doi: 10.17221/284/2014-CJFS

Chemometric Classification of Citrus Juices of Moroccan Cultivars by Infrared Spectroscopy

Aziz HIRRI[1], Michele DE LUCA[2], Giuseppina IOELE[2], Abdessamad BALOUKI[1], El Mahfoud El BASSBASI[1], Fouzia KZAIBER[1], Abdelkhalek OUSSAMA[1] and Gaetano RAGNO[2]

[1]Laboratory of Applied and Environmental Chemical Spectroscopy (LSCAE), Faculty of Sciences and Techniques, University Sultan Moulay Slimane, Beni-Mallat, Morocco; [2]Department of Pharmacy, Health and Nutritional Sciences, University of Calabria, Rende (CS), Italy

Abstract

Hirri A., De Luca M., Ioele G., Balouki A., Bassbasi M., Kzaiber F., Oussama A., Ragno G. (2015): **Chemometric classification of citrus juices of Moroccan cultivars by infrared spectroscopy**. Czech J. Food Sci., 33: 137–142.

Fourier transform – infrared (FTIR) spectroscopy in connection with chemometric methodologies was used as a fast and direct analytical approach to classify citrus cultivars by the measurements on their juice. Modern multivariate analysis responds to the current needs for economic, simple, and fast methods able to classify new unknown samples with great accuracy. A set of 135 samples of citrus juice, representative of three cultivars (Hamlin, Muska, and Valencia), all picked in the same geographical area of Morocco, were analysed. Chemometric discrimination of the juice samples was achieved by principal component analysis (PCA) performed on the FTIR spectral data from the juice samples, showing an explained variance of 97.84% by considering only 2 PCs. A fully correct classification of the three Moroccan cultivars was then obtained by using Partial Least Square-Discriminant Analysis (PLS-DA) modelling procedure.

Keywords: Fourier transform infrared; Multivariate analysis; cultivar classification; Principal Component Analysis; Partial Least Squares-Discriminant Analysis

Citrus fruits are grown throughout the world and are highly appreciated for their juice and for the numerous benefits for human health. Many therapeutic properties have been attributed to citrus fruits, such as antiviral, anticancer, anti-inflammatory activities (Ejaz *et al*. 2006; Codoñer-Franch & Valls-Bellés 2010). Beneficial effects on cardiovascular diseases have been also described (Mulvihill & Huff 2012). The high amounts of available bioactive compounds in citrus fruits, such as flavonoids, carotenoids, vitamins, and minerals are considered responsible for the numerous health benefits. Moreover, these fruits are a primary source of vitamin C (Martí *et al* 2009).

It is well known that chemical composition of the citrus juice varies with the cultivars (Fallahi *et al*. 1990; Bade *et al*. 1991; Bermejo *et al*. 2011, 2012). The chemical composition of the Citrus fruits is the result of the combined influence of genetic mechanisms and chemical, biological, and environmental factors (Gattuso *et al* 2007; Ebere Okwu 2008). Nowadays, the identification of the citrus cultivars is becoming a topic of technological and economic interest. The requirements for citrus juice are increasing and the commercial interest in high quality products is growing (Lanfranchi 2012; Zhao *et al*. 2012; Sidana *et al*. 2013). Moreover, the proper cultivar identification can be a useful tool for the growers who need to certify and guarantee their plant products.

The high biodiversity of citrus requires appropriate analytical methods for the correct discrimination of different cultivars. Until recently, the citrus cultivars have been mainly discriminated and classified on the basis of morphological and agronomic characteristics (Koehler-Santos *et al*. 2003). The identification based on the analysis of the inter-simple sequence repeat (ISSR) markers has also been cited (Fang &

Supported by the M.U.R.S.T. (Italy) and by European Commission, FSE fund and Calabria Region.

doi: 10.17221/284/2014-CJFS

Roose 1997). The greater part of analysis known for the identification of cultivars is often time and money consuming and involves a considerable sample pre-treatment, in which complex chemical procedures and sophisticated instruments are required.

Recently, Fourier transform infrared (FT-IR) spectroscopy has been well-accepted as an emerging analytical technique with advantages from the point of view of global cost per sample. This technique allows a rapid analysis and needs little or no sample pre-treatment. FT-IR has been widely used for the analysis of food and pharmaceuticals in both academic and industrial sectors (Galtier *et al*. 2008; Wu *et al*. 2008; Hennessy *et al*. 2009; Blanco *et al*. 2010; Terouzi *et al*. 2011; De Luca *et al*. 2012; Bassbassi *et al*. 2014). Some attempts at discriminating different citrus varieties by using the FTIR technique have been proposed (Zhou *et al*. 2012; Zhao *et al* 2014).

In literature, it is possible to find many research works where FTIR data have been processed by chemometric approaches in order to define the methods of classification or characterisation via multivariate data modelling. This methodology has shown to be very useful, because of the capability in extracting specific and/or unspecific information from FTIR signals (Zagonel *et al*. 2004).

In this research, we investigated a new method to classify different varieties of citrus according to the analytical information obtained from their juices by means of Attenuated Total Reflection coupled with Fourier Transform Infrared Spectroscopy (ATR-FTIR). Recently, the ATR technique has renovated the procedures in the analysis of solid and liquid samples. It greatly simplifies the sample pre-treatment and allows satisfactory spectral reproducibility. The details about this argument have been described in literature (Mirabella 1993; Hind *et al*. 2001; Kazarian & Chan 2006). The application of ATR has been recently reported for the analysis of food matrices, correlated with rich information on the use of this technique and its great potential in various research fields (Anjos *et al*. 2015; Jaiswal *et al*. 2015).

The goal of this work was the chemometric classification of three Moroccan citrus cultivars by ATR FT-IR data from juice samples. This approach could represent a real novelty in the chemical characterisation of the citrus. Spectroscopic data were collected from different citrus cultivars: Hamlin, Muska, and Valencia, all from the Moroccan region of Beni-Mellal. The first spectra exploration, after appropriate data pre-processing procedure, was performed by PCA, allowing a rapid and simple visualisation of the cultivar samples in three classes. The supervised method PLS-DA using the FTIR data was then applied to achieve the chemometric classification of the unknown samples.

Practical application. The identification of citrus cultivars is a topic of the food sector, both technological and economic. The great demand for the commercial interest in citrus juice are focused on the level of quality and value for money products. This research presents a chemometric discrimination of some Moroccan citrus cultivars by using the ATR FTIR data obtained out from direct analysis of the juice samples. This analytical procedure can depict an effective innovation in the chemical characterisation and classification of the citrus.

MATERIAL AND METHODS

Citrus samples. The experiments were carried out on the samples of citrus fruit of three Moroccan cultivars: Hamlin, Muska, and Valencia. All trees were cultivated in the same experimental field (Ouled Gnaou, Beni Mellal, Morocco). With each of the three cultivars, the samples were harvested from 45 different trees and from the four orientations (North (N), South (S), East (E), and West (W)). The investigation was focused on 135 samples of citrus juice: 39 samples on j-Hamlin, 48 samples of j-Muska, and 48 samples of j-Valencia.

Fourier transform mid-infrared (FTIR) analysis. Infrared spectra were acquired on a Vector 22 Bruker Fourier transform infrared (FTIR) spectrometer (Bruker, Casablanca, Morocco) equipped with a deuterated triglycine sulphate (DTGS). The measurements were performed using a single-reflection diamond attenuated total reflection (ATR) accessory with a sample pressure applicator (DuraSamplIR, Vitry-sur-Seine, France).

ATR measures the changes that occur in an infrared beam when it comes into contact with a sample. The IR beam is directed onto an optically dense crystal with a high refractive index and uses the internal reflection resulting in an evanescent wave that extends beyond the surface of the crystal into the sample held in contact with the crystal. This evanescent wave penetrates only a few microns (0.5–5 μ) into the sample surface and will be attenuated or altered when the sample absorbs energy. This energy variation is recorded for all the wavelengths, thus providing so the corresponding infrared spectrum. The crystal of the ATR accessory has a high refractive index, generally greater than those of most of the solid and liquid samples.

doi: 10.17221/284/2014-CJFS

Analysis conditions. Juice citrus samples were deposited on the ATR cell equipped with a diamond crystal prism (monoreflexion). Mid-Infrared spectra were recorded between 4000 and 700 cm^{-1}. 98 scans were co-added with a nominal resolution of 4 cm^{-1}. Air was taken as reference for the background spectrum before the collection of each sample spectrum. Before each analysis, the ATR plate was cleaned in situ by scrubbing it with ethanol solution, enabling the ATR to dry. Cleanliness was verified by collecting the background spectrum and comparing it with the previous background spectra in order to evaluate the instrumental conditions and laboratory interferences from H_2O and CO_2.

Chemometric treatment of FTIR spectra. With the aim to obtain more information from the FTIR analysis, the spectra were firstly subjected to the data pre-processing (De Luca *et al.* 2014). Figure 1 shows the FTIR mean spectra of the three citrus cultivars. In consideration of the high signal contribution due to H_2O and CO_2 in the juice samples, the spectral regions 4000–3030, 1900–1500, and 2400–2200 cm^{-1} were respectively removed because they are rich in instrumental noise and bring useless information (Iñón *et al.* 2003).

On the native data, Multiplicative Scatter Correction (MSC), Standard Normal Variate (SNV), and Derivative data pre-treatments were applied (Iñón *et al.* 2003; Maggio *et al.* 2009). The classification algorithm PLS-DA (Kowalski 1984) performs a PCA in both X and Y matrices. Therefore, each PC is extracted from the independent variables and correlated with the variance of the dependent variable (Wold & Sjöström 1998). PCA allows to reduce the analytical data to a few latent variables (PCs) that represent the main information from the original data. The first PCs carry most information whereas, after a number of PCs, the variance can be discarded because of the noise.

PLS-DA was performed with the purpose of creating a classification model able to determine the cultivar of new samples. During the calibration process, the PLS-DA method was trained to compute the three "membership values" (Y variables) relative to each cultivar.

The spectral data matrix, consisting of 135 spectra, was randomly divided into two groups: the first one, including 108 samples (30 samples for j-Hamlin, 39 samples for j-Muska, and 39 samples for j-Valencia), used for the modelling step, and the second one, including 27 samples (9 samples for each cultivar class), used for external validation procedure. The calibration model was validated by applying the leave-1-out cross as internal validation procedure. The assessment of errors in calibration was estimated by calculating the standard error of cross validation (RMSECV) for comparison with the actual values, calculated for each cultivar. The acceptable models had to have a low RMSECV and a high coefficient of correlation R^2 and good classification of the prediction samples.

Software for FTIR data treatment. ATR FTIR spectra were recorded by the instrument software OPUS 4.0 MSD (Bruker, Casablanca, Morocco). Unscrambler software version 10.3 from CAMO (Computer Aided Modelling, Trondheim, Norway) was used for the chemometric treatment of ATR FTIR spectral data.

RESULTS AND DISCUSSION

Analysis of FTIR citrus juice. The differences between the spectra, showed in Figure 1, were oviously small and manifested just in limited zones of the spectra. Data analysis by using single peaks or limited wavelength ranges for extracting the information about the cultivars seemed very hard. These data were so elaborated by multivariate data approaches, using the spectral range between 1400 and 700 cm^{-1}.

The performance in PCA data elaboration was used as a criterion to select the data pre-treatment procedure. The highest explained variance, considering the first two PCs, appeared to be 97.84% when the derivative function through the Savitzky-Golay algorithm was used (Candolfi *et al.* 1999). The derivative parameters were optimised as follows: first order, number of smoothing points 3, and polynomial order 2.

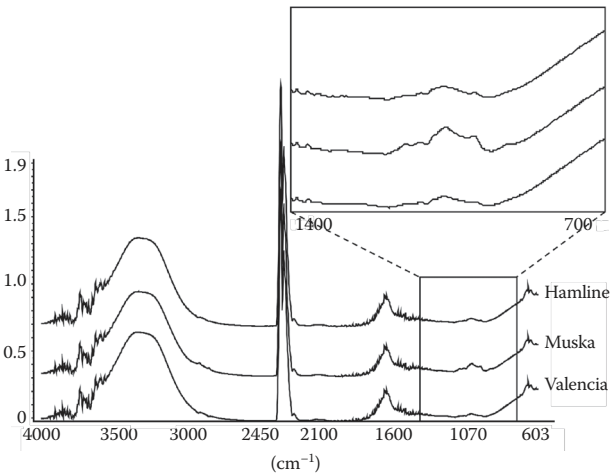

Figure 1. Mean FTIR-ATR spectra of citrus juice samples recorded for the three different cultivars

doi: 10.17221/284/2014-CJFS

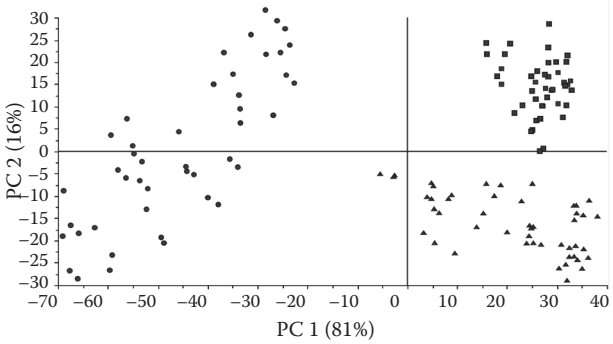

Figure 2. Score plot of citrus juices for PCA by using 2 PCs (●Muska; ▲Valencia; ■ Hamline)

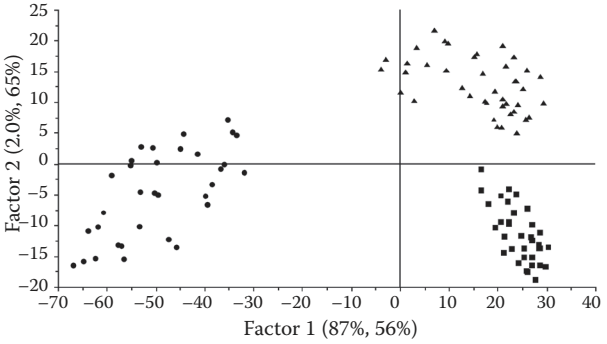

Figure 3. Score plot of PLS-DA model for citrus juice of different cultivars (●Muska; ▲Valencia; ■ Hamline)

Chemometric analysis. Figure 2 shows the PC1 vs. PC2 score plot, accounting for 97.84% of the data variance. All the classes proved to be perfectly separated from the others.

In the classification step, PLS-DA modelling was performed to discriminate the juices of samples picked at different times. The procedure used a PLS2 algorithm, fixing a Y-variable for each citrus cultivar. The performance of the model was evaluated in terms of the correlation coefficient between the calibrated and reference values (R^2) and RMSECV. The results of calibration and validation model are presented in Table 1. The statistical parameters resulted acceptable with a low RMSECV (0.0798–0.1211) and high R^2 values (below 0.9878).

In order to show the differences between the citrus cultivars, the score plot using the first two PCs was used (Figure 3). Three regions (groups) were clearly identified. Each group represented one type of citrus juice.

These results demonstrated the ability of the proposed technique to discriminate between the citrus juice cultivars used in this study. Figure 4 shows the *x*-variable loadings resulting from the calibration model. As can be seen, the spectral region between 1150 and 900 cm^{-1} seemed most abundat in analytical information useful to discriminate the classes.

A further selection of the wavelength regions to be used in calibration, so to ensure the acquisition of the most useful information and the removal of redundant zones or noisy data, was performed by using the Martens' Uncertainty Test (Martens & Martens 2000). Unfortunately, this approach was unable to improve the performance of the classification model. Consequently, it appeared necessary to consider all the signals within the range 1400–700 cm^{-1} in the construction of the model.

Classification of new samples. The results obtained from the external validation performed on the prediction set of samples are listed in Table 2. For each sample, a Y_n value lower than 0.45 was considered as not belonging to the cultivar class Y_n, while a value higher than 0.55 recognised the sample as belonging to the Y_n class. The results between 0.45 and 0.55 could be considered suspicious because these limits express 10% of error in the results. By adopting these limits, 100% of the samples from the cultivars were well predicted and no sample was detected as having suspicious origin. The rapidity of analysis, the availability of citrus all

Table 1. Statistical parameters by chemometric elaboration of citrus cultivars FTIR spectra in calibration and internal validation step

Cultivar	Full cross validation (PCs = 3)	
	R^2	RMSECV
Hamline	0.9878	0.0948
Muska	0.9885	0.1211
Valencia	0.9901	0.0798

RMSECV – root mean square error of cross validation

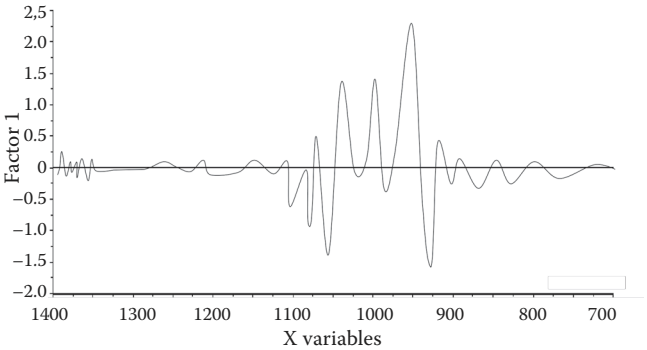

Figure 4. X-variables loading plot of PLS-DA classification model

doi: 10.17221/284/2014-CJFS

Table 2. Prediction of citrus juice cultivars by chemometric analysis of IR spectra

Cultivars	Samples	Hamlin		Muska		Valencia	
		predicted	reference	predicted	reference	predicted	reference
Hamlin	j-Hamlin-1	0.8763	1.0000	–0.0876	0.0000	0.3728	0.0000
	j-Hamlin-2	0.9855	1.0000	–0.0909	0.0000	–0.0463	0.0000
	j-Hamlin-3	0.9777	1.0000	–0.0378	0.0000	–0.0674	0.0000
Muska	j-Muska-1	0.2524	0.0000	0.9861	1.0000	–0.0563	0.0000
	j-Muska-2	0.1265	0.0000	0.9567	1.0000	–0.0768	0.0000
	j-Muska-3	0.0827	0.0000	0.9645	1.0000	–0.0382	0.0000
Valencia	j-Valencia-1	0.2146	0.0000	0.3810	0.0000	0.8947	1.0000
	j-Valencia-2	–0.1072	0.0000	–0.0284	0.0000	0.9476	1.0000
	j-Valencia-3	0.0311	0.0000	–0.0657	0.0000	0.9456	1.0000

through year, and the feasibility of the technique suggest the chemometric treatment of FTIR spectra of citrus as a promising approach for citrus cultivar discrimination and identification.

CONCLUSIONS

The discrimination between three Moroccan cultivars of citrus was performed by a chemometric modelling procedure using the FTIR data recorded by analysis of citrus juices. Each variety was identified on the basis of different spectral information spread along the selected spectral range of 1400–700 cm^{-1}. ATR FTIR analysis proved to be rapid and simple, requiring no chemical pre-treatment of the samples. The application of the PLS-DA algorithm on the samples of a prediction s*et* allowed a classification with an accuracy of 100%. This study demonstrates the great potential in the application of chemometric tools in infrared spectroscopy for the correct classification of food.

References

Anjos O., Campos M.G., Ruiz P.C., Antunes P. (2015): Application of FTIR-ATR spectroscopy to the quantification of sugar in honey. Food Chemistry, 169: 218–223.

Bade J.B., Gmitter F.G., Bowman K.D. (1991): Simplified method of volatile leaf oil analysis for identification of citrus cultivars. Horticultural Sciences, 26: 186–188.

Bassbasi M., De Luca M., Ioele G., Oussama A., Ragno G. (2014): Prediction of the geographical origin of butters by partial least square discriminant analysis (PLS-DA) applied to infrared spectroscopy (FTIR) data. Journal of Food Composition and Analysis, 33: 210–215.

Bermejo A., Llosá M.J., Cano A. (2011): Analysis of bioactive compounds in seven citrus cultivars. Food Science and Technology International, 17: 55–62.

Blanco M., Cueva-Mestanza R., Peguero A. (2010): Controlling individual steps in the production process of paracetamol tablets by use of NIR spectroscopy. Journal of Pharmaceutical and Biomedical Analysis, 51: 797–804.

Bourgou S., Rahali F., Ourghemmi I., Tounsi M.S. (2012): Changes of peel essential oil composition of four Tunisian citrus during fruit maturation. Science World Journal, 2012, Article ID 528593. doi: 10.1100/2012/528593

Candolfi A., de Maesschalck R., Jouan-Rimbaud D., Hailey P.A., Massart D.L. (1999): The influence of data pre-processing in the pattern recognition of excipients near-infrared spectra. Journal of Pharmaceutical and Biomedical Analysis, 21: 115–132.

Codoñer-Franch P., Valls-Bellés V. (2010): Citrus as functional foods. Current Topics in Nutraceutical Research, 8: 173–184.

De Luca M., Terouzi W., Kzaiber F., Ioele G., Oussama A., Ragno G. (2012): Classification of moroccan olive cultivars by linear discriminant analysis applied to ATR-FTIR spectra of endocarps. International Journal of Food Science and Technology, 47: 1286–1292.

De Luca M., Ioele G., Ragno G. (2014): Cumulative area pre-processing (CAP): a new treatment of UV data for the analysis of complex pharmaceutical mixtures. Journal of Pharmaceutical and Biomedical Analysis, 90: 45–51.

Ebere Okwu D. (2008): Citrus fruits: a rich source of phytochemicals and their roles in human health. International Journal of Chemical Society, 6: 451–471.

Ejaz S., Ejaz A., Matsuda K., Chae W.L. (2006): Limonoids as cancer chemopreventive agents. Journal of the Science of Food and Agriculture, 86: 339–345.

Fallahi E., Rodney D.R., Mousavi Z. (1990): Growth yield and fruit quality of eight lemon cultivars in Arizone.

doi: 10.17221/284/2014-CJFS

Journal of the American Society for Horticultural Science, 115: 6–8.

Fang Q., Roose L. (1997): Identification of closely related citrus cultivars with inter-simple sequence repeat markers. Theoretical and Applied Genetics, 95: 408–417.

Galtier O., Le Dréau Y., Ollivier D., Kister J., Artaud J., Dupuy N. (2008): Lipid compositions and French registered designations of origins of virgin olive oils predicted by chemometric analysis of mid-infrared spectra. Applied Spectroscopy, 62: 583–590.

Gattuso G., Barreca D., Gargiulli C., Leuzzi U., Caristi C. (2007): Flavonoid composition of citrus juices. Molecules, 12: 1641–1673.

Hennessy S., Downey G., O'Donnell C.P. (2009): Confirmation of food origin claims by Fourier transform infrared spectroscopy and chemometrics extra virgin olive oil from Liguria. Journal of Agricultural and Food Chemistry, 57: 1735–1741.

Hind A.R., Bhargava S.K., McKinnon A. (2001): At the solid/liquid interface FTIR/ATR – the tool of choice. Advances in Colloid and Interface Science, 93: 91–114.

Iñón F.A., Garrigues J.M., Garrigues S., Molina A., de la Guardia M. (2003): Selection of calibration set samples in determination of olive oil acidity by partial least squares-attenuated total reflectance – Fourier transform infrared spectroscopy. Analytica Chimica Acta, 489: 59–75.

Jaiswal P., Jha S.N., Borah A., Gautam A., Grewal M.K., Jindal G. (2015): Detection and quantification of soymilk in cow-buffalo milk using Attenuated Total Reflectance Fourier Transform Infrared spectroscopy (ATR-FTIR). Food Chemistry, 168: 41–47.

Kazarian S.G., Chan K.L.A. (2006): Applications of ATR-FTIR spectroscopic imaging to biomedical samples. BBA Biomembranes, 1758: 858–867.

Koehler-Santos P., Cugna Dornelles A.L., Brandão de Freitas L. (2003): Characterization of mandarin citrus germplasm from southern Brazil by morphological and molecular analyses. Pesquisa Agropecuária Brasileira, 7: 797–806.

Kowalski B.R. (1984): Chemometrics: Mathematics and Statistics in Chemistry. Dordrech, D. Reidel Publishing Co.

Lanfranchi M. (2012): Economic analysis on the enhancement of citrus waste for energy production. Journal of Essential Oil Research, 24: 583–591.

Maggio R.M., Kaufman T.S., Del Carlo R.M., Cerretani L., Bendini A., Cichelli A., Compagnone D. (2009): Monitoring of fatty acid composition in virgin olive oil by Fourier transformed infrared spectroscopy coupled with partial least squares. Food Chemistry, 114: 1549–1554.

Martens H., Martens M. (2000): Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). Food Quality and Preference, 11: 5–16.

Martí N., Mena P., Cánovas J.A., Micol V., Saura D. (2009): Vitamin C and the role of citrus juices as functional food. Natural Product Communications, 4: 677–700.

Mirabella F.M. (1993): Principles Theory and Practice of Internal Reflection Spectroscopy. Internal Reflection Spectroscopy-Theory and Applications. New York, Dekker.

Mulvihill E.E., Huff M.W. (2012): Citrus flavonoids and the prevention of atherosclerosis. Cardiovascular and Hematological Disorders – Drug Targets, 12: 84–91.

Sidana J., Saini V., Dahiya S., Nain P., Bala S. (2013): A review on citrus – 'the boon of nature'. International Journal of Pharmaceutical Sciences Review and Research, 18: 20–27.

Terouzi W., De Luca M., Bolli A., Oussama A., Patumi M., Ioele G., Ragno G. (2011): A discriminant method for classification of Moroccan olive varieties by using direct FT-IR analysis of the mesocarp section. Vibrational Spectroscopy, 56: 123–128.

Wold S., Sjöström M. (1998): Chemometrics, present and future success. Chemometrics and Intelligent Laboratory Systems, 44: 3–14.

Wu Y.W., Sun S.Q., Zhou Q., Leung H.W. (2008): Fourier transform mid-infrared (MIR) and near-infrared (NIR) spectroscopy for rapid quality assessment of Chinese medicine preparation Honghua oil. Journal of Pharmaceutical and Biomedical Analysis, 46: 498–504.

Zagonel G.F., Peralta-Zamora P., Ramos L.P. (2004): Multivariate monitoring of soybean oil ethanolysis by FTIR spectrometry. Talanta, 63: 1021–1025.

Zhao J.H., Zhao Y.H., Li S.Y. (2012): American brand development of citrus industry and the enlightenment to China three Gorges Reservoir area developments based on SRCT. Advances in Intelligent and Soft Computing, 143: 403–408.

Zhao X., Liu G., Li W., Wang X., Hao J., Zhou X.-P. (2014): Discrimination of five citrus diseased leaves by FTIR spectroscopy. Asian Agricultural Research, 6: 87–90.

Zhou X., Sun S.Q., Huang Q.H. (2012): Identification of different *Citrus sinensis* (L.) Osbeck trees varieties using Fourier transform infrared spectroscopy and hierarchical cluster analysis. Guang Pu Xue Yu Guang Pu Fen Xi, 32: 3006–3009. (in Chinese)

Received: 2014–05–28
Accepted after correction: 2014–10–13

Corresponding author:

Dr Michele De Luca, Department of Pharmacy, Health and Nutritional Sciences, Università della Calabria, 87036 Rende (CS), Italy; E-mail: michele.deluca@unical.it