ACADEMIA ROMÂNĂ
*Revue Roumaine de Chimie*
http://web.icf.ro/rrch/

*Rev. Roum. Chim.*,
**2018**, *63*(5-6), 489-496

*Dedicated to Dr. Maria Zaharescu*
*on the occasion of her 80th anniversary*

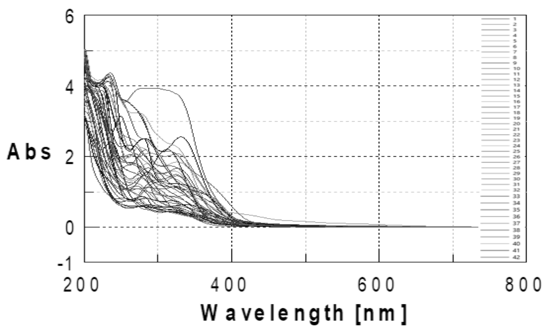
# SPECTROPHOTOMETRIC CHARACTERIZATION OF ROUMANIAN MEDICINAL HERBS ASSISTED BY ROBUST CHEMOMETRICS EXPERTISE

Ileana Maria SIMION,[a] Horia F. POP[b] and Costel SÂRBU[a]*

[a] Babeş-Bolyai University, Faculty of Chemistry and Chemical Engineering, 11 Arany Janos, RO-400028, Cluj-Napoca, Roumania
[b] Babeş-Bolyai University, Faculty of Mathematics and Computer Science, 1 Mihail Kogălniceanu, RO-400084 Cluj-Napoca, Roumania

*Received July 31, 2017*

This work describes a comparative study concerning the classical principal component analysis (PCA) and two fuzzy principal component analysis (FPCA) methods applied on digitized and normalized UV spectra of 42 Roumanian medicinal plant supplements. The hydro-alcoholic extracts were commercially acquired from specialized stores. The score plot of the samples onto the planes defined by the first principal components grouped the supplements according to their therapeutic effects. Best results were obtained for the two robust and much more compressive fuzzy methods by comparing with the classical PCA. This work indicates the feasibility and interest of chemometrics studies based on molecular spectroscopic data for the understanding of relationships between medicinal herbs and their healing properties.




## INTRODUCTION

According to World Health Organization (WHO) in many developing countries the main system used by the people to treat or to prevent diseases is traditional medicine, based on medicinal herbs, even though the modern medicine is available.[1,2] The complex chemical composition proved to be responsible for their curative properties and for this reason they have begun to be used also in the food, beverage and cosmetics industry.[3]

The main analytical methods used to evaluate herbs supplements, recommended by WHO, American Food and Drug Administration (FDA) and European Agency for Medicine (EAM) are the chromatographic and spectrophotometric methods witch lead to the chemical fingerprints of the analyzed vegetal material.[4,5] During the last decades, many experiments were conducted in order to obtain the composition of plants and their therapeutic effects.[6,7]

The huge amount of spectrophotometric and chromatographic data resulting in these studies are efficiently processed and realistic interpreted using multivariate analysis methods as Cluster Analysis (CA), Principal Component Analysis (PCA), Discriminant Analysis (DA).[4,8,9] PCA is a favorite tool in chemometrics and other fields for data compression and information extraction. PCA finds linear combinations of the original

* Corresponding author: csarbu@chem.ubbcluj.ro

measurement variables that describe the significant variations in the data. However, it is well-known that PCA, as with any other multivariate statistical method, is sensitive to outliers, missing data, and poor linear correlation between variables due to poorly distributed variables. As a result, data transformations have a large impact upon PCA. In this regard one of the most powerful approach to improve PCA appears to be the fuzzification of the matrix data, thus diminishing the influence of the outliers.[10,11]

Primary let us consider the outliers with respect to the first principal component only. In this case we will need to determine the best fuzzy membership degrees for each data item according to its distance to the first component. Here, by "best" we mean those fuzzy membership degrees such that the largest fuzzy data scatter value (corresponding to the first fuzzy principal component) is the largest (or close to that, if computationally better). By mathematical reasoning we observe that the linear model obtained by the fuzzy linear regression algorithm[11] is exactly the eigenvector corresponding to the largest eigenvalue of the fuzzy covariance/correlation matrix. We remember here that the fuzzy linear regression algorithm uses a predefined constant, $\alpha$, with values in the open interval (0,1), which pre-sets the fuzzy membership degree of the farthest outlier. Therefore, in order to find the best fuzzy weights for the first fuzzy principal component, we will loop for all $\alpha$ in (0,1) with a predefined step, let us say 0.01. At each run we will determine the fuzzy linear model and the largest eigenvalue. This eigenvalue is the fuzzy scatter value for the given value of $\alpha$. Therefore, we have determined the value of $\alpha$ that gives us the largest fuzzy scatter value among those tested values. With this value of $\alpha$ and the computed fuzzy membership degrees associated to the fuzzy linear model, we construct the fuzzy covariance/correlation matrix. This procedure has the major advantage that the first fuzzy principal component will accounts the merits of each data item according the squared distance to the fuzzy principal component. However, there is one problem with the above method. The outliers are studied with respect to the first principal component only. When computing the second principal component, instead of considering outliers with respect to it, the same fuzzy membership values as for the first component are used. Therefore we are going to use a different approach, in order to be able to determine fuzzy membership degrees related to each principal component as needed. After determining the first fuzzy principal component as it was described above, all data are projected onto the hyperplane orthogonal on it, which is a space sized one less than the size of the original space. The eigenvectors corresponding to the projected data will be orthogonal to the eigenvector determined above. This procedure has the advantage that the computation of the other fuzzy components is reduced to the computation of the first fuzzy component of a smaller-sized matrix.[12-16]

## RESULTS AND DISCUSSION

The UV spectra of the 42 medicinal plant extracts considered in this study (Table 1) are presented in Fig. 1. The digitized data obtained from spectra were normalised between 0 and 1 using the min-max normalization method**.** The min-max normalization approach, which is a linear transformation of the data, reduces the numbers of misclassification errors compared with the other normalization methods.[20] The rows from the data matrices are represented by the samples (42 medicinal plants hydro-alcoholic extracts) and the columns (230 values) by the recorded absorbances in the UV range. In Table 2 are presented the eigenvalues and the corresponding cumulative proportions of the first five PCs for all investigated methods. The digitized and normalized data obtained from the spectra were first analyzed by PCA. In this case 39 principal components explain 100 % of the total variance. When PCA analysis was performed there were taken into consideration the combinations of the first five PC's, in order to find (dis)similarities of the samples. In the score scatterplot corresponding to PC1 (64.88 % of the total variance) and PC2 (13.67 % of the total variance) appear three distinct groups including a different number of samples with similar therapeutic properties (Fig. 2). The first group includes plants with curative effects for respiratory system and/or as sedative and the smallest groups derived from the big one are differentiated by another common disease. The second group includes plants extracts with various therapeutic effects. The third group is more spread along the two PC's and includes small groups of two samples. In all representations corresponding to the combinations of the first five PC's, the samples 26 and 31 that correspond to the liquorice and quaking aspen appear as outliers in a good agreements to their spectral profiles quite different from others.

According to data in the literature, the principal bioactive compounds in quaking aspen form a group of four phenolic glicosides, amongst the most important is salicin, respective glycyrrhizin for liquorice.[21, 22] A better spreading and grouping of the samples was obtained for the PC3 (9.96 %) vs. PC5 (3.06 % ) score scatterplot applying again classical PCA approach. In this case, 6 groups are well separated and include plants extracts in good agreement with their similar medicinal properties. In addition, the samples 26 and 31 (in the other representations they appeared as extreme points as it was mentioned above) are well included in the groups with similar curative effects (Fig. 3).

*Table 1*

Scientific name and therapeutic properties of the 42 medicinal plants extracts

| No | Name | Scientific name | Therapeutic properties |
|---|---|---|---|
| 1 | Artichoke | Cynara scolymus | Cholesterol, antitoxic, diabetes, hepatitis, respiratory system |
| 2 | Blueberry | Vaccinium myrtillus | Diabetes, diuretic, internal disinfectant, antihemoragic, gout, rheumatism |
| 3 | Breckland thyme | Thymus serpyllum | Respiratory system, repealing, cough, nervous system, insomnia, anxiety[17] |
| 4 | Burdock | Arctium lappa | Cleansing, diabet, antimicrobial, hair tonic, skin, eczema, cancer[18] |
| 5 | Celery | Apium graveolens | Diuretic, astm nervous system, anxiety, diabetes, rheumatism[17] |
| 6 | Chilli pepper | Capsicum annuum | Toothache, increase blood circulation, rheumatism, muscle spasms, arthritis, antibacterial, anticancer, antioxidant[19] |
| 7 | Comfrey | Symphytum officinale | Cancer, diarrhea, cough, burns, respiratory system, antiinflammatory, healing |
| 8 | Common juniper | Juniperus communis | Diuretic, analgesic, gout, rheumatism, anorexia |
| 9 | Dill | Anethum graveolens | Digestiv/ respiratory system, blood circulation, sleeping pill, diabetes[17] |
| 10 | Echinacea | Echinacea purpurea | Flu, respiratory system, |
| 11 | Elder | Sambuctus nigra | Respiratory system, rheumatism, gout, diuretic[18] |
| 12 | Garlic | Allium sativum | Diabetes, hypertension, circulatory/digestive system, bactericidal, arterosclerosis |
| 13 | Gentian | Gentiana asclepiadea | Reduce fever, intestinal worms, anorexia, nervous system [17] |
| 14 | Ginger | Zingiber officinale | Analgesic, antiarthritis, anticancer, antidiabetic, antiulcer, anxiolytic, insect repelent, cold, mushroom poisoning, baldness, tootache[19] |
| 15 | Great celandine | Caelidonium majus | Tuberculosis, antiseptic, cardiac affections, cough, cancer, antitumor, nervous system (anxiaty) |
| 16 | Hawthorn | Crataegus monogyma | Cardiovascular system, artherial tension, insomnia, cerebral accidents, obesity, rheumatism |
| 17 | Heart's ease | Viola tricolor | Rheumatism, nervous system, respiratory system, diuretic, tonic, antiallergic |
| 18 | Hoary willowherb | Epilobium parviflorum | Hepatitis, gastritis, dygestive system, cancer[17] |
| 19 | Hogweed | Heracleum sphondylium | Afrodisiac, nervous system, sedative, digestive[18] |
| 20 | Horsetail | Equisetum arvense | Diuretic, antisweating, antiseptic, gout, rheumatism, bronchitis, gastric ulcer, cardiac affections, tooth caries |
| 21 | Lady's bedstraw | Galium verum | Gout, epilepsy, nervous system (anxiety, insomnia), diabetes |
| 22 | Lady's mantel | Alchemilla vulgaris | Diarrhea, bleadings, insomnia, dygestive system |
| 23 | Lavander | Lavandula augustifolia | Nervous sistem, antimicrobial, diuretic, heart problems, rheumatism, insomnia, |
| 24 | Lemon balm | Melissa officinalis | Bacteriostatic, antispasmodic, diabetes, sedative, antiseptic, healing, headache, anxiety |
| 25 | Lingon berry | Vaccinium vitis-idaea | Diuretic, diabet disinfectant, diarrhea, inflammatory, gout, rheumatism, „doctor" of kidney |
| 26 | Liquorice | Glycyrrhiza glabra | Laxative, inflamatory, diuretic, sedative, artritis, rheumatism |
| 27 | Milk thistle | Silybum marianum | Metal/ medicines poisoning, alcohool dependence, circulatory sistem, tension, |
| 28 | Mistletoe | Viscum album | Cardiac diseases, hypertension, astm, cough, epilepsia, diuretic, hysteria, cancer, circulatory system |

*Table 2 (continued)*

| | | | |
|---|---|---|---|
| 29 | Motherwort | Leonurus cardiaca | Nervous/respiratory system, sedative, depressive states, antiinflamatory, healing, diuretic, hypertension |
| 30 | Nettle | Urtica dioica | Diuretic, antiseptic, cough, tonic, healing, dandruff, gout, psoriasis, diabetes |
| 31 | Quaking aspen | Plopus nigra | Healing, antiinflamatory, expectorant, diuretic, antiseptic, respiratory/ digestive system |
| 32 | Ramson | Allium ursinum | Diuretic, cleansing, tonic, respiratory system, sedative, cancer |
| 33 | Rosemary | Rosmarinus officinalis | Rheumatism, anemia, fatigue, cardiovascular affections, amnesia, nervous system, astm, dandruff, diabetes |
| 34 | Sage | Salvia officinalis | Expectorant, carminative, febrifuge, tonic, antibacterial, antiinflamatory, sedative, diabetes, dental problems, |
| 35 | Saint John's wort | Hypericum perforatum | Astm, tension, intestinal worms, hair loss, sedative, antibiotic, diarrhea, rheumatism, gout |
| 36 | Shepherd's purse | Capsella bursa-pastoris | Hypotensive, analgesic, astringent, bleeding |
| 37 | Silver brich | Betula pendula | Hypercholesterolaemia, rheumatism, gout, edema cardiac, hypertension, circulatory system, |
| 38 | Spinycockle-bur | Xanthium spinosum | Inflammator, sedative, disinfectant, gout, respiratory, anxiety, system, cancer |
| 39 | Sweet flag | Acorus calamus | Antiseptic, analgesic, staomachic, anxiety, digestive system |
| 40 | Valerian | Valeriana officinalis | Sedative, nervous system, insomnia, vomiting, cardiac neurosis[17] |
| 41 | Wolf's-foot clubmoss | Lycopodium clavatum | Alcohol nicotine addiction, rheumatism, paralysis, anxiety[18] |
| 42 | Yarrow | Achillea millefolium | Dental abcesses, burns, headache (nervous sistem), scars healing, inflamatory, antiseptic, bronchodilators, anxiety[17] |

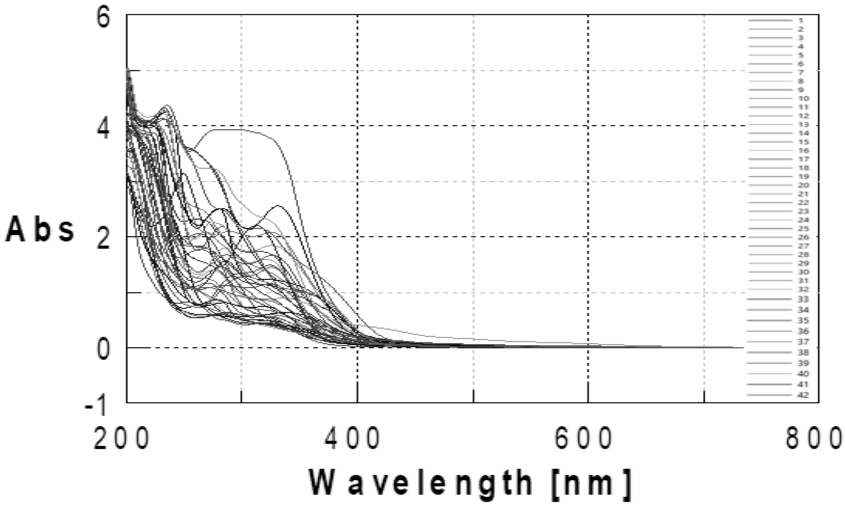



Fig. 1 – The UV-Vis spectra of 42 medicinal herbs hydro-alcoholic extracts.

*Table 2*

Eigenvalues of correlation matrix and related statistics for classical PCA, fuzzy PCA with the only first component fuzzified ($F_1$PCA) and fuzzy PCA with all components fuzzified (FoPCA)

| PC | PCA | | | $F_1$PCA | | | FoPCA | | |
|---|---|---|---|---|---|---|---|---|---|
| | Eigen-value | Prop. % | Cum. Prop. % | Eigen-value | Prop. % | Cum. Prop. % | Eigen- value | Pop. % | Cum. Prop. % |
| 1 | 14.59 | 64.89 | 64.89 | 1.695 | 84.08 | 84.08 | 1.6947 | 93.51 | 93.51 |
| 2 | 31.33 | 13.68 | 78.56 | 0.135 | 6.68 | 90.76 | 0.1005 | 5.55 | 99.06 |
| 3 | 22.82 | 9.96 | 88.53 | 0.071 | 3.54 | 94.31 | 0.0133 | 0.73 | 99.79 |
| 4 | 10.38 | 4.53 | 93.06 | 0.048 | 2.37 | 96.68 | 0.0031 | 0.17 | 99.96 |
| 5 | 7.02 | 3.06 | 96.13 | 0.032 | 1.57 | 98.25 | 0.0005 | 0.03 | 99.99 |

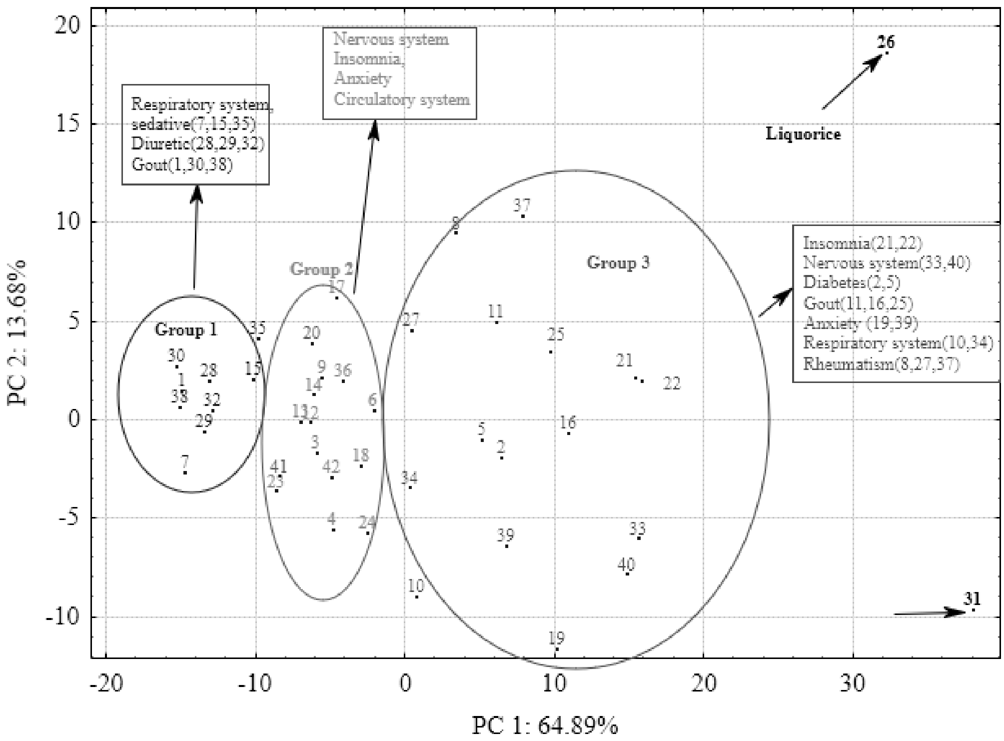



Fig. 2 – PC1-PC2 score scatterplot of 42 medicinal plants extracts.

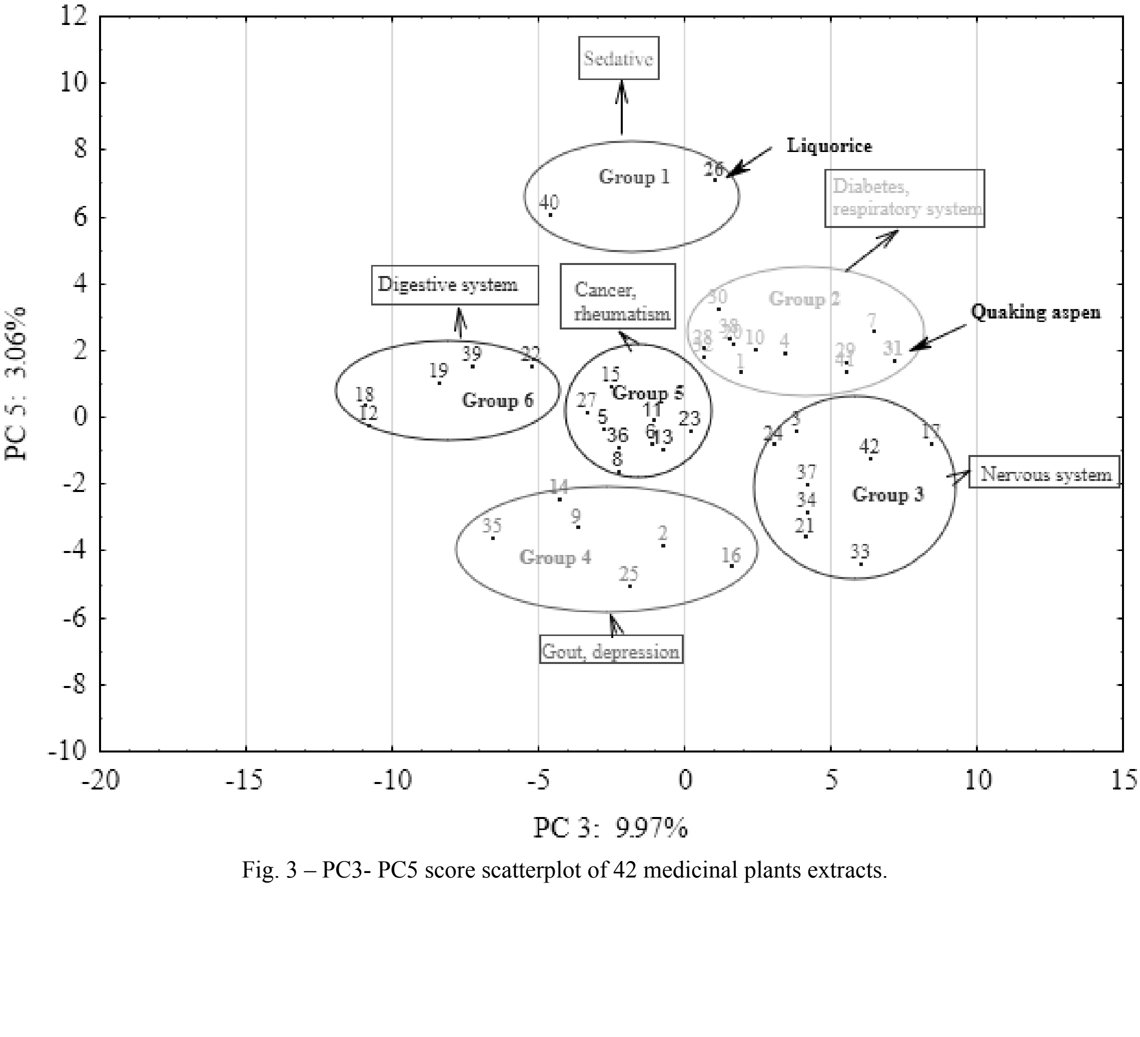



Fig. 3 – PC3- PC5 score scatterplot of 42 medicinal plants extracts.

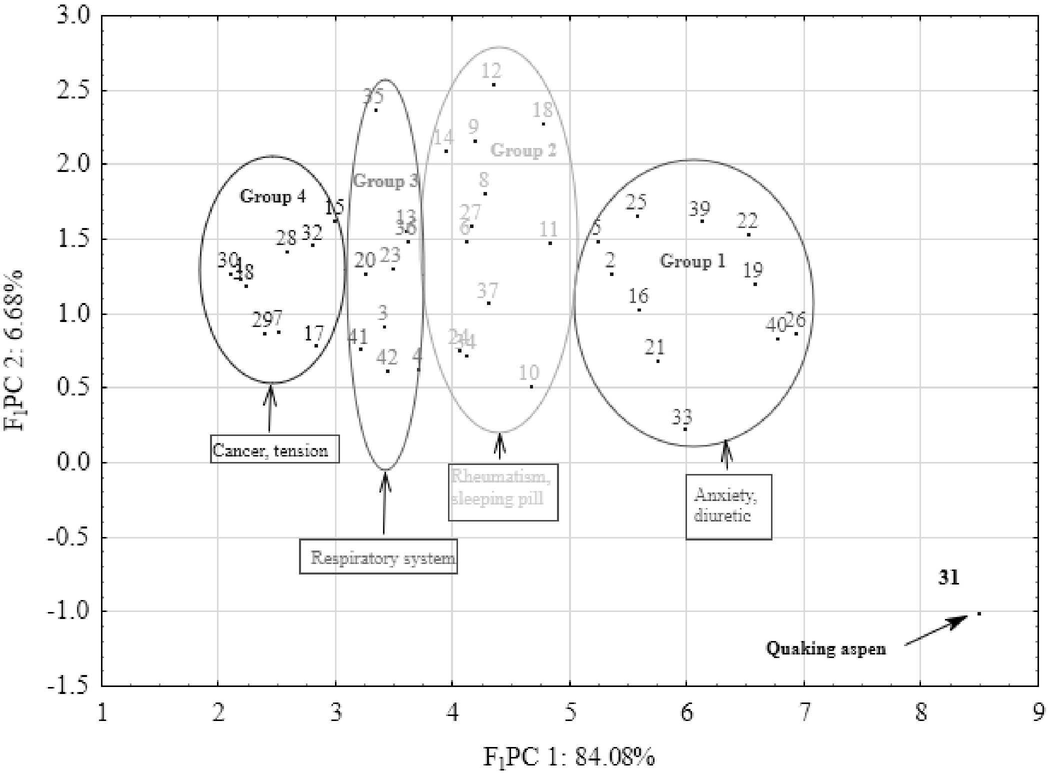



Fig. 4 – $F_1$PC1- $F_1$PC2 score scatterplot of 42 medicinal plants extracts.

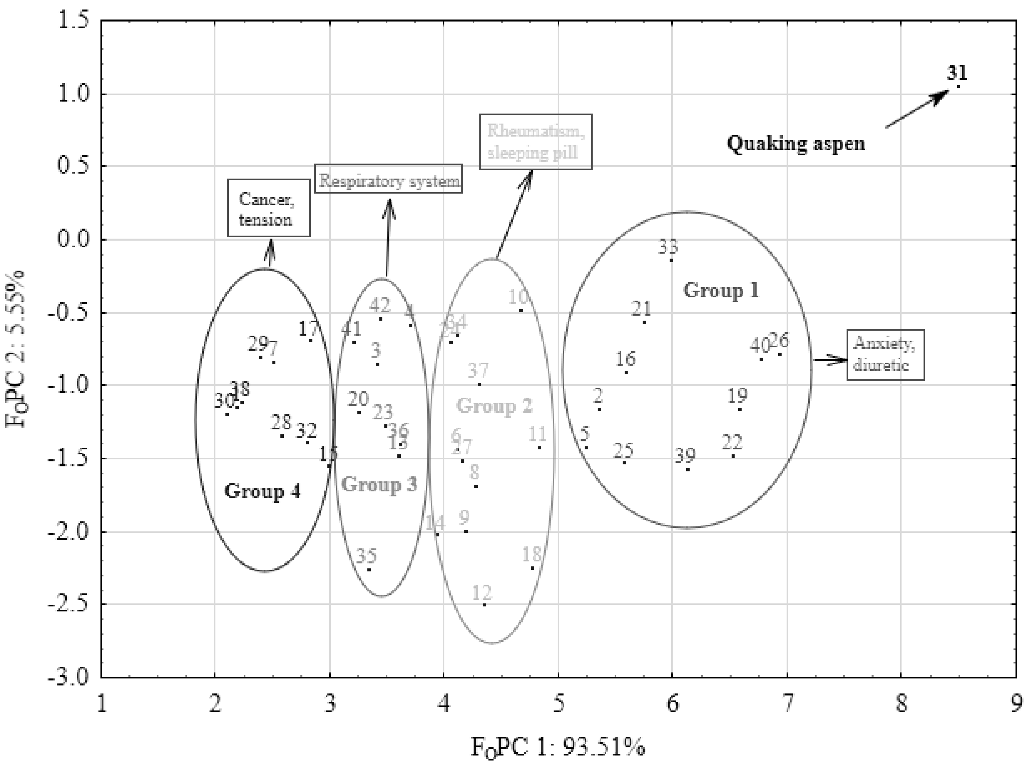



Fig. 5 – FoPC1- FoPC2 score scatterplot of 42 medicinal plants extracts.

Results obtained using $F_1$PCA algorithm described above are also shown in Table 2. The first component explains 84.08% of the total variance and the second one only 6.68%: the first 37 components accounts 100 % of the variance. Hence, the $F_1$PCA-two components, for example, account for significantly more of the variance than their classical PCA counterpart. As a direct consequence, the resultant pattern corresponding to the $F_1$PC1-$F_1$PC2 score scatterplot clearly discriminates four representative groups with similar therapeutic properties. The fourth group is more compact compared with the others indicating a more similar chemical composition excepting quaking aspen which appears again as an outlier. Considering the results obtained applying FoPCA we have to remark a quite difference from the classical PCA. We can see that, for example, the first principal component explains 93.51% of the total variance and the second one 5.55%: a two component model thus accounts for 99.06% as compared to 78.56 % for PCA and 90.76% for $F_1$PCA (Table 2): ten components account 100% of the variance. The pattern obtained by scatterplot of scores in the plane described FoPC1 and FoPC2 (Fig. 5) is more or less similar with that obtained using $F_1$PCA (Fig. 4). The main differences between the samples are underlined by the FoPC2 direction. Again, quaking aspen appears as an outlier.

## EXPERIMENTAL

### Materials and Reagents

The samples subjected to the analysis were procured from a local herbal store and consist of hydro-alcoholic extracts obtained from 42 medicinal plants. The mass/volume ratio between dried vegetal material and the extraction solvent was 1/(2-10) g/mL. A number of 41 of the samples are produced by Dacia plant manufacturer (S.C. Dacia Plant S.R.L., Brasov, Roumania) and one by Fares (Fares Orastie, Hunedoara, Romania). All the extracts are obtained from plants that are characteristic to the Roumanian flora, and the directions are specified in the prospect that accompanies the products. For the spectrophotometric measurements the initial hydroalcoholic samples were properly diluted with ethanol/water mixture (40/60 v/v). The involved solvents were ethanol (Chemical Company, Roumania) and distilled water produced in the laboratory using the Multilab GFL-2008.

### Equipments and Software

The UV spectra were registered with a UV-VIS Jasco double beam spectrophotometer, V-50 model, (Jasco Corporation) equipped with a deuterium lamp for the UV and a halogen lamp for visible domain. The slit was fixed at 0.5 nm. The 10 mm path length quartz cells were used to obtain the spectra of all the solutions. Other characteristics of the system are: the registering speed (400 nm min-1), wavelength precision (± 0.3 nm), photometric accuracy (± 0.004), and the wavelength reproducibility (± 0.1 nm). The spectra were registered in duplicate in the UV range, from 200 nm to 430 nm. The software package used for the spectra aquisition control, smoothing process, storage and spectral data digitization was Spectra Manager for Windows 95/NT version 1.53.04 (1995–2002, Jasco Corporation), further the data were processed by Statistica 8 software (StatSoft, Tulsa, USA) and SADIC (personal software package).

## CONCLUSIONS

This work indicates the feasibility and interest of chemometrics studies based on molecular spectroscopic data for the understanding of relationships between medicinal herbs and their healing properties. The robust fuzzy methods applied on the data obtained from UV spectra bring more information because the groups found are better highlighted and well-defined in a good agreement to therapeutic effects of the medicinal plant extracts. These remarkable facts (greater accounting for total variance and shaper delineation of principal components) should encourage the application of fuzzy principal component analysis methodology to other data mining efforts.

## REFERENCES


1. WHO Monographs on Selected Medicinal Plants, **1999**, *1*, 4.
2. M. Sadat-Hosseini, M. Farajpour, N. Boroomand and F. Solaimani-Sardou, *J. Ethnopharmacol.*, **2017**, *199*, 194–204.
3. D. Komes, A. Belščak-Cvitanović, D. Horžić, G. Rusak, S. Likić and M. Berendika, *Phytochem. Anal.*, **2011,** *22*, 172-180.
4. I. A. Sima and C. Sârbu, *Int. J. Food Sci. Technol*., **2016**, *51*, 1433–1440.
5. D. L. Giokas, N. C. Thanasoulias and A. G. Vlessidis, *J. Hazard. Mater.,* **2011**, *185*, 86–92.
6. D. A. Frommenwiler, Adulteration of St. John's Wort Products, CAMAG Laboratory: (http://www.camag.com/en/tlc_hptlc/camag_laboratory/methods.cfm).
7. D. Gîtea, M. Şipoş, T. Mircea and B. Paşca, *Tom.*, **2010**, *17*, 111-115.
8. C. Sârbu, R. D. Naşcu-Briciu, A. Kot-Wasik, S. Gorinstein, A. Wasik and J. Namieśnik, *Food Chem*., **2012**, *130*, 994–1002.
9. A. Medvedovici, F. Albu, R. D. Naşcu-Briciu and C. Sârbu, *Talanta*, **2014**, *119*, 524–532.
10. C. Sârbu and H. F. Pop, *Talanta*, **2001**, 40, 513-520.
11. H. F. Pop and C. Sârbu, *Anal. Chem*., **1996**, *68*, 771-778.
12. T. R. Cundari, C. Sârbu and H. F. Pop, *J. Chem. Inf. Model*., **2002**, *42*, 1363-1369.
13. C. Sârbu and H. F. Pop, *Talanta*, **2005**, *65*, 1215-1220.
14. H. F. Pop, J. W. Einax and C. Sârbu, *Chemometr. Intell. Lab. Syst*., **2009**, *97*, 25-32.
15. C. Sârbu and H. F. Pop, "Fuzzy Soft-Computing Methods and Their Applications in Chemistry" in "Reviews in Computational Chemistry", K.B. Lipkowitz, R. Larter and T. R. Cundari (Eds.), Wiley-VCH, 2004, Chapt. 5, p. 249-332.
16. C. Sârbu, "Fuzzy Clustering of Environmental Data" in "Current Applications of Chemometrics", M. Khanmohammadi (Ed.), NOVA Science Publishers, 2015, Chapt. 3, p. 33-56.

17. E. Mihăescu, “Dicționarul Plantelor de Leac”, Ed. Calin, București, 2008, p. 5-155.
18. O. Bojor, “Ghidul Plantelor Medicinale şi Aromatice de la A la Z”, Ed. Fiat Lux, București, 2003, p. 53-237.
19. K. H. Ling, C. T. Kian and T. C. Hoon, “A Guide to Medicinal Plants An Illustrated, Scientific and Medicinal Approach”, Ed. World Scientific Publishing, Singapore, 2009.
20. C. Saranya and G. Manikandan, *Int. J. Eng. Technol.*, **2013**, *5*, 2701-2704.
21. R.L. Lindroth and S.B. St.Clair, *For. Ecol. Manage.*, **2013**, *299*, 14–21.
22. G.R. Fenwick, J. Lutomski and C. Nieman, *Food Chem.,* **1990**, *38,*119-143.