Contents lists available at ScienceDirect

# Postharvest Biology and Technology

# Prediction of chemical and sensory properties in strawberries using Raman spectroscopy

Petter Vejle Andersen [a],[*], Nils Kristian Afseth [a], Kjersti Aaby [a], Mari Øvrum Gaarder [a],
Siv Fagertun Remberg [b], Jens Petter Wold [a]

[a] Norwegian Institute for Food, Fisheries and Aquaculture Research, Muninbakken 9-13, Breivika, Tromsø 9291, Norway
[b] Faculty of Biosciences, Norwegian University of Life Sciences, 1430 Ås, Norway

## ARTICLE INFO

## ABSTRACT

Non-destructive methods for assessing the chemical composition of fruit and berries are being developed for use in e.g., sorting, storage management or as part of decision making in autonomous harvesting systems. In this study, Raman spectroscopy (RS) was used to estimate a selection of chemical components using partial least squares regression (PLSR) from early (June) and late (September) in the Norwegian strawberry harvesting season. Satisfactory PLSR models were made for total soluble solids (TSS), fructose, glucose, sum of sugars (SS), citric acid and sum of acids (SA) with coefficient of determination ($R^2$) ranging from 0.81 to 0.92 when evaluated by cross validation. PLSR models for total acid content, sucrose and malic acid did not perform as well, with $R^2$ ranging from 0.42 to 0.68, when evaluated by cross validation. Strawberries harvested in September showed significant difference between samples in sweet, sour, and acidic taste ($p < 0.001$). Results from RS demonstrated that the method can be used to determine sensory properties, where e.g., the correlation between predicted values of TSS and SS/SA with sensory sweet taste were 0.80 and 0.87, respectively. In conclusion, RS performed very well for characterization of both chemical and sensory properties in fresh strawberries.

## 1. Introduction

Strawberry (*Fragaria × ananassa Duchesne ex Rozier*) is a fruit characterized by its red color, sweetness, and distinct berry flavor. It is a non-climacteric fruit, meaning that minimal maturing takes place after harvest, which emphasizes the importance of correct harvest time to ensure desired strawberry quality. Decision to pick is normally made by assessing red color development, i.e., when a specified percentage of surface area is red. To ensure that sensory attributes are acceptable, information about chemical composition. e.g. sugars and acids, is also needed in the decision making process (Kader, 2002). Strawberry fruit sweetness and acidity are important for consumer acceptance, and the sugar-acid ratio can be the most important factor for perceived sweetness in strawberries (Ikegaya et al., 2019).

There is a need for quantitative information to be used by autonomous harvesting systems (Zhou et al., 2022) in postharvest sorting and during storage (Lewers et al., 2020). Spectroscopic sensors can be used to assess chemical composition in single strawberries without the need for destructive chemical analyses. Several studies have been performed using near infrared (NIR) spectroscopy to determine a range of physicochemical properties in strawberries the last twenty years, including total soluble solids (TSS) (Agulheiro-Santos et al., 2022), TSS and individual sugars (Nishizawa et al., 2009), acidity (Shao & He, 2008) and both TSS and acidity (Amodio et al., 2017; Mancini et al., 2020; Saad et al., 2022; Sánchez et al., 2012). While many have reported good results for estimation of TSS, a proxy measurement of total sugar content, with $R^2$ values of 0.79–0.96 using NIR spectroscopy, results for acidity were generally not good enough for reliable use, with $R^2$ ranging from 0.54 to 0.91. Raman spectroscopy (RS) is a strong candidate method for assessment of TSS and acidity in strawberries, and given the high chemical resolution of RS, it may be possible to assess the content of individual sugars and acids as well. Other benefits of RS for analysis of strawberries are that the method is insensitive to water and has high sensitivity for different pigments (e.g., anthocyanins and carotenoids), which can be useful for determining strawberry maturity. There is currently no published scientific literature on the use of RS for analysis of strawberry quality, but the technique has shown promise when analyzing other fruit and berries, such as determination sugar content in

* Corresponding author.
*E-mail address:* petter.andersen@nofima.no (P.V. Andersen).

apples (Monago-Maraña et al., 2021), polysaccharide distribution in apples (Szymańska-Chargot et al., 2016) and evaluation of maturity in tomatoes (Qin et al., 2012). Monago-Maraña et al. (2021) also showed that the wide area Raman probe used for apple analysis could penetrate at least 8 mm into an intact apple, and it is reasonable to assume that penetration depth is comparable in strawberries, meaning that internal chemistry will contribute to the Raman spectrum. This asset could also be an additional benefit in strawberry analysis.

The main aim of this study was to use RS to predict chemical composition, including TSS, total acidity (TA), individual sugars and organic acids of strawberries harvested ten weeks apart in one growing season. Harvesting at different times induces changes in strawberry chemistry, and this will potentially give a good indication of performance of developed models over a season. A separate aim was to investigate if RS can be used to characterize sensory taste properties in strawberries.

## 2. Materials and methods

### 2.1. Samples

Fresh strawberries (*Fragaria* × *ananassa* Duchesne ex Rozier) were harvested in two batches in the summer of 2021, the first batch in week 26 (B1, n = 50) and the second batch in week 36 (B2, n = 100). The berries were selected to span different degrees of ripeness, visually assessed based on color development and color uniformity. In B1, the berries were selected to span from unripe to overripe, exemplified by all degrees of ripeness shown in Fig. 1A, and in B2 the ripeness varied from ripe to overripe, exemplified by the three degrees of ripeness to the right in Fig. 1A. Strawberries were grown in coconut coir in a table-top system with automatic watering and nutrient in an open polytunnel (Haygrove Ltd., UK) located in the south-eastern part of Norway. All strawberries were first analyzed with RS before chemical and sensory analysis. In B1, only RS and chemical analysis were performed, and the entire strawberry was used for chemical analysis. In B2, RS, chemical and sensory analysis were performed. Then a quarter of the berry, cut lengthwise, was used for chemical analysis, while the remaining three quarters were used for sensory analysis (Fig. 1B).

### 2.2. Spectroscopic analysis

Raman spectra were collected in duplicates from the surface of the
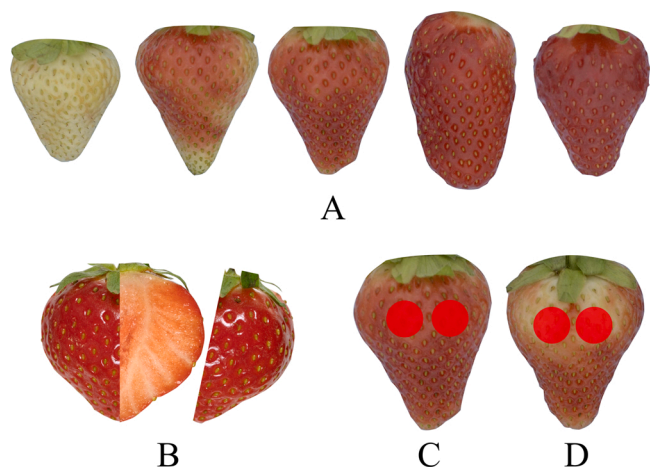


**Fig. 1.** Example of front side of a strawberry from each ripeness class, from unripe, on the left side, through overripe, on the right side (A). Front view of strawberry for sensory and chemical analyses (B), where ¾ was used for sensory analysis and ¼ was used for chemical analysis (in B2). Strawberry with indicated Raman laser spots shown from front (C) and back (D).

strawberry on the same side as the solar radiation (front, Fig. 1. C) and on the opposite side (back, Fig. 1. D). A Kaiser RamanRXN2™ Multi-channel Raman analyzer (Kaiser Optical Systems, Inc., Ann Arbor, MI, USA) with a spectral resolution of 5 cm$^{-1}$ was used. The spectrometer was equipped with a 785 nm laser and PhAT probe with a laser spot size diameter of 6 mm. The spectra were recorded with a laser power set to 400 mW in the range 300–1890 cm$^{-1}$ with 1.0 cm$^{-1}$ intervals. Exposure for each spectrum was set to 6 times 10 s. Instrument was controlled using iC Raman version 4.1.917 SP2 software (Kaiser Optical Systems, Inc., Ann Arbor, MI, USA). Raman spectra from DL-Malic acid (Sigma—Aldrich, MO, USA) and citric acid monohydrate (Merck, Darmstadt, Germany) were acquired in a water solution of 330 mg/mL using an immersion ballprobe (Matrix Solutions, Bothell, WA) suitable for liquids. Instrument settings were the same as for analysis of strawberries, with exception of the acquisition time which was changed to 4 times 20 s. Raman spectra of glucose, fructose and sucrose with the same system were recorded and published by Monago-Maraña et al. (2021).

### 2.3. Chemical reference analyses

The berry samples were homogenized, then centrifuged at 39200*g* for 10 min (Avanti J-26 XP, Beckman Coulter, USA) and the supernatant collected. The supernatant was used for analyses of TSS, TA and individual sugars and acids.

TSS was determined using a pocket Brix-acidity meter (PAL-BX| ACID1, Atago Co., Ltd., Tokyo, Japan). TA was measured in supernatant diluted with purified water (1/49, w/w) by the pocket Brix-acidity meter. The concentration of TA was expressed as g citric acid equivalents per 100 g.

Individual sugars and acids were determined using an Agilent 1100 series HPLC system (Agilent Technologies, Waldbronn, Germany) equipped with a diode array detector (DAD) and a refractometer index (RI) detector (Model 132; Gilson, Villiers-le-Bel, France) as previously described (Woznicki et al., 2017). Supernatant diluted with purified water (1/2, v/v) filtered through Millex HA 0.45 μm filters (Millipore Corp., MA, USA) was injected (20 μL) and analytes were separated on a Rezex ROA-Organic acid H+ (8%) column (300 × 7.8 mm; Phenomenex, CA, USA) at 45 °C with a mobile phase of 7.2 mmol L$^{-1}$ H$_2$SO$_4$ and a flow rate of 0.5 mL min$^{-1}$. External standards of glucose, sucrose and fructose (Chem Service Inc., West Chester, PA, USA), citric, malic and shikimic acids (Sigma—Aldrich, MO, USA) and quinic acid (Merck, Darmstadt, Germany) were used for quantification. The sugars were detected with the RI detector and the organic acids were detected with DAD at 210 nm. In B2 there were different number of samples analysed for TSS, TA and with HPLC (Table 1.) because there was too little sample left after sensory analysis.

### 2.4. Sensory analysis

The sensory properties of the strawberries from B2 were assessed by a sensory panel of 10 assessors, trained according to ISO 8586 at Nofima (Ås, Norway). A generic descriptive analysis was performed as described by Lawless and Heymann (2010) and in accordance with ISO 13299 (2016). Prior to sensory assessment, the hundred strawberries from B2 were combined to make 20 aggregated samples, consisting of five strawberries each, to have enough of each sample for all sensory assessors. Aggregated samples were put together based on predictions of TSS and the ratio between sum of sugars (SS) and sum of acids (SA) from Raman spectra and partial least squares regression (PLSR) models developed on strawberries from B1. The goal of sorting was to include strawberries with similar chemical composition in each aggregated sample as well as making the aggregated samples span the entire chemical range. This was accomplished by sorting strawberries by increasing TSS content, and then grouping the ones with similar SS/SA together. Three quarters of each strawberry was cut into smaller pieces and mixed so that each sample for the assessors contained pieces from all

**Table 1**
Summary of chemical reference data from Batch 1 and Batch 2.

| Property | Batch 1 | | | | | Batch 2 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | n | Mean | Min | Max | SD | n | Mean | Min | Max | SD |
| TSS (%) | 50 | 10.12 | 7.25 | 15.60 | 2.23 | 97 | 8.33 | 4.30 | 15.30 | 1.41 |
| TA (%) | 50 | 0.88 | 0.66 | 1.19 | 0.14 | 96 | 0.72 | 0.53 | 1.02 | 0.10 |
| Sucrose (g L$^{-1}$) | 50 | 13.17 | 1.10 | 27.40 | 5.25 | 91 | 8.46 | 0.80 | 18.70 | 3.63 |
| Fructose (g L$^{-1}$) | 50 | 38.64 | 23.61 | 68.21 | 12.21 | 91 | 32.37 | 15.15 | 70.58 | 7.01 |
| Glucose (g L$^{-1}$) | 50 | 35.51 | 19.40 | 63.50 | 11.76 | 91 | 28.73 | 10.90 | 62.10 | 6.69 |
| SS (g L$^{-1}$) | 50 | 87.32 | 52.91 | 144.16 | 24.08 | 91 | 69.56 | 26.90 | 136.10 | 14.35 |
| Malic acid (g L$^{-1}$) | 50 | 1.91 | 1.21 | 2.88 | 0.38 | 91 | 1.43 | 0.74 | 2.81 | 0.32 |
| Citric acid (g L$^{-1}$) | 50 | 8.82 | 6.40 | 13.30 | 1.74 | 91 | 6.75 | 4.60 | 10.30 | 1.08 |
| SA (g L$^{-1}$) | 50 | 10.72 | 7.90 | 15.90 | 2.06 | 91 | 8.19 | 5.90 | 12.20 | 1.21 |
| SS / SA | 50 | 8.60 | 4.11 | 15.90 | 3.20 | 91 | 8.74 | 3.49 | 15.39 | 2.40 |

SD: standard deviation; TSS: total soluble solids; TA: total acidity; SS: sum of sugars; SA: sum of acids.

the five strawberries included in the aggregated sample. In a pre-test session before the main session, the assessors were calibrated on samples that were considered the most different on the selected attributes typical for the samples to be tested. All products were coded with a three-digit number in a full balanced design (ISO 8589, 2007). Relevant product taste attributes: acidic, sweet, sour, bitter and total taste intensity were evaluated at individual speed on an unstructured 15 cm line scale with labelled end points from "no intensity" (1) to "high intensity" (9) and registered on a computer system for direct recording of data (EyeQuestion, Software Logic8 BV, Utrecht, Netherlands). Tap water and unsalted crackers were available for palate cleansing during the assessment.

### 2.5. Statistics, pre-processing of spectra and data analysis

Correlations between the different chemical and sensory parameters were calculated as Pearson correlation coefficient (r). Raman spectra were first smoothed using the Savitzky-Golay algorithm with polynomial order one and eleven points. Spectra were then base-line corrected and fluorescence background was removed using a modified iterative 5th order polynomial curve fitting procedure (Modpoly) as described by Lieber and Mahadevan-Jansen (2003). The basis for the Modpoly method is a least-squares based polynomial curve-fitting function, where after the first iteration all Raman peaks with higher intensity than the polynomial curve are removed, and a new polynomial curve is calculated based on the remaining spectrum. This process is repeated until, ideally, there are no points that need reassignment, and the resulting baseline spectrum is subtracted from the raw Raman spectrum.

PLSR was used for determining linear relationships between reference measurements and Raman spectra. PLSR emphasizes information in the spectra that is important for explaining variation in the reference measurements when making models (Martens & Martens, 2001). PLSR calibration models were evaluated by 10-fold venetian blind cross-validation. In addition, PLSR models calibrated only on strawberries from B1 were used to predict values for strawberries in B2, and the other way around, using each trial as independent test-set for assessment of model performance. For calibration models on taste attributes the average Raman spectrum from the five strawberries included in the aggregated sensory sample were used, and mean intensities for taste attributes were used as reference value. Because the sensory dataset contained only 20 samples, leave-one-sample out cross-validation was used for evaluation of RS PLSR models.

The descriptive sensory data were analyzed with analysis of variance (ANOVA) using a linear mixed model comprising the factors: strawberry sample, assessors, replica, and the second-order interactions. Assessors and interactions involving assessors were considered random, whereas the other factors were fixed. Mean intensities were calculated, and significant differences were checked using TUKEY's HSD test (p < 0.05). The descriptive sensory analyses were done with EyeOpenR in the

software EyeQuestion (Logic8 BV, Utrecht, Nederland).

Baseline correction of Raman spectra were carried out using MATLAB R2016b (The MathWorks, Natick, MA). Smoothing of spectra and PLSR analysis were conducted using The Unscrambler® X version 10.4 (CAMO Analytics AS, Norway). Statistical testing was done in Minitab® Statistical Software (Minitab LLC, Pennsylvania, USA).

## 3. Results and discussion

### 3.1. Reference analysis of TSS, sugars and acids

Concentrations of sugars and acids decreased significantly (p < 0.05) from B1 to B2 (Table 1). This seasonal effect on chemical composition has also been documented by Schwieterman et al. (2014), where authors cited environmental changes or plant maturity as main reasons for the change. Standard deviations were relatively large in the dataset because samples were selected to span from unripe to overripe and from ripe to overripe in B1 and B2, respectively. Correlations between chemical parameters are provided in table S1-S3. Note the high correlations between fructose, glucose and TSS (0.93 < r < 0.99), making them almost indistinguishable for modelling purposes. On the other hand, acids and sugars were not highly correlated (−0.26 < r < 0.08), making it theoretically possible to model these independently based on Raman spectra.

### 3.2. Raman spectra and PLSR

All raw spectra and two selected pre-processed Raman spectra from B1 are provided in Fig. 2. The raw Raman spectra show high fluorescence background, largely caused by the seeds on the outside of the strawberries, which were impossible to avoid because of the large laser spot size, as indicated in Fig. 1. After pre-processing, resolved peaks related to chemical composition of the strawberries were apparent and many of the peaks could be assigned to sugars, organic acids, and anthocyanin.

Results from cross validated PLSR models were very good to excellent for all chemical components, except for malic acid and sucrose (Table 2). Reasons for the poor performance for malic acid and sucrose could be that they were present in relatively low concentrations. Cross-validation statistics were better for the trial in B1 than in B2, most likely due to the larger range in the reference data in B1. When the two datasets were merged PLSR results remained good, but in many cases extra latent variables (LV) were needed. PLSR models developed individually from one of the four spots where Raman spectra were recorded resulted in slightly lower accuracy for the single spots compared with average of two spots on the frontside or average of two spots on the backside, while all spots merged resulted in the best accuracy (Table S4). Glucose and fructose were best modeled from Raman spectra recorded on the front of strawberries, while all other chemical parameters were slightly better modeled from Raman spectra recorded on the backside.

Interpretation of regression coefficients of selected models (Fig. 3)
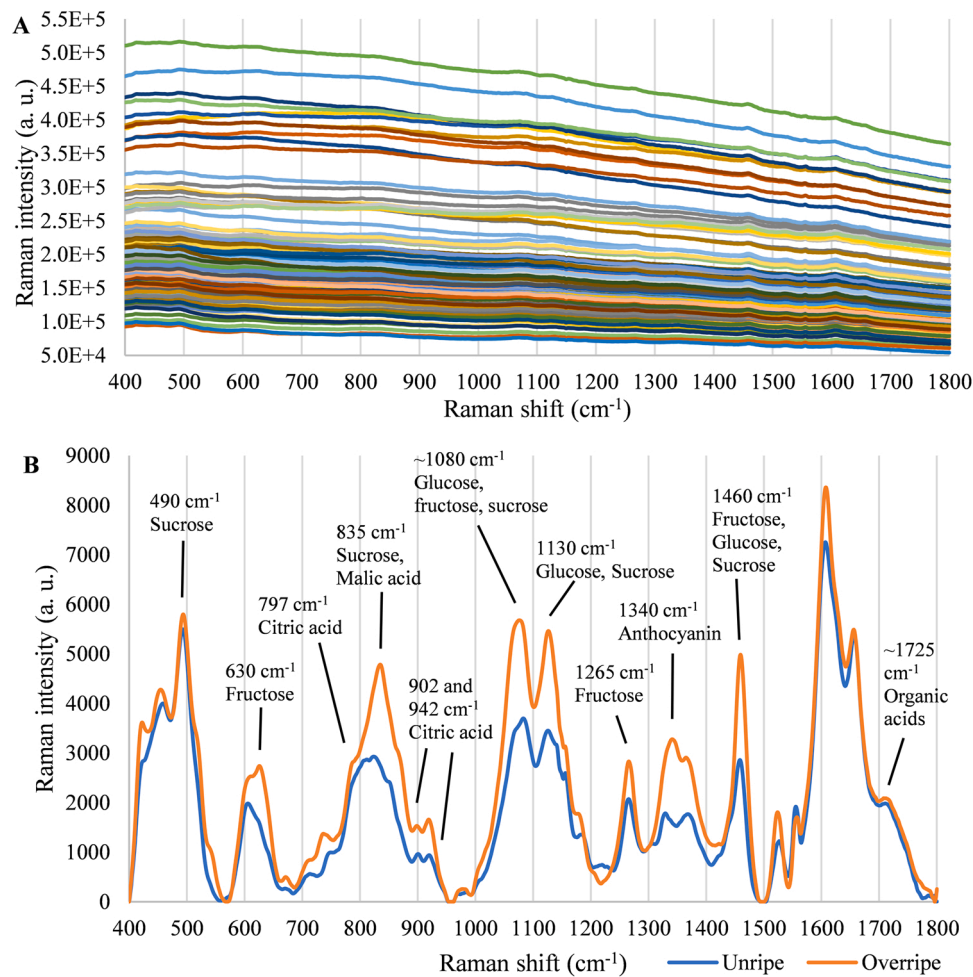
**Fig. 2.** Raw Raman spectra (A) and two baseline corrected spectra with assignments of peaks (B), where the blue spectrum is from an unripe strawberry (leftmost panel in Fig. 1.A) with low total soluble solids and orange spectrum from an overripe strawberry (rightmost panel in Fig. 1.A) with high total soluble solids. Assignments are made from spectra in Fig. S1 and S2.

**Table 2**
Summary of PLSR RS model performance for chemical reference analyses from Batch 1, Batch 2 and the full data set. $R^2$ and RMSECV were calculated from 10-fold venetian blind cross-validation.

| Property | Batch 1 | | | Batch 2 | | | Batch 1 and Batch 2 | | |
|---|---|---|---|---|---|---|---|---|---|
| | LV | $R^2$ | RMSECV | LV | $R^2$ | RMSECV | LV | $R^2$ | RMSECV |
| TSS (%) | 5 | 0.94 | 0.54 | 6 | 0.84 | 0.58 | 5 | 0.92 | 0.56 |
| TA (%) | 5 | 0.62 | 0.09 | 6 | 0.58 | 0.06 | 7 | 0.68 | 0.08 |
| Sucrose (g L$^{-1}$) | 6 | 0.41 | 4.09 | 6 | 0.22 | 3.23 | 7 | 0.45 | 3.62 |
| Fructose (g L$^{-1}$) | 3 | 0.91 | 3.75 | 3 | 0.78 | 3.27 | 4 | 0.88 | 3.35 |
| Glucose (g L$^{-1}$) | 3 | 0.91 | 3.48 | 3 | 0.79 | 3.11 | 4 | 0.89 | 3.10 |
| SS (g L$^{-1}$) | 3 | 0.90 | 7.57 | 6 | 0.80 | 6.44 | 4 | 0.88 | 6.98 |
| Malic acid (g L$^{-1}$) | 1 | 0.36 | 0.30 | 1 | 0.16 | 0.29 | 6 | 0.42 | 0.32 |
| Citric acid (g L$^{-1}$) | 5 | 0.85 | 0.69 | 8 | 0.61 | 0.68 | 10 | 0.84 | 0.68 |
| SA (g L$^{-1}$) | 5 | 0.84 | 0.83 | 7 | 0.55 | 0.82 | 10 | 0.81 | 0.86 |
| SS / SA | 6 | 0.90 | 1.01 | 6 | 0.77 | 1.16 | 6 | 0.82 | 1.17 |

LV: Latent Variables; $R^2$: Coefficient of determination for cross validated values; RMSECV: Root Mean Square Error of Cross Validation; TSS: total soluble solids; TA: total acidity; SS: sum of sugars; SA: sum of acids.

revealed strong association between coefficients and Raman spectra of pure chemical components (Fig. S1 and S2). Regression coefficients for TSS show a strong link to individual sugars, as seen by peaks at approx. 630, 1080, 1130 and 1460 cm$^{-1}$, but were still different from the regression vectors for glucose and fructose, even with high underlying correlations. The most prominent peak in the regression vector of TSS was at approx. 1460 cm$^{-1}$, assigned to C-H, and this may indicate importance of combined intensity from all individual sugars, with

minimal interference from other chemical components in strawberries, for the regression vector. Due to their very high correlation (r = 0.99), fructose and glucose regression coefficients were almost identical, even though the pure component spectra are quite different (e.g., strong peaks at 629 cm$^{-1}$ and 1265 cm$^{-1}$ for fructose, but not for glucose). Hence, it is impossible to model these components independently in this dataset, and robustness of the models will also most likely be compromised when used for prediction of strawberries where fructose and
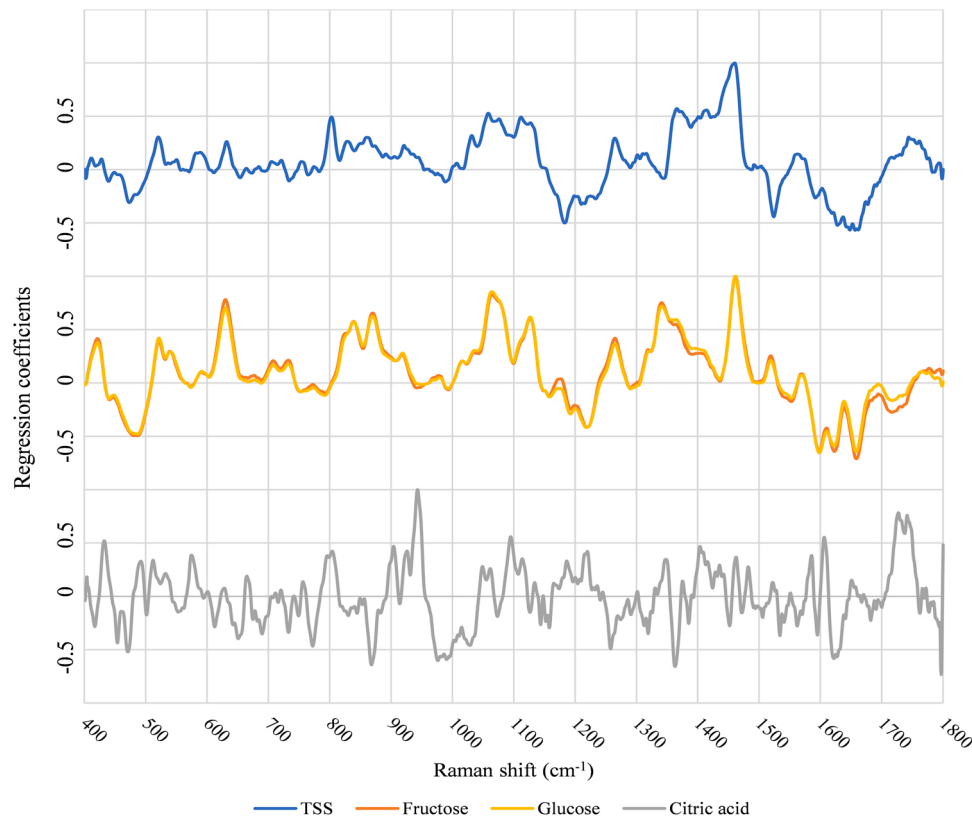
**Fig. 3.** Regression coefficients for total soluble solids, fructose, glucose, and citric acid from calibration models for the full dataset (Batch 1 and Batch 2). For easier comparison, regression coefficients were scaled between − 1 and 1 by dividing all values by the maximum value for each regression coefficient.

glucose are less correlated. Two distinct peaks were visible in citric acid regression coefficients, a peak at 945 cm$^{-1}$, assigned to C-C-C skeletal vibration, and a peak at 1725 cm$^{-1}$, assigned to carboxylic acid residues, closely mimicking the pure spectrum of citric acid.

To validate performance of PLSR models independently, performance metrics were calculated using PLSR models made from one harvest batch to predict the other harvest batch (Table 3). Calibrations on samples from B2 performed better than those on samples from B1 judged by R$^2$ values. This was contrary to what was expected, because the reference range was smaller in B2, leading to extrapolation in the predictions of B1. On the other hand, there were more samples in B2, possibly contributing to increased robustness of the models. Results for TSS and individual sugars were very good, except for sucrose, exhibiting low RMSEP and bias for each validation set. Models for fructose,

glucose, TSS and SS had large r$^2$ and R$^2$ values, showing that these chemical components can be predicted with little to no bias correction. In general, results for individual acids had large biases and as a result large RMSEP and low R$^2$, but the r$^2$-values for citric acid and SA were good when using B1 as validation set, indicating that it could be possible to adjust models (i.e., bias correction) for improved model performance. Many of the models for acids had negative R$^2$ values, which may seem contradictory, but is made possible by the fact that the large bias contributes to large residuals for predicted values. The reason for better models for sugars than for acids may be that sugar content was about one order of magnitude higher than acid content, thus giving the largest chemical contribution to the Raman spectra. It is noteworthy that even though predictions for SA were affected by large biases, prediction of SS/SA was still acceptable, because of the high correlation between

**Table 3**
Summary of Raman spectroscopy partial least squares regression model prediction performance on chemical reference analyses from Batch 1 and Batch 2. Prediction statistics from models calibrated on the other batch (i.e., partial least squares regression model only calibrated on samples from Batch 1 used to predict samples from Batch 2 and the other way around).

| Property | Batch 1 | | | | Batch 2 | | | |
|---|---|---|---|---|---|---|---|---|
| | r$^2$ | R$^2$ | RMSEP | Bias | r$^2$ | R$^2$ | RMSEP | Bias |
| TSS (%) | 0.95 | 0.91 | 0.66 | -0.37 | 0.84 | 0.78 | 0.66 | 0.34 |
| TA (%) | 0.44 | < 0 | 0.17 | -0.13 | 0.19 | < 0 | 0.11 | 0.06 |
| Sucrose (g L$^{-1}$) | 0.36 | < 0 | 5.49 | -3.50 | 0.25 | < 0 | 5.17 | 3.62 |
| Fructose (g L$^{-1}$) | 0.91 | 0.89 | 4.07 | 0.21 | 0.76 | 0.72 | 3.72 | -0.85 |
| Glucose (g L$^{-1}$) | 0.92 | 0.89 | 3.87 | -0.88 | 0.77 | 0.74 | 3.41 | 0.13 |
| SS (g L$^{-1}$) | 0.92 | 0.85 | 9.15 | -4.84 | 0.78 | 0.72 | 7.64 | 3.03 |
| Malic acid (g L$^{-1}$) | 0.36 | < 0 | 0.60 | -0.51 | 0.20 | < 0 | 0.52 | 0.43 |
| Citric acid (g L$^{-1}$) | 0.82 | 0.18 | 1.59 | -1.33 | 0.35 | < 0 | 1.54 | 1.26 |
| SA (g L$^{-1}$) | 0.80 | < 0 | 2.32 | -2.10 | 0.34 | < 0 | 1.93 | 1.64 |
| SS /SA | 0.86 | 0.66 | 1.86 | 1.38 | 0.74 | 0.65 | 1.42 | -0.54 |

r$^2$: squared correlation between reference and predicted values; R$^2$: Coefficient of determination for predicted values; RMSEP: Root Mean Square Error of Prediction; TSS: total soluble solids; TA: total acidity; SS: sum of sugars; SA: sum of acids.

predicted and reference value.

### 3.3. Sensory analysis and RS PLSR

ANOVA analysis of the results from sensory assessment revealed that for the three taste attributes acidic, sweet, and sour taste, there were significant difference between some of the samples, while there was no statistically significant difference between any of the samples for bitterness and total taste intensity (Table S5). The general trend was that strawberries with high intensity of sweet taste, had lower taste intensity of acidic and sour.

Correlation between taste attributes and chemical parameters showed that sugar content, TSS/TA and SS/SA were highly positively correlated with sweet taste, while acid content was highly negatively correlated with sweet taste (Table S6). The correlation was negative between TSS and SS/SA for acidic and sour taste, and the absolute value for correlations were lower for acidic taste than for sour taste. Correlation between TSS and sweetness was lower than correlation between glucose, fructose or SS/SA and sweetness, indicating that a method capable of detailed chemical characterization is superior when predicting taste of strawberries. It should also be noted that RS PLSR prediction of SS/SA gave higher correlation (correlation coefficient (r) = 0.87) than reference measurement of TSS (r = 0.81) and RS PLSR predicted values for TSS (r = 0.80). This relation is visualized in Fig. 4, where it is evident that the results from sensory assessment of sweetness follows the trend of SS/SA more closely than for TSS. Even though the $R^2$ and RMSEP were not great for TSS and SS/SA for the B2 dataset, the $R^2$ Pearson values were good, and for sorting purposes within one sample set, this is the most important metric.

RS PLSR models for acidic and sweet taste were poor ($0.68 < R^2 < 0.70$), and the model for sour taste was fair ($R^2 = 0.85$). Nevertheless, these models are of sufficient quality to be used for rough screening to differentiate strawberries based on their taste and designate them to different consumer groups or processing methods. RS PLSR model for bitterness and total taste intensity were very poor ($R^2 < 0$), there were no significant difference between samples for these two taste parameters from the sensory assessment, making modelling almost impossible. These models can be seen as indicative of performance at best because they are based on a limited number of samples, which were combined from five strawberries, making direct connections between RS and sensory attributes more indistinct. (Table 4).
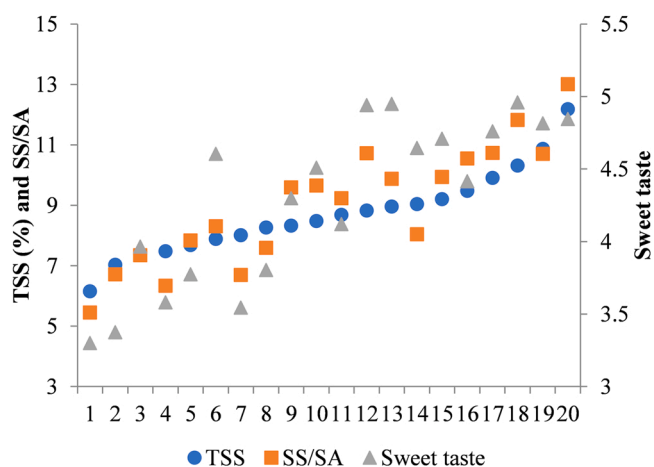


**Fig. 4.** Strawberry sample sorting in Batch 2 for sensory assessment. Average values for total soluble solids (TSS) (dots) and sum of sugars divided by sum of acids (SS/SA) (squares) from Raman predictions using models calibrated with samples from Batch 1, shown on the left y-axis, while the sweetness scores (triangles) were from sensory assessment, values shown on the right x-axis.

**Table 4**

Summary of Raman spectroscopy partial least squares regression model performance on sensory taste attributes from Batch 2 (n = 20). $R^2$ and RMSECV from leave-one-out cross-validation.

| Model metric | Acidic | Sweet | Sour | Bitter | Total intensity |
|---|---|---|---|---|---|
| LV | 9 | 2 | 2 | 1 | 1 |
| $R^2$ | 0.68 | 0.70 | 0.85 | < 0 | < 0 |
| RMSECV | 0.29 | 0.32 | 0.21 | 0.16 | 0.23 |

LV: latent variables; $R^2$: coefficient of determination; RMSECV: root mean square error of cross-validation.

### 4. Discussion

When comparing our PLSR results from RS with NIRS for chemical composition from the literature it seems that RS performs on par or better than NIRS. For TSS both methods performed very well, with NIRS giving $R^2$ values of 0.79 on the lower end (Sánchez et al., 2012) and 0.96 on the high end (Saad et al., 2022). RS performed better than NIRS for estimating glucose and fructose, while results were similar for sucrose (Nishizawa et al., 2009). For glucose and fructose, this better performance can possibly be attributed to the specificity of Raman signals.

For TA, titratable acidity or acidity results are more varied, but RS seems to perform better than NIRS based on most studies. Using NIRS for acidity resulted in no models (Mancini et al., 2020), $R^2$ values of 0.54 (Sánchez et al., 2012) and 0.58 (Amodio et al., 2017), and the peculiar outlier with $R^2$ of 0.91 (Saad et al., 2022). The fundamental reason for RS performing better for prediction of acids than NIRS is that the functional carboxylic group has very low absorptivity in the NIR region and that acids are typically present in low concentrations (Williams et al., 2006). This increases the requirement for the spectroscopic method to be specific for functional groups to distinguish these molecules from others present in the sample. Of note from the current study is that the RS PLSR model for citric acid and SA are better than the one for TA, indicating that the reference method for TA is not precise enough for good PLSR models.

One factor that can disturb Raman measurements is ambient light, which could be solar radiation, artificial lighting, or a combination of both. To accommodate for this, measurements could be carried out in a dark chamber or by using defined light sources (e.g., LED bulbs emitting light at lower wavelengths than the Raman excitation laser) so that it does not interfere with the Raman signal, or that it can be easily removed by dark subtraction.

Another issue for implementation of RS is the lengthy acquisition times often associated with the technique. In the present study, 60 s was used to collect one spectrum, which is too long for practical use in the field. However, a study by Lindtvedt et al. (2022) shows that acquisition times for RS can be reduced significantly, as low as one to two seconds for each Raman spectrum, and still contain relevant chemical information of minor chemical components. Four spots were analyzed on each strawberry, and results indicate that model performance increased when averaging Raman spectra from more spots. It is unclear if this is caused by an increase in the signal to noise (S/N) when combining more spectra or if the improvement comes from a more representative sample being included when more than one spot is used for PLSR models. If this is an S/N issue, then changing the measurement scheme will not help much in reducing the measurement time or robustness of PLSR models. On the other hand, if this is a sampling issue, the analysis time could be reduced by moving the strawberry in relation to the Raman laser to cover more of the strawberry surface in less time. It is also possible that the sampling spots used for RS in the present study are not the best suited for analysis, maybe S/N is better in parts of the strawberry where TSS content is higher or there are certain parts of the strawberry that are more representative of the whole.

Some of the chemical components exhibited a very high degree of covariance, so much so that they could not be modeled independently.

This is a difficult challenge to address in biological systems, as we need to work with the actual variation present at the time of analysis, but it is something to keep in mind when analyzing new data to confirm if the covariance structures are intact as a part of model maintenance.

## 5. Conclusion

RS was successfully used to make PLSR models for TSS, SS, SA, fructose, glucose, citric acid, and SS/SA. PLSR models for TA, sucrose and malic acid were not satisfactory. This indicates that RS can be used to estimate all major chemical components in strawberries. In addition, RS could be used to sort strawberries according to TSS content and SS/SA giving high correlation with sweetness as assessed by a trained sensory panel.

## CRediT authorship contribution statement

**Petter Vejle Andersen:** Writing – original draft, Conceptualization, Methodology, Investigation, Validation, Formal analysis, Visualization, Data curation. **Nils Kristian Afseth:** Writing – review & editing, Conceptualization. **Kjersti Aaby:** Writing – review & editing, Investigation, Resources, Data curation. **Mari Øvrum Gaarder:** Writing – review & editing, Methodology, Resources. **Siv Fagertun Remberg:** Writing – review & editing, Resources. **Jens Petter Wold:** Writing – review & editing, Conceptualization, Methodology, Supervision, Project administration, Funding acquisition.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data Availability

Data will be made available on request.

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.postharvbio.2023.112370.

## References

Agulheiro-Santos, A.C., Ricardo-Rodrigues, S., Laranjo, M., Melgao, C., Velazquez, R., 2022. Non-destructive prediction of total soluble solids in strawberry using near infrared spectroscopy. J. Sci. Food Agric. 102 (11), 4866–4872.

Amodio, M.L., Ceglie, F., Chaudhry, M.M.A., Piazzolla, F., Colelli, G., 2017. Potential of NIR spectroscopy for predicting internal quality and discriminating among strawberry fruits from different production systems. Postharvest Biol. Technol. 125, 112–121.

Ikegaya, A., Toyoizumi, T., Ohba, S., Nakajima, T., Kawata, T., Ito, S., Arai, E., 2019. Effects of distribution of sugars and organic acids on the taste of strawberries. Food Sci. Nutr. 7 (7), 2419–2426.

Kader, A.A., 2002. Postharvest technology of horticultural crops (3rd ed.). Oakland, Calif: University of California, Agriculture and Natural Resources.

Lewers, K.S., Newell, M.J., Park, E., Luo, Y., 2020. Consumer preference and physiochemical analyses of fresh strawberries from ten cultivars. Int. J. Fruit. Sci. 20 (sup2), 733–756.

Lieber, C.A., Mahadevan-Jansen, A., 2003. Automated method for subtraction of fluorescence from biological Raman spectra. Appl. Spectrosc. 57 (11), 1363–1367.

Lindtvedt, T.A., Andersen, P.V., Afseth, N.K., Marquardt, B., Gidskehaug, L., Wold, J.P., 2022. Feasibility of In-Line Raman Spectroscopy for Quality Assessment in Food Industry: How Fast Can We Go? Appl. Spectrosc., Press.

Mancini, M., Mazzoni, L., Gagliardi, F., Balducci, F., Duca, D., Toscano, G., Mezzetti, B., Capocasa, F., 2020. Application of the non-destructive NIR technique for the evaluation of strawberry fruits quality parameters. Foods 9 (4), 441.

Martens, H., Martens, M., 2001. Analysis of two data tables X and Y: partial least squares regression (PLSR). In: Martens, H., Martens, M. (Eds.), Multivariate Analysis of Quality: An Introduction. John Wiley & Sons Ltd, Chichester, UK, pp. 111–126.

Monago-Maraña, O., Afseth, N.K., Knutsen, S.H., Wubshet, S.G., Wold, J.P., 2021. Quantification of soluble solids and individual sugars in apples by Raman spectroscopy: a feasibility study. Postharvest Biol. Technol. 180, 111620.

Nishizawa, T., Mori, Y., Fukushima, S., Natsuga, M., Maruyama, Y., 2009. Non-destructive analysis of soluble sugar components in strawberry fruits using near-infrared spectroscopy. Nippon Shokuhin Kagaku Kogaku Kaishi 56 (4), 229-229.

Qin, J., Chao, K., Kim, M.S., 2012. Nondestructive evaluation of internal maturity of tomatoes using spatially offset Raman spectroscopy. Postharvest Biol. Technol. 71, 21–31.

Saad, A., Azam, M.M., Amer, B.M.A., 2022. Quality analysis prediction and discriminating strawberry maturity with a hand-held vis–nir spectrometer. Food Anal. Methods 15 (3), 689–699.

Sánchez, M.-T., De la Haba, M.J., Benítez-López, M., Fernández-Novales, J., Garrido-Varo, A., Pérez-Marín, D., 2012. Non-destructive characterization and quality control of intact strawberries based on NIR. Spectr. data. J. Food Eng. 110 (1), 102–108.

Schwieterman, M.L., Colquhoun, T.A., Jaworski, E.A., Bartoshuk, L.M., Gilbert, J.L., Tieman, D.M., Odabasi, A.Z., Moskowitz, H.R., Folta, K.M., Klee, H.J., Sims, C.A., Whitaker, V.M., Clark, D.G., 2014. Strawberry flavor: diverse chemical compositions, a seasonal influence, and effects on sensory perception. PLoS One 9 (2), e88446.

Shao, Y., He, Y., 2008. Nondestructive measurement of acidity of strawberry using Vis/NIR spectroscopy. Int. J. Food Prop. 11 (1), 102–111.

Szymańska-Chargot, M., Chylińska, M., Pieczywek, P.M., Rösch, P., Schmitt, M., Popp, J., Zdunek, A., 2016. Raman imaging of changes in the polysaccharides distribution in the cell wall during apple fruit development and senescence. Planta 243 (4), 935–945.

Williams, P., Saranwong, S., Kawano, S., Isaksson, T., Segtnan, V.H., 2006. Applications to agricultural and marine products. In: Ozaki, Y., McClure, W.F., Christy, A.A. (Eds.), Near-Infrared Spectroscopy in Food Science and Technology. John Wiley & Sons, Inc, New Jersey, pp. 163–277.

Woznicki, T.L., Sønsteby, A., Aaby, K., Martinsen, B.K., Heide, O.M., Wold, A.B., Remberg, S.F., 2017. Ascorbate pool, sugars and organic acids in black currant (*Ribes nigrum* L.) berries are strongly influenced by genotype and post-flowering temperature. J. Sci. Food Agric. 97 (4), 1302–1309.

Zhou, H., Wang, X., Au, W., Kang, H., Chen, C., 2022. Intelligent robots for fruit harvesting: recent developments and future challenges. Precis. Agric.