ASSESSMENT OF TOMATO QUALITY CHARACTERISTICS USING VIS/NIR HYPERSPECTRAL IMAGING AND CHEMOMETRICS

S.J. Ramos-Infante, V. Suárez-Rubio, P. Luri-Esplandiu, M.J. Sáiz-Abajo

National Center for Food Technology and Safety (CNTA), San Adrián, Spain

ABSTRACT

The aim of this study was to assess the applicability of VIS/NIR hyperspectral imaging in intelligently detecting postharvest quality of tomatoes. The postharvest quality of tomatoes was evaluated in terms of firmness, soluble solids content (SSC), pH and color (L*a*b*C*h). VNIR and NIR hyperspectral imaging (400-1700 nm) were used to build regression models using Partial Least Square (PLS). The effects of different preprocessing techniques, including moving weighted average smoothing, first- and second-derivative Savitzky Golay (S-G) and standard normal variate (SNV) on prediction performance were also evaluated. NIR spectra-based models performed the best at estimating the quality characteristics in tomatoes. Excellent prediction for SSC, firmness and pH (RPD > 3.0), and good prediction for color (RPD > 2.0) were obtained. Consequently, NIR spectra-based systems could be introduced in the manufacturing process of tomato industries as nondestructive method to monitor quality characteristics of tomatoes.

Keywords— tomato, VNIR hyperspectral imaging, NIR hyperspectral imaging, Partial Least Square, quality characteristics

1. INTRODUCTION

Tomato is the most important fruit grown worldwide with approximately 170.8 million tons produced in 2014, which are widely consumed in either fresh or processed form [1]. The tomato sector is particularly important in the EU and specifically in the Mediterranean area. It possesses a number of beneficial nutrients such as vitamins A, C, and E; lycopene; β-carotene; and other bioactive components [2-3]. As tomato is a climacteric fruit, quality parameters change rapidly because of respiration and metabolic activities during postharvest storage period. Firmness, soluble solids content (SSC), pH and color (L*a*b*C*h) are the primary quality attributes of tomatoes which are associated with changes in structure and composition of the cells, and they directly affect consumer acceptance [4-5]. These quality attributes of tomatoes are often obtained by instrumental measurement. However, instrumental methods of quality evaluation are highly time-consuming, needs to destroy the fruit, and sometimes require specific instrumentation to be performed [6]. Because of these limitations, increasing number of researches are focused on developing sensor-based technologies and nondestructive methods to evaluate the quality attributes from whole samples, which will be very useful for producers, processors and distributors to ascertain the quality of tomatoes.

Hyperspectral imaging (HSI), a new advance technique, provides both spectra and spatial information for samples by integrating the principles of spectroscopic and imaging technologies in a system [7-10]. The hyperspectral image is a three-dimensional (3-D) cube, normally called a hypercube, in which the spectral information is provided for each pixel in the image. HSI can also provide more detailed or complete information, including internal structure characteristics, morphological information, and chemical composition of the sample, compared with a single-machine vision technology [11]. HSI has been applied well as nondestructive and rapid analytical method to predict internal quality attributes of postharvest fruit and vegetables [12-13], such as apple [14], pear [15], and kiwifruit [16].

In a previous study, the HSI technique using PLS regression analysis was developed for the rapid and nondestructive measurement of moisture content, pH and SSC in intact tomatoes [17]. In addition, an image processing-based algorithm has been also developed to evaluate the potential of HSI combined with artificial neural networks to spatial mapping of moisture content in tomato fruits in the spectral range of 400–1000 nm [18]. Another study [19] investigated the feasibility of hyperspectral fluorescence imaging to detect cuticle crack defects on cherry tomatoes. However, to the best of our knowledge, there is no report on determining all the primary quality attributes of intact tomatoes (SSC, Firmness, pH and color (L*a*b*C*h)).

The overall objective of this study was to assess the feasibility of VNIR (400-1000 nm) and NIR (900-1700 nm) hyperspectral imaging using PLS regression analysis for rapid and nondestructive measurement of SSC, Firmness, pH and color (L*a*b*C*h) in intact tomatoes. The specific purposes of our study were (1) to acquire VNIR and NIR hyperspectral images of intact tomatoes; (2) to reduce noise and scattering in selected region of interests (ROIs) and extract the corresponding spectral data and (3) to compare the performance of VNIR and NIR hyperspectral imaging to

build PLS-based regression models to predict the primary quality attributes of intact tomatoes.

2. MATERIALS AND METHODS

2.1. Tomato samples

The tomatoes were supplied by a local industry in Navarra (Spain) and rapidly delivered to the laboratory (within 2 h) every week during two months. From the original 10 kg of tomatoes received, 12 tomatoes were selected randomly. The sampling process included two stages: hyperspectral analysis and quality parameters measurement. Prior to the experiment, the samples had been equilibrated at laboratory temperature (20 ± 1 ºC) to avoid any effects of temperature on the measurements. The key steps of the experimental and data analysis procedures are outlined in Fig.1.

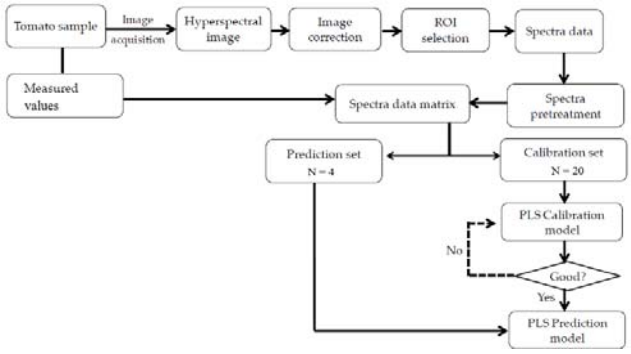

Figure 1: Key steps in the full procedure for predicting the soluble solids content (SSC), firmness, pH and color (L*a*b*C*h) in tomatoes

2.2 Hyperspectral imaging system

A commercial laboratory-based push-broom HSI system was used to acquire the hyperspectral images of 12 tomatoes. Each spectrum was the average of 4 scans (4 tomatoes). The system comprised a VNIR hyperspectral camera (Fx10e, Specim, Oulu, Finland) that covered the spectral range 400-1000 nm with 5.5 nm spectral resolution and a NIR hyperspectral camera (Fx17e, Specim, Oulu, Finland) working at 900-1700 nm with 8 nm spectral resolution, a scanning system (LabScanner 40x20, Specim, Oulu, Finland) and data acquisition software (LUMO v1.2, Specim, Oulu, Finland).

The tomatoes were illuminated by three halogen lamps (DECOSTAR 51 ALU 35 W 12 V, Apeldoorn, Netherlands). Tomato samples were placed on a tray and transferred to the conveyer belt to be scanned line by line. Both spectral and spatial information were obtained from the samples when it was in the range of the camera during movement through the conveyor unit. The hyperspectral images of tomato samples were saved in a raw format as a three-dimensional (3-D) hypercube consisting of two spatial dimensions and one spectral dimension. The dimensions of the hypercube were 1024 pixels in the x-direction, n pixels in the y-direction (based on the length of the sample), and 224 bands in the k-direction. The raw intensity of the samples was corrected using two reference standards: the white reference intensity and the dark reference intensity, obtained under the same condition as sample intensity. The white reference intensity was obtained using a white Teflon board of nearly 100% reflectance, and the dark reference intensity was acquired by turning off the light source and completely covering the lens with its opaque cap. Then the corrected intensity was calculated by the following (Eq. (1)):

$$R = \frac{R_0 - R_d}{R_w - R_d} \quad (1)$$

where R_0 is the recorded intensity, R_d is the dark reference intensity (with 0% reflectance), and R_w is the white reference intensity.

2.3 Measurements of quality parameters

The quality parameters of tomatoes were measured after hyperspectral images collection. Each reference value was the average of 4 tomatoes. The evaluated quality parameters were SSC, firmness, pH and color. Firmness of tomatoes was determined at the fruit equator using a portable puncture tester (PCE-PTR 200, PCE, Albacete, Spain). Then, tomato juice was obtained by pulping, homogenizing, and filtering as reported in the literature [20]. The SSC of the juice was measured using a refractometer (Bellingham & Stanley RFM340+, ThermoFisher Scientific, Madrid, Spain). The result was expressed as %Brix. pH value was measured by a pH meter (Crison, Barcelona, Spain). Finally, the color of tomatoes was measured using a colorimeter (X·rite SP64, X·rite, Barcelona, Spain).

2.4 Data analysis

2.4.1 Image correction

The calibrated hyperspectral image was segmented by using a simple threshold value as the average value of the background and tomato pixels to remove the effect of the background and to visualize only the tomato pixels. Furthermore, the region-of-interest (ROI) step was performed on the segmented image to extract the spectral signature and statistical information (Fig.2). The pixel spectra were averaged for each band within the entire ROI of each tomato separately before further analysis. In total, 24 average spectra representing the 96 scanned tomatoes were calculated and saved. The image correction and ROI selection steps were incorporated using MATLAB software (Version 2019a, The Mathworks Inc., Natick, MA, USA).

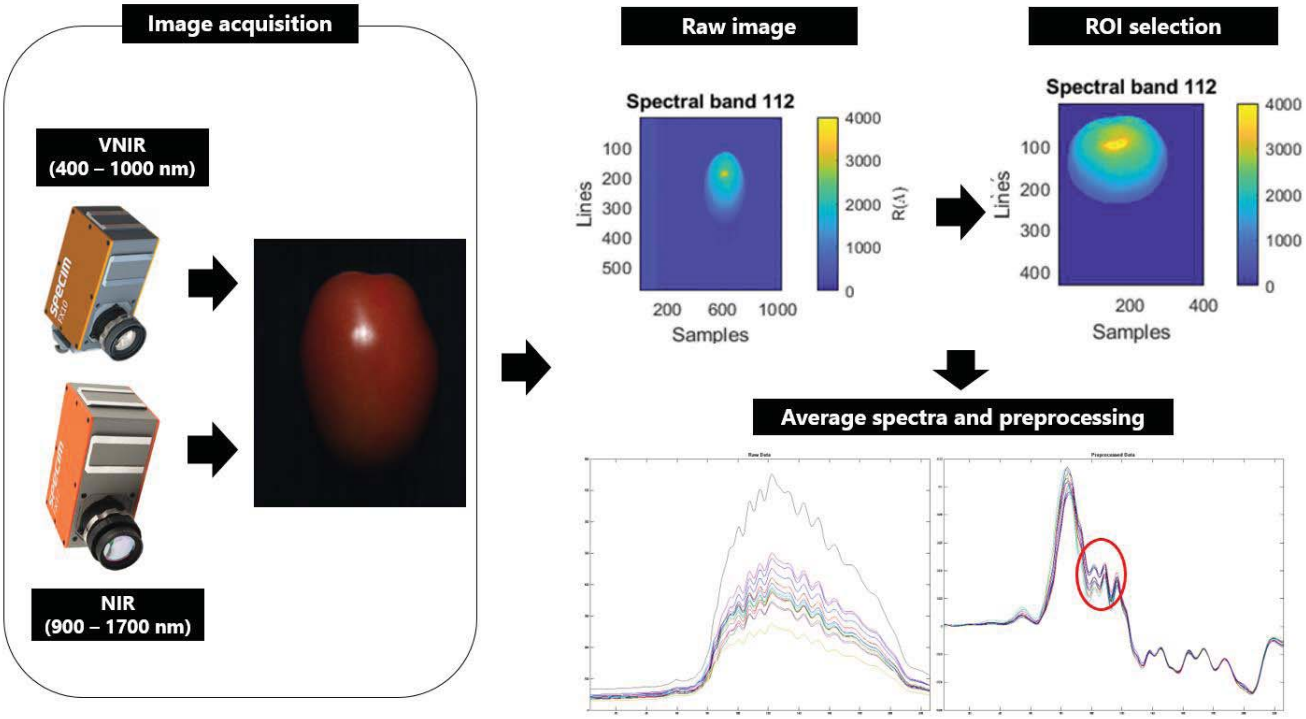

Figure 2: Workflow for image and spectra preprocessing

2.4.2 Spectral preprocessing

The average spectra were processed independently using different preprocessing methods in order to optimize the calibration performance. Moving weighted average smoothing, Savitzky-Golay (S-G) (Fig.2) and standard normal variate (SNV) were used to preprocess VNIR and NIR spectra data obtained from hyperspectral images. The averaging technique is used to reduce the number of wavelengths or to smooth the spectrum of tomatoes. S-G is a smoothing technique to remove random noise and improvement the visual aspect of spectra [21]. SNV is a common algorithm to correct the effects of scattering and remove the slope variation by transforming each spectrum to a zero mean-intensity value [22].

2.4.3 PLS model

After preprocessing of the spectra, the samples were randomly divided into two sample sets following Kennard-Stone algorithm. A calibration sample set consisted of 20 samples. This set was used for developing the calibration model. A prediction sample set, consisting of 4 samples, was used for prediction purposes. In this study, partial least squares (PLS) regression multivariate analysis was used to develop a calibration model for SSC, firmness, pH and color (L*a*b*C*h) of tomatoes prediction. All types of spectral preprocessing and the development of the calibration and prediction models for SSC, firmness, pH, and color of tomatoes prediction were performed using MATLAB software (Version 2019a, The Mathworks Inc., Natick, MA, USA).

2.4.4 Evaluation of the calibration and prediction models

Performance of the calibration and prediction models were evaluated using several statistical parameters, including coefficient of correlation (r) between measured and predicted value, root mean square error of calibration (RMSEC), root mean square error of prediction (RMSEP), number of latent variables and residual predictive deviation (RPD) [23-24]. A RPD value greater than 3.0 indicates that the model has an excellent prediction accuracy, between 2.5 and 3.0 indicates that the model has a good precision, and between 2.0 and 2.5 indicates that the model has an approximate precision [25]. Generally, high R_p^2 and RPD values and low RMSEP indicate the good performance for model prediction. They are defined in Eq. (2), (3) and (4) as follows:

$$R_c^2, R_p^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - y_m)^2} \qquad (2)$$

Table 1: Partial least squares regression results for the prediction of soluble solid content (SSC), firmness, pH and color (L*a*b*C*h) using a) VNIR and b) NIR spectra

a)

VNIR (400-1000 nm)													
Parameter	Preprocessing method	N LVs	Calibration					Prediction					
			Mean	N cal	rcal	RMSEC	RMSEC (%	Mean	N val	rpred	RMSEP	RMSEP (%	RPD
SSC (BRIX)	*Smoothing*	7,00	4,73	20,00	0,69	0,34	7,17	4,66	4,00	0,86	0,21	**4,43**	**3,10**
FIRMNESS	*SNV*	15,00	0,83	20,00	1,00	0,01	0,80	0,87	4,00	0,77	0,13	**14,40**	**2,41**
pH	*Raw (mean center)*	11,00	0,31	20,00	0,89	0,02	5,75	0,26	4,00	0,70	0,04	**15,56**	**2,12**
L	*Smoothing*	6,00	45,53	20,00	0,64	2,28	5,00	47,13	4,00	0,88	0,41	**0,88**	**3,34**
a	*S-G First Derivatives*	14,00	24,74	20,00	0,98	0,31	1,26	23,72	4,00	0,86	1,35	**5,68**	**3,09**
b	*Smoothing*	13,00	22,78	20,00	0,87	0,46	2,02	24,04	4,00	0,67	1,11	**4,61**	**2,00**
C	*S-G Second Derivatives*	11,00	33,74	20,00	0,94	0,58	1,72	33,91	4,00	0,83	1,50	**4,43**	**2,81**
h	*S-G First Derivatives*	13,00	42,82	20,00	0,93	0,71	1,66	45,66	4,00	0,64	1,74	**3,82**	**1,93**

b)

NIR (900-1700 nm)													
Parameter	Preprocessing method	N LVs	Calibration					Prediction					
			Mean	N cal	rcal	RMSEC	RMSEC (%	Mean	N val	rpred	RMSEP	RMSEP (%	RPD
SSC (BRIX)	*Smoothing*	6,00	4,60	20,00	1,00	0,04	0,85	4,67	4,00	0,96	0,11	**2,30**	**5,77**
FIRMNESS	*Raw (mean center)*	5,00	0,81	20,00	0,91	0,05	5,64	0,85	4,00	0,86	0,10	**11,64**	**3,13**
pH	*S-G Second Derivatives*	3,00	0,30	20,00	0,76	0,04	11,99	0,31	4,00	0,92	0,02	**5,40**	**4,21**
L	*SNV*	1,00	44,22	20,00	0,02	2,91	6,57	47,57	4,00	0,73	3,52	**7,40**	**2,10**
a	*Smoothing*	5,00	23,74	20,00	0,77	1,14	4,80	24,44	4,00	0,90	1,03	**4,21**	**3,69**
b	*S-G Second Derivatives*	4,00	22,38	20,00	0,98	0,23	1,01	23,16	4,00	0,96	0,43	**1,86**	**6,04**
C	*Smoothing*	5,00	32,73	20,00	0,82	1,08	3,29	33,78	4,00	0,94	0,88	**2,59**	**4,82**
h	*SNV*	2,00	43,42	20,00	0,76	1,12	2,58	43,52	4,00	0,93	0,68	**1,55**	**4,39**

$$RMSEC, RMSEP = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (3)$$

$$RPD = \frac{SD}{RMSEP} \quad (4)$$

where n is the number of calibration or prediction subset; y_i and $\hat{y}_i$ represent the measured and predicted values of the *i*th sample, respectively; y_m is the mean values of the measured values of calibration or prediction subset; and SD is the standard deviation of prediction subset.

3. RESULTS AND DISCUSSION

Table 1 listed the results of calibration and prediction of PLS models and the best spectra preprocessing for SSC, firmness, pH and color. As shown in Table 1b), NIR spectra yielded models with the highest prediction performances which had the maximum R_p^2, the lowest RMSEP and the highest RPD (R_p^2 = 0.96, RMSEP = 0.11 and RPD = 5.77 for SSC; R_p^2 = 0.86, RMSEP = 0.10 and RPD = 3.13 for firmness; R_p^2 = 0.92, RMSEP = 0.02 and RPD = 4.21 for pH; R_p^2 = 0.73, RMSEP = 3.52 and RPD = 2.10 for L (Lightness) color parameter, RPD values are above 3.00 for the rest of color parameters (a*, b*, C*, h*) being a* directly related to red/green value and b* with blue/yellow value). The prediction performance improves with different spectra preprocessing when treated hyperspectral images with different spectra preprocessing. The developed algorithm allowed selecting the best preprocessing to reach the highest prediction in terms of RMSEP and RPD values.

4. CONCLUSIONS

Calibration models computed on the hyperspectral NIR range predicted better the tomato quality parameters due to the high correlation coefficients, low RMSEC and RMSEP values, minimal number of latent variables and high RPD values obtained. According to the results yielded and considering further validation studies must be performed, hyperspectral image technology and PLS models can be used as a tool for the nondestructive estimation of quality parameters in the tomato industry.

5. REFERENCES

[1] FAOSTAT: Food and Agriculture Organiztion of the United Nations Statistics Division. Available online: http://www.fao.org/faostat/en/#data/QC/visualize.

[2] W. Du, C.W. Olsen, R.J. Avena-Bustillos, T.H McHugh, C.E. Levin and R. Friedman. Antibacteriale of allspice,

garlic, and oregano essential oils in tomato films determined by overlay and vapor-phase methods. *J Food Sci* 74(7):390–397, 2009.
[3] J. Yun, X. Fam, X. Li, T.Z. Jin, X. Jia and J.P. Matttheis. Natural surface coating to inactivate Salmonella enterica serovar Typhimurium and maintain quality of cherry tomatoes. *Int J Food Microbiol* 193:59–67, 2015.
[4] X. Hong, J. Hang and G. Qi. E-nose combined with chemometrics to trace tomato-juice quality. *J Food Eng* 149:38–43, 2015.
[5] Y. Shao, Y. He, A.H. Gómez, A.G. Pereir, Z. Qiu and Y. Zhang. Visible/near infrared spectrometric technique for nondestructive assessment of tomato 'Heatwave' (Lycopersicum esculentum) quality characteristics. *J Food Eng* 81(4):672–678, 2007.
[6] K.Flores, M. Sánchez, D. Pérez-Marín, J. Guerrero and A. Garrido-Varo. Feasibility in NIRS instruments for predicting internal quality in intact tomato. *J Food Eng* 91(2):311–318, 2009.
[7] J.M. Amigo, I. Martí and A. Gowen. Hyperspectral imaging and chemometrics: A perfect combination for the analysis of food Structure, composition and quality. *Data Handl. Sci. Technol*, 28, 343–370, 2013.
[8] J. Qin, K. Chao, M.S. Kim, R. Lu and T.F Burks. Hyperspectral and multispectral imaging for evaluating food safety and quality. *J. Food Eng*. 118, 157–171, 2013.
[9] T. Mohammadi Moghaddam, S.M.A. Razavi and M. Taghizadeh. Applications of hyperspectral imaging in grains and nuts quality and safety assessment: A review. *J. Food Meas. Charact*., 7, 129–140, 2013.
[10] D. Wu and D.W. Sun. Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review—Part I: Fundamentals. *Innov. Food Sci. Emerg. Technol*. 19, 1–14, 2013.
[11] M. Huang, X. Wan, M. Zhang and Q. Zhu. Detection of insect-damaged vegetable soybeans using hyperspectral transmittance image. *J. Food Eng.* 116, 45–49, 2013.
[12] M.L. Amodio, F. Ceglie, M.M.A. Chaudhry, F. Piazzolla and G. Colelli. Potential of NIR spectroscopy for predicting internal quality and discriminating among strawberry fruits from different production systems. *Postharvest Biol Technol* 125:112–121, 2017.
[13] S. Escribano, W.V. Biasi, R. Lerud, D.C. Slaughter and E.J. Mitcham. Nondestructive prediction of soluble solids and dry matter content using NIR spectroscopy and its relationship with sensory quality in sweet cherries. *Postharvest Biol Technol* 128:112–120, 2017.
[14] E. Bobelyn, A. Serban, M. Nicu, J. Lammertyn, B.M. Nicolai and W. Saeys. Postharvest quality of apple predicted by NIR-spectroscopy: study of the effect of biological variability on spectra and model performance. *Postharvest Biol Technol* 55(3):133–143, 2010.
[15] J. Wang, J. Wang, Z. Chen and D. Han. Development of multi-cultivar models for predicting the soluble solid content and firmness of European pear (Pyrus communis L.) using portable vis-NIR spectroscopy. *Postharvest Biol Technol* 129:143–151, 2017.
[16] M. Li, R.R. Pullanagari, T. Pranamornkith, I.J. Yule and A.R. East. Quantitative prediction of post storage 'Hayward' kiwifruit attributes using at harvest vis-NIR spectroscopy. *J Food Eng* 202:46–55, 2017.
[17] A. Rahman, L. Kandpal, S. Lohumi, M. Kim, H. Lee, C. Mo and B.K. Cho. Nondestructive estimation of moisture content, pH and soluble solid contents in intact tomatoes using hyperspectral imaging. *Applied Sciences*, *7*(1), 109, 2017.
[18] K. Mollazada, M. Omid, F.A. Tab, S.S. Mohtasebi and M. Zude. Spatial mapping of moisture content in tomato fruits using hyperspectral imaging and artificial neural networks. *In Proceedings of the CIGR-AgEng2012: IV International workshop on Computer Image Analysis in Agriculture*, Valencia, Spain, 8-12 July 2012.
[19] B.K. Cho, M.S. Kim, I.S. Baek, D.Y. Kim, W.H. Lee, J. Kim and Y.S. Kim. Detection of cuticle defects on cherry tomatoes using hyperspectral fluorescence imagery. *Postharvest biology and technology*, *76*, 40-49, 2013.
[20] C. Carner, M.S. Aday and M. Demir. Extending the quality of fresh strawberries by equilibrium modified atmosphere packaging. *Eur Food Res Technol* 227(6):1575–1583, 2008.
[21] B.M. Nicolaï, K. Beullens, E. Bobelyn, A. Peirs, W. Saeys, K.I. Theron and J. Lammertyn. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. *Postharvest Biol Technol* 46(2):99–118, 2007.
[22] Z. Xiaobo, Z. Jiewen, M. Holmes, M. Hanpin, S. Jiyong, Y. Xiaopin and L. Yanxiao. Independent component analysis in information extraction from visible/near-infrared hyperspectral imaging data of cucumber leaves. *Chemom Intell Lab Syst* 104(2):265–270, 2010.
[23] M.M. Cascant, M. Sisouane, S. Tahiri, M. El Krati, M.L. Cervera, S. Garrigues and M. de la Guardia. Determination of total phenolic compounds in compost by infrared spectroscopy. *Talanta* 153:360–365, 2017.
[24] L. Feng, M. Zhang, B. Adhikari and Z. Guo. Nondestructive Detection of Postharvest Quality of Cherry Tomatoes Using a Portable NIR Spectrometer and Chemometric Algorithms. *Food Analytical Methods*, 1-12, 2019.
[25] Z. Wu, E. Xu, J. Long, X. Pan, X. Xu, Z. Jin and A. Jiao. Comparison between ATR-IR, Raman, concatenated ATR-IR and Raman spectroscopy for the determination of total antioxidant capacity and total phenolic content of Chinese rice wine. *Food Chem* 194:671–679, 2016.