



Experimental variability and data pre-processing as factors affecting the discrimination power of some chemometric approaches (PCA, CA and a new algorithm based on linear regression) applied to (+/−)ESI/MS and RPLC/UV data: Application on green tea extracts

E. Iorgulescu^a, V.A. Voicu^{b,c}, C. Sârbu^d, F. Tache^a, F. Albu^e, A. Medvedovici^{a,*}

^a University of Bucharest, Faculty of Chemistry, Department of Analytical Chemistry, Panduri Ave., no. 90, Bucharest 050663, Romania

^b Romanian Academy, Medical Science Section, Calea Victoriei no. 125, Bucharest 010071, Romania

^c University of Medicine and Pharmacy "Carol Davila", Department of Pharmacology, Toxicology and Clinical Psychopharmacology, #8 Floreasca St., Bucharest 014461, Romania

^d Babeş-Bolyai University, Faculty of Chemistry and Chemical Engineering, Department of Chemistry, Arany Janos Street, no. 11, Cluj-Napoca 400028, Romania

^e Analytical Application Laboratory, Agilrom, # 40S Th. Pallady Ave., Bucharest 032266, Romania

ARTICLE INFO

Article history:

Received 17 March 2016

Received in revised form

15 April 2016

Accepted 19 April 2016

Available online 21 April 2016

Keywords:

Multivariate methods

(+/−) ESI-MS

RPLC-UV

Influence of the experimental variability

Influence of data pre-processing

Linear regression analysis

Green teas

ABSTRACT

The influence of the experimental variability (instrumental repeatability, instrumental intermediate precision and sample preparation variability) and data pre-processing (normalization, peak alignment, background subtraction) on the discrimination power of multivariate data analysis methods (Principal Component Analysis -PCA- and Cluster Analysis -CA-) as well as a new algorithm based on linear regression was studied. Data used in the study were obtained through positive or negative ion monitoring electrospray mass spectrometry (+/−ESI/MS) and reversed phase liquid chromatography/UV spectrometric detection (RPLC/UV) applied to green tea extracts. Extractions in ethanol and heated water infusion were used as sample preparation procedures. The multivariate methods were directly applied to mass spectra and chromatograms, involving strictly a holistic comparison of shapes, without assignment of any structural identity to compounds. An alternative data interpretation based on linear regression analysis mutually applied to data series is also discussed. Slopes, intercepts and correlation coefficients produced by the linear regression analysis applied on pairs of very large experimental data series successfully retain information resulting from high frequency instrumental acquisition rates, obviously better defining the profiles being compared. Consequently, each type of sample or comparison between samples produces in the Cartesian space an ellipsoidal volume defined by the normal variation intervals of the slope, intercept and correlation coefficient. Distances between volumes graphically illustrates (dis) similarities between compared data. The instrumental intermediate precision had the major effect on the discrimination power of the multivariate data analysis methods. Mass spectra produced through ionization from liquid state in atmospheric pressure conditions of bulk complex mixtures resulting from extracted materials of natural origins provided an excellent data basis for multivariate analysis methods, equivalent to data resulting from chromatographic separations. The alternative evaluation of very large data series based on linear regression analysis produced information equivalent to results obtained through application of PCA and CA.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Extracts from tea leaves (*Camellia sinensis*) contain a large variety of compounds belonging to various chemical classes [1,2]. Different analytical investigation techniques were used for characterization of

such complex mixtures. This includes gas and liquid chromatography, as single or multi-dimensional approaches [3–6], coupled to various detection techniques for unravelling organic compounds patterns as well as atomic absorption or optical emission/mass spectrometry inductively coupled plasma techniques for determination of the multi-elemental profiles [7–9]. Electrochemistry [10], anti-oxidant activity and chelating capacity [11], olfactory properties [12,13], texture properties (through image processing) [14], high resolution melting assay [15], UV [16], NIR [17] and DART-MS [18] spectrometric

* Corresponding author.

E-mail address: avmedved@yahoo.com (A. Medvedovici).

the cumulative effects of the sample preparation, intra and interday instrumental variability made impossible the distinction between the compared data. Background subtraction of the (+)ESI/MS spectra produced an increase of the measured true distances, while for the (–)ESI/MS spectra the differences are minor. However, one can conclude that the data pre-processing modes do not add any additional discrimination power to the proposed evaluation algorithm.

4. Conclusions

Mass spectra obtained after ionization from the liquid state, under atmospheric pressure working conditions (either with positive or negative ion monitoring) of infused bulk solutions of complex mixtures (resulting from extraction of natural origin materials) are confirmed to represent an excellent data basis for making discrimination among samples, when applying chemometric algorithms, such as PCA or CA. Separation of compounds via a chromatographic method using a “classical” detection approach (such as UV spectrometry) are equivalent to mass spectra obtained from the bulk mixture, from the point of view of the chemometric algorithms. Comparison of chromatograms and mass spectra can be made with respect to the experimental shapes only, without any structural attribution of the compounds in the mixtures (holistic approaches). Grouping of the samples according to different chemometric approaches may be influenced by the sample preparation algorithm or the analytical investigational technique.

Instrumental repeatability and sample preparation induced variability are affecting to a less extent the chemometric algorithms used for discrimination between samples, compared to the instrumental intermediate precision. Mass spectrometric data are more influenced by the instrumental intermediate precision than chromatographic data. Variability issues should be considered and included in the experimental design of the analytical stages, to better assess and validate the dissimilarities between compared samples. It is recommended that data acquisition should be made in a single experimental session. MS experiments have the advantage to be performed faster compared to LC approaches.

An alternative data processing algorithm for discriminating against the samples was proposed. This algorithm is based on the linear regression model applied mutually to experimental data series. Each sample or comparison between two samples is thus represented as an ellipsoidal volume in the 3D Cartesian space, having the slope, the intercept and the correlation coefficient plotted on the three coordinates. These volumes are determined by the normal variation intervals of the slopes, intercepts and correlation coefficients. In fact, these normal variation intervals are retrieving the information contained in very large data series (better reproducing the experimental profiles being acquired), compared to the limited ability of the chemometric PCA and CA approaches to accept data series higher than 1000 items. The proposed algorithm does not require a specialized statistic software.

Data pre-processing options, as normalization, peak alignment and baseline subtraction modes, are increasing the absolute distances in the principal components space, but not necessarily increase the discrimination power of the PCA and CA algorithms. For the LRA algorithm, normalization reduces the absolute measured distances, while background subtraction produced similar or enhanced distances. However, data pre-processing modes are not certainly improving the discrimination power of the LRA approach.

The new algorithm based on Linear Regression illustrates the way the holistic evaluations may contribute to characterization of very complex mixtures of natural origins. This may represent a straightforward issue for evaluation of the quality characteristics

of complex mixtures, with a potential impact on the dietary supplements and functional food markets, representing a fast and relatively cheap way for the quality control & evaluation over production batches.

Acknowledgements

This work was possible with the financial support offered by the Romanian Ministry of Education and Research, through research grants PN-II-ID-PCE-2011-3-152 and PN-II-ID-PCE-2011-3-0366.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.talanta.2016.04.042>.

References

- [1] Tea in Health and Diseases Prevention, in: V.R. Preedy (Ed.), Elsevier Inc. Academic Press, London, 2013.
- [2] J. Zhao, J.W. Deng, Y.W. Chen, S.P. Li, Advanced phytochemical analysis of herbal tea in China, *J. Chromatogr. A* 1313 (2013) 2–23.
- [3] C.T. Scoparo, L.M. de Souza, N. Dartora, G.L. Sasaki, P.A.J. Gorin, M. Iacomini, Analysis of Camellia sinensis green and black teas via ultra high performance liquid chromatography assisted by liquid–liquid partition and two-dimensional liquid chromatography (size exclusion \times reversed phase), *J. Chromatogr. A* 1222 (2012) 29–37.
- [4] L. Zhang, Z. Zeng, C. Zhao, H. Kong, X. Lu, G. Xu, A comparative study of volatile components in green, oolong and black teas by using comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry and multivariate data analysis, *J. Chromatogr. A* 1313 (2013) 245–252.
- [5] L.F. Wang, J.Y. Lee, J.O. Chung, J.H. Baik, S. So, S.K. Park, Discrimination of teas with different degrees of fermentation by SPME–GC analysis of the characteristic volatile flavour compounds, *Food Chem.* 109 (2008) 196–206.
- [6] Y. Zhao, P. Chen, L. Lin, J.M. Harnly, L. Yu, Z. Li, Tentative identification, quantitation, and principal component analysis of green pu-erh, green, and white teas using UPLC/DAD/MS, *Food Chem.* 126 (2011) 1269–1277.
- [7] A. Szymczycha-Madeja, M. Welna, P. Pohl, Determination of essential and non-essential elements in green and black teas by FAAS and ICP OES simplified – multivariate classification of different tea products, *Microchem. J.* 121 (2015) 122–129.
- [8] A. Szymczycha-Madeja, M. Welna, P. Pohl, Elemental analysis of teas and their infusions by spectrometric methods, *Trends Anal. Chem.* 35 (2012) 165–181.
- [9] P.H.G.D. Diniz, M.F. Pistonesi, M.B. Alvarez, B.S.F. Band, M.C.U. de Araujo, Simplified tea classification based on a reduced chemical composition profile via successive projections algorithm linear discriminant analysis (SPA-LDA), *J. Food Comp. Anal.* 39 (2015) 103–110.
- [10] C. Bardpho, P. Rattanarat, W. Siangproh, O. Chailapakul, Ultra-high performance liquid chromatographic determination of antioxidants in teas using ink jet-printed graphene-polyaniline electrode, *Talanta* (2015), <http://dx.doi.org/10.1016/j.talanta.2015.05.020>.
- [11] E. Damiani, T. Bacchetti, L. Padella, L. Tiano, P. Carloni, Antioxidant activity of different white teas: comparison of hot and cold tea infusions, *J. Food Comp. Anal.* 33 (2014) 59–66.
- [12] B. Tudu, S. Ghosh, A.K. Bag, D. Ghosh, N. Bhattacharyya, R. Bandyopadhyay, Incremental FCM technique for black tea quality evaluation using an electronic nose, *Fuz. Inf. Eng.* 7 (2015) 275–289.
- [13] V. Kraujalyte, E. Pelvan, C. Alasalvar, Volatile compounds and sensory characteristics of various instant teas produced from black tea, *Food Chem.* 194 (2016) 864–872.
- [14] Z. Tang, Y. Su, M.J. Er, F. Qi, L. Zhang, J. Zhou, A local binary pattern based texture descriptors for classification of tea leaves, *Neurocomputing* 168 (2015) 1011–1023.
- [15] A. Xanthopoulou, I. Ganopoulos, A. Kalivas, M. Osathanunkul, P. Chatzopoulou, A. Tsaftaris, P. Madesis, Multiplex HRM analysis as a tool for rapid molecular authentication of nine herbal teas, *Food Control* 60 (2016) 113–116.
- [16] P.H.G.D. Diniz, M.F. Barbosa, K.D.T. de Melo Milanez, M.F. Pistonesi, M.C.U. de Araujo, Using UV–vis spectroscopy for simultaneous geographical and varietal classification of tea infusions simulating a home-made tea cup, *Food Chem.* 192 (2016) 374–379.
- [17] Q. Chen, J. Zhao, C.H. Fang, D. Wang, Feasibility study on identification of green, black and Oolong teas using near-infrared reflectance spectroscopy based on support vector machine (SVM), *Spectrochim. Acta A* 66 (2007) 568–574.
- [18] K. Fraser, G.A. Lane, D.E. Otter, S.J. Harrison, S.Y. Quek, Y. Hemar, S. Rasmussen, Monitoring tea fermentation/manufacturing by direct analysis in real time