Heliyon 5 (2019) e02122

Contents lists available at ScienceDirect

Heliyon

www.elsevier.com/locate/heliyon

Determination of the superficial citral content on microparticles: An application of NIR spectroscopy coupled with chemometric tools

Ives Yoplac [a,b,*], Himer Avila-George [c], Luis Vargas [d], Paz Robert [e], Wilson Castro [f]

[a] Facultad de Ingeniería Zootecnista, Agronegocios y Biotecnología, Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, Chachapoyas, Peru
[b] Escuela de Posgrado, Programa Doctoral en Ciencia de Alimentos, Universidad Nacional Agraria La Molina, Lima, Peru
[c] Departamento de Ciencias Computacionales e Ingenierías, Universidad de Guadalajara, Ameca, Jalisco 46600, Mexico
[d] Facultad de Industrias Alimentarias, Universidad Nacional Agraria La Molina, Lima, Peru
[e] Departamento de Ciencia de Alimentos y Tecnología Química, Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, Santiago, Chile
[f] Facultad de Ingeniería, Universidad Privada del Norte, Cajamarca 06002, Peru

ARTICLE INFO

Keywords:
Food science
Spectroscopy
Chemometrics
Food analysis
Food composition
Food chemistry
ANN
MLR
PCA
PLSR
Prediction

ABSTRACT

This work evaluates near-infrared (*NIR*) spectroscopy coupled with chemometric tools for determining the superficial content of citral (SC_{Ct}) on microparticles. To perform this evaluation, using spray drying, citral was encapsulated in a matrix of dextrin using twelve combinations of citral:dextrin ratios (*CDR*) and inlet air temperatures (*IAT*). From each treatment, six samples were extracted, and their SC_{Ct} and *NIR* absorption spectral profiles were measured. Then, the spectral profiles, pretreated and randomly divided into modeling and validation datasets, were used to build the following prediction models: principal component analysis-multilinear regression (*PCA-MLR*), principal component analysis-artificial neural network (*PCA-ANN*), partial least squares regression (*PLSR*) and an artificial neural network (*ANN*). During the validation stage, the models showed R^2 values from 0.73 to 0.96 and a root mean squared error (*RMSE*) range of [0.061–0.140]. Moreover, when the models were compared, the full and optimized ANN models showed the best fits. According to this study, *NIR* coupled with chemometric tools has the potential for application in determining SC_{Ct} on microparticles, particularly when using *ANN* models.

1. Introduction

Citral (3,7-dimethyl-2,6-octadienal), whose chemical structure is shown in Fig. 1, is widely used in beverages, soft drinks, breads, sweets, pharmaceuticals, perfumes, and cigarettes, among others [1, 2]; mainly due to its broad antimicrobial spectrum in vitro and its antioxidant capacity [3, 4, 5, 6].

However, the citral structure is susceptible to degradation and loss of activity [6], which is a continuing concern for food scientists. In this scene, microencapsulation appears to be an alternative for protecting citral [4, 6], and spray drying is the most commonly used method for citral microencapsulation [2, 7].

However, as with other techniques, the efficiency of this process must be measured to evaluate the process and determine the amount of active compound trapped within the matrix (efficiency of microencapsulation *EM*) [1]. The *EM* calculation, according to [8], uses Eqs. (1) and (2), which require knowing the superficial content of the encapsulated material as an intermediate value.

$$SP_{ME}(\%) = 100\left(\frac{SC_{ME}}{TC_{ME}}\right) \quad (1)$$

$$EM(\%) = 100 - SP_{ME}(\%) \quad (2)$$

where SP is the superficial percentage; SC is the superficial content; TC is the theoretical content; and ME denotes the material to be encapsulated.

Several methods have been used to evaluate SP and EM for citral, among which are gas chromatography (*GC*) [9, 10, 11, 12], solid phase microextraction-gas chromatography (*SPME-GC*) [13], GC-mass spectrometry (*GC-MS*) [14, 15], high-efficiency liquid chromatography (*HPLC*) [16], and ultraviolet spectrophotometry (*UVS*) [4, 17].

* Corresponding author at: Facultad de Ingeniería Zootecnista, Agronegocios y Biotecnología, Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, Chachapoyas, Peru.
E-mail address: ives.yoplac@untrm.edu.pe (I. Yoplac).

https://doi.org/10.1016/j.heliyon.2019.e02122
Received 25 May 2019; Received in revised form 7 July 2019; Accepted 18 July 2019

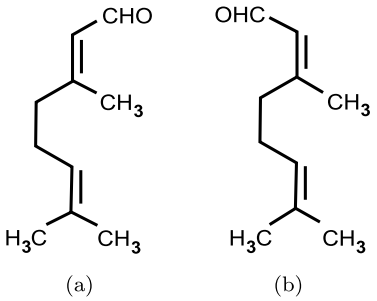

Fig. 1. Chemical structures of citral: (a) geranial (citral *a* or *cis*-citral) and (b) neral (citral *b* or *trans*-citral). Adapted from [2] and [18].

However, all the above methods require time, sample preparation, use of reagents and generally carry high costs, which provides an interesting field for research in the application of non-invasive techniques. Therefore, *NIR* spectroscopy is a technique with high potential in food analysis. Its main advantages are as follows: (1) multiple components can be measured simultaneously; (2) it is a nondestructive technique; (3) its cost is relatively low; and (4) it does not require chemical reagents or sample preparation [19].

NIR spectroscopy measures the response (mainly the vibrations of bonds such as -CH, -OH, -SH and -NH) of a sample when it is subjected to electromagnetic energies between 800 to 2500 nm, thus forming traces or spectral profiles [20]. To date, *NIR* has shown potential for analyses in polymer industries, textiles, industrial process control, biomedical science [20] and is widely used in the agricultural and food sectors [19]. However, the large amount of information involved in these profiles requires the use of special tools capable of analyzing such information; consequently, chemometric tools such as principal component analysis (*PCA*), correlation coefficients (*CC*), artificial neural networks (*ANN*), multiple linear regression (*MLR*), and partial least squares regression (*PLSR*), among others [20, 21], must be explored.

A review of the literature shows that relatively few studies exist on the use of *NIR* spectroscopy to quantify citral, as in the case of lemon and orange essential oils [18]. Therefore, the objective of this research was to evaluate the feasibility of using *NIR* spectroscopy and *chemometric* tools to determine the superficial citral (SC_{Ct}) content in microparticles.

2. Materials & methods

2.1. Reagents

The reagents used in the production and analysis of the microparticles were soy lecithin (Epikuron® 145 V, Cargill, Germany), dextrin (Amisol® 4810, Ingredion, Peru), a citral mixture of cis- and trans-isomers ≥ 96%(Sigma-Aldrich, Germany) and the citral analytical standard (Sigma-Aldrich, Germany).

2.2. Experimental procedure

The experimental procedure performed for the present study is shown in Fig. 2 and detailed in the following subsections.

2.3. Encapsulation

Citral encapsulation by spray drying was performed by varying the citral:dextrin ratio *(CDR)* (from 1:5 to 1:20) and the inlet air temperature *(IAT)* to the spray dryer (from 120 to 200 °C). A central composite design (face centered) with a factorial arrangement of 2^2 add star was used (four experimental points, four axial points and four central points, see Table 1).

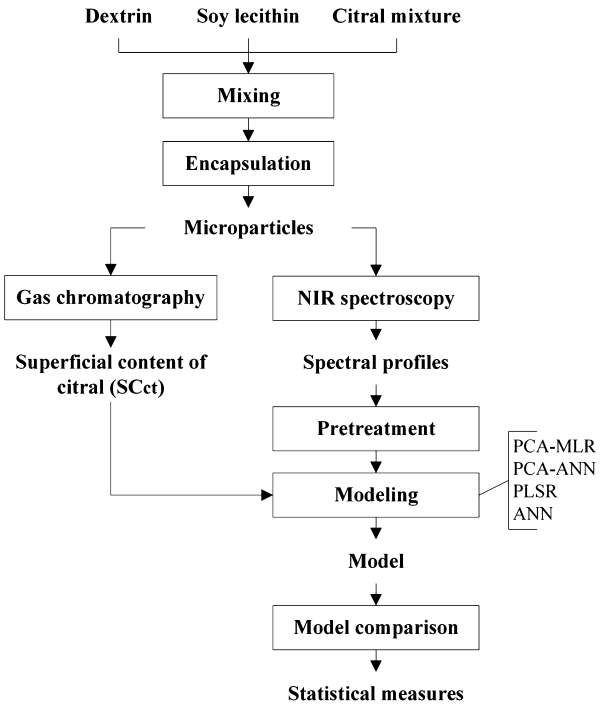

Fig. 2. Experimental procedure.

Table 1
Experimental treatments for microparticle preparation.

	IAT (°C)	CDR	Emulsion (g)	Dx (g)	Wt (g)
1	200	1:20.0	32.0	20.0	48.0
2	200	1:12.5	32.0	12.5	55.5
3	120	1:20.0	32.0	20.0	48.0
4	160	1:20.0	32.0	20.0	48.0
5	160	1:12.5	32.0	12.5	55.5
6	160	1:12.5	32.0	12.5	55.5
7	120	1:5.0	32.0	5.0	63.0
8	120	1:12.5	32.0	12.5	55.5
9	160	1:5.0	32.0	5.0	63.0
10	200	1:5.0	32.0	5.0	63.0
11	160	1:12.5	32.0	12.5	55.5
12	160	1:12.5	32.0	12.5	55.5

IAT: Inlet air temperature to the spray dryer.
CDR: citral:dextrin ratio.
Emulsion = Soy lecithin (SL) + Water (Wt) + Citral (Ct).
Dx: Dextrin.
Wt: Water.

For each treatment, 100 g of feed solution was considered. The emulsions (citral and soy lecithin - *SL*) were prepared by dissolving 1 g of *SL* in 30 mL of distilled water at 50 °C, stirring at 350 rpm for 15 min, cooling to 30 °C, and then adding 1 g of citral and homogenizing for 3 minutes at 15,000 rpm with a Polytron PT 2100 homogenizer (Kinematica AG, Switzerland). In parallel, dextrin [5 - 20] g was dissolved in distilled water [48 - 63] g at 50 °C, stirred at 350 rpm for 3 hours, cooled to 30 °C and then mixed with the emulsion. The resulting mixture was homogenized at 15,000 rpm for 3 minutes.

The obtained dispersions were fed to a mini spray dryer, model B-290 (Buchi, Switzerland); the feed and drying air flow were arranged in a co-current configuration. The process conditions in the dryer were as follows: the feed temperature was 42 ± 2.5 °C; the feed rate was 5 mL/min (5%); the atomization pressure was 50 mbar; the air flow was 600 L/h; and the IAT of the dryer varied from 120–200 °C. The resulting microparticles were stored at -20 °C in 45 mL Falcon tubes until analysis.

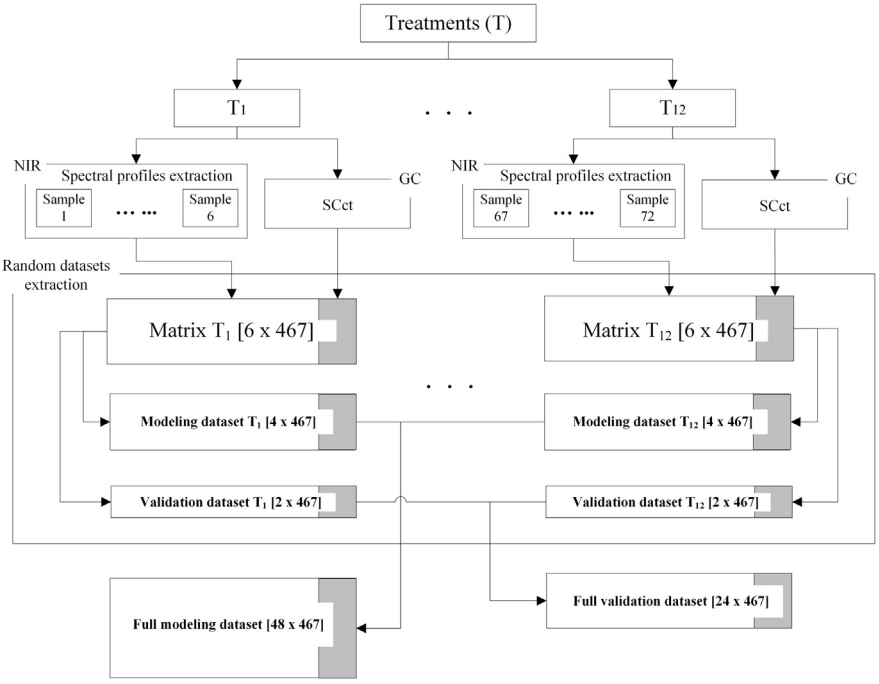

Fig. 3. Flowchart of dataset building.

2.4. Determination of superficial content of citral

To determine the superficial content of citral (SC_{Ct}), 0.2 g of microparticles (Mp) were weighed, and 2 mL of hexane was added. The mixture was then manually shaken with three inversion movements and centrifuged at 2,000 rpm for 1 minute. Finally, the supernatants were transferred to 2 mL amber vials for analysis by *GC* [22, 23].

The *GC* analysis was performed following the methodology proposed by [14] in a gas chromatograph (7890A, Agilent Technologies, USA) equipped with a *DB-5MS* fused silica capillary column (30 m x 0.25 mm x 0.25 μm film, J & W Scientific, USA) and a flame ionization detector (*FID*). The carrier gas was nitrogen at a flow rate of 1.11 mL/min. The injected sample volume was 1 μL. The column temperature was initially set to 60 °C for 5 min and then heated to 120 °C at a rate of 5 °C/min and finally, to 250 °C at a rate of 10 °C/min.

The quantification of SC_{Ct} was performed using a calibration curve (1 - 2,000 μg/mL, $R^2 = 0.999$) calculated according to Eq. (3) [8].

$$SC_{Ct} = \frac{CC_{St}}{W_{Mp}} \tag{3}$$

where SC_{Ct} is the superficial content of citral (mg Ct/g Mp); *St* is solvent; *Mp* are the microparticles; *CC* is the concentration of superficial citral in solvent (mg Ct/2 mL hexane); and W is the weight (g).

2.5. NIR spectroscopy

To determine the absorbance spectra we followed the methodology reported by [24] and [25]. In this study, we used a Unity Scientific NIR spectrometer (SpectraStar 2500XL, USA) equipped with a tungsten halogen lamp as a light source and an InGaAs (indium-gallium-arsenic) detector with a range from 1,100 to 2,500 nm, a resolution of 3 nm, and 467 wavelengths.

The measurements were performed in the reflectance mode and applied directly to the microparticles without pretreatment or manipulation using a quartz cuvette with an internal diameter of 3.5 cm and a thickness of 1.0 cm, to which 2.3 ± 0.5 g of microparticles were added.

Additionally, dilutions of citral with petroleum ether were prepared at ratios of 0.25, 0.50 and 1.0 w/w. The spectral profile for each dilution was measured according to the previously reported procedure for microparticles, with the difference that 1.2 mL of each citral dilution was added to the cuvette.

2.6. Pretreatment of spectral profiles

As discussed by ElMasry and Nakauchi [26], in most cases, extracted spectral profiles contain noise and variability and thus require the application of spectral enhancements such as spectral filtering, smoothing, normalization, mean centering, and auto-scaling. In our case, similar way to [27] and [28], the spectra were smoothed using a second-order Savitzky-Golay filter with eleven frames according to the following equation (4):

$$Y^o = \frac{\sum_{i=-m}^{m} C_i . y_{j+i}}{N} \tag{4}$$

where Y^o is the smoothed profile; y_j is the original profile; C is the coefficient for the i^{th} term of the profile; and N is an integer number of convolutions.

2.7. Dataset construction

Fig. 3 depicts the procedure performed to obtain the modeling and validation datasets, using both SC_{ct} and spectral profiles. First, the SC_{ct} used in each treatment were determined; then, the microcapsules obtained for each treatment were divided into six parts, and their spectral profiles were determined.

For each obtained treatment matrix with dimensions [6 x 467]; in a sample of a row, the columns contain both an absorbance spectral profile and SC_{ct}. Then, the matrices was randomly divided: four were used for modeling and two for validation. Finally, the two datasets for each treatment were concatenated into two full datasets (forty-eight profiles for modeling and twenty-four profiles for validation) for use in model building, validation, and comparison.

2.8. Model construction

To model the relationships between the *NIR* absorption spectral profiles and SC_{ct}, we applied different chemometric tools, such as *PCA*, *MLR*, *PLSR* and *ANNs*. Models were implemented using MATLAB v9.2

scripts (Mathworks, Natick, MA), based on the following works [20, 21, 27]. In all cases, both calibration and validation datasets were used.

2.8.1. PCA

The *PCA* algorithm creates new variables or identifies principal components and sorts them according to their capacity to explain variance. In this sense, the first principal component (PC_1) has the highest variance, the second principal component (PC_2) has the second highest variance, and so on until a pre-established percentage [29] is reached (99.5% in our case).

Likewise, to calculate components, Eq. (5) is commonly used:

$$PC_i = \phi_i^1.X^1 + \phi_i^2.X^2 + \phi_i^3.X^3 + \cdots + \phi_i^n.X^n + \quad (5)$$

where PC_i is the i^{th} principal component, ϕ_i^j is the loading vector for the i^{th} principal component, X^j are normalized predictors, and j is the number of variables (wavelengths).

2.8.2. MLR

MLR is the oldest of the methods used here; however, its use in applications has steadily diminished due to improvements in computing power. This regression allows the establishment of a link between a reduced number of predictors (the previously calculated *PCs*) and a sample property [20]. The prediction y_i for the target property can then be described by Eq. (6):

$$y_i = b_0 + \sum_{i=1}^{k} b_i PC_i + e_{i,j} \quad (6)$$

where b_i is the computed coefficient; PC_i, i^{th} principal component; and $e_{i,j}$ is the error.

2.8.3. PLSR

PLSR is a statistical method that transforms input variables X into output Y variables. In this work, X is an absorbance matrix [n x m], where n is the wavelength and m is the number of observations, and Y are the response variable values (SC_{Ct}) of the microparticles. *PLSR* decomposes X and Y by projecting them in new directions with the restriction that the decomposition must describe how variables change together to the greatest extent possible. After the variable decomposition, a regression step is performed in which the decomposed X and Y are used to calculate a regression model called the full model (see Eq. (7)):

$$\underbrace{\begin{bmatrix} Y_1 \\ \vdots \\ Y_{n-1} \\ Y_n \end{bmatrix}}_{SC_{Ct}} = \underbrace{\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{n-1} \\ \beta_n \end{bmatrix}}_{\text{Beta coefficients}} * \overbrace{\begin{bmatrix} X_{1,1} & \cdots & X_{1,m-1} & X_{1,m} \\ \vdots & \ddots & \vdots & \vdots \\ X_{n-1,1} & \cdots & \ddots & X_{n-1,m-1} \\ X_{n,1} & \cdots & X_{n,m-1} & X_{n,m} \end{bmatrix}}^{m \text{ samples absorbance profiles with } n \text{ wavelengths}} + \underbrace{e}_{\text{Error}} \quad (7)$$

This equation can be rewritten as shown in Eq. (8):

$$Y = \beta.X + e \quad (8)$$

where Y is the SC_{Ct}; β is the beta coefficient matrix; X is the absorbance data matrix; and e is the model error.

Full *PLSR* models were constructed using all the wavelengths in the spectral profiles. The relevance of each wavelength in the model was evaluated by the regression coefficients or β coefficients [30]. The wavelengths were linked to the absolute β coefficient value for the full *PLSR* model; then, the wavelengths were selected according to their values, which reflects their ability to explain R^2 (coefficient of determination) and root mean square error (*RMSE*). Finally, simplified or optimized models were constructed using only the most relevant wavelengths [27, 31].

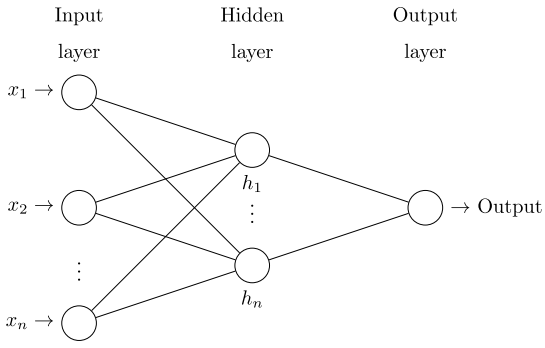

Fig. 4. Multilayer neural network example.

2.8.4. ANN

An *ANN* is a nonlinear model inspired by brain neural architecture and was developed in an attempt to model the learning capacity of biological neural systems [32]. *ANNs* have been widely used for prediction (multilayer perceptron - *MLP*) and classification (radial basis neuronal networks - *RBN*). An artificial neuron can calculate the weighted sum of its inputs and then apply an activation function to obtain a signal that is transmitted to the next neuron [33, 34].

Typical *ANN* models consist of three layers, as shown in Fig. 4. The first layer (input layer), receives the input values and, employing a transfer function, distributes the input attribute values to the processing elements in the second layer. The second layer is a hidden layer, in which the data are calculated using a nonlinear sigmoidal transfer function. The third layer is the output layer whose number of processing elements depends on the number of prediction categories [35].

We built three different *ANN* models: a) *PCA-ANN* which uses *PCs* as input, b) A full *ANN* that uses all the wavelengths in the spectral profiles, and c) an optimized *ANN* that uses the most relevant wavelengths according to the β coefficient analysis.

2.9. Model comparison criteria

All the models were compared in terms of statistical metrics such as the coefficient of determination (R^2), adjusted coefficient of determination (R^2_{adj}), mean square error (*MSE*), and root mean square error (*RMSE*). These measures can be calculated for both the calibration (R^2_c, $R^2_{adj,c}$, MSE_c and $RMSE_c$) and validation datasets (R^2_v, $R^2_{adj,v}$, MSE_v and $RMSE_v$) using Eqs. (9)–(12) [28, 36, 37]. In addition, the residual prediction deviation (*RPD*) (Eq. (13)) and range error ratio (*RER*) (Eq. (14)) were used to evaluate the quality of the models on the validation data set [38]:

$$R^2_{c,v} = \frac{\sum_{i=1}^{n} \left(\hat{y}_i - y_i\right)^2}{\sum_{i=1}^{n} \left(\hat{y}_i - y\right)^2} \quad (9)$$

$$R^2_{adj,c,v} = 1 - \left(1 - R^2\right)\left(\frac{n-1}{n-p-1}\right) \quad (10)$$

$$MSE_{c,v} = \frac{1}{n}\sum_{i=1}^{n} \left(\hat{y}_i - y\right)^2 \quad (11)$$

$$RMSE_{c,v} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(\hat{y}_i - y\right)^2} \quad (12)$$

$$RPD = \frac{SD_r}{RMSE_v} \quad (13)$$

$$RER = \frac{R_r}{RMSE_v} \quad (14)$$

where $\hat{y}_i$ and y_i are the quality attribute values of the i_{th} sample for prediction and reference, respectively; y is the average value; p is the number of predictors; n is the number of samples; SD_r is the standard deviation of the reference values of the samples in the validation set;

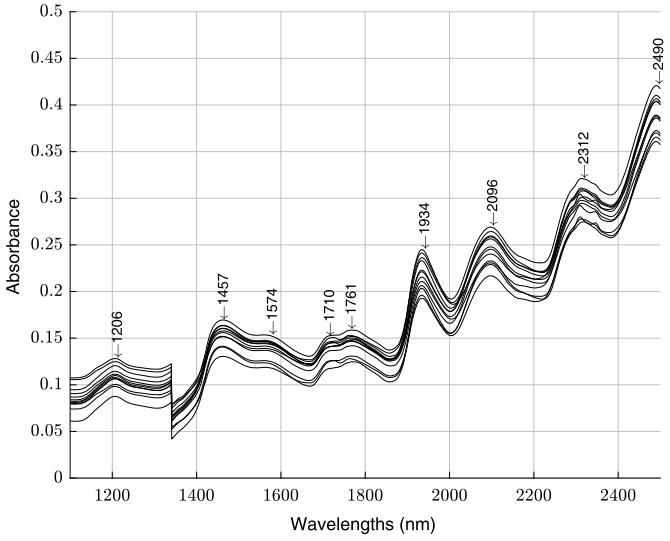

Fig. 5. Evolution of mean absorbance according to SC_{Ct}.

and R_r is the difference between the largest and smallest reference values (range) in the validation set. In this sense and according to [39], the model must have high R^2, low $RMSE_c$ and $RMSE_v$, an RPD greater than three and RER greater than ten.

3. Results & discussion

3.1. Reference analysis

According to the reference analyses of the microparticles made by *GC*, the SC_{Ct} varied between 0.17 and 1.28 mg Ct/g Mp.

3.2. Spectral profiles of superficial citral content

Fig. 5 presents the average spectral profiles that illustrate the evolution of absorbance. This figure shows peaks related to the absorbance bands in the overtone region (from 1,206 to 1,761 nm) and the combination region (from 1,934 to 2,490 nm). In the first region, an absorbance drop was observed that averaged 0.042 units between 1,340 and 1,341 nm, mainly due to the detector in the NIR equipment. This drop in absorbance does not affect the prediction calculation because it depends on the chemometric model used.

The absorption peaks at 2,096 and 2,312 nm are within the ranges corresponding to *C-H bending-stretching* and $C=O$ *stretching* interactions—possibly due to the presence of an aldehyde group [25, 40]. The peaks at 1,206, 1,710 and 1,761 nm correspond to *C-H stretching* interactions, probably due to the presence of dextrin, the polymer used as an encapsulant in this study. The obtained results were similar to those reported by [41] and [42].

The peaks at 1,457 and 1,934 correspond to vibration bands of *O-H stretching* and *O-H combinations*, respectively, which could be due to the presence of water. In addition, a small peak was observed at 1574 nm, which is within the vibration band of *C-H stretching* associated with lipids, probably due to the presence of the soy lecithin used as an emulsifier during microparticle preparation [19].

These profiles indicate a correspondence between the spectral changes and the SC_{Ct} in the microparticles, which supports pattern analysis and the implementation of prediction models.

3.3. Modeling

In the following subsections, the principal results of the modeling stage are shown.

3.3.1. PCA-MLR and PCA-ANN

Fig. 6a shows that four PCs can explain over 99.5% of the variance; these first two PCs account for 86.4 and 10.8%of the spectral variance, respectively. These results are similar to those provided by [43], who reported 88.9 to 94.4%, or the 98% reported by [44] when citrus fruits and their juices were evaluated. Then, using the PC matrix, the $PCA-MLR$ and $PCA-ANN$ models were built (see Fig. 6b and c). In both models, on the calibration and validation datasets, R^2 ranged between 0.72 and 0.77.

3.3.2. PLSR modeling

PLSR model construction was divided into two stages. In the first stage, all the wavelengths were used as inputs for the model. In this case, using the calibration and validation datasets, the R^2 varied between 1.0 and 0.898 (Fig. 7a). According to these results, the full *PLSR* validation showed a major dispersion concerning full *PLSR* calibration in the results. In this case, according to [45, 46, 47], the result may be due to high collinearity and the overlapping of spectral profiles, which introduce noise to the model, making it necessary to determine the relevance of each wavelength previously used in the model.

Then, in an approach similar to [31] through cross-validation, PLSR models of the effects of different numbers of latent variables (LVs, wavelengths) were found and plotted in Fig. 7b; the number of relevant LVs or wavelengths was set to seventeen.

Fig. 8 shows the most relevant wavelength according to its absolute β coefficient value, the mean profiles of the microparticles, and the average citral profiles. As shown, the relevant wavelengths (RW) were distributed mainly in the range [2,282–2,390] nm and in two other ranges ([1,190–1,514] nm and [1,664–2,000] nm). The β coefficients resulting from the SC_{Ct} calibration procedure show a similarity between the citral and microparticle profiles, with higher predominance overtones and bands of *C-H bending-stretching* in the range from 2,282 to 2,354 nm.

Likewise, a high β coefficient was observed at 2,303 nm for citral caused by the combination of the $C=O$ and *C-H stretching* frequencies of the aldehyde group present in citral. The negative peaks (2,282 and 2,333 nm) occur because the absorbance peaks are reversed in the β coefficient calculation [40]. These results are within the ranges obtained by [25] when they evaluated the *NIR* spectrum of pure limonene, which showed predominant peaks between 2,250 and 2,360 nm. Meanwhile, the results are slightly superior to those reported by [40] for their evaluation of the *NIR* spectrum of citral in lemongrass (*Cymbopogon citratus*) and lemon (*Citrus limon*) oils, in which they observed that the aldehydic *C-H* group (present in citral) exhibits characteristic combination bands between 2,210 nm and 2,250 nm. Additionally, they noted an increase in absorbance at approximately 2,212 nm with an increase in citral content. This difference in the results could be due to the presence of dextrin and soy lecithin as part of the microparticle formulation.

Then, the relevant wavelengths were used to build an optimized *PLSR* model (see Fig. 9). As shown in this figure, the R^2 values are 0.96 and 0.94 for calibration and validation, respectively. These results are similar to those obtained by [48] in their evaluation of *NIR* correlation statistics for aldehyde composition in cinnamon leaf (*Cinnamomum zeylanicum*), clove leaf and clove bud (*Syzygium aromaticum*) essential oils. Additionally, these results are consistent with those obtained by [25] and [49] when they evaluated limonene in citrus oil and with those reported by [40] and [18], who quantified citral in lemongrass and lemon orange oils. Finally, these results are consistent with [50], who evaluated the chemical compositions of thyme, oregano, and chamomile essential oils. All these studies reported R^2 values above 0.90.

3.4. ANN modeling

Fig. 10 summarizes the results of this stage. Fig. 10a shows the full *ANN* model, where the values for both the R^2 calibration and validation

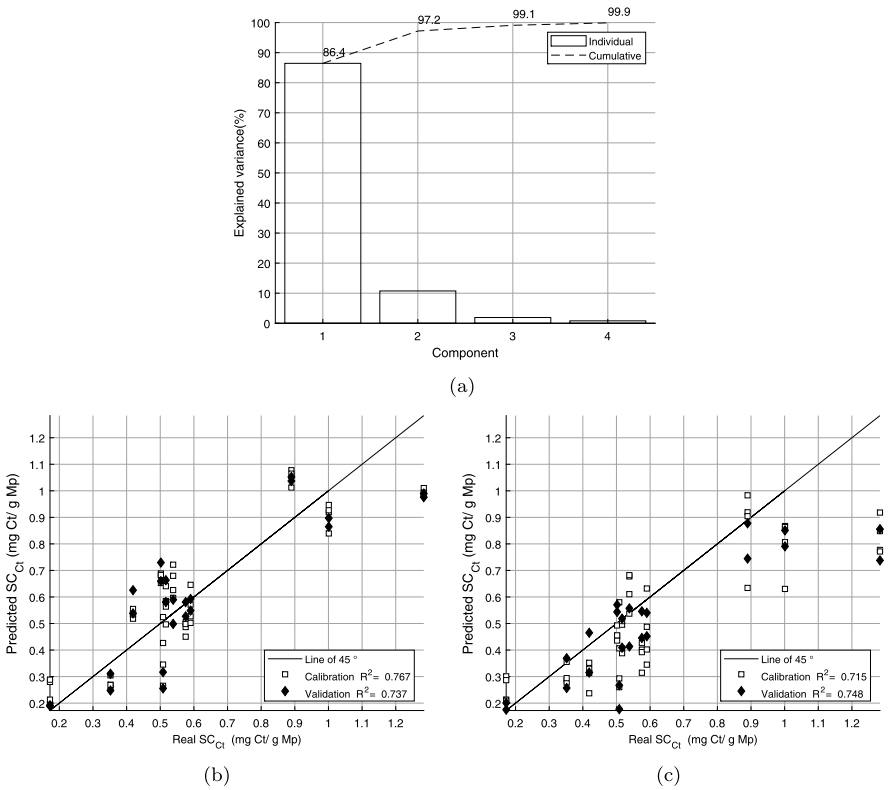

Fig. 6. (a)*PC vs Variance*, (b) *PCA-MLR*, and (c) *PCA-ANN*.

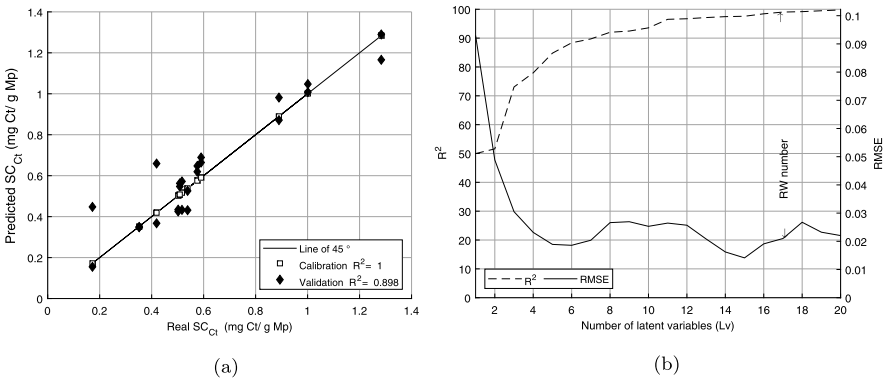

Fig. 7. (a) Full PLSR, (b) R_c^2 and $RMSE_c$ vs LVs.

(0.97 and 0.95, respectively) were similar. Fig. 10b shows the optimized *ANN* model, which was built from the selected LV, where the R^2 values for calibration and validation were 0.98 and 0.96, respectively. Consistent with these results, in both figures, little data dispersion was observed. In this case, similar to [27], the ANN models are shown to be more stable.

3.5. Model comparisons

The metrics for the chemometric models (full and optimized) on both the calibration and validation datasets are shown in Table 2. The obtained values show that the *PCA-MLR*, *PCA-ANN*, *PLSR* and *ANN* models produced acceptable results, achieving high R^2 values (above 0.71). Except for the full *PLSR* model, the *ANN* model achieved a slightly better fit than did the other models.

The superior performance of the *ANN* model can be partially attributed to its ability to adjust to nonlinear functions, an ability that the other models do not possess. These results agree with those reported by [27, 51, 52], who compared different models and found the best fit, with at least a slight advantage, for ANN models in all the studies.

4. Conclusions

This paper evaluated the feasibility of applying *NIR* spectroscopy coupled with chemometric tools (*PCA-MLR*, *PCA-ANN*, *PLSR* and *ANN*) to determine the superficial citral content on microparticles. The results indicated that all the models were able to predict the superficial content of citral on the microparticles, achieving a coefficient of determination over 0.71. The results of the *ANN* models (R^2 from 0.950 to 0.980) were slightly better than those of the PLSR models (0.935 - 0.970) and better than *PCA-MLR* (0.737 - 0.767) and *PCA-ANN* (0.715 - 0.748). Similarly, the ANN model achieved the best RPD (4.479) and RER (18.235). Consequently, this technique shows potential as a method for predicting the superficial content of citral on microparticles. However, a larger number of samples must be studied to improve the generalizability of the models.

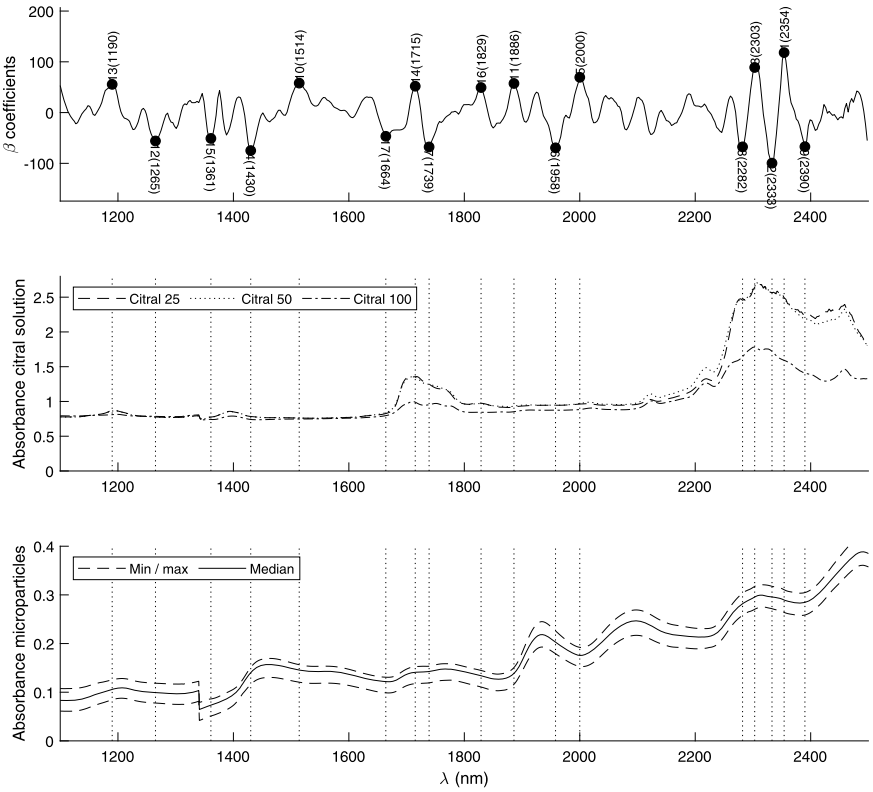

Fig. 8. From top to bottom are the β coefficients and RW (dark points), the spectral profiles of the citral solutions, and the spectral profiles of microparticles.

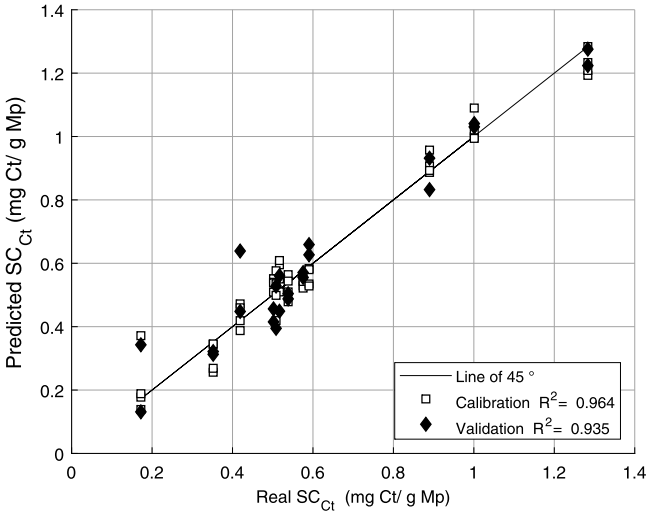

Fig. 9. Optimized PLSR models.

Table 2
Statistical measures for SC_{Ct} determination models.

Calibration dataset

Parameter	PCA		PLSR		ANN	
	MLR	ANN	Full	Optimized	Full	Optimized
R^2	0.767	0.715	1.000	0.964	0.970	0.980
R^2_{adj}	0.762	0.709	1.000	0.963	0.969	0.979
MSE	0.016	0.014	0.000	0.003	0.003	0.002
RMSE	0.126	0.120	0.000	0.055	0.052	0.043

Validation dataset

Parameter	PCA		PLSR		ANN	
	MLR	ANN	Full	Optimized	Full	Optimized
R^2	0.737	0.748	0.898	0.935	0.950	0.957
R^2_{adj}	0.725	0.737	0.893	0.932	0.948	0.955
MSE	0.020	0.013	0.009	0.006	0.004	0.004
RMSE	0.140	0.112	0.094	0.077	0.067	0.061
RPD	2.127	2.659	3.173	3.867	4.436	4.479
RER	7.948	9.936	11.857	14.451	16.579	18.235

4.1. Difficulties and future work

Chemometric tools have shown efficacy for selecting relevant wavelengths. However, this selection could be optimized by applying a different method of selecting characteristics and regression models, similar to [53, 54], who showed that the selection method influences the efficacy of prediction models. Despite having this background, we still do not have a standardized method for extracting variables.

Therefore, in subsequent studies, a combined method for variable selection and prediction properties must be applied for cases similar to the one exposed here. This approach, along with a larger number of samples, will allow us to propose this method as a referential method for analyzing compounds on the surfaces of microparticles.

Declarations

Author contribution statement

Ives Yoplac, Wilson Castro: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Himer Avila-George: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Luis Vargas: Conceived and designed the experiments; Performed the experiments; Contributed reagents, materials, analysis tools or data.

Paz Robert: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data.

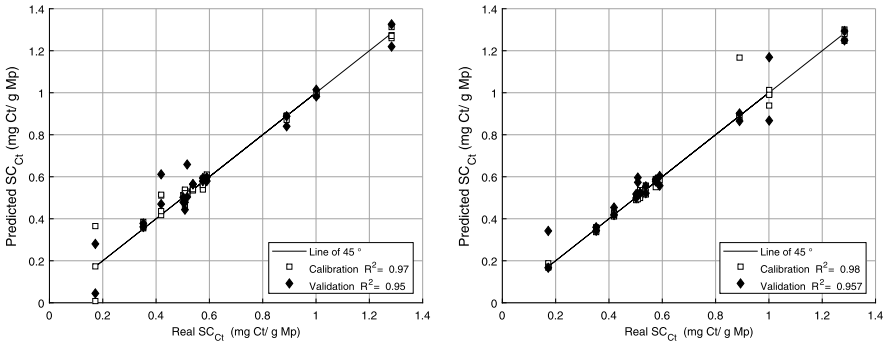

Fig. 10. ANN models: (a) full and (b) optimized.

Funding statement

This work was supported by FONDECYT CONCYTEC-Peru.

Competing interest statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

Acknowledgements

The authors thank Cristina Vergara, Estefanía González, Inés Cea and Guibeth Morello for their contributions, technical support and suggestions during the execution of this study.

References

[1] P. Campelo, E. do Carmo, R. Domingues, M. Yoshida, V.P. Ferraz, R.V. De Barros, D. Botrel, S. Vilela, Effect of dextrose equivalent on physical and chemical properties of lime essential oil microparticles, Ind. Crop. Prod. 102 (2017) 105–114.

[2] M. Maswal, A.A. Dar, Formulation challenges in encapsulation and delivery of citral for improved food quality, Food Hydrocoll. 37 (2014) 182–195.

[3] F. Fancello, G.L. Petretto, S. Zara, M. Sanna, R. Addis, M. Maldini, M. Foddai, J.P. Rourke, M. Chessa, G. Pintore, Chemical characterization, antioxidant capacity and antimicrobial activity against food related microorganisms of Citrus limon var. pompia leaf essential oil, LWT-Food Sci. Technol. 69 (2016) 579–585.

[4] W. Lu, D. Huang, C. Wang, C. Yeh, J. Tsai, Y. Huang, P. Li, Preparation, characterization, and antimicrobial activity of nanoemulsions incorporating citral essential oil, J. Food Drug Anal. 26 (1) (2018) 82–89.

[5] A. Amna, S. Khayyat, Chemical and antimicrobial studies of monoterpene: citral, Pestic. Biochem. Phys. 98 (1) (2010) 89–93.

[6] Y. Yang, S. Cui, J. Gong, S.S. Miller, Q. Wang, Y. Hua, Stability of citral in oil-in-water emulsions protected by a soy protein–polysaccharide Maillard reaction product, Food Res. Int. 69 (2015) 357–363.

[7] D. Botrel, R.V. De De Barros, S. Vilela, Chapter 12 - microencapsulation of essential oils using spray drying technology, in: L.M. Sagis (Ed.), Microencapsulation and Microspheres for Food Applications, Academic Press, San Diego, 2015, pp. 235–251.

[8] P. García, J. Vega, P. Jimenez, J. Santos, P. Robert, Alpha-tocopherol microspheres with cross-linked and acetylated inulin and their release profile in a hydrophilic model, Eur. J. Lipid Sci. Technol. 115 (7) (2013) 811–819.

[9] N. Sosa, M.C. Zamora, J. Chirife, C. Schebor, Spray-drying encapsulation of citral in sucrose or trehalose matrices: physicochemical and sensory characteristics, Int. J. Food Sci. Technol. 46 (10) (2011) 2096–2102.

[10] N. Sosa, C. Schebor, O.E. Pérez, Encapsulation of citral in formulations containing sucrose or trehalose: emulsions properties and stability, Food Bioprod. Process. 92 (3) (2014) 266–274.

[11] K. Strassburger, W. Startup, V. Levey, T. Mattingly, J. Briggs, J. Harrison, T. Wilson, Enhanced stability of citral in juice beverages by applying cyclodextrin micro emulsion technology, in: N.C. Da Costa, R.J. Cannon (Eds.), Flavors in Noncarbonated Beverages, American Chemical Society, Washington, DC, 2010, pp. 143–158, Ch. 11.

[12] B. Bhandari, E. Dumoulin, H.M. Richard, I. Noleau, A. Lebert, Flavor encapsulation by spray drying: application to citral and linalyl acetate, J. Food Sci. 57 (1) (1992) 217–221.

[13] X. Yang, H. Tian, C. Ho, Q. Huang, Inhibition of citral degradation by oil-in-water nanoemulsions combined with antioxidants, J. Agric. Food Chem. 59 (11) (2011) 6113–6119.

[14] U. Ruktanonchai, W. Srinuanchai, S. Saesoo, I. Sramala, S. Puttipipatkhachorn, A. Soottitantawat, Encapsulation of citral isomers in extracted lemongrass oil with cyclodextrins: molecular modeling and physicochemical characterizations, Biosci. Biotechnol. Biochem. 75 (12) (2011) 2340–2345.

[15] X. Yang, H. Tian, C. Ho, Q. Huang, Stability of citral in emulsions coated with cationic biopolymer layers, J. Agric. Food Chem. 60 (1) (2012) 402–409.

[16] V. Weisheimer, D. Miron, C. Silva, S. Guterres, E. Schapoval, Microparticles containing lemongrass volatile oil: preparation, characterization and thermal stability, Die Pharmazie Intern. J. Pharm. Sci. 65 (12) (2010) 885–890.

[17] D. Miron, F. Battisti, T.C. Schwengber, P. Mayorga, E. Scherman, Spectrophotometric simultaneous determination of citral isomers in cyclodextrin complexes with partial least squares supported approach, Current Pharm. Anal. 8 (4) (2012) 401–408.

[18] P. López-Mahía, J. Simal, P. Paseiro, Infrared spectrophotometric determination of citral corrected for limonene interference in lemon and orange essential oils, Food Chem. 46 (2) (1993) 193–197.

[19] S. Kawano, Application to agricultural products and foodstuff, in: H.W. Siesler, Y. Ozaki, S. Kawata, H.M. Heise (Eds.), Near-Infrared Spectroscopy: Principles, Instruments, Application, 3rd edition, Wiley-VCH Verlag GmbH, Weinheim, Germany, 2007, pp. 269–287.

[20] Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond, N. Jent, A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies, J. Pharm. Biomed. Anal. 44 (3) (2007) 683–700.

[21] H.M. Heise, R. Winzen, Chemometrics in near-infrared spectroscopy, in: H. Siesler, Y. Ozaki, S. Kawata, H.M. Heise (Eds.), Near-Infrared Spectroscopy: Principles, Instruments, Application, 3rd edition, Wiley-VCH Verlag GmbH, Weinheim, Germany, 2007, pp. 125–162.

[22] A. Bustamante, L. Masson, J. Velasco, J.M. del Valle, P. Robert, Microencapsulation of H. pluvialis oleoresins with different fatty acid composition: kinetic stability of astaxanthin and alpha-tocopherol, Food Chem. 190 (2016) 1013–1021.

[23] C. Vergara, J. Saavedra, C. Sáenz, P. García, P. Robert, Microencapsulation of pulp and ultrafiltered cactus pear (Opuntia ficus-indica) extracts and betanin stability during storage, Food Chem. 157 (2014) 246–251.

[24] O. Escuredo, M.I. González, G. Wells, S. Fischer, J.M. Hernández, Amino acid profile of the quinoa (Chenopodium quinoa Willd.) using near infrared spectroscopy and chemometric techniques, J. Cereal Sci. 60 (1) (2014) 67–74.

[25] B. Steuer, H. Schulz, E. Läger, Classification and analysis of citrus oils by NIR spectroscopy, Food Chem. 72 (1) (2001) 113–117.

[26] G.M. ElMasry, S. Nakauchi, Image analysis operations applied to hyperspectral images for non-invasive sensing of food quality – a comprehensive review, Biosyst. Eng. 142 (2016) 53–82.

[27] N. Vásquez, C. Magán, J. Oblitas, T. Chuquizuta, H. Avila-George, W. Castro, Comparison between artificial neural network and partial least squares regression models for hardness modeling during the ripening process of swiss-type cheese using spectral profiles, J. Food Eng. 219 (2018) 8–15.

[28] W. Castro, J. Prieto, R. Guerra, T. Chuquizuta, W. Medina, B. Acevedo-Juárez, H. Avila-George, Feasibility of using spectral profiles for modeling water activity in five varieties of white quinoa grains, J. Food Eng. 238 (2018) 95–102.

[29] P. Oliveri, R. Simonetti, Chemometrics for food authenticity applications, in: Advances in Food Authenticity Testing, Elsevier, 2016, pp. 701–728, Ch. 25.

[30] T. Mehmood, K. Hovde, L. Snipen, S. Sæbø, A review of variable selection methods in partial least squares regression, Chemom. Intell. Lab. Syst. 118 (2012) 62–69.

[31] D. Liu, D. Sun, X. Zeng, Recent advances in wavelength selection techniques for hyperspectral image processing in the food industry, Food Bioprocess Technol. 7 (2) (2014) 307–323.

[32] L.M. Salchenberger, E. Cinar, N.A. Lash, Neural networks: a new tool for predicting thrift failures, Decis. Sci. 23 (4) (1992) 899–916.

[33] W. Castro, J. Oblitas, R. Santa-Cruz, H. Avila-George, Multilayer perceptron architecture optimization using parallel computing techniques, PLoS ONE 12 (12) (2017) 1–17.
[34] L. Pan, Q. Zhang, W. Zhang, Y. Sun, P. Hu, K. Tu, Detection of cold injury in peaches by hyperspectral reflectance imaging and artificial neural network, Food Chem. 192 (2016) 134–141.
[35] Q. Dai, J. Cheng, D. Sun, H. Pu, X. Zeng, Z. Xiong, Potential of visible/near-infrared hyperspectral imaging for rapid detection of freshness in unfrozen and frozen prawns, J. Food Eng. 149 (2015) 97–104.
[36] Y. Sun, Y. Liu, H. Yu, A. Xie, X. Li, Y. Yin, X. Duan, Non-destructive prediction of moisture content and freezable water content of purple-fleshed sweet potato slices during drying process using hyperspectral imaging technique, Food Anal. Methods 10 (5) (2017) 1535–1546.
[37] M. Elskens, D. Vloeberghs, L. Liesbeth Van-Elsen, W. Baeyens, L. Goeyens, Multiple testing of food contact materials: a predictive algorithm for assessing the global migration from silicone moulds, Talanta 99 (2012) 161–166.
[38] T. Fearn, Assessing calibrations: *sep*, *rpd*, *rer* and r^2, NIR News 13 (6) (2002) 12–14.
[39] E. Furtado, A. Bridi, D. Barbin, A. Barbon, N. Andreo, B. Giangareli, D. Terto, J. Batista, Prediction of pH and color in pork meat using VIS-NIR Near-infrared spectroscopy (NIRS), Food Sci. Technol. 39 (1) (2019) 88–92.
[40] N.D. Wilson, M. Ivanova, R.A. Watt, A.C. Moffat, The quantification of citral in lemongrass and lemon oils by near-infrared spectroscopy, J. Pharm. Pharmacol. 54 (9) (2002) 1257–1263.
[41] L.S. Magwaza, U.L. Opara, H. Nieuwoudt, P.J. Cronje, W. Saeys, B. Nicolaï, Nir spectroscopy applications for internal and external quality analysis of citrus fruit—a review, Food Bioprocess Technol. 5 (2) (2012) 425–444.
[42] J.S. Shenk, J.J. Workman, M.O. Westerhaus, Chapter 17 - application of NIR spectroscopy to agricultural products, in: D.A. Burns, E.W. Ciurczak (Eds.), Handbook of Near-Infrared Analysis, 3rd edition, Taylor & Francis Group, New York, USA, 2008, pp. 347–386.
[43] S. Munera, J. Amigo, J. Blasco, S. Cubero, P. Talens, N. Aleixos, Ripeness monitoring of two cultivars of nectarine using VIS-NIR hyperspectral reflectance imaging, J. Food Eng. 214 (2017) 29–39.
[44] S. Shafiee, S. Minaei, Combined data mining/NIR spectroscopy for purity assessment of lime juice, Infrared Phys. Technol. 91 (2018) 193–199.
[45] V. Esposito, W. Chin, J. Henseler, H. Wang, Handbook of Partial Least Squares: Concepts, Methods and Applications, Springer, 2010.
[46] L. Van der Maaten, E. Postma, Dimensionality Reduction: A Comparative Review, Tech. rep., Tilburg centre for Creative Computing, 2009.
[47] J. Vega-Vilca, J. Guzmán, Regresión pls y pca como solución al problema de multicolinealidad en regresión múltiple, Rev. Mat. Teor. Apl. 18 (1) (2011) 9–20.
[48] H.R. Juliani, J. Kapteyn, D. Jones, A.R. Koroch, M. Wang, D. Charles, J.E. Simon, Application of near-infrared spectroscopy in quality control and determination of adulteration of African essential oils, Phytochem. Anal. 17 (2) (2006) 121–128.
[49] H. Schulz, B. Schrader, R. Quilitzsch, B. Steuer, Quantitative analysis of various citrus oils by ATR/FT-IR and NIR-FT Raman spectroscopy, Appl. Spectrosc. 56 (1) (2002) 117–124.
[50] H. Schulz, R. Quilitzsch, H. Krüger, Rapid evaluation and quantitative analysis of thyme, origano and chamomile essential oils by ATR-IR and NIR spectroscopy, J. Mol. Struct. 661–662 (2003) 299–306.
[51] G. Santos, I. Fernández-Olmo, A. Irabien, F. Ledoux, D. Courcot, PLSR and ANN estimation models for PM10-bound heavy metals in Dunkerque (Northern France), in: 1st International Conference on Atmospheric Dust, vol. 1, 2014, pp. 100–105.
[52] B. Kuang, Y. Tekin, A. Mouazen, Comparison between artificial neural network and partial least squares for on-line visible and near infrared spectroscopy measurement of soil organic carbon, pH and clay content, Soil Till. Res. 146 (2015) 243–252.
[53] R.M. Balabin, S.V. Smirnov, Variable selection in near-infrared spectroscopy: benchmarking of feature selection methods on biodiesel data, Anal. Chim. Acta 692 (1–2) (2011) 63–72.
[54] M. Prakash, J.K. Sarin, L. Rieppo, I.O. Afara, J. Töyräs, Optimal regression method for near-infrared spectroscopic evaluation of articular cartilage, Appl. Spectrosc. 71 (10) (2017) 2253–2262.