Chemical composition and antioxidant capacity in strawberry cultivars (*Fragaria x ananassa* Duch.) by FT-MIR spectroscopy and chemometrics

Patricia Minutti-López Sierra, Tzayhri Gallardo-Velázquez, Guillermo Osorio-Revilla and Ofelia Gabriela Meza-Márquez

**ABSTRACT**

FT-MIR spectroscopy coupled to chemometrics can be an alternative technique to conventional methods to determine the chemical composition of berries such as strawberry. This study developed chemometric models based on FT-MIR spectroscopy to identify strawberry cultivars at seven stages of ripening and quantify the main attributes involved in sensory quality of strawberry fruits (total soluble solids, total acidity, reducing sugars, pH) and the main bioactive compounds involved in antioxidant capacities (ascorbic acid, phenolics, flavonoids, anthocyanins). PCR, PLS1 and PLS2 algorithms were used to develop the prediction models. PLS1 algorithm developed excellent predictions with $R_c^2$ greater than 0.90. SIMCA model identified the cultivars with 99% confidence. Unlike conventional methods, FT-MIR allowed us to analyze simultaneously the attributes of quality, bioactive compounds, and antioxidant capacity of strawberries fruits in a fast and reliable way without the use of reagents and/or solvents, which considerably reduces the time and cost of the analyzes.

1. Introduction

Strawberries (*Fragaria x ananassa* Duch.) are cultivated and consumed worldwide. Strawberry fruits contain important amounts of phenolic compounds with beneficial health effects, such as antioxidant, antimutagenic, and anti-inflammatory, among others (Aaby, Mazur, Nes, & Skrede, 2012; Pinelli et al., 2011). Recently, it has been considered necessary to quantify the main attributes involved in the sensory quality of strawberry fruits (such as total soluble solids, total acidity, reducing sugars, pH), as well as the main bioactive compounds involved in the antioxidant capacity (ascorbic acid, phenolics, flavonoids, anthocyanins), because it is a fruit rich in antioxidants and its chemical and economic relevance requires higher quality standards (Aaby et al., 2012; Buendia et al., 2010; Pinelli et al., 2011; Rios de Souza et al., 2014; Young-Jun & Youngjae, 2015). Nevertheless, to analyze these compounds, spectrophotometric and chromatographic methods are used, which are laborious, costly, time consuming, and use toxic chemical products (Buendia et al., 2010).

Nowadays, FT-MIR spectroscopy coupled to chemometrics is an alternative used to analyze quality attributes, bioactive compounds, and antioxidant capacity in foods (Mohamed, Mohamed, Khalil, Hussein, & Kamil, 2011; Tamaki & Mazza, 2011). Multivariate statistical techniques exhibit advantages over conventional methods because this methodology does not use reagents nor solvents, it is fast and does not require pretreatment of the sample and extraction processes, which leads to a reduction in costs and time of the analyses. Also, FT-MIR spectroscopy coupled to multivariate analysis offers the possibility of analyzing several compounds simultaneously. There are several studies...
that use FT-MIR spectroscopy coupled to chemometrics to determine quality attributes and chemical composition of fruits, such as apricot, blueberries, grapes, blackberries and apple (Mohamed et al., 2011). However, in the revised literature, we did not find studies on strawberries that determine simultaneously the aforementioned compounds by FT-MIR spectroscopy coupled to multivariate analysis. FT-MIR spectroscopy coupled to chemometrics would allow to quickly select the most promising strawberry cultivars for the farmer based on their quality attributes, as well as their bioactive compounds.

Therefore, the objective of this study was to develop an analytical alternative to quantify quality attributes (total soluble solids, total acidity, reducing sugars, pH), bioactive compounds (ascorbic acid, phenolics, flavonoids, anthocyanins), and antioxidant capacity in ten strawberry cultivars in seven ripeness stages by means of FT-MIR spectroscopy coupled to chemometric tools. Also, a SIMCA model was developed to identify and classify the strawberry cultivars.

2. Materials and methods

2.1. Standards and chemicals

All chemicals were of analytical grade. Folin-Ciocalteu’s phenol reagent, 1,1-diphosphonyl-2-picyrhydral (DPPH), 2,2’-Azino-bis (3-ethylbenzthiazoline-6-sulfonic acid) (ABTS), 6-hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid (Trolox), 3,5-dinitrosalicylic acid (DNS), 3,4,5-Trihydroxybenzoic acid (gallic acid), 2,6-dichlorophenol, indophenol, ascorbic acid, glucose, and catechin were purchased from Sigma Chemical Company. All chemicals were of analytical grade. Folin-Ciocalteu colorimetric method at 760 nm; quantifications were calculated by a calibration curve prepared daily with known concentrations of gallic acid (GA) as a standard; the results were expressed in milligrams equivalent of gallic acid per 100 g of fresh weight (mg EqG/100 g FW). Total flavonoids content (TFC) was determined according to Singleton and Rossi (1965) with the Folin-Ciocalteau colorimetric method at 760 nm; quantifications were calculated by a calibration curve prepared daily with known concentrations of gallic acid (GA) as a standard; the results were expressed in milligrams equivalent of gallic acid per 100 g of fresh weight (mg EqG/100 g FW). Total flavonoids content (TFC) was determined according to Tulipani et al. (2011); catechin standard solutions were used for the calibration curve and the absorbance was measured at 510 nm; the results were expressed as milligrams equivalent of catechin per 100 g of fresh weight (mg EqC/100 g FW). The total anthocyanins content (TAC) was determined using a pH differential method (Giusti & Wrolstad, 2001). The absorbance at 520 nm and 700 nm was measured and the absorbance at 700 nm was subtracted from the absorbance at 520 nm; the results were expressed as milligrams equivalent of cyanidin-3-glucoside per 100 g of fresh weight (mg EqC-3-gluc/100 g FW).

2.5. Antioxidant capacity

Total antioxidant capacity was determined in triplicate using the Trolox equivalent antioxidant capacity (TEAC) by means of two methodologies: ABTS+ (2,2’-azino-bis (3-ethylbenzthiazoline-6-sulfonic acid) assay (Moura et al., 2002) and DPPH (1,1-diphenyl-2-picyrhydral) assay (Brand-Williams, Cuvelier, & Berzet, 1995). Aqueous standard dilutions of Trolox (6-hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid) were used for calibration, and the results were expressed as micromoles equivalent of Trolox per 100 g of fresh weight (µEqT/100 g FW).

2.6. Statistical analysis

The values of TSS, TA, RS, pH, AAC, TPC, TFC, TAC, TEAC-ABTS and TEAC-DPPH were analyzed through one-way ANOVA and Tuckey multiple comparisons of means (p ≤ 0.05). Before developing the chemometric models, a principal component analysis (PCA) was performed to know the relation among variables. Moreover, a Pearson correlation was
performed to find a correlation between the analyzed compounds and the ripeness stage of the analyzed samples. For the analysis, the Minitab version 16.1.0 software (State College, PA, USA) was used.

2.7. FT-MIR spectra

FT-MIR spectra were recorded in a model GX FTIR spectrophotometer (PerkinElmer®, Massachusetts, USA), equipped with an attenuated total reflectance (ATR) and a selenide zinc (ZnSe) crystal (model 022-12xx, PerkinElmer®, Massachusetts, USA) with a 45° incidence angle and a deuterated triglycerine sulfate (DTGS) detector. Data acquisition was achieved with the Spectrum software (version 5.3.1. PerkinElmer®, Massachusetts, USA).

FT-MIR spectra were obtained in the 4000-550 cm⁻¹ region with 64 scans, resolution of 4 cm⁻¹, in absorbance (A) units. Before obtaining the FT-MIR spectra, the reference spectrum was recorded by reading the empty ZnSe crystal under the same conditions as those of the samples. To obtain the FT-MIR spectra, approximately 3 g of macerated strawberries was placed on the ZnSe crystal, covering the whole crystal and avoiding the formation of bubbles. The FT-MIR spectra were recorded in triplicate with the Spectrum software (version 5.3.1. PerkinElmer®, Massachusetts, USA). A total of 210 FT-MIR spectra were recorded, which were used for the multivariate analysis. After ending each reading, the ZnSe crystal was washed with liquid detergent (Extran®), then, it was rinsed with distilled water and left to air dry.

2.8. Multivariate analysis

2.8.1. Quantitative models

For these models, the Spectrum Quant+ version 4.51.02 software (PerkinElmer®, Massachusetts, USA) was used, which contains the partial least squares (PLS1 and PLS2) and the principal components regression (PCR) algorithms to establish the correlation between the spectral data and the chemical composition values (TSS, TA, RS, pH, AAC, TPC, TFC, TAC) and antioxidant capacity (ABTS and DPPH assays). The regression model was created with 80% of the samples (168 FT-MIR spectra) and the remainder 20% (42 FT-MIR spectra) was used to validate the model.

The spectral pretreatments to improve the prediction of the chemometric model were: spectral smoothing (Savitzky-Golay filter, 19 points), base line correction to avoid changes and/or displacements in it, second derivative (Savitzky-Golay filter, 13 points), and normalization of the variable normal standard (SNV)-type to eliminate or minimize interferences caused by dispersion effects.

The selection of the best model depended on: 1) number of factors (latent variables) for PLS or principal components for PCR, 2) coefficient of determination of calibration (R²c, must be as close to 1 as possible) and, 3) standard error of calibration (SEC, must be as low as possible). To assess the external validation stage, we analyzed: 1) coefficient of determination of calibration (R²c, must be as close to 1 as possible), 2) the standard error of prediction (SEP, must be as low as possible), 3) Mahalanobis distance (must be lower than 1), 4) residuals relation (must be lower than 3), 5) residual error (must be as low as possible), 6) the relative difference percentage (% RD, must be lower than 10%), and 7) residual predicted deviation (RPD) was calculated for each parameter, RPD is defined as the ratio between the standard deviation of the population (SD) and the SEP, high values of RPD are desirable.

2.8.2. Soft independent modeling of class analogies (SIMCA)

SIMCA is one of the most used analyses in chemometrics because it is a supervised classification method based on PCA that is used to reduce the dimensionality of the set of multivariate data. Thus, SIMCA helps to analyze the similarity or difference among groups of spectra (classes) by constructing a reliable region (confidence level ≥99%) around each class (Kumar, Bansal, Sarma, & Rawal, 2014).

With the Assure ID Method Explorer version 4.0.2.0175 software (PerkinElmer®), 10 classes were established: Albión, Camino Real, Celin, Dorado, Dulce Ana, Festival, Fortuna, Osceola, San Andreas, and Spartan. To develop the SIMCA model, 210 FT-MIR spectra were obtained (21 of each class), of these, 168 FT-MIR spectra (80% of the samples) were used to calibrate the model and 42 spectra (20% of samples) were used to validate it. The model was optimized by applying different mathematical pretreatments: baseline correction, spectral smoothing (Savitzky-Golay filter, 19 points), ambient filters (CO₂ and H₂O), second derivative (Savitzky-Golay filter, 9 points), SNV-type normalization.

To evaluate the performance of the SIMCA model, we analyzed: 1) 3D projection of the classes that allows for the visualization of groupings among classes using the first three principal components; 2) interclass distance between the ten classes, which is an indicator of the similirude of classes, and it should be greater than 3; and, 3) percentage of recognition (sensitivity) and rejection (specificity) of samples, a value of 100% describes a model that differentiates the classes correctly.

3. Results and discussion

3.1. Chemical composition

The results for TSS, TA, RS, and pH in the 10 strawberry cultivars at seven ripeness degrees are presented on Supplementary Figure 1S. TSS content varied from 5.59% in the DAN0 cultivar to 10.73% in SPA6. Flavor is one of the most important attributes in strawberries and is due by TSS, like saccharine, fructose, and glucose, the main sugars contained in this fruit. TA varied from 0.53% in OSC6 to 0.94% in OSC0. The RS content varied from 1.10% in FES0 to 10.73% in SPA6. The pH, represented as glucose percentage, are the most important attributes in strawberries and provide energy for the metabolic changes in the fruit, aside from influencing markedly the flavor of the fruit. The pH varied from 2.87 in CAR0 to 3.89 in DOR6. Values of TSS, TA, RS, and pH showed a significant difference (p ≤ 0.05), because, during ripening of the fruit, the TA diminished, whereas TSS, AR, and pH increased due to enzymatic unfolding of starch, hence, the fruit is sweeter. Likewise, during the ripening process, the concentration of organic acids diminishes, particularly that of the citric acid, which leads to a diminution in TA and an increase in pH. Results of TSS, TA, RS, and pH in the 10 strawberry cultivars under different ripeness degrees coincide with data reported by other authors (Pineli et al., 2011; Ríos de Souza et al., 2014).
3.2. Bioactive compounds

The results for AAC, TPC, TFC, and TAC in the 10 strawberry cultivars at seven ripening degrees are presented on Supplementary Figure 2S. The AAC varied from 31.87 to 143.30 mgAA/100 g FW in SPA6 and FES0, respectively. The AAC was higher in green (0–1 ripening degree) and rose-colored (2–5 ripening degree) fruits, but AAC diminished in the red fruit (6 ripening degree), this behavior is similar to that observed by Pineli et al. (2011). The aforementioned indicates that in green (0 ripeness) and rose-colored (2–5 ripening degree) fruits, a higher synthesis of AA occurs than in ripe fruits. The TPC varied from 46.30 to 358.02 mg EqGA/100 g FW in FOR6 and CEL0, respectively. The TFC increased from 1.35 to 5.56 mg EqC/100 g FW in FES6 and FES0, respectively. The TPC and TFC is higher in green (0–1 ripeness) and rose-colored (2–5 ripening degree) fruits. The least content of these compounds is found in red fruits (6 ripeness degree). This behavior is due to changes in the composition of the fruit occurring during the ripening process. Ferreya, Viña, Mugridge, and Chaves (2007) observed that the TPC is higher in rose-colored fruits due to the accumulation of anthocyanins which have a typical flavonoid structure, and they are responsible for the characteristic red to purple color of berries (Haminiuk, Maciel, Plata-Oviedo, & Peralta, 2012).

Color is an important attribute of strawberries and it is partially defined by the anthocyanins content. TAC in the 10 strawberry cultivars varied from 0.04 to 30.33 mg EqC-3-glc/100 g FW in SPA0 and CAR6, respectively. TAC increased as the degree of ripeness increased, thus, the highest anthocyanins content will be found in ripe fruits. This behavior is in agreement with Pineli et al. (2011) and Aaby et al. (2012), who observed the same tendency in Brazilian and Norwegian strawberries, respectively. The aforementioned is a behavior posed by the ripening process of strawberries, as there is an increase in anthocyanins content as the fruit ripens and gets an intense red color. Kalt, Lawand, Ryan, McDonald, and Donner (2003) suggest that this increment is because anthocyanins are synthesized from the phenolic compounds, leading to a diminution of the latter.

3.3. Antioxidant capacity

Total antioxidant capacity varied, according to the ABTS method, from 553.82 to 1729.57 µmoles EqT/100 g FW in FES6 and SPA0, respectively. According to the DPPH radical, it varied from 345.11 to 1371.60 µmoles EqT/100 g FW in DOR6 and FES0, respectively (Supplementary Figure 3S). The ABTS assay yielded a higher antioxidant capacity, probably because this radical measured the antioxidant capacity of hydrophilic and lipophilic compounds, whereas the DPPH assay only measures that of hydrophobic compounds. Antioxidant capacity is higher in green (0–1 ripeness) and rose-colored (2–5 ripeness) fruits than in red fruits (6 ripeness degree). The latter is because green and rose-colored fruits need a higher amount of antioxidants as they are more susceptible to pests and diseases. Results on the antioxidant capacity agree with those reported by Pineli et al. (2011).

Results of TSS, TA, RS, pH, AAC, TPC, TFC, TAC, TEAC-ABTS and TEAC DPPH in the 10 strawberry cultivars at seven ripeness degrees showed variability (Tukey test, p ≤ 0.05), which is fundamental to build the proposed chemometric models.

3.4. Principal components analysis (PCA)

The PCA allowed to examine sample grouping and detect outliers (Figure 1). The PCA showed that three PC explained 81.80% of the variability of the samples (PC1 53.90%, PC2 18.80% and PC3 9.10%). The 3D scatter plot (Figure 1(a)) shows the distribution of the ten strawberry cultivars at seven ripeness stages. The 3D graph does not show clustering effect and the data showed variability, which is fundamental to construct the prediction models. On the other hand, the principal components loading plot (Figure 1(b)) shows the correlation between the variables. The PCA developed a graph that represents the loads obtained with the first two principal components, explaining 72.70% of the variability. The first component explains 53.90% of the variability and is associated with TSS, pH and RS; these parameters are related because as the ripening of the fruit progresses, their content increases. The TPC, TFC, TEAC-
ABTS and TEAC DPPH are related in such a way that the antioxidant capacity of the fruits can be attributed to TPC and TFC. Some antioxidant compounds (phenolic compounds, flavonoids, and anthocyanins) can be synergistic and even inhibitory. Young-Jun and Youngjae (2015) found a relation of phenolic compounds and flavonoids with the antioxidant capacity, this agrees with the present work because, based on the PCA, the antioxidant capacity of strawberries is related to TPC and TFC. The second component explains 18.80% of the variability and is associated with TA, AAC, and TAC. The TA is located opposed to the pH, because as the fruit ripens, the pH increases and the acidity decrease.

### 3.5. Pearson’s correlation analysis

Table 1 reveals the positive correlation of TSS, RS, pH and TAC with the degree of ripeness (r = 0.6561, r = 0.6777, r = 0.3644, and r = 0.9108 respectively), because these compounds increase with the advancing ripeness degree of the fruit. TA shows a negative correlation with the degree of ripeness (r = −0.4478), because, during ripening, there is an increase in TSS, RS, and pH, and a diminution in the TA. The AAC, TPC, TFC, and antioxidant capacity (ABTS and DPPH assays) have a negative correlation with the ripeness degree (r = −0.0759, r = −0.2297, r = −0.3438, r = −0.1659, and r = −0.3486, respectively), suggesting that these compounds diminish with the increasing ripeness degree. Finally, TAC presents a positive correlation coefficient with the ripeness degree (r = 0.9108), because anthocyanins content increase as the fruit ripens. The aforementioned agrees with the results reported by Pineli et al. (2011) and Aaby et al. (2012).

### 3.6. FT-MIR spectra

The chemical composition of the strawberry cultivars can be correlated with the FT-MIR spectra of the samples. Figure 2 depicts the FT-MIR spectra of the ten strawberry cultivars at seven ripeness stages. Interpretation of the FT-MIR spectra were made based on the literature (Mohamed et al., 2011). The broad band at 3640-3200 cm⁻¹ is due to tension vibrations of the OH bonds of phenolic compounds and water. The band at 3380 cm⁻¹ corresponds to tension vibrations of OH bonds of carbohydrates, organic acids (citric acid) or amino acids (tyrosine). Peaks at 2900-2800 cm⁻¹ are due to tension vibrations of CH bonds of methyl and methylene groups. The band at 1730 cm⁻¹ indicates the presence of ester-type bonds (C = O) such as the ethyl hexanoate, and methyl and ethyl butanoate, which influence the fragrance of the fruit. The region at 1622 cm⁻¹ is due to the deformation movements of the hydroxyl group (OH). The peak at 1415 cm⁻¹ corresponds to the deformation movements of groups OCH, COH, and CCH characteristic of fructose. The region at 1270-1180 cm⁻¹ is due to flexion vibrations of the OH of carboxylic acids and phenolic compounds. The region at 1200-950 cm⁻¹ corresponds to vibrations of the C-O-C bonds present in disaccharides, polysaccharides, and glycosylated anthocyanins. The region at 1045 cm⁻¹ corresponds to vibrations of C-O, C-C-C and the pyranose ring of glucose. Finally, at 927 cm⁻¹, pectin-characteristic balancing vibrations occur.

### 3.7. Multivariate analyses

#### 3.7.1. Quantitative models

Based on Table 2, PLS1 developed the best predictive model by obtaining higher Rc² values (0.999) and lower values of SEC (0.03–1.05) and SEP (0.03–1.07) as compared to models developed with PLS2 (Rc²: 0.635–0.897, SEC: 0.04–315.73, SEP: 0.05–670.01) and PCR (Rc²: 0.108–0.971, SEC: 0.04–207.51, SEP: 0.05–516.63). Values of Rc² higher than 0.90 suggest excellent quantitative information, whereas values of 0.81–0.66 represent poorly reliable quantitative predictions (Tamaki & Mazza, 2011). SEC and SEP are statistical parameters that evaluate the accuracy with which the calibration and prediction samples adjust to regression; hence, low values, like those obtained for PLS1, correspond to a better adjustment of data. Also, the standard error (SE) of the reference methods was calculated in order to compare the calibrated models obtained. The SE is an important indicator of how accurate an estimate of the population parameter of the statistical sample is; a lower standard error value indicates a more accurate estimate. As seen in Table 2, the standard error values were from 0.06 to 1.69, these results confirmed the accuracy of the developed model, because the standard error values were low, very similar to the values of SEC and SEP.

![Figure 2. FT-MIR spectra of the ten strawberry cultivars at seven ripeness stages.](image-url)
Table 2. Calibration data of the chemometric models to predict the chemical composition, bioactive compounds, and antioxidant capacity in ten strawberry cultivars at seven ripeness stages.

Tabla 2. Dados de calibración de los modelos quimiométricos para predecir la composición química, los compuestos bioactivos y la capacidad antioxidante en diez cultivares de fresa en siete etapas de madurez.

| Algorithm | Parameter (units) | UV-Vis spectroscopy | Calibration (n = 168) | Validation (n = 42) |
|-----------|------------------|---------------------|----------------------|---------------------|
|           | Range            | Mean                | SD                | SEC               | Rx   | SDP  | MD    | RR    | RE    | % RD | RPD  |
| PLS1      | TSS (%)          | 5.59–10.73          | 8.18              | 0.22              | 0.13  | 11   | 0.999 | 0.03  | 0.25–0.31 | 0.58–1.82 | 0.98–1.08 | 0.69–3.78 | 7.33 |
|           | TA (% of citric acid) | 0.53–0.94          | 0.80              | 0.49              | 0.06  | 10   | 0.999 | 0.06  | 0.07     | 0.15–0.34 | 1.38–1.79 | 1.08–1.13 | 1.41–4.93 | 7.01 |
|           | RS (% glucose)   | 1.10–2.89           | 2.06              | 0.54              | 0.14  | 11   | 0.999 | 0.06  | 0.08     | 0.08–0.37 | 1.55–2.78 | 1.64–1.78 | 0.64–6.17 | 6.75 |
|           | pH               | 2.87–3.89           | 3.40              | 0.78              | 0.09  | 11   | 0.999 | 0.08  | 0.09     | 0.13–0.31 | 1.45–1.79 | 1.31–1.89 | 0.31–5.77 | 8.66 |
|           | AAC (mg/gFW)     | 31.87–143.30        | 90.68             | 0.89              | 0.11  | 12   | 0.999 | 0.06  | 0.07     | 0.11–0.33 | 1.58–1.92 | 0.91–0.99  | 0.61–3.34 | 12.71 |
|           | TPC (mg EqC/100 g FW) | 46.30–338.02       | 169.64            | 14.11             | 1.69  | 13   | 0.999 | 0.08  | 0.95     | 0.15–0.37 | 1.55–2.51 | 0.38–1.12 | 0.98–8.17 | 13.18 |
|           | TFC (mg EqC/100 g FW) | 1.35–5.56          | 3.17              | 0.59              | 0.07  | 12   | 0.999 | 0.04  | 0.99     | 0.14–0.36 | 1.41–1.88 | 0.96–1.67 | 1.33–4.76 | 9.83 |
|           | TAC (mg EqC-3-gluc/100 g FW) | 0.04–30.33       | 10.38             | 6.87              | 0.82  | 11   | 0.999 | 0.77  | 0.972     | 0.13–0.35 | 1.69–2.09 | 1.12–1.54 | 0.22–5.46 | 8.09 |
|           | TEAC-ABTS (µmolEq/100 g FW) | 553.82–1729.57   | 1150.41           | 8.73              | 1.04  | 14   | 0.999 | 0.86  | 0.998     | 0.16–0.36 | 1.84–2.31 | 1.36–1.61 | 4.73–6.75 | 9.48 |
|           | TEAC-PPH (µmolEq/100 g FW) | 345.11–1371.60  | 775.12            | 9.45              | 1.13  | 15   | 0.999 | 1.05  | 0.999     | 0.17–0.38 | 1.46–2.49 | 1.08–1.34 | 0.95–3.89 | 9.01 |
| PLS2      | TSS               | 0.823              | 0.49              | -                 | -     | 8    | 0.861 | 0.04  | 0.05      | -         | -          | -         | -         | - |
|           | TA                | 0.888              | 0.12              | -                 | -     | 8    | 0.897 | 0.06  | 0.08      | -         | -          | -         | -         | - |
|           | RS                | 0.677              | 0.21              | -                 | -     | 4    | 0.677 | 0.21  | 2.11      | -         | -          | -         | -         | - |
|           | AAC               | 0.635              | 44.71             | -                 | -     | 4    | 0.824 | 3.72  | 3.98      | -         | -          | -         | -         | - |
|           | TPC               | 0.825              | 0.29              | -                 | -     | 8    | 0.825 | 0.29  | 0.34      | -         | -          | -         | -         | - |
|           | TFC               | 0.843              | 151.20            | -                 | -     | 8    | 0.843 | 151.20 | 187.41    | -         | -          | -         | -         | - |
|           | TAC               | 0.889              | 315.73            | -                 | -     | 6    | 0.889 | 315.73 | 670.01    | -         | -          | -         | -         | - |
| PCR       | TSS               | 0.695              | 0.63              | -                 | -     | 6    | 0.921 | 0.04  | 0.05      | -         | -          | -         | -         | - |
|           | TA                | 0.770              | 0.18              | -                 | -     | 9    | 0.929 | 0.05  | 0.06      | -         | -          | -         | -         | - |
|           | AAC               | 13.049             | 16.76             | -                 | -     | 13   | 0.849 | 16.76 | 21.44     | -         | -          | -         | -         | - |
|           | TPC               | 8.700              | 45.08             | -                 | -     | 8    | 0.700 | 45.08 | 50.71     | -         | -          | -         | -         | - |
|           | TFC               | 5.637              | 5.15              | -                 | -     | 5    | 0.637 | 5.15  | 5.75      | -         | -          | -         | -         | - |
|           | TAC               | 1.018              | 0.61              | -                 | -     | 1    | 0.108 | 0.61  | 0.63      | -         | -          | -         | -         | - |
|           | TEAC-ABTS        | 4.666              | 207.51            | -                 | -     | 4    | 0.666 | 207.51 | 223.61    | -         | -          | -         | -         | - |
|           | TEAC-PPH         | 12.571             | 175.92            | -                 | -     | 12   | 0.571 | 175.92 | 516.63    | -         | -          | -         | -         | - |

TSS: total soluble solids, TA: total acidity, RS: reducing sugars, pH, AAC: ascorbic acid content, TPC: total phenolic content, TFC: total flavonoids content, TAC: total anthocyanins content, TEAC-ABTS: trolox equivalent antioxidant capacity with ABTS assay, TEAC-PPH: trolox equivalent antioxidant capacity with DPPH assay. SD: Standard deviation; SE: Standard error; Factors (PLS) or principal components (PCR); SEC: determination coefficient in calibration, it must be as close as possible to 1; SD: standard error of calibration, must be as low as possible; RV: determination coefficient in validation, must be as close to 1 as possible; SEP: standard error of prediction, must be as low as possible; RPD: residual predicted deviation (SD/SEP), must be as high as possible.

TSS: sólidos solubles totales, TA: acidez total, RS: azúcares reductores, pH, AAC: contenido de ácido ascórbico, TPC: contenido fenólico total, TFC: contenido total de flavonoides, TAC: contenido total de antocianinas, TEAC-ABTS: capacidad antioxidante equivalente de trolox con ensayo ABTS, TEAC-PPH: capacidad antioxidante equivalente de trolox con ensayo DPPH. SD: desviación estándar; SE: errores de factores (PLS) o componentes principales (PCR); SEC: coeficiente de determinación en la calibración, debe ser lo más cercano a 1; SD: error estándar de calibración, debe ser lo más bajo posible; RV: coeficiente de determinación en la validación, debe ser lo más cercano a 1; SEP: error estándar de predicción, debe ser lo más bajo posible; RPD: desviación residual predicha (SD/SEP), debe ser lo más alta posible.
On the other hand, factors represent a significant source of data variations in FT-MIR spectra, so that small values will result in a model without enough information to predict adequately (under-estimated model) and too high values will make the model to have too much information (over-estimated model). Thus, the optimal number of factors is that which provides the lowest SEP (Table 2) and good prediction results in the validation samples.

There are two versions of PLS: PLS1 and PLS2. PLS1 calibrates properties one by one, and PLS2 calibrates properties simultaneously. In contrast to PLS, PCR uses the properties of decomposing in principal components, performing an inverse multiple regression of the property to be determined on the scores (which reveal differences or similarities among samples) instead of performing it on the original data. PLS is the preferred method in multivariate calibration, hence, it is one of the most common regression algorithms in the field of chemometrics (Karoui, Downey, & Blecker, 2010).

Table 2 presents the statistical predictions data with PLS1. According to Tamaki and Mazza (2011), excellent correlations ($R^2 > 0.91$) were obtained between the actual and the predicted values for each parameter. Values of the Mahalanobis distance ($\leq 1$), residuals relation ($\leq 3$), and residual error (must be as low as possible) comply with established criteria. Also, % RD for all validation samples were lower than 9%; according to Granato and Ares (2014), percentages of RD $\leq 10\%$ are considered excellent. Finally, RPD values ranged from 6.75 to 13.18 for all parameters, according to Williams (2001), RPD values between 3.1 and 4.98 are considered acceptable, between 5 and 6.39 are good, and between 6.4 and 8 are very good for analytical purposes. These results indicated that the model developed with PLS1 performed an excellent prediction.

### 3.7.2. SIMCA model

In the SIMCA model, the spectral region between 3500-2800 and 1750-950 cm$^{-1}$ showed adequate spectral variability to discriminate the ten classes of strawberry. Figure 3 presents the spatial distribution of the 3D modeled classes, consisting of a geometric shape known as hyperbox, hyperspace, or cloud. The limit of hyperboxes around each group represents the 99% confidence level and each point represents one FT-MIR spectrum (sample). According to Figure 3, the SIMCA model identified the classes in well-defined groups, no extreme or overlapped samples were presented, and therefore, the model was able to identify the 10 strawberry cultivars.

The above-mentioned is confirmed by the results depicted on Table 3, which present the interclass distance given by the SIMCA model. In the 10 classes, the interclass distance was greater than 3, which indicates that the populations had different spectral characteristics. The interclass distance is the geometrical distance that exists among the hyperboxes and represents the 99% confidence that the sample pertains or not to a class. An interclass distance greater than 3 indicates that the populations were well separated without overlap between the classes. In contrast, an interclass distance lower than 3 indicates that hyperboxes are similar, and a distance equal to 0 indicates that they are identical (Kumar et al., 2014).

Also, the SIMCA model determines the recognition and rejection rates. Recognition (sensitivity) indicates how many samples are recognized as pertaining to a determined class, whereas rejection (specificity) indicates how many samples do not pertain to a determined class. Recognition and rejection rates of 100% were obtained for the ten classes, i.e., if the class that was analyzed was the Albión cultivar, the FT-MIR spectra pertaining to this class were identified as Albion (100% recognition). In contrast, the FT-MIR spectra that do not pertain to the Albión class were rejected (100% rejection). Thus, the model classified samples in the correct category.

External validation results of the SIMCA model indicate that the samples were recognized and classified correctly (Table 4). The total distance ranged from 0.07 to 0.87, this value indicates whether the sample was correctly classified (must be lower than 1.0). Residual distance ranged from 0.34 to 1.87, a value higher than 3 indicates that the sample contains a variation source not found previously. External validation

### Table 2. Distancia interclases de las clases derivadas del modelo SIMCA.

|          | Albión | Camino real | Celin | Dulce Ana | Dorado | Festival | Fortuna | Osceola | San Andreas | Spartan |
|----------|--------|-------------|-------|-----------|--------|----------|---------|---------|-------------|---------|
| Albión   | 0      | 59.20       | 26.50 | 86.40     | 92.30  | 50.20    | 60.40   | 56.30   | 16.30       | 42.60   |
| Camino real | 0     | 58.70       | 32.00 | 70.90     | 55.70  | 23.50    | 49.60   | 50.10   | 23.50       | 42.30   |
| Celin    | 0      | 84.50       | 65.30 | 72.10     | 19.30  | 48.60    | 70.70   | 59.10   | 41.60       | 38.40   |
| Dulce Ana| 0      | 62.70       | 55.40 | 49.00     | 59.60  | 43.40    | 49.00   | 31.80   | 31.60       | 38.00   |
| Dorado   | 0      | 68.30       | 53.80 | 34.40     | 68.10  | 34.40    | 87.00   | 31.80   | 41.40       | 38.00   |
| Festival | 0      | 62.70       | 55.40 | 34.40     | 68.10  | 34.40    | 49.00   | 31.80   | 31.60       | 38.00   |
| Fortuna  | 0      | 41.60       | 51.50 | 38.50     | 41.60  | 41.60    | 51.50   | 38.50   | 31.60       | 38.00   |
| Osceola  | 0      | 52.40       | 41.40 | 41.40     | 38.50  | 38.50    | 41.40   | 38.00   | 38.00       | 38.00   |
| San Andreas| 0    | 31.60       | 31.60 | 31.60     | 38.00  | 38.00    | 38.00   | 38.00   | 38.00       | 38.00   |
| Spartan  | 0      | 0           | 0     | 0         | 0      | 0        | 0       | 0       | 0           | 0       |

Interclass distance must be as high as possible, minimum of 3.

Distancia interclase debe lo más alta posible, mínimo de 3.
confirmed that the SIMCA model is useful and reliable to identify the 10 strawberry cultivars with a 99% confidence.

4. Conclusions

Results indicated that the model developed with PLS1 is useful to predict simultaneously total soluble solids, total acidity, reducing sugars, pH, ascorbic acid, total phenolic content, total flavonoids content, total anthocyanins content, and antioxidant capacity in 10 strawberry cultivars at variable ripeness stages. Likewise, the SIMCA model identified different strawberry cultivars with a 99% confidence. The aforementioned demonstrates that FT-MIR spectroscopy coupled to multivariate analysis provides similar results to those obtained with conventional methods but less costly, more efficiently, and faster (5–10 min), without the use of solvents or long pretreatments of samples. The models developed in this study could be used to analyze in real time the quality attributes and bioactive compounds in different strawberry cultivars. This proposed methodology can be useful for farmers looking to produce strawberry cultivars with better quality attributes and bioactive compounds with health benefits.

Acknowledgments

Authors thank the Consejo Nacional de la Fresa (CONAFRESA), Consejo Nacional de Ciencia y Tecnología (CONACYT) and the Secretaria de Investigación y Posgrado of the Instituto Politécnico Nacional (SIP-IPN).

Data availability statement

The data used to support the findings of this study are included within the article.

Disclosure statement

No potential conflict of interest was reported by the authors.
ripeness stages. *Journal of Food Composition and Analysis*, 24, 11–16. doi:10.1016/j.jfca.2010.05.004

Ríos de Souza, V., Pimenta-Pereira, P., Da Silva, T., de Oliveira, L., Pio, R., & Queiroz, F. (2014). Determination of the bioactive compounds, antioxidant activity and chemical composition of Brazilian blackberry, red raspberry, strawberry, blueberry and sweet cherry fruits. *Food Chemistry*, 156, 362–368. doi:10.1016/j.foodchem.2014.01.125

Singleton, V. L., & Rossi, J. A. (1965). Colorimetry of total phenolics with phosphomolybdic–Phosphotungstic acid reagents. *American Journal of Enology and Viticulture*, 16, 144–158.

Tamaki, Y., & Mazza, G. (2011). Rapid determination of lignin content of straw using Fourier mid-infrared spectroscopy. *Journal of Agricultural and Food Chemistry*, 59, 504–519. doi:10.1021/jf1036678

Williams, P. C. (2001). Implementation of near-infrared technology. In P. C. Williams & K. H. Norris (Eds.), *Near infrared technology in the agricultural and food industries* (pp. 145–169). St. Paul, MN: American Association of Cereal Chemist.

Young-Jun, K., & Youngjae, S. (2015). Antioxidant profile, antioxidant activity, and physicochemical characteristics of strawberries from different cultivars and harvest locations. *Journal of the Korean Society for Applied Biological Chemistry*, 58, 587–595. doi:10.1007/s13765-015-0085-z

Tulipani, S., Marzban, G., Herndl, A., Laimer, M., Mezzetti, B., & Battino, B. (2011). Influence of environmental and genetic factors on health-related compounds in strawberry. *Food Chemistry*, 124, 906–913. doi:10.1016/j.foodchem.2010.07.018