12060.
7. Shen, F. – Yang, D. T. – Ying, Y. B. – Li, B. B. – Zheng, Y. F. – Jiang, T.: Discrimination between Shaoxing wines and other Chinese rice wines by near-infrared spectroscopy and chemometrics. Food and Bioprocess Technology, *5*, 2012, pp. 786–795. DOI: 10.1007/s11947-010-0347-z.
8. Martelo-Vidal, M. J. – Dominguez-Agis, F. – Vazquez, M.: Ultraviolet/visible/near-infrared spectral analysis and chemometric tools for the discrimination of wines between subzones inside a controlled designation of origin: a case study of *Rias Baixas*. Australian Journal of Grape and Wine Research, *19*, 2013, pp. 62–67. DOI: 10.1111/ajgw.12003.
9. Cozzolino, D. – Shah, N. – Cynkar, W. – Smith, P.: Can spectroscopy geographically classify Sauvignon Blanc wines from Australia and New Zealand? Food Chemistry, *126*, 2011, pp. 673–678. DOI: 10.1016/j.foodchem.2010.11.005.
10. Riovanto, R. – Cynkar, W. U. – Berzaghi, P. – Cozzolino, D.: Discrimination between Shiraz wines from different Australian regions: The role of spectroscopy and chemometrics. Journal of Agricultural and Food Chemistry, *59*, 2011, pp. 10356–10360. DOI: 10.1021/jf202578f.
11. Airado-Rodríguez, D. – Durán-Merás, I. – Galeano-Díaz, T. – Wold, J. P.: Front-face fluorescence spectroscopy: A new tool for control in the wine industry. Journal of Food Composition and Analysis, *24*, 2011, pp. 257–264. DOI: 10.1016/j.jfca.2010.10.005.
12. Wan, Y. – Pan, F. – Shen, M.: Identification of Jiangxi wines by three-dimensional fluorescence fingerprints. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, *96*, 2012, pp. 605–610. DOI: 10.1016/j.saa.2012.07.030.

13. Yin, Ch. – Li, H. – Ding, Ch. – Wang, H.: Preliminary investigation of variety, brewery and vintage of wines using three-dimensional fluorescence spectroscopy. Food Science and Technology Research, *15*, 2009, pp. 27–38. DOI: 10.3136/fstr.15.27.
14. Dufour, E. – Letort, A. – Laguet, A. – Lebecque, A. – Serra, J. N.: Investigation of variety, typicality and vintage of French and German wines using front-face fluorescence spectroscopy. Analytica Chimica Acta, *563*, 2006, pp. 292–299. DOI: 10.1016/j.aca.2005.11.005.
15. Mignani, A. G. – Ciaccheri, L. – Gordillo, B. – Mencaglia, A. A. – González-Miret, M. L. – Heredia, F. J. – Culshaw, B.: Identifying the production region of single-malt Scotch whiskies using optical spectroscopy and pattern recognition techniques. Sensors and Actuators B: Chemical, *171–172*, 2012, pp. 458–462. DOI: 10.1016/j.snb.2012.05.011.
16. Sádecká, J. – Uríčková, V. – Hroboňová, K. – Májek, P.: Classification of Juniper-flavoured spirit drinks by multivariate analysis of spectroscopic and chromatographic data. Food Analytical Methods, *8*, 2014, pp. 58–69. DOI: 10.1007/s12161-014-9869-8.
17. Adams, M. J.: Chemometrics in analytical spectroscopy. Letchworth: Royal Society of Chemistry, 1995. ISBN 0-85404-555-4.
18. Otto, M.: Chemometrics: statistics and computer application in analytical chemistry. Weinheim: Wiley-VCH, 1999. ISBN 3-527-29628-X.
19. Clifford, M. N. – Wu, W. G. – Kirkpatrick, J. – Kuhnert, N.: Profiling the chlorogenic acids and other caffeic acid derivatives of herbal chrysanthemum by LC-MS^n. Journal of Agricultural and Food Chemistry, *55*, 2007, pp. 929-936. DOI: 10.1021/jf062314x.
20. Liang, X. – Wu, H. – Su, W.: A rapid UPLC-PAD fingerprint analysis of *Chrysanthemum morifolium* ramat combined with cemometrics methods. Food Analytical Methods, *7*, 2014, pp. 197–204. DOI: 10.1007/s12161-013-9618-4.
21. Deming, S. N. – Michotte, Y. – Massart, D. L. – Kaufman, L. – Vandeginste, B. G. M.: Chemometrics: A textbook. 1st edition. Amsterdam: Elsevier Science, 1988. ISBN 978-0-444-42660-4.
22. Yang, J. – Yang, J. Y.: Why can LDA be performed in PCA transformed space? Pattern Recognition, *36*, 2003, pp. 563–566. DOI: 10.1016/S0031-3203(02)00048-1.
23. McLachlan, G. J.: Discriminant analysis and statistical pattern recognition. New York : John Wiley and Sons, 1992, 2004. ISBN 0-471-69115-1.
24. Sun, D. W.: Infrared spectroscopy for food quality analysis and control. 1st edition. Dublin : Elsevier, 2008. ISBN 978-0-12-374136-3.
25. Vapnik, V.: The nature of statistical learning theory. 2nd edition. New York : Springer-Verlag, 1995, 2000. ISBN 0-387-98780-0.
26. Arlot, S. – Celisse, A.: A survey of cross-validation procedures for model selection. Statistics Surveys, *4*, 2010, pp. 40–79. DOI: 10.1214/09-SS054.
27. He, W. – Zhou, J. – Cheng, H. – Wang, L. – Wei, K. – Wang, W. – Li, X.: Validation of origins of tea samples using partial least squares analysis and Euclidean distance method with near-infrared spectroscopy data. Spectrochimica acta. Part A, Molecular and Biomolecular Spectroscopy, *86*, 2012, pp. 399–404. DOI: 10.1016/j.saa.2011.10.056.
28. Wang, L. – Wei, K. – Cheng, H. – He, W. – Li, X. – Gong, W.: Geographical tracing of Xihu Longjing using high performance liquid chromatography. Food Chemistry, *146*, 2014, pp. 98–103. DOI: 10.1016/j.foodchem.2013.09.043.
29. Bridges, J. W. – Creaven, P. J. – Williams, R. T.: The fluorescence of some biphenyl derivatives. Biochemical Journal, *96*, 1965, pp. 872–878.
30. Mateo, C. R. – Prieto, M. – Micol, V. – Shapiro, S. – Willalain, J.: A fluorescence study of the interaction and location of (+)-totarol, a diterpenoid bioactive molecule, in model membranes. Biochimica et Biophysica Acta, *1509*, 2000, pp. 167–175. DOI: 10.1016/S0005-2736(00)00291-1.
31. Stassi, V. – Verykokidou, E. – Loukis, A. – Harvala, C.: Polyphenolic compounds from the leaves of *Juniperus oxycedrus* L**.** subsp. *macrocarpa* (Sm.) Ball. Pharmaceutica Acta Helvetiae, *72*, 1998, pp. 311–312. DOI: 10.1016/S0031-6865(97)00037-X.
32. Nunez, Y. O. – Salabarria, I. S. – Collado, I. G. – Hernandez-Galan, R.: Sesquiterpenes from the wood of *Juniperus lucayana*. Phytochemistry, *68*, 2007, pp. 2409–2414. DOI: 10.1016/j.phytochem.2007.05.030.
33. Chen, Y. – Yang, J. – Wang, Z. – Wu, X. – Wang, F.: Scopoletine as fluorescence probe for determination of protein. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, *66*, 2007, pp. 686–690. DOI: 10.1016/j.saa.2006.04.012.
34. Tomeckova, V. – Guzy, J. – Kusnir, J. – Fodor, K. – Marekova, M. – Chavkova, Z. – Perjesi, P.: Comparison of the effects of selected chalcones, dihydrochalcones and some cyclic flavonoids on mitochondrial outer membrane determined by fluorescence spectroscopy. Journal of Biochemical and Biophysical Methods, *69*, 2006, pp. 143–150. DOI: 10.1016/j.jbbm.2006.05.004.
35. Sharma, A. – Machwe, M. K. – Murti, V. V. S.: Fluorescence spectra of biflavones. Current Science, *52*, 1983, pp. 858–861.

Received 25 February 2015; revised 6 June 2015; accepted 11 June 2015; published online 17 September 2015.