

Leave-One-Day-Out Cross-Validation of NIR-PLSR Models for Simultaneous Prediction of Eight Banana Quality Attributes during 12-Day Storage

Pu Zhang

Xi'an Jiaotong University

Feng Zhang

China National Research Center for Food Safety Inspection and Testing

Jin Luo

Xi'an Jiaotong University

Wenlin Wang

Shaanxi Institute for Food and Drug Control

Xiaohe Wang

Shaanxi Institute for Food and Drug Control

Hanxia Liu

China National Research Center for Food Safety Inspection and Testing

Fang Lin

Shaanxi Institute for Food and Drug Control

Hong Zhao

Xi'an Jiaotong University

Jinfeng Peng

jfpeng@xjtu.edu.cn

Xi'an Jiaotong University

Research Article

Keywords: Postharvest banana, Near-infrared spectroscopy, Carbohydrates, Non-destructive prediction, Time generalization performance

Posted Date: September 10th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10785980/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

**Leave-One-Day-Out Cross-Validation of NIR-PLSR Models for Simultaneous
Prediction of Eight Banana Quality Attributes during 12-Day Storage**

*Pu Zhang^a, Feng Zhang^b, Jin Luo^a, Wenlin Wang^{c,d}, Xiaohe Wang^{c,d}, Hanxia Liu^b,
Fang Lin^{c,d}, Hong Zhao^{a,e}, Jinfeng Peng^{a,*}*

*a. School of Instrument Science and Technology, Xi'an Jiaotong University,
Xi'an, Shaanxi 710049, China.*

*b. China National Research Center for Food Safety Inspection and Testing,
Beijing 100045, China.*

*c. Shaanxi Institute for Food and Drug Control, Xi'an, Shaanxi 710065,
China.*

*d. Shaanxi Key Laboratory for Safety Monitoring of Food and Drug, Xi'an,
Shaanxi 710065, China.*

*e. State Key Laboratory for Manufacturing System Engineering, Xi'an
Jiaotong University, Xi'an, Shaanxi 710049, China.*

**Corresponding author. Tel.: +86-29-82667814*

E-mail address: jfpeng@xjtu.edu.cn

*Postal address: No. 28, Xianning West Rd., School of Instrument Science and
Technology, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China*

Abstract

This study evaluated the feasibility of simultaneously predicting eight banana quality attributes during 12-day storage using near-infrared (NIR) spectroscopy combined with partial least squares regression (PLSR), with a specific focus on temporal generalization assessed by leave-one-day-out cross-validation (LODO-CV). A total of 72 banana samples were stored at 21 ± 1 °C and $85 \pm 5\%$ RH for 12 days, and their fructose, glucose, sucrose, starch, total soluble solids (TSS), firmness, pH, and titratable acidity (TA) were measured over 12 days. Seven spectral preprocessing methods were compared to optimize the PLSR models. LODO-CV was used to evaluate the predictive performance of the models over time, and the results were compared with those obtained by repeated k-fold cross-validation. Under LODO-CV, glucose and TSS models showed the best performance ($R^2_P > 0.90$; $RPD > 3.5$), while the fructose, starch, and firmness models provided good predictions ($R^2_P = 0.85-0.88$; $RPD = 2.63-2.99$). Only approximate predictions were obtained for sucrose, and the predictive ability for pH and TA was limited. Preprocessing significantly affected the performance of models for specific attributes. LODO-CV generally yielded lower performance than repeated k-fold cross-validation, indicating that random sample splitting may overestimate predictive performance. Overall, NIR-PLSR combined with LODO-CV enables the simultaneous non-destructive evaluation of multiple banana quality attributes under controlled storage, particularly for TSS, monosaccharides, and firmness, providing a more realistic assessment of model generalization over time.

42 **Keywords:** Postharvest banana; Near-infrared spectroscopy; Carbohydrates; Non-
43 destructive prediction; Time generalization performance

1. Introduction

Banana (*Musa* spp.) is one of the most important fruit crops in tropical and subtropical regions. Its commercial value depends strongly on postharvest quality during storage and ripening (Chen et al. 2026). Nevertheless, bananas are highly perishable and postharvest deterioration causes significant quantitative and qualitative losses along the supply chain, imposing economic and food-security pressures on producers, retailers and consumers (Kikulwe et al. 2018). The banana ripening process involves coordinated biochemical and structural changes that lead to starch degradation, sugar accumulation, tissue softening and changes in acidity-related properties (Nandakumar 2024). Studies on banana storage have shown that starch content declines as ripening proceeds, whereas total soluble solids (TSS), reducing sugars, sucrose and pH generally increase, while peel or pulp firmness decreases significantly (Watharkar et al. 2021; Souripet 2026; Thakur et al. 2019; Fatima and Waheed 2024). These quality attributes not only indicate the ripening status but also support harvest management, storage control, shelf-life estimation, and market classification (Al-Dairi et al. 2023). Conventional analytical methods for determining sugars, starch, firmness, pH, titratable acidity (TA) and other related attributes are informative but destructive, labor-intensive and not well suited to repeated measurements on the same fruit during storage (Pornchaloempong et al. 2022). This limitation is particularly critical for climacteric fruits such as banana, because the same lot can show rapid temporal changes and high fruit-to-fruit heterogeneity during ripening (Sun et al. 2022). Therefore, nondestructive

65 optical methods are attractive because they can rapidly assess multiple quality attributes
66 in intact fruit without reagents and allow repeated measurements during storage (Wang
67 et al. 2025).

68 Near-infrared (NIR) spectroscopy, which covers a spectral range of approximately 750–
69 2500 nm, is arguably one of the most widely used approaches for fruit quality evaluation
70 because spectral responses can be related to pigments, water, sugars, organic acids, and
71 structural state (Zhao et al. 2022; Ping et al. 2023; Kasampalis et al. 2025). In fruit
72 systems modeling, TSS and dry matter have typically been modelled with considerable
73 success using NIR wavelengths. Furthermore, firmness, pH, and TA can often be
74 predicted with more variable success depending on cultivar, instrument configuration,
75 and sample heterogeneity (Mancini et al. 2020; Amoriello et al. 2025). NIR
76 spectroscopy can quantify sugars that are traditionally measured destructively and can
77 nondestructively predict chlorophyll and individual sugars in whole banana fruit (Zude
78 2003). Subsequent banana studies further showed that partial least squares (PLS)-based
79 spectroscopy models can estimate dry matter, pH, TSS, and acid–Brix ratio in intact
80 fruit, supporting the feasibility of spectral maturity assessment (Jaiswal et al. 2012).

81 Among chemometric methods, partial least squares regression (PLSR) is one of the
82 most widely used techniques for relating high-dimensional spectral data to reference
83 quality measurements, as it can handle multicollinearity and generate latent variables
84 (Sibuea et al. 2025). The application of PLSR has been successful in predicting TSS,
85 firmness, dry matter, pH, and acidity in various fresh fruits, including mango, pear,

pomegranate, grape, apple, kiwifruit, melon, and apricot (Cevoli et al. 2024; Hemmati et al. 2024; Kasampalis et al. 2025).

However, current NIR research on bananas has primarily focused on post-harvest ripeness grading and the detection of fixed ripeness stages, with few studies continuously and dynamically tracking attributes such as sugars, starch, TSS, firmness, and acidity during the transition of bananas from green to ripe. As a result, time-series datasets covering the entire ripening cycle—based on daily sampling of individual fruit clusters under strictly controlled constant temperature and humidity storage conditions—are relatively scarce. Although numerous studies have demonstrated the ability of NIR-PLSR to predict conventional attributes such as TSS, firmness, or dry-matter content (Guo et al. 2024; Elamshity and Alhamdan 2025), there is a lack of research on the simultaneous quantitative prediction of carbohydrate changes, which represent a core physiological process in banana postharvest ripening, and even fewer studies evaluate the predictability of multiple important quality attributes within a single research framework. Moreover, banana postharvest ripening is a dynamic time-series process, and samples from adjacent storage dates exhibit strong similarities in sugar and starch levels, water content, color, tissue structure, spectral shape, and ripeness. Therefore, if samples from different storage times are randomly assigned to training and validation sets, there may be a strong similarity in the temporal distribution between the two sets, leading to an overestimation of model performance (Amoriello et al. 2025). Most existing studies rely on random splitting or k-fold cross-validation,

methods that overlook the issue of data leakage in the temporal dimension, and model performance often declines significantly when predicting samples with unseen storage durations—a phenomenon rarely reported in the literature. Practical applications of postharvest quality monitoring constitute typical time-series forecasting problems; however, very few current studies use leave-one-day-out cross-validation (LODO-CV) to evaluate a model's time-series generalization ability, and there is a lack of systematic comparisons of performance differences between random splitting and time-series validation. Additionally, many machine learning-based studies on non-destructive fruit inspection lack mechanistic explanations for the relationship between feature wavelengths and the chemical bonds of target physicochemical attributes, and they lack a corresponding analysis of the correlation between spectral evolution trends during the storage period and the dynamic changes in multiple physicochemical attributes, making it difficult to fully reveal the intrinsic patterns of spectral responses during the banana ripening process.

To address these gaps, this study aimed to monitor changes in fructose, glucose, sucrose, starch, TSS, firmness, pH and TA during controlled postharvest ripening of bananas and to evaluate their nondestructive prediction using NIR spectroscopy and PLSR. The specific objectives were to specifically compare seven spectral preprocessing strategies through full-spectrum modeling; model parameters were optimized using repeated k-fold cross-validation (repeated k-fold CV), while LODO-CV was used to evaluate the model's generalization ability under time-series conditions, and the final model is

validated using an independent prediction set; compare the predictability of carbohydrates, texture and acidity-related attributes within a single experimental framework; select informative wavelengths using variable importance in projection (VIP) scores and interpret the selected regions based on PLS regression coefficients and the assigned NIR bands; rebuild the model using the selected wavelengths and compare the predictive performance of the full-spectrum model with that of the informative-wavelength model. This framework was developed to assess the potential and limitations of NIR spectroscopy for monitoring banana quality during ripening, while providing a basis for the development of simpler sensing systems.

2. Materials and methods

2.1. Sample preparation

Bananas were harvested at commercial maturity from the orchards in Hainan Province, China, and transported to the laboratory on the following day of harvest. Selected fruits were of medium size and free of obvious physical damage and defects. The fruits were stored individually in a constant temperature and humidity chamber maintained at 21 ± 1 °C and $85 \pm 5\%$ relative humidity for ripening. The storage period lasted 12 days covering the entire postharvest ripening process. Six fruits, each obtained from a different bunch, were analyzed on each storage day, resulting in a total of 72 fruits from 72 bunches. Before measurements, the banana samples were taken out and equilibrated at room temperature for 0.5 h. Next, the samples were washed thoroughly with ultrapure water and air-dried. On each sampling day, destructive analyses were performed on

individual fruits. For each sample, NIR spectra were first collected using a NIR spectrometer, after which the same fruit was subjected to destructive analysis of its physicochemical attributes.

2.2. Spectral data acquisition

As shown in Fig. 1, absorbance spectra in diffuse-reflection mode were acquired from each intact banana using a near-infrared spectrometer (SupNIR-1500, Focused Photonics Inc., Hangzhou, Zhejiang, China) with a wavelength range of 1000–1799 nm and a resolution of 1 nm. Spectra were collected at three equidistant positions along the longitudinal axis of each fruit, and the three spectra were averaged to obtain one representative spectrum per fruit.

Fig. 1

2.3. Reference measurements

2.3.1. Carbohydrates

Banana pulp contains the soluble sugars fructose, glucose, and sucrose. A high-performance liquid chromatography (HPLC) system (UltiMate 3000, Thermo Fisher Scientific Inc., Waltham, MA, USA) equipped with an evaporative light-scattering detector (ELSD) (ELSD 6000, Alltech, Inc., Nicholasville, KY, USA) was used to determine these compounds. A portion of the pulp was homogenized, and a sample was placed into a centrifuge tube for the determination of soluble sugars. Another portion of the sample was placed in 5% (w/w) hydrochloric acid and subjected to reflux under

169 a condenser in a 95 °C water bath for 2 hours for hydrolysis. After complete hydrolysis
170 of starch (indicated by the absence of a blue color when tested with iodine–potassium
171 iodide reagent), this portion was used for starch determination, and the starch content
172 was calculated using the modified formula (Eq. (1)) (European Commission 2008). The
173 soluble sugar and starch contents of the banana pulp were expressed as grams per 100
174 grams of material (g/100 g).

$$175 \quad Starch = 0.9 \times (P_1 - P_2 - P_3) \quad (1)$$

176 Where P_1 represents the total glucose content after starch hydrolysis; P_2 represents
177 the glucose content of the sample before hydrolysis; P_3 represents the original sucrose
178 content of the sample; and 0.9 is the conversion factor for converting glucose to starch.

179 For soluble sugar and starch determination, adequate amounts of zinc acetate
180 solution and potassium ferricyanide solution were added to the respective samples. The
181 mixtures were vortex-mixed, sonicated for 30 min, and brought to volume, followed by
182 centrifugation at $12,866 \times g$ (relative centrifugal force, RCF) for 20 min. The resulting
183 supernatant was filtered through a 0.45 μm hydrophilic membrane filter into a 2-mL
184 sample vial for HPLC-ELSD analysis. The instrumental conditions were as follows:

185 a) Column: NH_2 amino column, 4.6×250 mm, 5 μm ;

186 b) Mobile phase: Acetonitrile + water = 75 + 25 (v/v);

187 c) Flow rate: 1.0 mL/min;

188 d) Column temperature: 30°C;

e) Injection volume: 10 μ L;

f) ELSD conditions: Drift tube temperature: 85°C; gas flow rate: 2.1 L/min
supplied by a gas generator.

2.3.2. Firmness

The maximum puncture force required to break through the peel and underlying tissue was used as an attribute of firmness. Firmness, expressed in gram-force (gf) was determined with a texture analyzer (CT-3, AMETEK Brookfield, Middleboro, MA, USA). Measurements were taken at three equidistant positions along the longitudinal axis of each fruit, and the final result was calculated as the average. A probe with a diameter of 2 mm was used at a test speed of 1 mm/s.

2.3.3. pH, TA and TSS

Part of the fruit pulp was homogenized, centrifuged at $12,866 \times g$ (RCF) for 20 min, and allowed to equilibrate to room temperature. The pH of the supernatant was measured in triplicate using a pH meter (PB-10, Sartorius AG, Göttingen, Germany), and the results were averaged. TSS were determined using an Abbe refractometer (2WAJ, Shanghai Optical Instrument, Shanghai, China); the supernatant was measured three times, and the average was expressed in °Brix. TA was determined by titration. The supernatant was diluted with water at a 1:3 (v/v) sample-to-water ratio and mixed well. A 25 mL aliquot of the diluted sample was taken, and a few drops of phenolphthalein indicator were added. The mixture was titrated with 0.1 mol/L NaOH

until a pale pink color was observed. TA was then calculated according to Eq. (2) (Laryea et al. 2025).

$$TA (\%) = \frac{c \times (V_1 - V_0) \times K \times V_2 \times F}{m \times V_3 \times 1000} \times 100\% \quad (2)$$

Where c is the concentration of the NaOH solution; V_1 and V_0 are the volumes of NaOH solution consumed by the sample and the blank, respectively; F is the dilution factor ($F = 4$); V_3 is the volume of diluted solution used for titration; V_2 is the total volume of supernatant obtained after extraction; m is the mass of banana pulp sample; K is the conversion factor for malic acid ($K = 0.067$).

2.4. Statistical analysis of physicochemical attributes

Data for all physicochemical attributes on each storage day were subjected to descriptive statistics. A one-way analysis of variance (ANOVA) was performed to test the significance of differences among storage days, with storage time as the independent factor and each physicochemical attribute as the dependent variable. Before ANOVA, the normality assumption was evaluated using the Shapiro-Wilk test on the residuals and by inspecting Q-Q plots. Levene's test with median centering was used to assess homogeneity of variance. When the assumptions of normality and equal variance were met (Levene's test $P \geq 0.05$), classical one-way ANOVA was performed, followed by Tukey's honestly significant difference post hoc test for comparison of means. When equal variance was violated (Levene's test $P < 0.05$), Welch's one-way ANOVA was used, followed by the Games-Howell post hoc test, which does not assume equal variance. All statistical analyses were conducted at the significance level $\alpha = 0.05$.

Groups that share at least one letter are not significantly different, whereas groups that do not share letters are significantly different at $P < 0.05$. Error bars represent the mean $\pm$ standard deviation (SD).

2.5. Spectral preprocessing

NIR spectra often contain various types of interference, such as baseline drift, light scattering, and instrumental artifacts. To obtain meaningful chemical information and improve model robustness, appropriate spectral preprocessing is required (Li et al. 2026). To enhance spectral features and suppress systematic noise, seven preprocessing strategies were applied to the raw NIR spectra: (1) raw spectra without preprocessing (Raw), (2) Savitzky-Golay smoothing (SG0), (3) Savitzky-Golay first derivative (SG1), (4) standard normal variate (SNV) transformation, (5) SNV followed by SG1 (SNV+SG1), (6) multiplicative scatter correction (MSC), and (7) MSC followed by SG1 (MSC+SG1).

Savitzky-Golay filtering was performed using a window length of 11 data points and a second-order polynomial. SG0 smoothed the spectra while preserving peak shapes. For SG1, the first derivative was calculated using the same parameters, which effectively removed baseline shifts and enhanced spectral resolution by emphasizing subtle differences in absorption.

SNV transformation was applied to individual spectra to correct for light scattering effects and path length variations. For each spectrum, the mean absorbance was subtracted, and the result was divided by the standard deviation:

$$SNV(x_i) = (x_i - \bar{x})/\sigma_x \quad (3)$$

Where x_i represents the absorbance at wavelength i , $\bar{x}$ is the mean absorbance across all wavelengths, and σ_x is the standard deviation.

MSC was applied to minimize the effects of light scattering caused by variations in particle size and surface properties (Zhu et al. 2025). Each spectrum was corrected by fitting a linear regression against the mean spectrum of the entire dataset. The corrected spectrum was then obtained by applying the fitted slope (a) and intercept (b) to the original spectrum:

$$MSC(x) = (x - b)/a \quad (4)$$

where a and b are the slope and intercept from the linear regression of each spectrum against the mean spectrum.

The SNV+SG1 method corrected scatter effects and baseline drift simultaneously while improving spectral features. Similarly, MSC+SG1 corrected multiplicative light-scattering effects and removed baseline drift, thereby enhancing spectral features.

2.6. Model and evaluation

2.6.1. Prediction model

PLSR was used to establish quantitative relationships between the NIR spectra and the physicochemical attributes. PLSR is particularly suitable for spectroscopic data because it can handle collinearity among spectral variables and extract LVs that maximize the covariance between spectra and the response variable (Zeng et al. 2024). Using the

SPXY partitioning method, the samples were divided into calibration and prediction sets with a ratio of 3:1. In order to optimize model parameters and evaluate the model's time generalization ability, two cross validation strategies were implemented within the calibration set.

The repeated k-fold cross-validation procedure involved randomly partitioning the data into five folds, with four folds used for calibration and one fold used for validation. To minimize the variance in performance estimates, this process was repeated 20 times with different random partitions. This method evaluated model performance on data with temporal characteristics similar to those of the training set (Mancini et al. 2020).

The one-SE rule was used to select the optimal number of LVs. In the LODO-CV approach, samples from each storage day were iteratively reserved as the validation set, while all samples from the remaining 11 days constituted the calibration set. This procedure was repeated 12 times, with each storage day used once as the validation set. LODO-CV mimics real-world scenarios in which data from different days must be used to predict fruit quality, thus providing a stronger test of model generalization to unseen storage days (Chen et al. 2024).

2.6.2. Variable selection methods

To identify the most informative spectral regions and reduce model complexity, informative wavelengths were selected based on VIP scores, and the selected regions were interpreted using PLS regression coefficients and the assigned NIR bands.

VIP scores quantify the contribution of each wavelength to the PLSR model by

weighting the squared PLS loadings according to the explained variance of each latent variable (Yu et al. 2020). Wavelengths with $VIP > 1$ were considered to have above-average importance for predicting the target attribute, because VIP values greater than 1 indicate variables that contribute more than the average contribution of all variables. The magnitude and sign of a regression coefficient in the final PLSR model indicate the strength and direction of the effect of a wavelength on the physicochemical attribute. Wavelengths with large absolute regression coefficients correspond to spectral regions with high predictive capacity. To ensure biological plausibility, the selected wavelengths were cross-referenced with known NIR absorption bands of the relevant chemical constituents. Spectral regions corresponding to overtones and combination bands of O–H and C–H stretching vibrations were prioritized. Wavelengths were considered informative if they satisfied $VIP > 1$, had a significant absolute regression coefficient, and corresponded to chemically interpretable absorption features.

2.6.3. Model evaluation

The performance of the models was assessed based on six complementary metrics computed from predicted versus measured values in both cross-validation frameworks. The coefficient of determination (R^2) measures the proportion of variance in the measured values that is explained by the model; higher R^2 values indicate a better fit of the data to the model. Root mean square error (RMSE) is the standard deviation of the prediction errors and is expressed in the same units as the measured attribute. Lower RMSE values indicate better prediction quality. Mean absolute error (MAE) reflects the

313 average magnitude of the prediction errors regardless of their sign. Because it is less
 314 sensitive to outliers than RMSE and provides a direct measure of the average prediction
 315 error, MAE is considered a robust estimator. Bias represents the average prediction
 316 error and indicates systematic over- or under-prediction. A bias close to zero indicates
 317 unbiased predictions, whereas positive values indicate over-prediction and negative
 318 values indicate under-prediction. The ratio of performance to deviation (RPD) evaluates
 319 model utility by comparing the standard deviation of the measured values with the
 320 prediction error. RPD values between 1.5 and 2.0 indicate that the model can distinguish
 321 between high and low values but has weak predictive capability; values between 2.0
 322 and 2.5 indicate that the model is suitable for approximate quantitative prediction;
 323 values between 2.5 and 3.0, or greater than 3.0, indicate good to excellent predictive
 324 performance and can be used for accurate prediction of the target attribute (Lin et al.
 325 2025). The ratio of performance to interquartile range (RPIQ) is an alternative
 326 robustness measure that uses the interquartile range instead of the standard deviation
 327 and is therefore less sensitive to outliers. The interpretation of RPIQ is similar to that
 328 of RPD, with higher values indicating a more robust model. The calculation formulas
 329 for these metrics (Eqs. (5) – (10)) are as follows:

$$330 \quad R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

$$331 \quad RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (6)$$

$$332 \quad MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (7)$$

$$333 \quad Bias = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)}{n} \quad (8)$$

$$RPD = SD/RMSE \quad (9)$$

$$RPIQ = (Q_3 - Q_1)/RMSE \quad (10)$$

Where y_i is the measured value, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the mean of measured values, n is the number of samples, Q_3 and Q_1 are the third and first quartiles of measured values, respectively.

2.7. Software

Python 3.13 was used for statistical analysis, spectral preprocessing, modeling, and visualization.

3. Results and discussion

3.1. Spectra

The average NIR absorbance spectra of banana samples collected over the 12-day storage period is shown in Fig. 2a. The spectra exhibit characteristic absorption features across the 1000–1799 nm range, arising from overtones and combination bands of fundamental molecular vibrations in organic matter. The most prominent feature is a strong, broad absorption band around 1450 nm, which is mainly attributed to the first overtone of O–H stretching vibrations of water and carbohydrates, as well as O–H stretching and bending combination bands. This band is dominant because of the high moisture content of banana fruit. A weaker peak near 1750 nm is attributed to the first overtone of O–H stretching vibrations associated with carbohydrates such as starch and sugars. The spectral region between 1150 and 1200 nm is associated with the second

overtone of O–H stretching and combination bands of O–H vibrations. (Golic et al. 2003; Goodarzi and Saeys 2016; Walsh et al. 2020; Heise et al. 2021). Therefore, to enhance the precision of the spectra and ensure the accuracy and reliability of the data, this study employed seven different preprocessing techniques as illustrated in Fig. 2b–h. After preprocessing, the baseline drift in the reflectance spectra was significantly alleviated, the smoothness of the spectral curves was further improved, and the curves became more compact.

Fig. 2

The spectral data exhibited a clear temporal trend, with absorbance intensity increasing steadily from day 1 to day 12 across almost all wavelength regions. The increase in absorbance during ripening likely reflected changes in the optical properties of banana tissue caused by a broad range of physical and biochemical alterations. The observed spectral variation is probably driven by the conversion of starch to sugars, redistribution of water, and tissue softening, which alters light scattering. At the same time, cell wall breakdown and tissue softening may modify light scattering properties, increasing the effective path length and overall absorbance. The spectral differences in O–H absorption regions could also result from changes in moisture distribution within the tissue, including shifts between bound water in cell walls and free water in the cytoplasm. The spectral progression closely parallels the physicochemical changes shown in Fig. 3, with the most pronounced spectral differences occurring between days 6 and 10, coinciding with the climacteric period when starch-to-sugar conversion and

cell wall degradation are most active. The wavelength regions showing the greatest temporal variation, approximately 1400–1500 nm and 1700–1799 nm, correspond to absorption bands of water and carbohydrates. This suggests that changes in water status and carbohydrate composition contributed substantially to the observed spectral variation.

3.2. Physicochemical attributes

3.2.1. Carbohydrates

As storage time increased, starch degradation and the formation of soluble sugars (glucose, fructose, and sucrose) in banana fruits followed distinctly different temporal patterns (Fig. 3). Overall, the increase in soluble sugars coincided with the decrease in starch content between days 5 and 7, followed by a plateau in carbohydrate levels between days 7 and 9. Unripe banana pulp has a very high starch content (approximately 22%). During ripening, this starch is rapidly hydrolyzed by α - and β -amylase and further converted to soluble sugars through the action of enzymes such as UDP-glucose pyrophosphorylase, sucrose synthase, and invertase (Wang et al. 2026).

Fig. 3

The fructose content (Fig. 3a) was virtually undetectable during the first 5 days of storage, suggesting that the bananas were at the pre-climacteric stage with low metabolic activity. After day 6, a sudden increase was recorded, marking the start of the climacteric ripening. This accumulation pattern suggested the activation of invertase

and sucrose synthase, which hydrolyzed sucrose into monosaccharides (Li et al. 2024). The temporal profile of glucose (Fig. 3b) was almost identical to that of fructose, suggesting equimolar release of the two reducing sugars from sucrose. Sucrose content showed a more complex biphasic pattern (Fig. 3c). Initial values were low, followed by a sharp increase beginning on day 6 and peaking on day 8. The content fluctuated between days 8 and 10 with no clear directional trend. These fluctuations, together with the large standard deviations, indicated that individual fruits metabolized sucrose asynchronously during the later storage days, reflecting fruit-to-fruit variation in enzyme activities. Starch (Fig. 3d) showed the expected inverse pattern, as starch serves as the primary carbon source. Starch content initially remained high and relatively constant until day 4. From day 5 onward, starch was rapidly depleted, declining significantly to 11.4 ± 3.3 g/100 g by day 6, and the decline continued until day 8. From day 8 to day 12, starch content remained very low and relatively stable, corresponding to approximately 20–30% of the initial value on day 1. This loss pattern is typical of climacteric fruit ripening, during which amylolytic enzymes (α -amylase and β -amylase) are activated at the onset of the respiratory climacteric and ethylene rise (Do Nascimento et al. 2006). Even at full ripeness (days 10–12), residual starch may persist in bananas; this residual starch may consist of amylopectin-rich fractions that resist complete amylolysis, or of starch granules that are physically inaccessible to degradative enzymes (Gao et al. 2016).

3.2.2. TSS, firmness, pH and TA

416 TSS mainly reflects soluble sugars, as well as organic acids, vitamins, and minerals,
417 and is a key attribute for assessing banana ripeness and quality. The increase became
418 rapid after day 6 and stabilized after day 8 (Fig. 3e). The direct conversion of starch
419 into sugars results in an increase in TSS, which improves fruit sweetness. Moreover,
420 the breakdown of starch into soluble sugars is also responsible for the loss of firmness.
421 The rapid decrease in firmness on day 5 further supports this observation (Fig. 3f). The
422 early-stage increase in firmness may have occurred due to slight dehydration and
423 shrinkage of the fruit skin, among other factors. As the fruit ripens, ethylene production
424 and the respiration rate increase. This promotes the breakdown of complex
425 carbohydrates into simple sugars, making the fruit sweeter and softer (Chen et al. 2022).
426 The decrease in firmness is also attributed to the breakdown of cellulose, hemicellulose,
427 and pectin in the cell walls. As degradation progresses, the activities of pectinases and
428 cellulases increase, thereby degrading the middle lamella and primary cell wall and
429 ultimately reducing the firmness of the peel (Paniagua et al. 2014).

430 The composition of organic acids changes significantly during ripening, ultimately
431 affecting the flavor, palatability, and postharvest quality of banana fruit (Huang et al.
432 2024). TA primarily refers to the total amount of acidic substances that can neutralize a
433 strong base; its value increased gradually over time, then decreased and stabilized (Fig.
434 3h). The pH of banana fruit is a measure of the concentration of free hydrogen ions and
435 provides a direct measure of fruit acidity. During the whole storage period, pH generally
436 decreased first, then increased, and finally leveled off (Fig. 3g). TA and pH exhibited

opposite trends over time. Both underwent significant changes from day 5 onward, with day 7 serving as a turning point, after which values remained relatively stable until day 10. In the early stages of ripening, the concentration of organic acids in bananas gradually increased, resulting in a sour and astringent taste. As ripening progressed, the total organic acid content decreased significantly, primarily because these acids were utilized as respiratory substrates in the tricarboxylic acid (TCA) cycle. Once bananas entered the senescence stage, metabolic activity changed; because of the decrease in respiratory rate, the demand of the TCA cycle for organic acids declined (Sheng et al. 2023; Wang et al. 2026).

In summary, the firm, starchy banana (days 1–5) first underwent a rapid climacteric phase (days 6–9), during which starch was converted to sugars, softening occurred, and organic acids accumulated. Fully ripe bananas (days 10–12) were characterized by high sugar content, soft texture, and moderate acidity.

3.3. Full spectra regression model

The statistics for the two datasets are summarized in Table 1. Calibration and prediction results of the PLSR models for predicting the eight physicochemical attributes are reported in Table 2-4. Calibration results were obtained by first optimizing the parameters associated with each model via repeated k-fold CV on the calibration set, and then applying the optimal parameters on the entire calibration set for the final model calibration. LODO-CV validation results were obtained by cross-validating the calibration model within the calibration set. The prediction results were achieved by

using the calibrated model on the prediction set for the final evaluation. The optimization process during the cross validation was performed using the combinations of different preprocessing methods and LVs. Taking the optimal preprocessing method for firmness as an example, significant changes in RMSEC and R^2_C were observed as the number of LVs varied (Fig. 4). It can be seen that, rather than directly selecting the LVs yielding the smallest RMSEC, the smallest possible number of LVs was chosen within an acceptable cross-validation-error range to reduce the risk of model overfitting. Beyond the optimal LVs, the curve shows an upward trend, and the error gradually increases. Table 5 reports additional prediction-performance metrics for the prediction set for each physicochemical attribute under the optimal model with the optimal preprocessing method. The accuracy of the models varied considerably among physicochemical attributes, reflecting differences in their spectral responses and in the degree of compositional variation during ripening.

Table 1

Table 2

Table 3

Table 4

Fig. 4

Table 5

As shown in Table 2, the regression models indicated that, with the exception of pH and

478 TA, the R^2_c values for all other physicochemical attributes were greater than 0.70.

479 Using the calibration models, the validation results for LODO-CV are shown in Table

480 3. For all physicochemical attributes under identical preprocessing conditions, model

481 performance was, to varying degrees, poorer than that obtained from repeated k-fold

482 CV. This suggests that traditional random-split validation methods tend to overestimate

483 the performance of the matured models. The calibration models for soluble sugars and

484 TSS still maintain high performance in terms of temporal generalizability; however, the

485 calibration model for starch showed weaker predictive ability for unknown storage

486 durations. This may be because the relationship between spectral response and changes

487 in soluble sugars and TSS is relatively stable, whereas the relationship between starch

488 content and spectral response lacks sufficient temporal robustness. In particular, the

489 pronounced nonlinear behavior of pH and TA was one of the key factors affecting the

490 temporal generalizability of the calibration model (Fig. 3g-h).

491 With the calibrated model, prediction results were reported as shown in Fig. 5. The

492 carbohydrate components showed good internal predictive performance. Glucose (Fig.

493 5b) exhibited the best predictive ability when SG0 preprocessing was used ($R^2_p = 0.92$

494 and $RPD = 3.57$). For fructose (Fig. 5a), SG0 preprocessing yielded $R^2_p = 0.88$ and

495 $RPD = 2.99$. The sucrose (Fig. 5c) model with MSC+SG1 preprocessing achieved R^2_p

496 $= 0.75$ and $RPD = 2.05$ was obtained. For starch (Fig. 5d), SG0 preprocessing produced

497 an RPD of 2.63. The RMSEP values for fructose, glucose, sucrose, and starch were 0.65,

498 0.54, 1.97, and 2.89 g/100 g, respectively.

Fig. 5

With SG1 preprocessing, TSS (Fig. 5e) showed excellent predictive performance, with $R^2_P = 0.92$, RPD = 3.54, and RPIQ = 6.91. Models with RPIQ values above 5.0 indicate robust performance even in the presence of outliers. TSS was expected to perform better because it represents the total amount of all soluble sugars and other dissolved solids, thereby providing a stronger and more stable spectral signal than individual sugars. Firmness (Fig. 5f) prediction yielded good results with raw spectra ($R^2_P = 0.90$; RPD = 2.95).

The models for pH and TA did not reach the threshold for acceptable accuracy, indicating that they are not suitable for quantitative prediction due to their very low R^2_P and RPD values (Fig. 5g-h). As shown in Fig. 3g-h, this failure can be attributed to the narrow variation range and small absolute changes in both pH and TA during the 12-day storage period. These small changes approach the noise level of NIR measurements, making it difficult to extract meaningful spectral signals. In addition, organic acids in banana have relatively weak and overlapping NIR absorption features compared with carbohydrates and water, further limiting the discriminative power of the models.

As shown in Table 4, there were significant differences in the performance of prediction models for various physicochemical attributes when different preprocessing methods were applied. This indicates that different physicochemical attributes manifest differently within spectra and respond distinctly to preprocessing methods; certain preprocessing methods can effectively enhance spectral features for specific attributes,

yet they may prove ineffective for other attributes or even exert negative effects. This may be due to differences in the types and degrees of spectral interference (such as baseline drift, scattering effects, and noise) experienced by different physicochemical attributes. Therefore, it is essential to conduct a systematic comparison and selection of preprocessing methods in spectral modeling.

For physicochemical attributes with good predictability ($RPD > 2.0$), both the MAE and bias are close to 0 within their respective orders of magnitude (Table 5). This indicates that the prediction models exhibit high accuracy, minimal systematic bias, and good generalization ability, further demonstrating the strong correlation between these physicochemical attributes and spectral features. Therefore, they are valuable for further wavelength variable selection.

3.4. Selection of informative wavelength variables

To reduce model complexity and identify informative spectral regions, VIP scores were used to select informative wavelengths for six physicochemical attributes (fructose, glucose, sucrose, starch, soluble solids, and firmness) that performed well in full-spectrum prediction. The selected regions were then interpreted based on PLS regression coefficients and the assigned NIR bands (Fig. 6). Attributes with weak full-spectrum models (TA and pH) were excluded from variable selection, since dimensionality reduction would not address their fundamental predictability issues.

Fig. 6

Fructose and glucose (Fig. 6a-b) exhibited similar informative wavelength regions because of their structural similarity as monosaccharides and comparable accumulation patterns. The region from 1350 to 1425 nm consistently showed high VIP scores and strongly negative PLS regression coefficients for both sugars. This region, associated with O–H overtones and sugar–water interactions, reflects the combined effects of soluble sugar accumulation, changes in free water content, and altered hydrogen-bonding networks (Walsh et al. 2020; Zhao et al. 2024). The negative-coefficient pattern likely resulted from changes in water-carbohydrate hydrogen-bonding, as increasing amounts of free water become structured around newly formed sugar molecules during ripening, thereby altering the characteristics of O–H vibrations. The core predictive region for monosaccharide quantification was identified in the 1430–1520 nm range, which was characterized by the highest VIP scores and strong positive PLS regression coefficients. The first overtone of O–H stretching is strongly affected by the hydrogen-bonding states of water and hydroxyl groups in sugars (Golic et al. 2003; Walsh et al. 2020). The positive coefficients indicated that increased sugar concentration enhanced absorbance within this region, consistent with the accumulation of sugar-associated O–H groups and the transition of tissue water from bound to free states during cell wall degradation. The 1690–1790 nm region displayed moderately elevated VIP scores with positive coefficients, corresponding to the first overtone of O–H stretching. This sensitive region is associated with O–H bonds in glucose and fructose molecules, as well as in structural carbohydrates (cellulose and hemicellulose) undergoing degradation (Goodarzi and Saeys 2016; Walsh et al. 2020; Xu et al. 2025). This

additional O–H region complements the other O–H features, providing a more complete spectral picture of monosaccharide chemistry.

Sucrose (Fig. 6c) showed a more complex pattern of informative wavelengths across two distinct regions. The 1120–1150 nm range was dominated by high VIP scores, mostly negative coefficients. This O–H second-overtone region is sensitive to the skeletal structures of carbohydrates, including sugars and other dissolved organic matter contributing to TSS (Golic et al. 2003; Walsh et al. 2020; Yang et al. 2021). The prominence of this region suggested that sugar accumulation induced characteristic changes in O–H vibrational modes. Within the 1380–1510 nm range, several localized VIP peaks were observed with positive coefficients, and the PLS regression coefficient profile showed a clear alternating pattern. This O–H overtone region reflects complex hydrogen-bonding environments, as sucrose concentration varies during its dynamic accumulation and hydrolysis. The alternating coefficient pattern may reflect overlapping O–H absorption from sucrose, monosaccharides, and water.

The spectral signature of starch, a polymeric carbohydrate undergoing depletion, was characterized by three informative wavelength regions (Fig. 6d). The 1000–1020 nm region exhibited high VIP scores and strong positive coefficients. This band is likely attributable to the second or higher overtone of O–H stretching, with additional contributions from O–H combination bands associated with the helical glycosidic structure of starch (Golic et al. 2003; Theanjumpol et al. 2013; Walsh et al. 2020). The positive coefficient indicated that higher starch content enhanced absorbance within

this region, possibly because of the organized water molecules within the amylose helix and crystalline regions of amylopectin. The 1350–1550 nm region showed the highest VIP scores across the entire spectrum, with a strong positive peak and alternating positive coefficients. This strong first O–H overtone absorption region is highly sensitive to hydrogen-bonding changes, as starch granules undergo hydrolysis and water transitions from a structured state (bound to starch granules) to a free state (cytoplasmic). The alternating coefficient pattern indicates that multiple overlapping O–H absorption features are associated with different water populations, such as granule-bound water, free water, and sugar-bound water, which change during starch-to-sugar conversion. The 1770–1799 nm region also showed high VIP values and strong negative coefficients. This narrow band at the long-wavelength end corresponds to the first O–H stretching overtone and associated combination modes involving O–H stretching and structural deformation vibrations (Subedi and Walsh 2011; Heise et al. 2021). The strong negative coefficients indicate that starch depletion leads to decreased absorbance in this region. This reflects the loss of organized O–H environments in crystalline starch granules and the emergence of more disordered O–H environments from sugars and degraded cell wall polysaccharides.

TSS (Fig. 6e) showed one informative wavelength region. In the 1490–1520 nm region, VIP values showed two high peaks, and the coefficients displayed sharp alternating positive–negative transitions. These sharp changes suggest that this narrow region captures spectral features that can distinguish among different TSS levels.

Banana firmness is a physical texture parameter that reflects the structure of the fruit pulp cell walls, the state of pectin, starch granules, moisture content, tissue density, and ripeness. It is primarily predicted based on three informative wavelength regions, specifically near the three VIP score peaks at 1400 nm, 1460 nm, and 1720 nm (Fig. 6f). The 1350–1500 nm range contains the two highest VIP score peaks and very strong positive and negative coefficients, indicating that the model primarily utilizes spectral changes in this region to characterize firmness. This region corresponds to the strong absorption band of the first harmonic of the O–H stretching vibration, which is closely related to changes in water content, pectin, hydroxyl groups in cell wall polysaccharides, and hydrogen bonding networks. During banana ripening, the degradation of cell wall pectin, the hydrolysis of starch, and changes in the water-binding state cause the pulp to soften. The model utilizes a combination of positive and negative PLS regression coefficients near 1400 nm and 1460 nm in this region to extract spectral differences. The region near 1720 nm, corresponding to the first harmonic of C–H, is typically associated with carbohydrate structures. The VIP score is slightly higher than 1, and the coefficients are negative, indicating that this region makes a certain auxiliary contribution to firmness prediction. This may be related to starch degradation and the degradation of cell wall polysaccharides.

Common patterns across all six attributes provide important insights into banana fruit chemistry and NIR spectral interpretation (Table 6). The O–H absorption regions (1350–1550 nm) were selected by all models, with high VIP scores indicating that water

and hydrogen bonding are the main sources of spectral changes during ripening. The PLS regression coefficients exhibited alternating positive–negative patterns within broad absorption bands, indicating that the fine spectral features within these bands contain useful compositional information. This indicates that high-resolution spectroscopic data can provide more detailed information about fruit quality. The overlap between the wavelengths selected for monosaccharides and TSS indicates that reducing sugars are the main source of TSS in banana fruit.

Table 6

3.5. Regression models using informative wavelength variables

Calibration and prediction results of the PLSR models for predicting the eight physicochemical attributes are reported in Table 7 and Table 8. The LODO-CV validation results showed that, for all physicochemical attributes, the model performance was significantly weaker than the results obtained from repeated k-fold CV. This indicates that VIP feature wavelength screening does not fundamentally address the limitation of traditional random split validation methods, which tend to overestimate model performance. However, compared to the full-spectrum model, the model based on screened wavelengths showed virtually no change in temporal generalization performance. This suggests that, for post-harvest time-series dynamic banana samples, VIP feature wavelength screening not only reduces the model's dimensionality but also has a negligible impact on temporal generalization performance. This indicates that this screening method can capture true spectral-chemical

relationships and provides a basis for reducing model complexity in models that similarly rely on time-series spectral analysis.

Table 7

Table 8

With the calibrated model, the prediction results based on the filtered wavelength variables were shown in Fig. 7. For all physicochemical attributes, the predictive performance of the model declined slightly. The RPD for fructose (Fig. 7a) decreased from 2.99 to 2.71 (R^2_P decreased to 0.86), and for glucose (Fig. 7b) from 3.57 to 3.23 (R^2_P decreased to 0.90), and firmness (Fig. 7f) from 2.95 to 2.60 (R^2_P decreased to 0.84). Although the performance of the prediction models for these physicochemical attributes declined somewhat, it remained on par with that of the full-spectrum model. Their RMSEPs were 0.71, 0.60 g/100 g, and 176.6 gf, respectively. For sucrose, starch, and TSS, however, the predictive performance was one level lower than that of the full-spectrum model. However, the prediction models for starch and TSS still achieved good predictive performance, with $R^2_P = 0.78$ and RPD = 2.19, and $R^2_P = 0.85$ and RPD = 2.64, respectively (Fig. 7c-e). The prediction model based on the selected wavelength variables did not perform well in predicting sucrose content ($R^2_P = 0.72$ and RPD = 1.93).

Fig. 7

As shown in Table 8, for fructose, glucose, starch, TSS, and firmness, using only 38 %,

40 %, 33 %, 20 %, and 56 % of all available wavelengths, respectively, lowered prediction complexity without substantially compromising predictive performance.

Although the MAE increased slightly, the bias decreased for fructose, glucose, TSS, and firmness, while for other parameters, the bias increased only slightly. This indicates that, although the model based on the filtered wavelengths incurred a slight cost in terms of random error, it reduced systematic bias. For TSS, the predictive model still maintains an RPIQ > 5.0, indicating that even though only 20% of the wavelength variables were used, its performance remains robust in the presence of outliers.

Overall, the models were simplified by selecting informative wavelengths while maintaining predictive performance comparable to that of the full-spectrum models. The number of wavelength variables was reduced by 44–80%, depending on the target attribute, compared with the full-spectrum models, indicating that most of the spectral information was concentrated in a few key absorption regions. In addition to improving computational efficiency, dimensionality reduction facilitated the chemical interpretation of the selected spectral regions by linking them to known absorption bands of relevant molecular constituents.

4. Conclusions

This study demonstrated that NIR spectroscopy combined with PLSR is capable of tracking the main biochemical and structural changes during postharvest banana ripening under controlled storage, particularly starch-to-sugar conversion, tissue softening, and the increase in total soluble solids. The spectral responses were

dominated by these ripening-related compositional changes, confirming the feasibility of using NIR signals to reflect the dynamic evolution of banana quality. The predictive performance varied considerably among quality attributes. Carbohydrate-related attributes such as glucose, fructose, starch, and total soluble solids, as well as firmness, could be predicted with acceptable to good accuracy, whereas sucrose, pH, and titratable acidity remained challenging. These differences can be attributed to the magnitude and spectral relevance of the chemical changes associated with each attribute, providing useful guidance for selecting target indicators in practical applications. A key methodological finding was that leave-one-day-out cross-validation generally produced lower performance than repeated k-fold cross-validation, indicating that conventional random-split validation tends to overestimate the generalization ability of models built on time-series postharvest data. This highlights the importance of temporal validation strategies for fruit ripening studies and offers a more realistic protocol for evaluating NIR models intended for continuous monitoring. VIP-based feature wavelength selection substantially reduced model dimensionality while maintaining comparable predictive performance under time-series validation. The retained wavelengths corresponded mainly to O–H and C–H absorption bands in water and carbohydrates, suggesting that the selected variables captured the actual compositional changes during ripening rather than spurious spectral variations. This supports the development of simplified, low-cost NIR sensing systems using a limited number of wavelengths, especially for monitoring total soluble solids, monosaccharides, and firmness.

708 The main limitation is that the dataset came from a single batch and one controlled
709 storage condition, so the reported prediction levels should be viewed as internal
710 estimates. External validation across different cultivars, seasons, and commercial
711 handling conditions is needed before practical use. Even so, the results support the
712 development of simplified NIR sensing systems targeting a small number of
713 informative wavelengths for rapid, nondestructive assessment of banana quality,
714 especially TSS, monosaccharides and firmness. These also provide new insights into
715 future validation methods for NIR models for the dynamic storage of fruits.

Acknowledgements

The authors gratefully acknowledge the financial support provided by the National Key Research and Development Program of China (Grant No. 2022YFF1100900), the Shaanxi Key Laboratory for Safety Monitoring of Food and Drug (Grant No. SXKLFD202603).

Author contribution

Pu Zhang: Writing – original draft, Visualization, Writing – review & editing. Feng Zhang: Methodology, Funding acquisition. Jin Luo: Investigation, Writing – review & editing. Wenlin Wang: Resources, Validation. Xiaohe Wang: Resources, Data curation. Hanxia Liu: Formal analysis, Investigation. Fang Lin: Methodology. Data curation. Hong Zhao: Writing – review & editing, Validation. Jinfeng Peng: Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Funding

This study was financially supported by the National Key Research and Development Program of China (2022YFF1100900), the Shaanxi Key Laboratory for Safety Monitoring of Food and Drug (No. SXKLFD202603).

Declarations

Competing interests

The authors declare no competing interests.

Declaration of generative AI in scientific writing

737 The authors did not use generative artificial intelligence (AI) nor AI-assisted
738 technologies in the writing process.

References

- Al-Dairi M, Pathare PB, Al-Yahyai R, et al (2023) Postharvest quality, technologies, and strategies to reduce losses along the supply chain of banana: A review. *Trends Food Sci Technol* 134:177–191. <https://doi.org/10.1016/j.tifs.2023.03.003>
- Amoriello T, Ciorba R, Ruggiero G, et al (2025) Vis/NIR spectroscopy and vis/NIR hyperspectral imaging for non-destructive monitoring of apricot fruit internal quality with machine learning. *Foods* 14:196. <https://doi.org/10.3390/foods14020196>
- Cevoli C, Iaccheri E, Fabbri A, Ragni L (2024) Data fusion of FT-NIR spectroscopy and vis/NIR hyperspectral imaging to predict quality parameters of yellow flesh “jintao” kiwifruit. *Biosyst Eng* 237:157–169. <https://doi.org/10.1016/j.biosystemseng.2023.12.011>
- Chen H, Fan L, Wang T, et al (2026) Banana preservation-valorization synergies: advancing sustainable systems. *Crit Rev Food Sci Nutr* 66:2676–2706. <https://doi.org/10.1080/10408398.2025.2579775>
- Chen J, Li Y, Li F, et al (2022) Effects of procyanidin treatment on the ripening and softening of banana fruit during storage. *Sci Hortic* 292:110644. <https://doi.org/10.1016/j.scienta.2021.110644>
- Chen K, Yao C, Sun M, et al (2024) Raman spectroscopic analysis for osteoporosis identification in humans with hip fractures. *Spectrochim Acta, Part A* 314:124193. <https://doi.org/10.1016/j.saa.2024.124193>
- Do Nascimento JRO, Júnior AV, Bassinello PZ, et al (2006) Beta-amylase expression and starch degradation during banana ripening. *Postharvest Biol Technol* 40:41–47.

<https://doi.org/10.1016/j.postharvbio.2005.11.008>

Elamshity M, Alhamdan A (2025) Development and prediction of a non-destructive quality index (qi) for stored date fruits using VIS–NIR spectroscopy and artificial neural networks.

Foods 14:3060. <https://doi.org/10.3390/foods14173060>

European Commission. Commission Regulation (EC) No 900/2008 of 16 September 2008

laying down the methods of analysis and other technical provisions necessary for the application of the arrangements for imports of certain goods resulting from the processing

of agricultural products (Codified version), 2008.

<http://data.europa.eu/eli/reg/2008/900/oj>. Accessed 8 August 2026.

Fatima T, Waheed F (2024) Investigating the impact of different postharvest treatments on shelf-life extension and quality attributes of banana. J Food Process Nutr Res 33–41.

<https://doi.org/10.22194/fpn.2024.1007>

Gao H, Huang S, Dong T, et al (2016) Analysis of resistant starch degradation in postharvest ripening of two banana cultivars: Focus on starch structure and amylases. Postharvest Biol

Technol 119:1–8. <https://doi.org/10.1016/j.postharvbio.2016.03.022>

Golic M, Walsh K, Lawson P (2003) Short-wavelength near-infrared spectra of sucrose, glucose, and fructose with respect to sugar concentration and temperature. Appl Spectrosc 57:139–

145. <https://doi.org/10.1366/000370203321535033>

Goodarzi M, Saeys W (2016) Selection of the most informative near infrared spectroscopy wavebands for continuous glucose monitoring in human serum. Talanta 146:155–165.

<https://doi.org/10.1016/j.talanta.2015.08.033>

Guo Z, Zou Y, Sun C, et al (2024) Nondestructive determination of edible quality and watercore

- degree of apples by portable vis/NIR transmittance system combined with CARS-CNN. *J Food Meas Charact* 18:4058–4073. <https://doi.org/10.1007/s11694-024-02476-z>
- Heise HM, Delbeck S, Marbach R (2021) Noninvasive monitoring of glucose using near-infrared reflection spectroscopy of skin—constraints and effective novel strategy in multivariate calibration. *Biosensors* 11:64. <https://doi.org/10.3390/bios11030064>
- Hemmati A, Mahmoudi A, Jamshidi B, Ghaffari H (2024) Assessment of persian export pomegranate quality: a reliable non-destructive method based on spectroscopy and chemometrics. *J Food Compos Anal* 131:106202. <https://doi.org/10.1016/j.jfca.2024.106202>
- Huang P-H, Cheng Y-T, Lu W-C, et al (2024) Changes in nutrient content and physicochemical properties of cavendish bananas var. Pei chiao during ripening. *Horticulturae* 10:384. <https://doi.org/10.3390/horticulturae10040384>
- Jaiswal P, Jha SN, Bharadwaj R (2012) Non-destructive prediction of quality of intact banana using spectroscopy. *Sci Hortic* 135:14–22. <https://doi.org/10.1016/j.scienta.2011.11.021>
- Kasampalis DS, Tsouvaltzis P, Siomos AS (2025) Assessment of melon fruit nutritional composition using VIS/NIR/SWIR spectroscopy coupled with chemometrics. *Horticulturae* 11:658. <https://doi.org/10.3390/horticulturae11060658>
- Kikulwe EM, Okurut S, Ajambo S, et al (2018) Postharvest losses and their determinants: a challenge to creating a sustainable cooking banana value chain in Uganda. *Sustainability* 10:2381. <https://doi.org/10.3390/su10072381>
- Laryea D, Supapvanich S, Pinsirodom P, et al (2025) Evaluating variety and ethephon effect on banana fruit quality at different ripening stages: A comparative and multivariate approach.

- J Agric Food Res 19:101569. <https://doi.org/10.1016/j.jafr.2024.101569>
- Li J, Hu Y, Hu J, et al (2024) Sucrose synthase: an enzyme with multiple roles in plant physiology. J Plant Physiol 303:154352. <https://doi.org/10.1016/j.jplph.2024.154352>
- Li Y, Qiao Y, Zhu R, et al (2026) Non-destructive prediction of SSC and storage days classification of kiwifruit during postharvest storage using near-infrared hyperspectral imaging. Food Anal Methods 19:194. <https://doi.org/10.1007/s12161-026-03108-6>
- Lin S, Yan K, Sun Z, et al (2025) Non-destructive prediction of soluble solids content and firmness and maturity determination of guava fruit using color and fluorescence imaging. Postharvest Biol Technol 227:113620. <https://doi.org/10.1016/j.postharvbio.2025.113620>
- Mancini M, Mazzoni L, Gagliardi F, et al (2020) Application of the non-destructive NIR technique for the evaluation of strawberry fruits quality parameters. Foods 9:441. <https://doi.org/10.3390/foods9040441>
- Nandakumar P (2024) Biochemical challenges in post-harvest banana storage and their role in food security. Int J Clin Biol Biochem 6:48–51. <https://doi.org/10.33545/26646188.2024.v6.i2a.100>
- Paniagua C, Posé S, Morris VJ, et al (2014) Fruit softening and pectin disassembly: An overview of nanostructural pectin modifications assessed by atomic force microscopy. Ann Bot 114:1375–1383. <https://doi.org/10.1093/aob/mcu149>
- Ping F, Yang J, Zhou X, et al (2023) Quality assessment and ripeness prediction of table grapes using visible–near-infrared spectroscopy. Foods 12:2364. <https://doi.org/10.3390/foods12122364>
- Pornchaloempong P, Sharma S, Phanomsophon T, et al (2022) Non-destructive quality

- evaluation of tropical fruit (mango and mangosteen) purée using near-infrared spectroscopy combined with partial least squares regression. *Agriculture* 12:2060. <https://doi.org/10.3390/agriculture12122060>
- Sheng O, Yin Z, Huang W, et al (2023) Metabolic profiling reveals genotype-associated alterations in carotenoid content during banana postharvest ripening. *Food Chem* 403:134380. <https://doi.org/10.1016/j.foodchem.2022.134380>
- Sibuea N, Syamsudhuha S, Adnan A, Silalahi DD (2025) Robust method with cross-validation in partial least square regression. *J Math Comput Stat* 8:69–79. <https://doi.org/10.35580/jmathcos.v8i1.4766>
- Souripet A (2026) Conversion of starch in to sugar during ripening of tongka langit banana (*musa troglodytarum*), a local mollucan banana commodity, at different harvesting time. *Trans Chin Soc Agric Mach*. <https://doi.org/10.62321/issn.1000-1298.2025.12.1>
- Subedi PP, Walsh KB (2011) Assessment of sugar and starch in intact banana and mango fruit by SWNIR spectroscopy. *Postharvest Biol Technol* 62:238–245. <https://doi.org/10.1016/j.postharvbio.2011.06.014>
- Sun X, Deng D, Liu J, Feng S (2022) Model development and update of portable NIRS instrument for assessment of internal quality attributes of two navel orange varieties. *Front Nutr* 9:976178. <https://doi.org/10.3389/fnut.2022.976178>
- Thakur R, Pristijono P, Bowyer M, et al (2019) A starch edible surface coating delays banana fruit ripening. *LWT* 100:341–347. <https://doi.org/10.1016/j.lwt.2018.10.055>
- Theanjumpol P, Self G, Rittiron R, et al (2013) Selecting variables for near infrared spectroscopy (NIRS) evaluation of mango fruit quality. *J Agric Sci* 5:p146.

<https://doi.org/10.5539/jas.v5n7p146>

Walsh KB, Blasco J, Zude-Sasse M, Sun X (2020) Visible-NIR ‘point’ spectroscopy in postharvest fruit and vegetable assessment: the science behind three decades of commercial use. *Postharvest Biol Technol* 168:111246. <https://doi.org/10.1016/j.postharvbio.2020.111246>

Wang C, Li X, Zhang Z, et al (2025) Nondestructive quality detection of characteristic fruits based on vis/NIR spectroscopy: principles, systems, and applications. *Agriculture* 15:2167. <https://doi.org/10.3390/agriculture15202167>

Wang L, Jing T, Cai B, et al (2026) Comparative analysis of physicochemical quality and metabolite profiles during fruit ripening in three banana cultivars using multivariate approaches. *Postharvest Biol Technol* 231:113847. <https://doi.org/10.1016/j.postharvbio.2025.113847>

Watharkar RB, Chakraborty S, Srivastav PP, Srivastava B (2021) Physicochemical and mechanical properties during storage-cum maturity stages of raw harvested wild banana (musa balbisiana, BB). *J Food Meas Charact* 15:3336–3349. <https://doi.org/10.1007/s11694-021-00907-9>

Xu Y, Luo J, Xue S, et al (2025) A multi-task deep attention network for simultaneous rapid quantification of sucrose, glucose, and fructose contents in pumpkin using FT-NIR spectroscopy. *Food Chem* 496:146648. <https://doi.org/10.1016/j.foodchem.2025.146648>

Yang X, Liu G, He J, et al (2021) Determination of sugar content in lingwu jujube by NIR–hyperspectral imaging. *J Food Sci* 86:1201–1214. <https://doi.org/10.1111/1750-3841.15674>

- Yu H-D, Yun Y-H, Zhang W, et al (2020) Three-step hybrid strategy towards efficiently selecting variables in multivariate calibration of near-infrared spectra. *Spectrochim Acta, Part A* 224:117376. <https://doi.org/10.1016/j.saa.2019.117376>
- Zeng S, Zhang Z, Cheng X, et al (2024) Prediction of soluble solids content using near-infrared spectra and optical properties of intact apple and pulp applying PLSR and CNN. *Spectrochim Acta, Part A* 304:123402. <https://doi.org/10.1016/j.saa.2023.123402>
- Zhao F, Du G, Huang Y (2022) Exploring the use of near-infrared spectroscopy as a tool to predict quality attributes in prickly pear (*rosa roxburghii* tratt) with chemometrics variable strategy. *J Food Compos Anal* 105:104225. <https://doi.org/10.1016/j.jfca.2021.104225>
- Zhao Y, Zhou L, Wang W, et al (2024) Visible/near-infrared spectroscopy and hyperspectral imaging facilitate the rapid determination of soluble solids content in fruits. *Food Eng Rev* 16:470–496. <https://doi.org/10.1007/s12393-024-09374-6>
- Zhu H, Qin S, Liang S, et al (2025) Hyperspectral imaging and machine learning for quality assessment of apples with different bagging types. *Spectrochim Acta, Part A* 343:126443. <https://doi.org/10.1016/j.saa.2025.126443>
- Zude M (2003) Non-destructive prediction of banana fruit quality using VIS/NIR spectroscopy. *Fruits* 58:135–142. <https://doi.org/10.1051/fruits:2003001>

Table 1 Statistics of the reference values of the eight physicochemical attributes of the calibration and prediction dataset

Attribute	Dataset	N	Min	Max	Mean	SD
Fructose	Calibration	54	0.02	6.22	2.21	1.89
	Prediction	18	0.04	4.97	1.47	1.93
Glucose	Calibration	54	0.04	6.61	2.22	1.93
	Prediction	18	0.05	4.97	1.47	1.93
Sucrose	Calibration	54	0.10	14.13	5.06	4.35
	Prediction	18	0.20	10.41	3.20	3.72
Starch	Calibration	54	0.82	23.95	11.03	7.02
	Prediction	18	1.53	22.17	14.09	7.61
TSS	Calibration	54	3.25	24.30	14.80	8.42
	Prediction	18	3.00	24.20	11.14	9.10
Firmness	Calibration	54	503.8	1971.5	1175.3	448.5
	Prediction	18	552.5	1950.2	1296.3	459.7
pH	Calibration	54	4.45	5.34	4.88	0.24
	Prediction	18	4.51	5.33	5.01	0.24
TA	Calibration	54	0.29	0.77	0.46	0.12
	Prediction	18	0.26	0.70	0.40	0.12

Table 2 Predictive performance of full-spectrum PLSR models for eight physicochemical attributes using different preprocessing methods, evaluated via repeated k-fold CV on the calibration set. Repeated k-fold CV metrics are reported as means over 20 repetitions

Preprocessing	Fructose		Glucose		Sucrose		Starch		TSS		Firmness		pH		TA	
	R ² _C	RMSEC	R ² _C	RMSEC	R ² _C	RMSEC	R ² _C	RMSEC	R ² _C	RMSEC	R ² _C	RMSEC	R ² _C	RMSEC	R ² _C	RMSEC
Raw	0.82	0.79	0.79	0.87	0.85	1.69	0.80	3.09	0.98	1.05	0.74	226.8	0.48	0.17	0.08	0.12
SG0	0.82	0.79	0.79	0.87	0.84	1.70	0.80	3.11	0.91	2.55	0.74	226.9	0.48	0.17	0.08	0.12
SG1	0.80	0.83	0.78	0.90	0.90	1.36	0.73	3.58	0.98	1.05	0.73	230.8	0.70	0.13	0.17	0.11
SNV	0.81	0.82	0.78	0.90	0.85	1.69	0.84	2.78	0.99	0.98	0.70	242.7	0.65	0.14	0.24	0.11
MSC	0.81	0.82	0.78	0.90	0.85	1.69	0.84	2.80	0.93	2.20	0.70	242.7	0.64	0.14	0.24	0.11
SNV+SG1	0.79	0.86	0.76	0.94	0.91	1.31	0.68	3.91	0.98	1.10	0.71	240.3	0.59	0.15	0.42	0.09
MSC+SG1	0.79	0.86	0.76	0.94	0.91	1.32	0.69	3.90	0.98	1.10	0.71	240.0	0.59	0.15	0.43	0.09

Units of RMSEC: g/100 g (starch, sucrose, glucose, and fructose), °Brix (TSS), % (TA) and gf (firmness).

Table 3 Predictive performance of full-spectrum PLSR models for eight physicochemical attributes using different preprocessing methods, evaluated by LODO-CV within the calibration set. LODO-CV metrics were calculated from pooled out-of-fold predictions by holding out each of the 12 storage days sequentially

Preprocessing	Fructose		Glucose		Sucrose		Starch		TSS		Firmness		pH		TA	
	R ² _{LO}	RMSEL	R ² _{LO}	RMSEL	R ² _{LO}	RMSEL	R ² _{LO}	RMSEL	R ² _{LO}	RMSEL	R ² _{LO}	RMSEL	R ² _{LOD}	RMSEL	R ² _{LOD}	RMSEL
	DO	ODO	DO	ODO	DO	ODO	DO	ODO	DO	ODO	DO	ODO	o	ODO	o	ODO
Raw	0.78	0.88	0.73	0.99	0.77	2.06	0.59	4.44	0.85	3.24	0.62	272.1	0.12	0.22	-0.13	0.13
SG0	0.78	0.88	0.73	0.99	0.77	2.06	0.60	4.42	0.83	3.41	0.62	272.2	0.12	0.22	-0.13	0.13
SG1	0.67	1.08	0.65	1.13	0.72	2.28	0.62	4.28	0.86	3.15	0.54	301.6	0.16	0.22	-0.25	0.14
SNV	0.78	0.89	0.74	0.97	0.78	2.02	0.65	4.12	0.87	3.04	0.60	282.6	0.35	0.19	-0.09	0.13
MSC	0.78	0.89	0.74	0.97	0.78	2.02	0.64	4.19	0.86	3.08	0.59	282.8	0.33	0.19	-0.09	0.13
SNV+SG1	0.75	0.93	0.72	1.00	0.76	2.13	0.59	4.46	0.85	3.22	0.59	286.0	0.12	0.22	-0.21	0.13
MSC+SG1	0.75	0.93	0.72	1.00	0.75	2.13	0.59	4.44	0.85	3.18	0.59	284.2	0.13	0.22	-0.21	0.13

Units of RMSELODO: g/100 g (starch, sucrose, glucose, and fructose), °Brix (TSS), % (TA) and gf (firmness).

Table 4 Predictive performance of full-spectrum PLSR models for eight physicochemical attributes using different preprocessing methods on the prediction set

Preprocessing	Fructose		Glucose		Sucrose		Starch		TSS		Firmness		pH		TA	
	R _P	RMSEP	R _P	RMSEP	R _P	RMSEP	R _P	RMSEP	R _P	RMSEP	R _P	RMSEP	R _P	RMSEP	R _P	RMSEP
Raw	0.88	0.65	0.92	0.54	0.70	1.97	0.85	2.90	0.90	2.75	0.88	156.0	0.29	0.20	-0.12	0.12
SG0	0.88	0.65	0.92	0.54	0.70	1.97	0.85	2.89	0.91	2.65	0.88	156.0	0.29	0.20	-0.12	0.12
SG1	0.79	0.86	0.82	0.79	0.74	1.85	0.80	3.27	0.92	2.57	0.76	217.6	0.31	0.20	0.19	0.11
SNV	0.83	0.77	0.87	0.69	0.72	1.90	0.80	3.27	0.90	2.74	0.71	239.5	0.33	0.19	0.23	0.10
MSC	0.83	0.77	0.87	0.69	0.72	1.90	0.79	3.43	0.89	2.96	0.71	239.5	0.29	0.20	0.23	0.10
SNV+SG1	0.77	0.89	0.80	0.84	0.75	1.82	0.75	3.71	0.91	2.72	0.69	249.6	0.43	0.18	0.29	0.10
MSC+SG1	0.77	0.89	0.80	0.84	0.75	1.81	0.75	3.70	0.91	2.69	0.69	249.5	0.43	0.18	0.30	0.10

Units of RMSEP: g/100 g (starch, sucrose, glucose, and fructose), °Brix (TSS), % (TA) and gf (firmness).

Table 5 Performance metrics on the prediction set for full-spectrum PLSR models for eight physicochemical attributes using the optimal preprocessing method

Attribute	Preprocessing	MAE	Bias	RPD	RPIQ
Fructose	SG0	0.57	0.05	2.99	4.11
Glucose	SG0	0.47	0.02	3.57	4.96
Sucrose	MSC+SG1	1.46	0.05	2.05	3.40
Starch	SG0	2.35	0.23	2.63	4.69
TSS	SG1	2.22	0.81	3.54	6.91
Firmness	Raw	118.0	-36.4	2.95	4.10
pH	MSC+SG1	0.13	0.04	1.36	1.78
TA	MSC+SG1	0.08	-0.03	1.23	1.30

Units of MAE and Bias: g/100 g (starch, sucrose, glucose, and fructose), °Brix (TSS), % (TA) and gf (firmness).

Table 6 Tentative spectral assignments of the informative wavelength variables

Wavelength range (nm)	Chemical assignment	Potential spectral contributors	Reference
1000–1040	Region associated primarily with higher-order O–H overtones, with possible weak contributions from sugar-related C–H vibrations	Water and O–H-containing environments in starch and soluble sugars	(Golic et al. 2003; Theanjumpol et al. 2013; Walsh et al. 2020)
1100–1250	Region associated mainly with the second overtone of C–H stretching vibrations	Sucrose and other soluble carbohydrates	(Yang et al. 2021)
1290–1425	O–H overtone region potentially reflecting sugar–water interactions and variations in hydrogen bonding	Soluble sugars, TSS, and sample water status	(Zhao et al. 2024)
1440–1550	Strong absorption region commonly assigned to the first overtone of O–H stretching	Water and sugar-related changes in the hydrogen-bonding environment	(Walsh et al. 2020)
1640–1740	Region commonly assigned to the first overtone of C–H stretching vibrations	Glucose, fructose, sucrose, and other carbohydrate structures	(Goodarzi and Saeys 2016; Xu et al. 2025)
1780–1799	Long-wavelength end of the first C–H overtone region, approaching the onset of more complex overlapping absorptions	Starch, bound water, and structural carbohydrates	(Subedi and Walsh 2011; Heise et al. 2021)

The assignments are tentative interpretations of the combined effects of the analytes and sample matrix because NIR absorption bands are broad and highly overlapping.

Table 7 Prediction performance on the calibration set of informative-wavelength PLSR models for six physicochemical attributes using their respective optimal preprocessing methods. Repeated k-fold CV metrics are reported as means over 20 repetitions. LODO-CV metrics were calculated from pooled out-of-fold predictions by holding out each of the 12 storage days sequentially

Attribute	Preprocessing	Repeated k-fold CV		LODO-CV	
		R^2_C	RMSEC	R^2_{LODO}	RMSELODO
Fructose	SG0	0.80	0.83	0.73	0.96
Glucose	SG0	0.78	0.90	0.73	1.00
Sucrose	MSC+SG1	0.81	1.87	0.71	2.31
Starch	SG0	0.82	2.97	0.62	4.27
TSS	SG1	0.91	2.48	0.83	3.40
Firmness	Raw	0.73	230.9	0.61	279.2

Units of RMSEC, RMSELODO: g/100 g (starch, sucrose, glucose, and fructose), °Brix (TSS), and gf (firmness).

Table 8 Predictive performance on the prediction set of informative-wavelength PLSR models for six physicochemical attributes using their respective optimal preprocessing methods

Attribute	Preprocessing	Wavelength number	R^2_P	RMSEP	MAE	Bias	RPD	RPIQ
Fructose	SG0	305	0.86	0.71	0.63	0.03	2.71	3.73
Glucose	SG0	317	0.90	0.60	0.51	0.01	3.23	4.50
Sucrose	MSC+SG1	158	0.72	1.92	1.58	0.12	1.93	3.20
Starch	SG0	265	0.78	3.48	2.81	-0.29	2.19	3.90
TSS	SG1	157	0.85	3.45	2.98	0.42	2.64	5.16
Firmness	Raw	445	0.84	176.6	140.0	-34.4	2.60	3.62

Units of RMSEP, MAE and Bias: g/100 g (starch, sucrose, glucose, and fructose), °Brix (TSS), % (TA) and gf (firmness).

Figure Captions

Graphical abstract

Fig. 1 Schematic diagram of the NIR detection system

Fig. 2 Original NIR absorbance spectra of bananas and spectra after different preprocessing. (a) Mean NIR absorbance spectra of banana samples over a 12-day storage period ($n = 6$); (b) Raw absorbance spectra without preprocessing; (c) Absorbance spectra after SG0 preprocessing; (d) Absorbance spectra after SG1 preprocessing; (e) Absorbance spectra after SNV preprocessing; (f) Absorbance spectra after MSC preprocessing; (g) Absorbance spectra after SNV+SG1 preprocessing; (h) Absorbance spectra after MSC+SG1 preprocessing

Fig. 3 Changes in (a) fructose, (b) glucose, (c) sucrose, (d) starch, (e) TSS, (f) firmness, (g) pH, and (h) TA during banana ripening ($n = 6$)

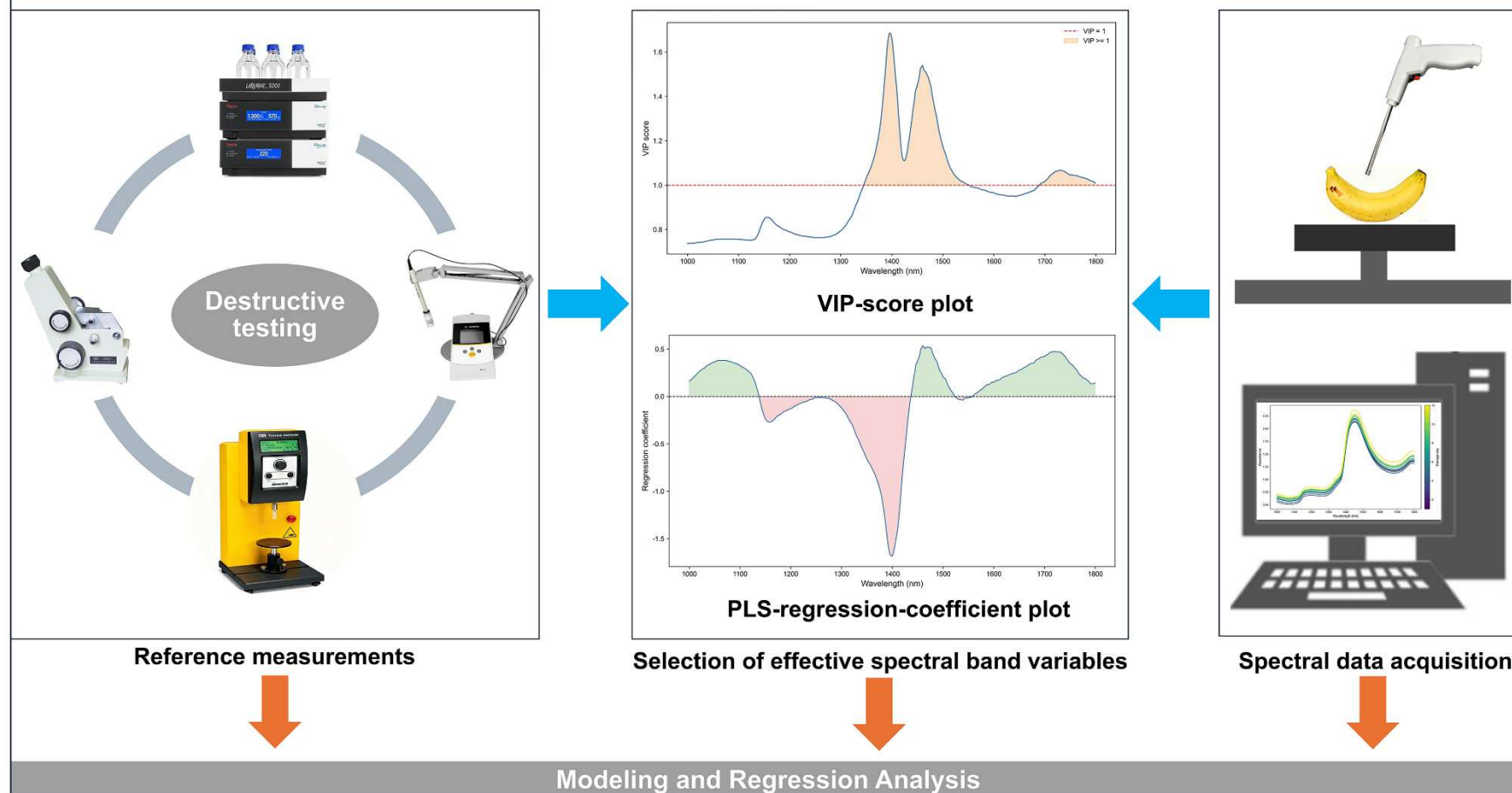
Fig. 4 Variation in RMSEC and R^2_c with the number of LVs, exemplified using the optimal preprocessing (raw) for firmness

Fig. 5 Scatter plots of reference (x) versus predicted (y) values for eight physicochemical attributes obtained from the full-spectrum PLSR models with optimal prediction performance. (a) Fructose; (b) Glucose; (c) Sucrose; (d) Starch; (e) TSS; (f) Firmness; (g) pH; (h) TA. Blue symbols denote calibration-set samples; orange symbols denote prediction-set samples

Fig. 6 VIP scores and PLS regression coefficients for wavelength selection and interpretation. The left panels show VIP scores, and the right panels show PLS regression coefficients. (a) Fructose; (b) Glucose; (c) Sucrose; (d) Starch; (e) TSS; (f) Firmness. The dashed line in the PLS-regression-coefficient panels denotes a coefficient value of 0

Fig. 7 Scatter plots of reference (x) versus predicted (y) values for eight physicochemical attributes obtained from the informative-wavelength PLSR models with optimal prediction performance. (a) Fructose; (b) Glucose; (c) Sucrose; (d) Starch; (e) TSS; (f) Firmness. Blue symbols denote calibration-set samples; orange symbols denote prediction-set samples

Non-Destructive Prediction of Banana Quality Under Controlled Storage Conditions Using Near-Infrared Spectroscopy



Graphical abstract

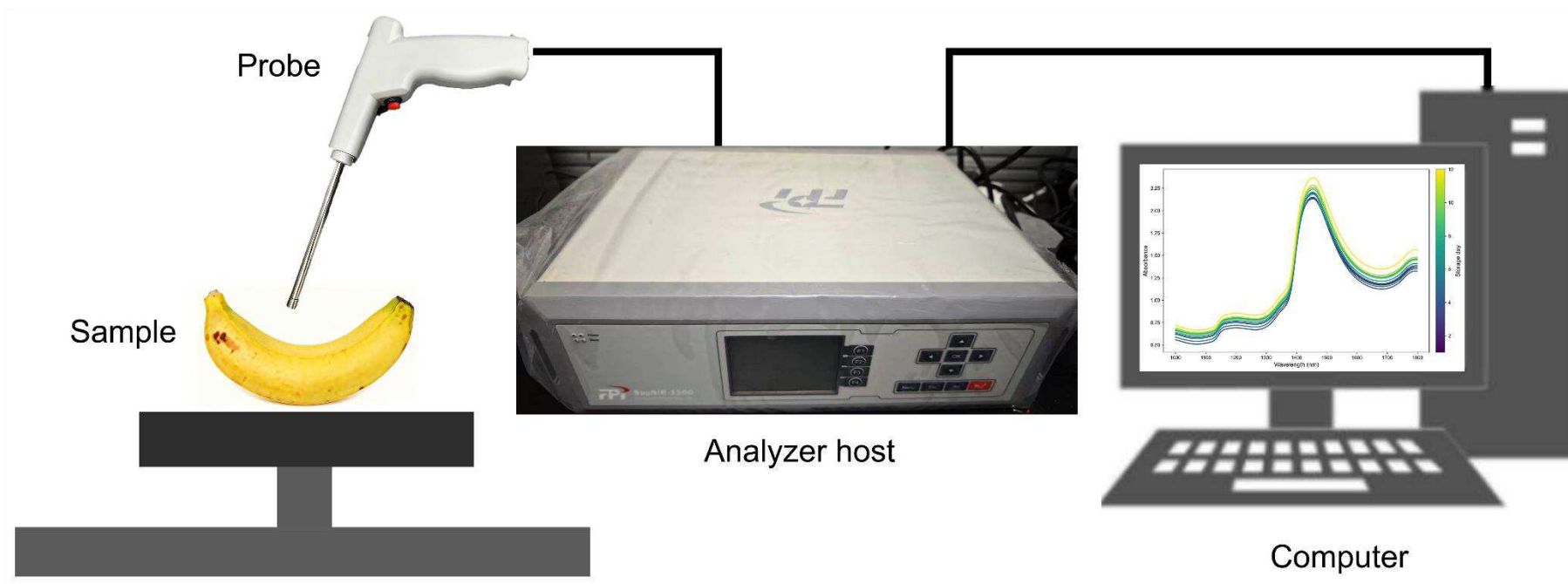
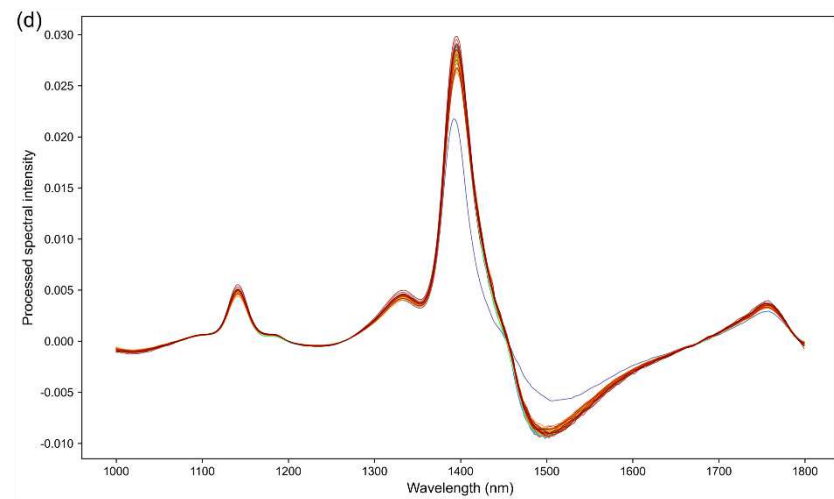
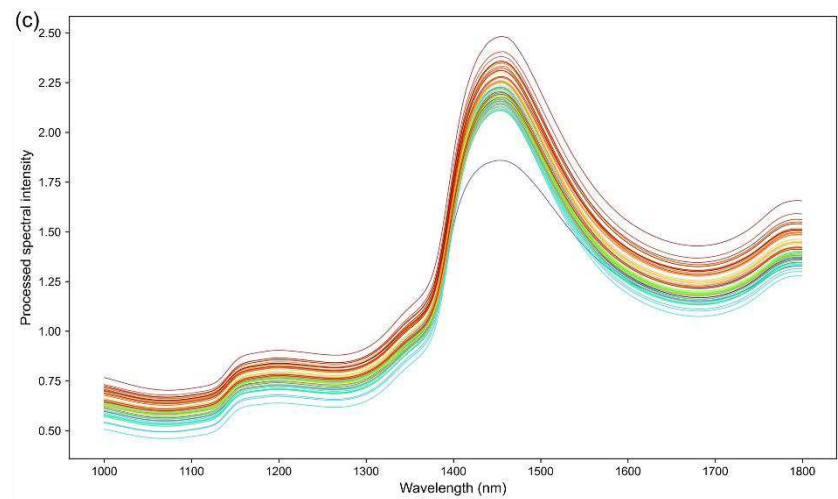
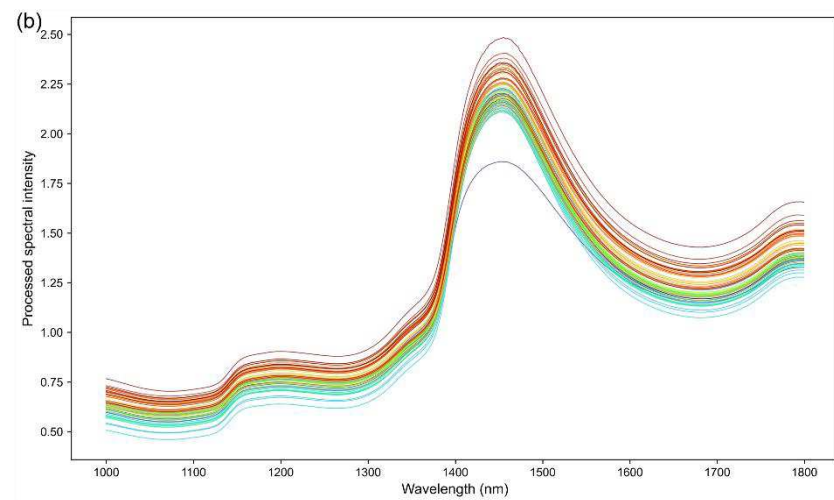
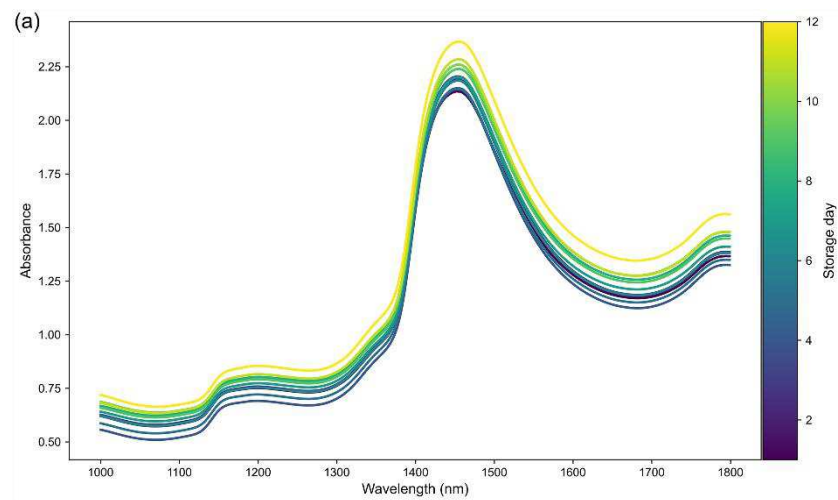


Fig. 1 Schematic diagram of the NIR detection system



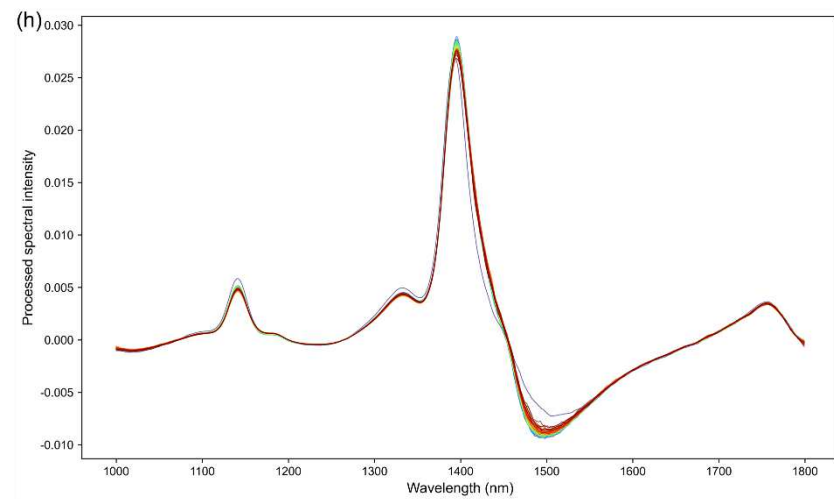
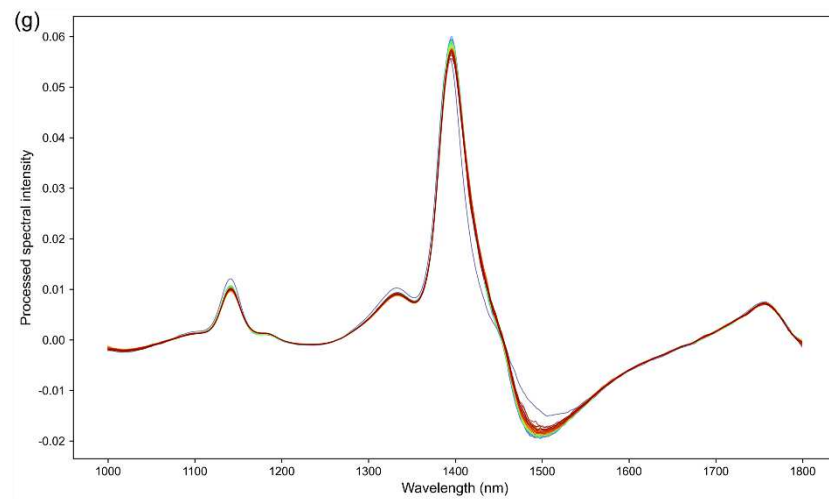
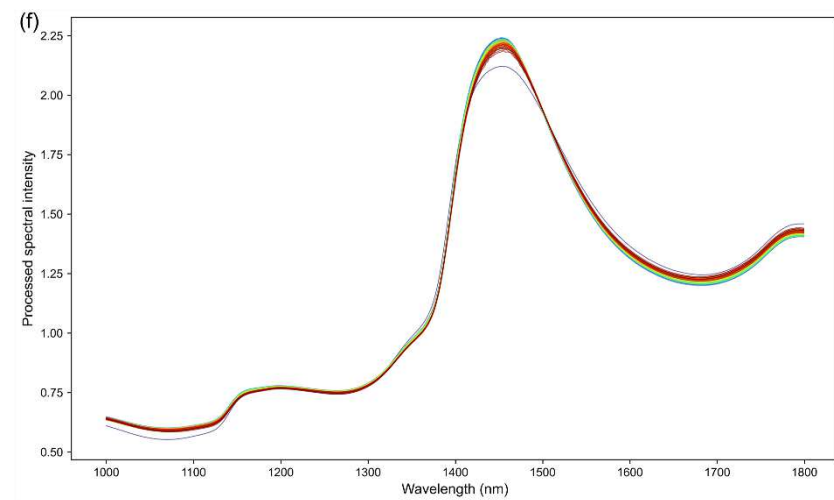
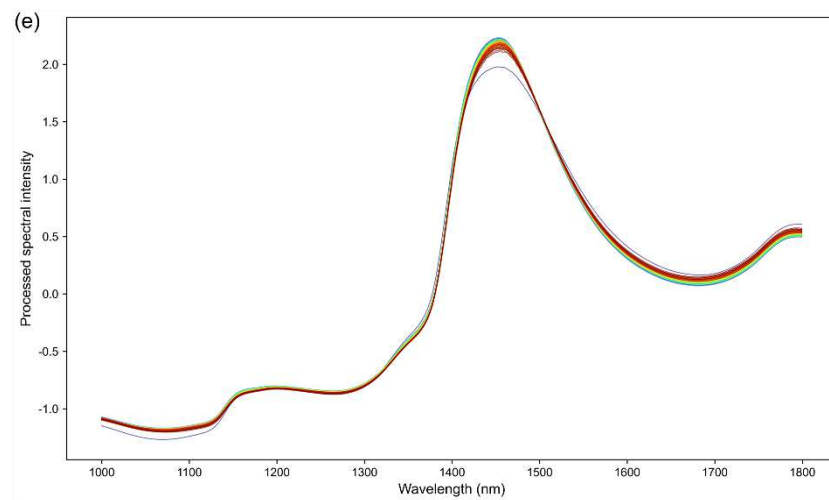


Fig. 2 Original NIR absorbance spectra of bananas and spectra after different preprocessing. (a) Mean NIR absorbance spectra of banana samples over a 12-day storage period ($n = 6$); (b) Raw absorbance spectra without preprocessing; (c) Absorbance spectra after SG0 preprocessing; (d) Absorbance spectra after SG1 preprocessing; (e) Absorbance spectra after SNV preprocessing; (f) Absorbance spectra after MSC preprocessing; (g) Absorbance spectra after SNV+SG1 preprocessing; (h) Absorbance spectra after MSC+SG1 preprocessing

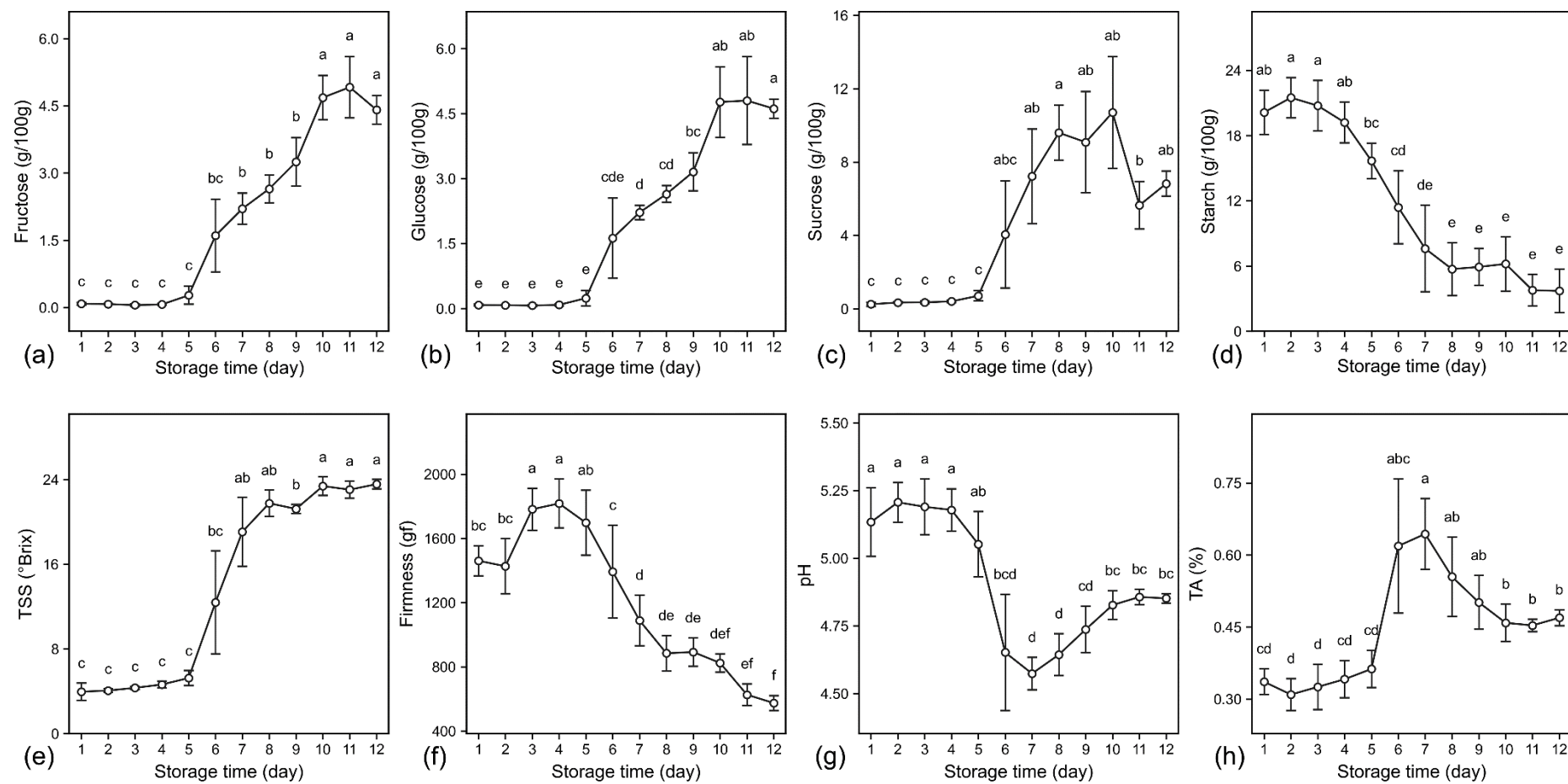


Fig. 3 Changes in (a) fructose, (b) glucose, (c) sucrose, (d) starch, (e) TSS, (f) firmness, (g) pH, and (h) TA during banana ripening (n = 6)

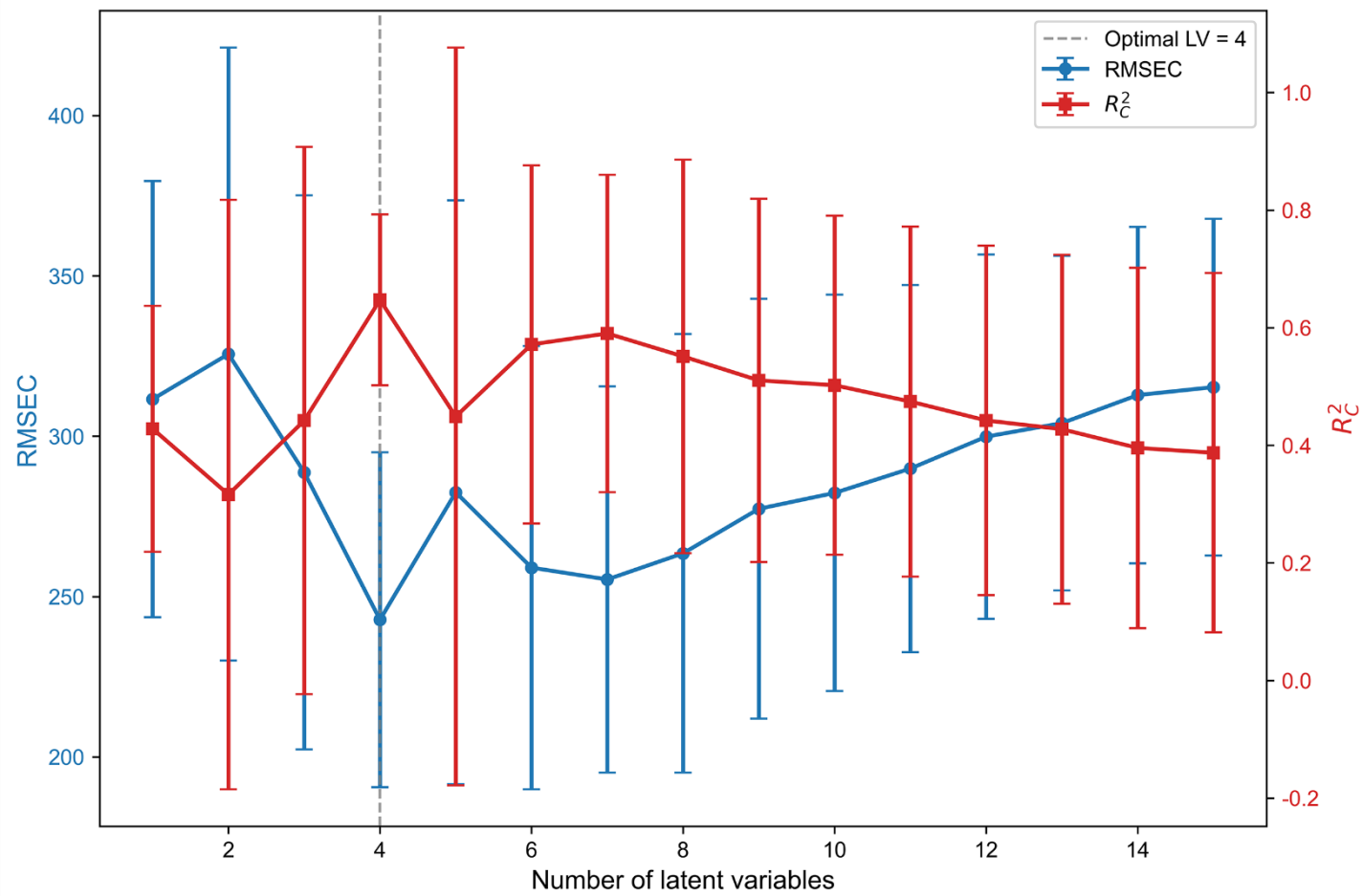
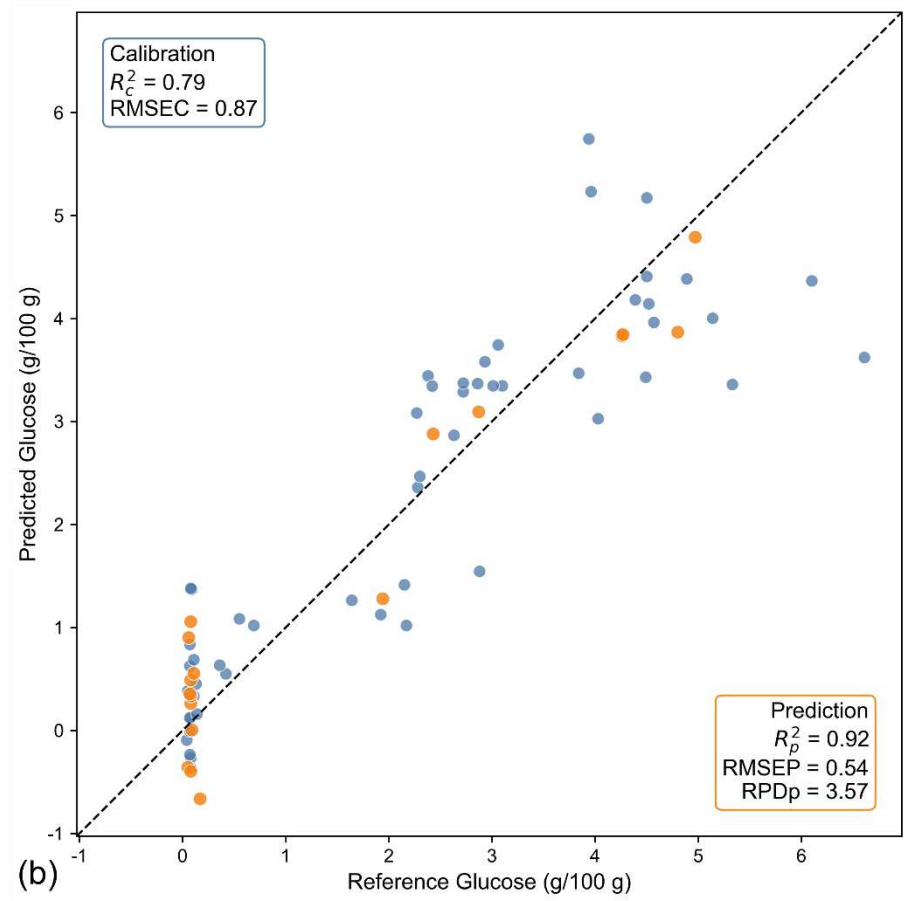
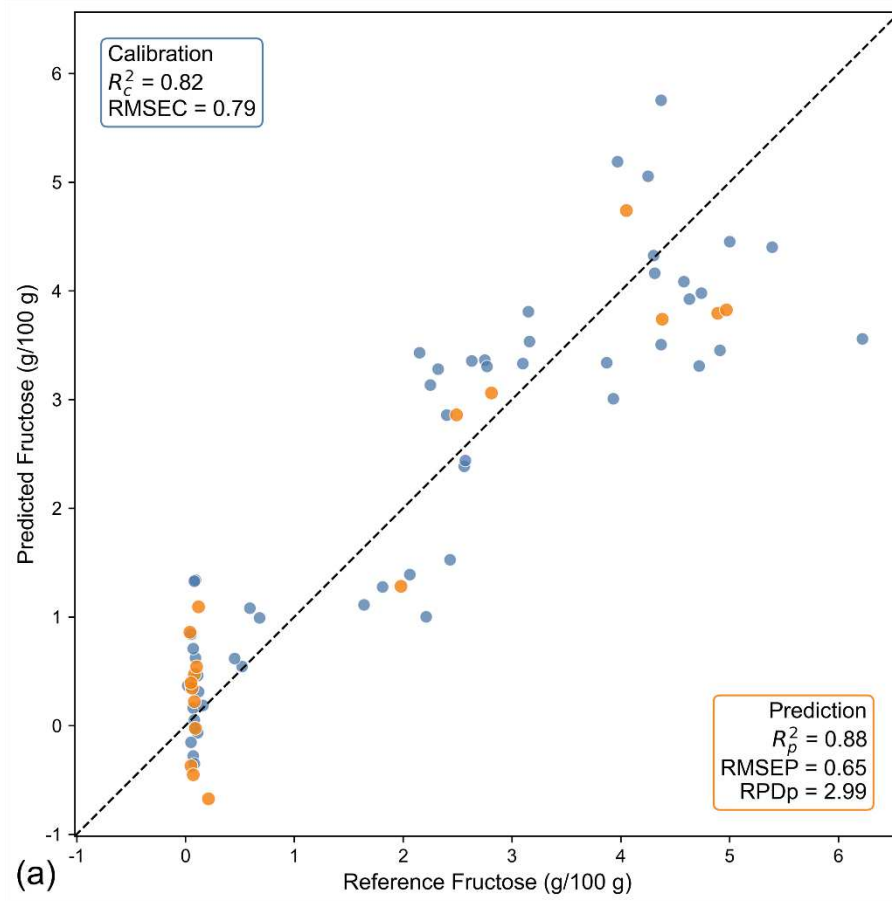
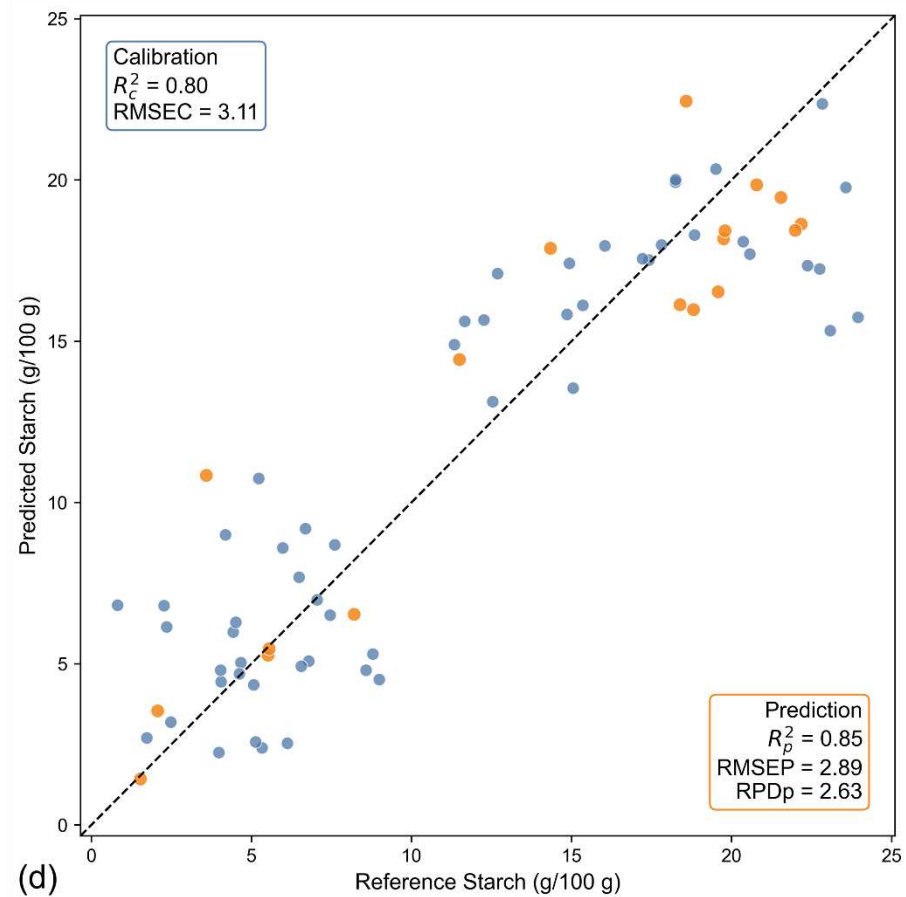
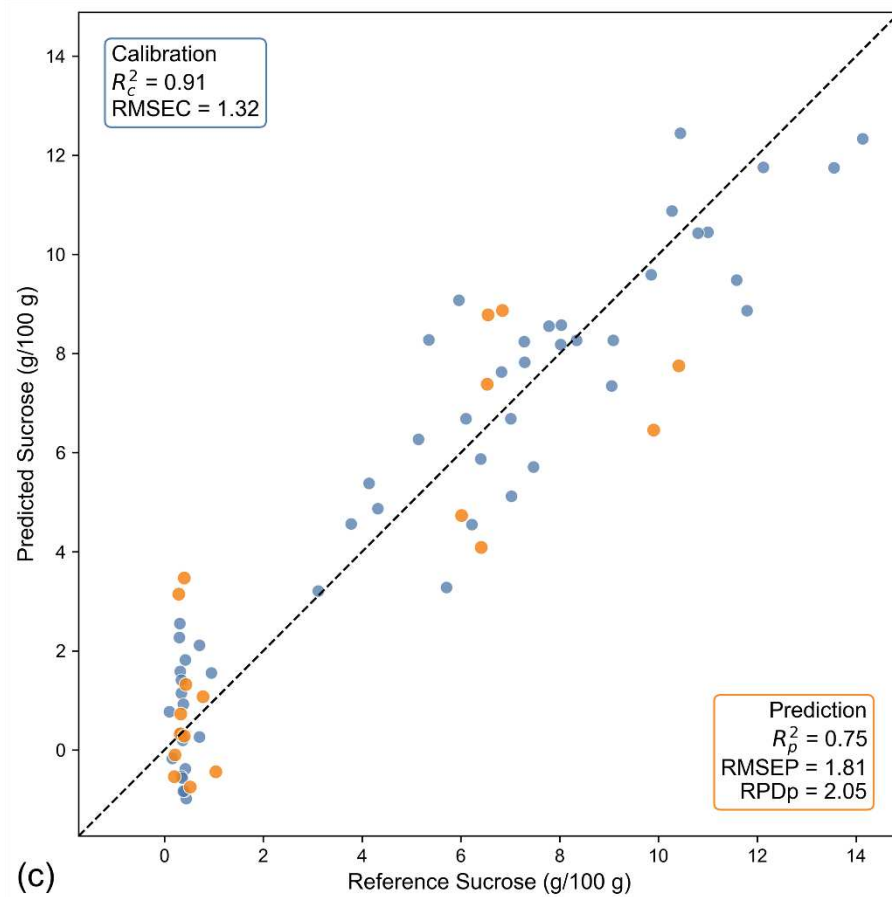
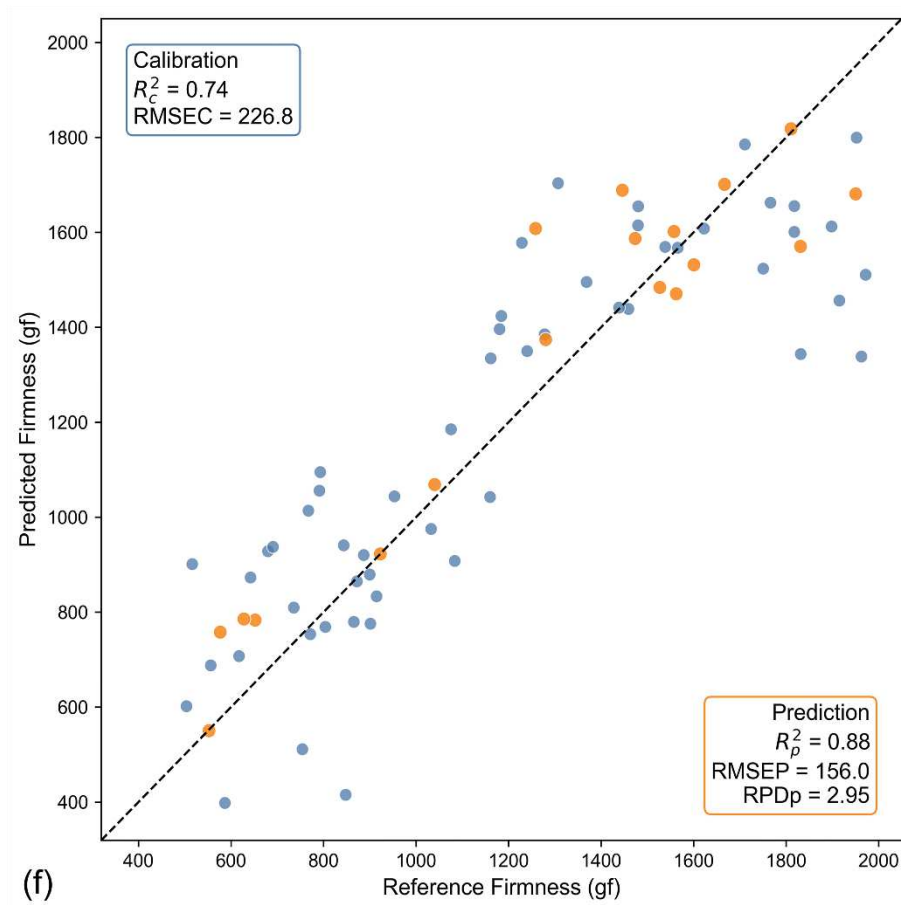
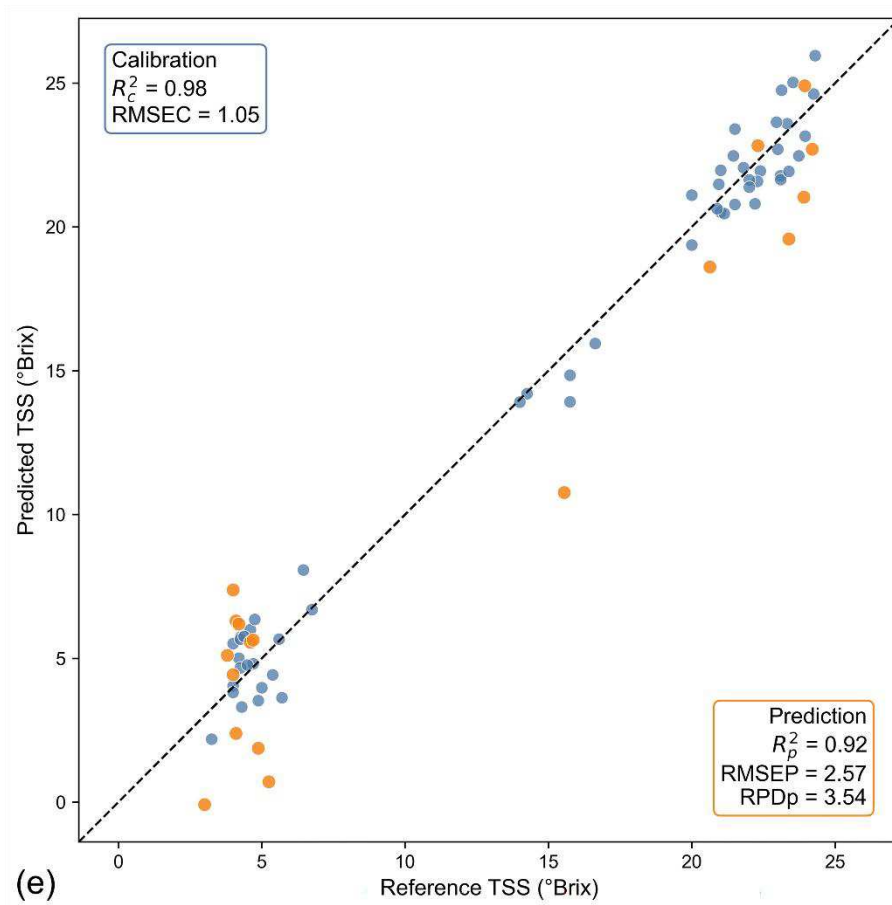


Fig. 4 Variation in RMSEC and R^2_C with the number of LVs, exemplified using the optimal preprocessing (raw) for firmness (gf)







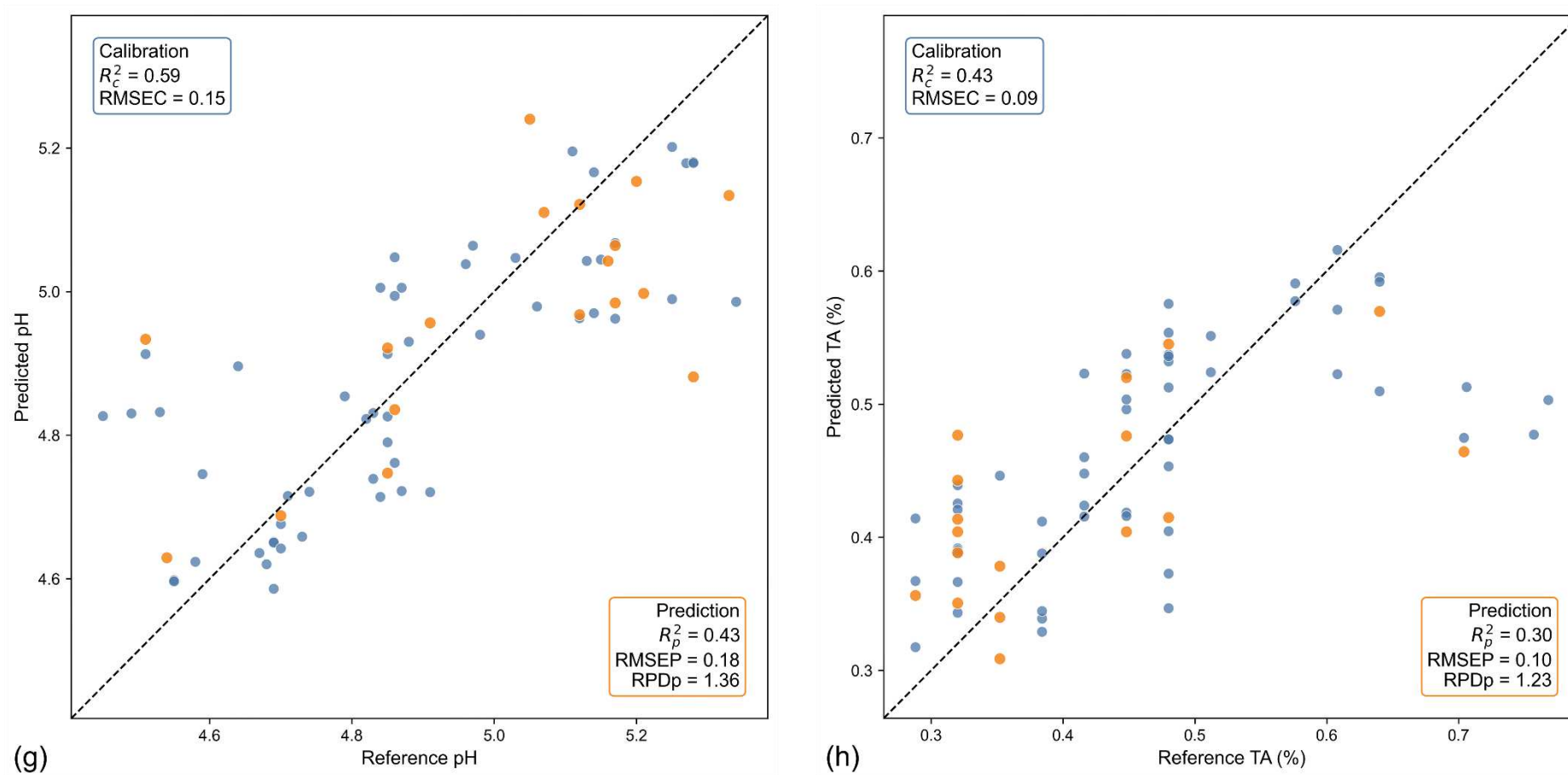
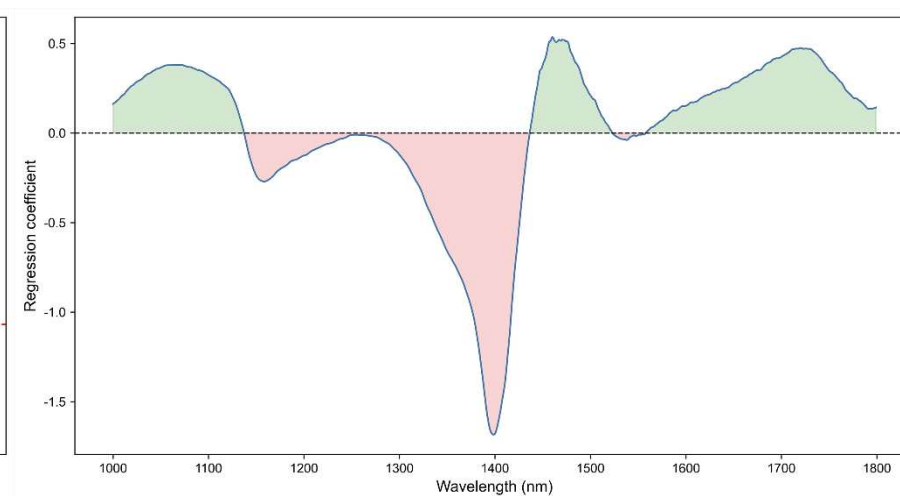
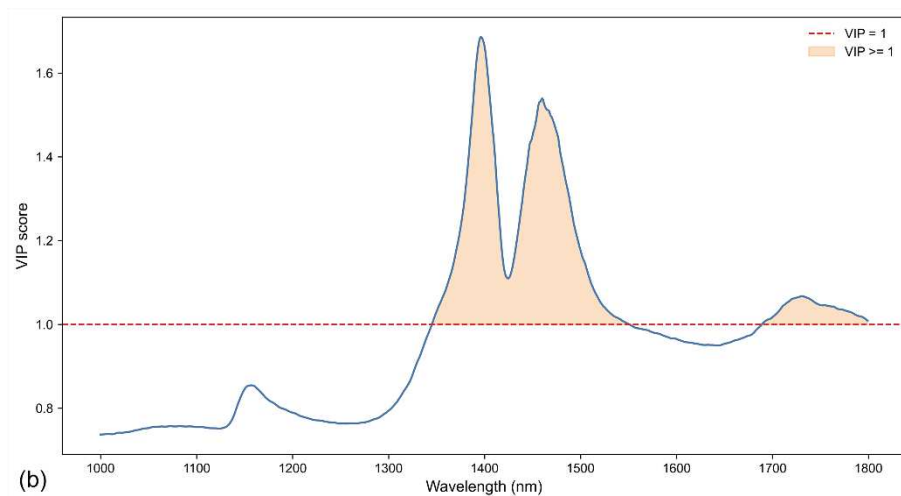
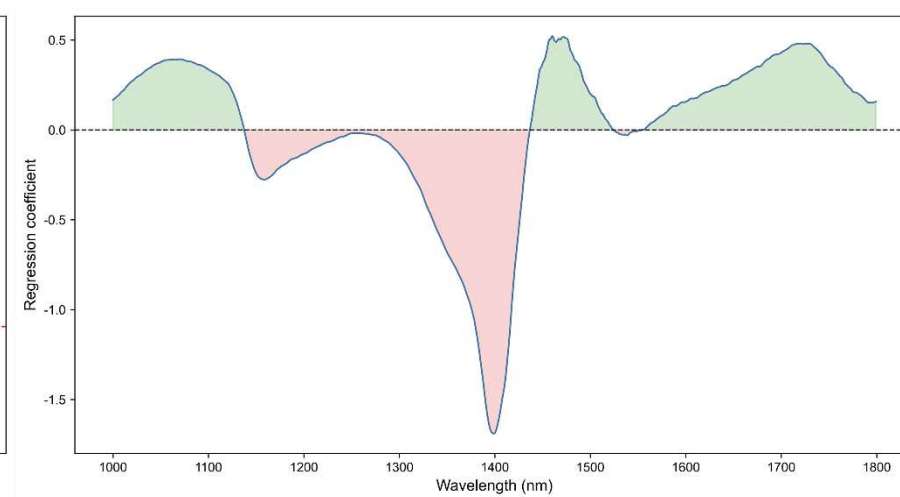
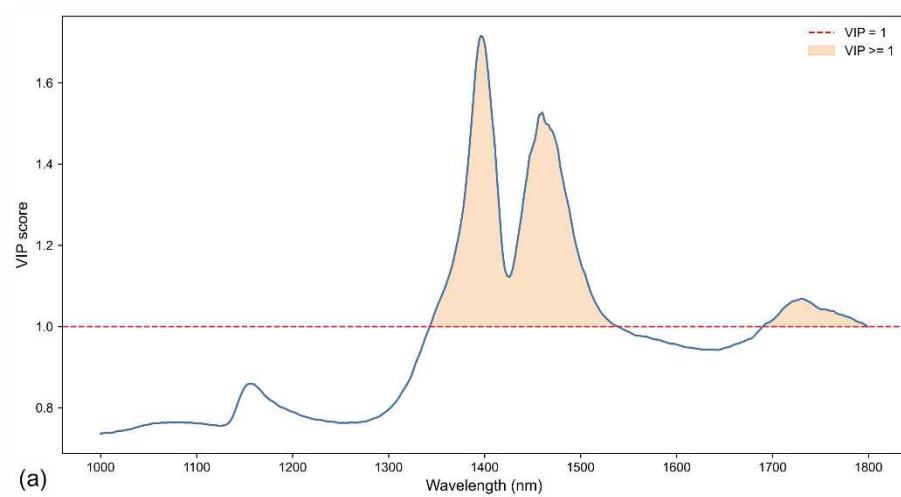
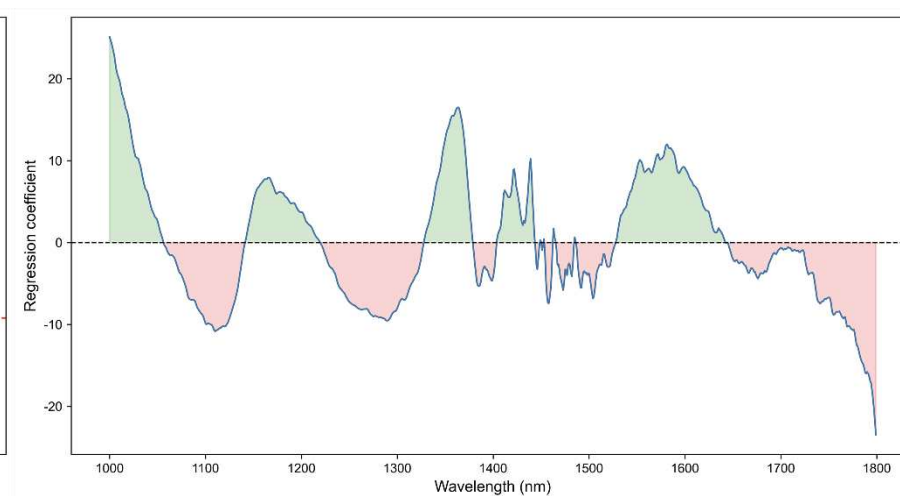
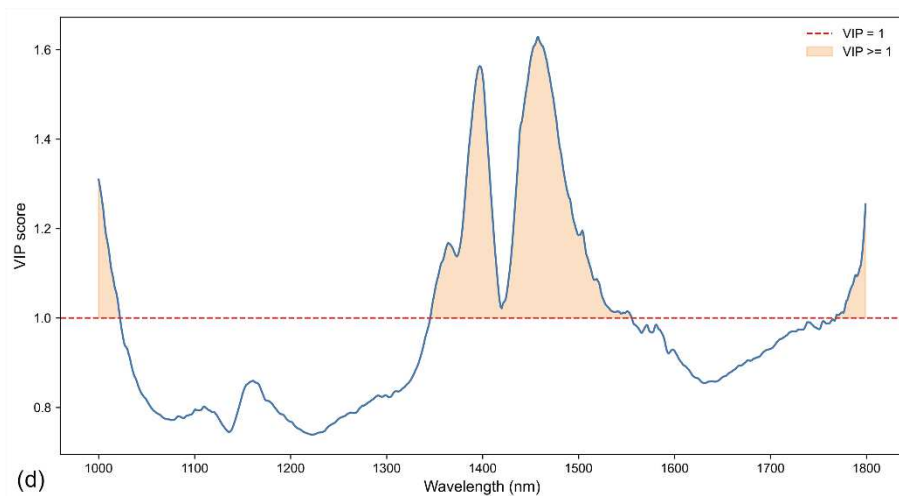
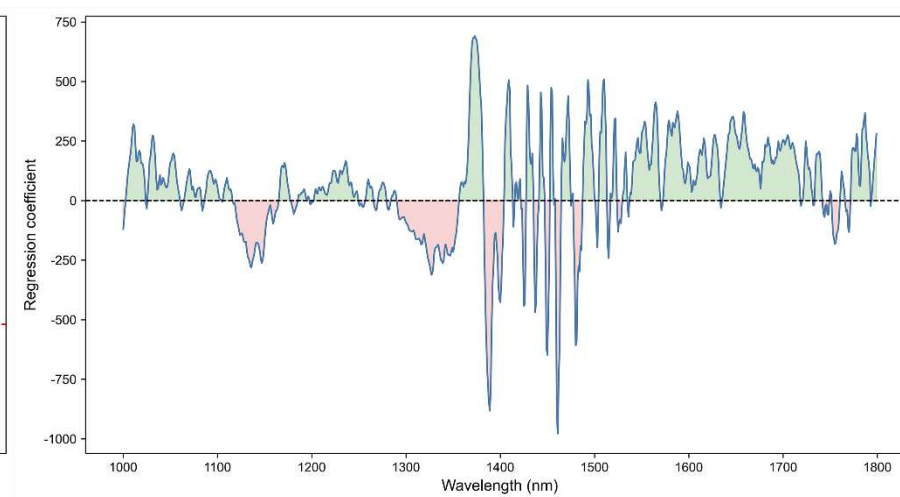
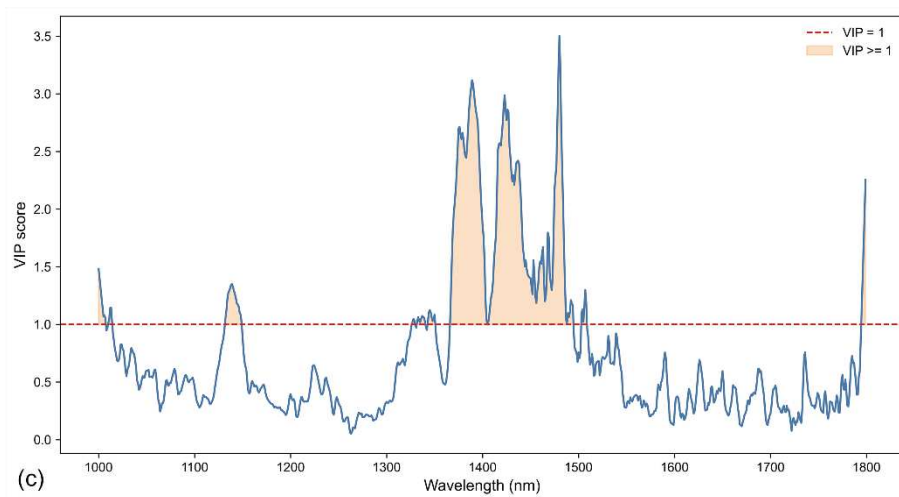


Fig. 5 Scatter plots of reference (x) versus predicted (y) values for eight physicochemical attributes obtained from the full-spectrum PLSR models with optimal prediction performance. (a) Fructose; (b) Glucose; (c) Sucrose; (d) Starch; (e) TSS; (f) Firmness; (g) pH; (h) TA. Blue symbols denote calibration-set samples; orange symbols denote prediction-set samples





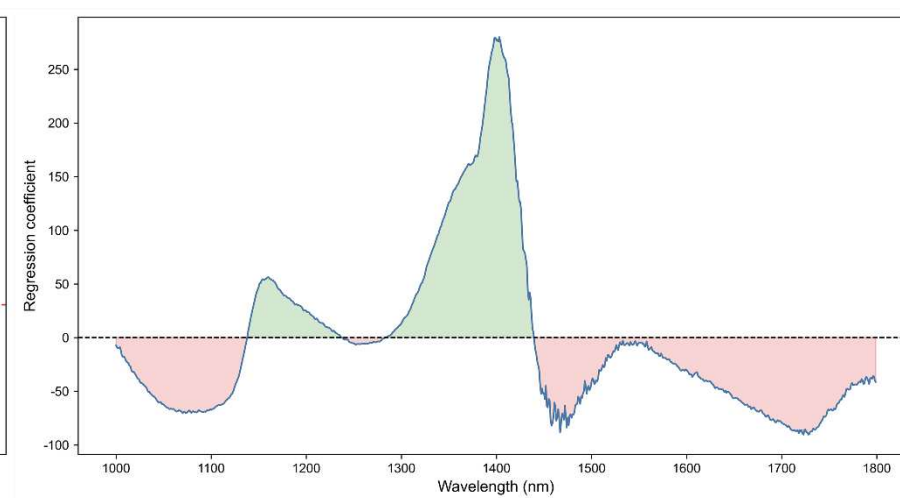
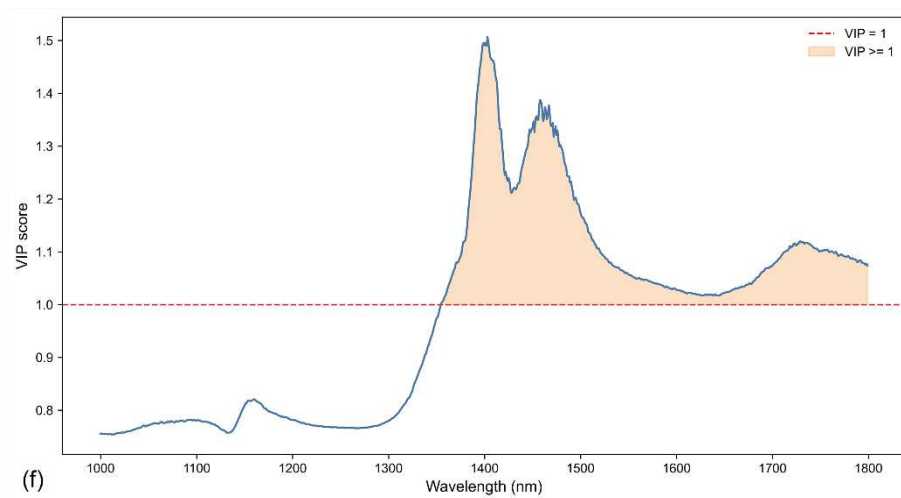
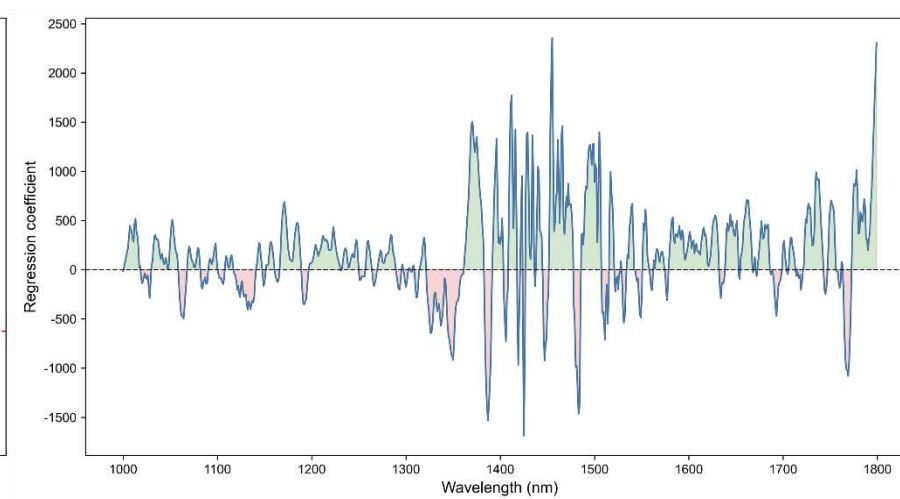
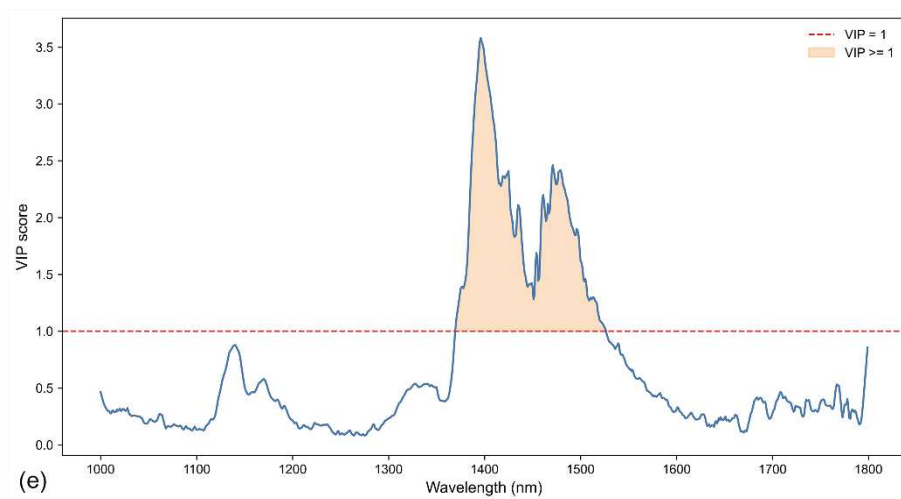
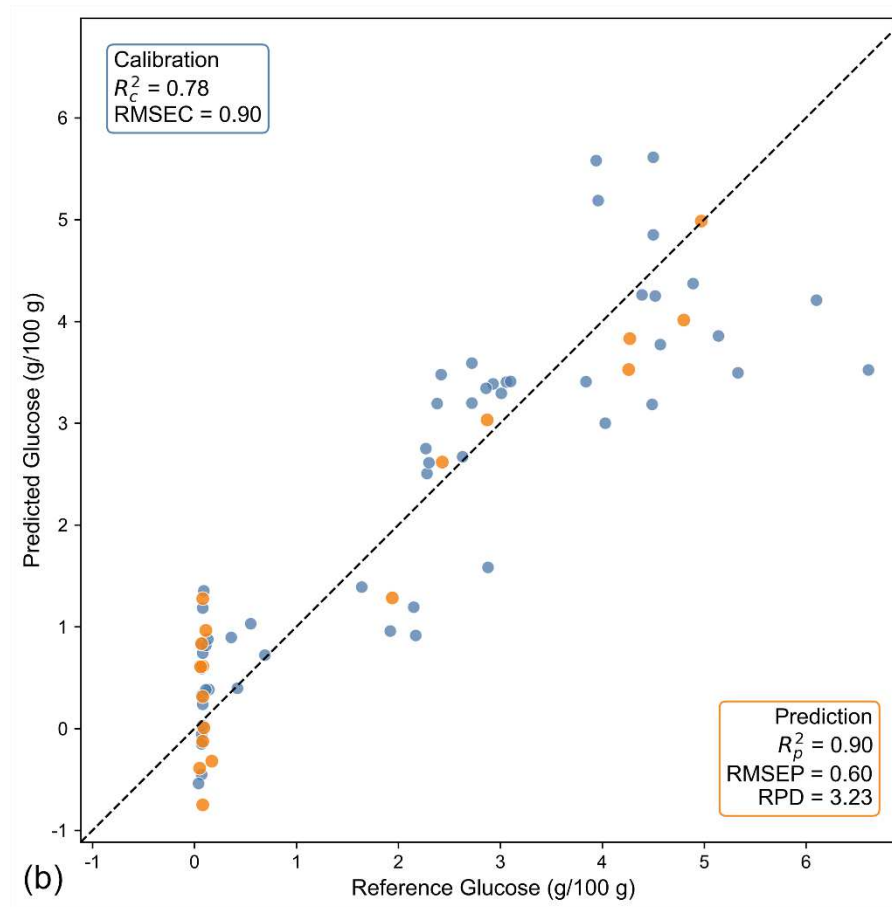
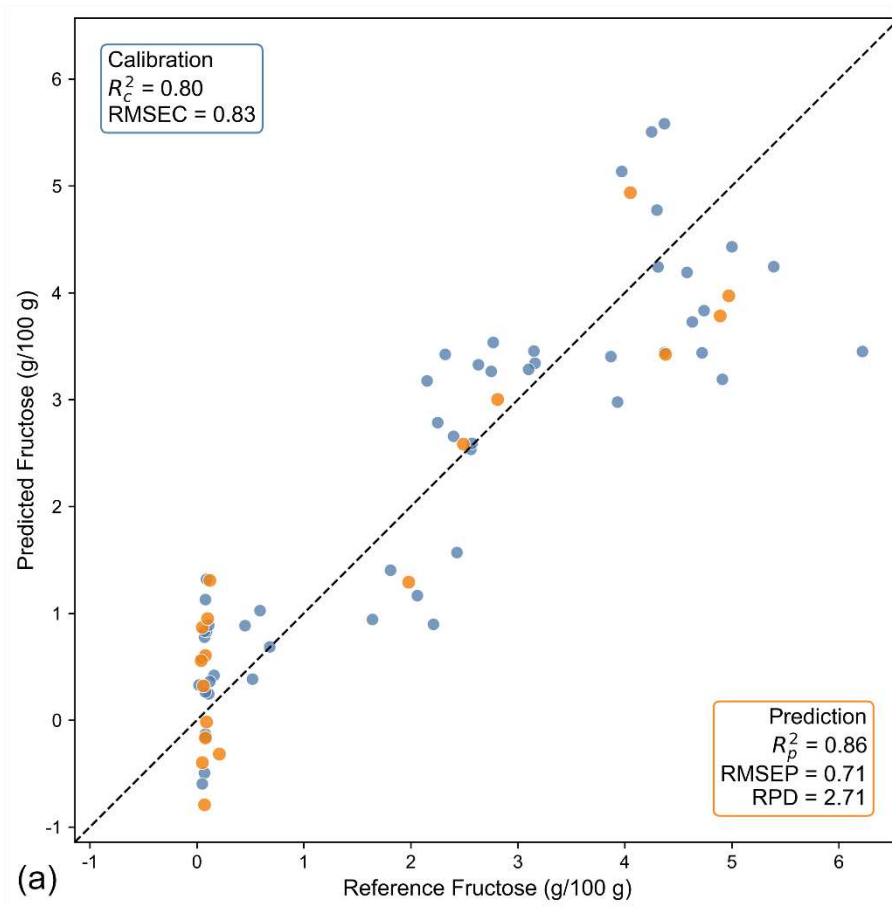
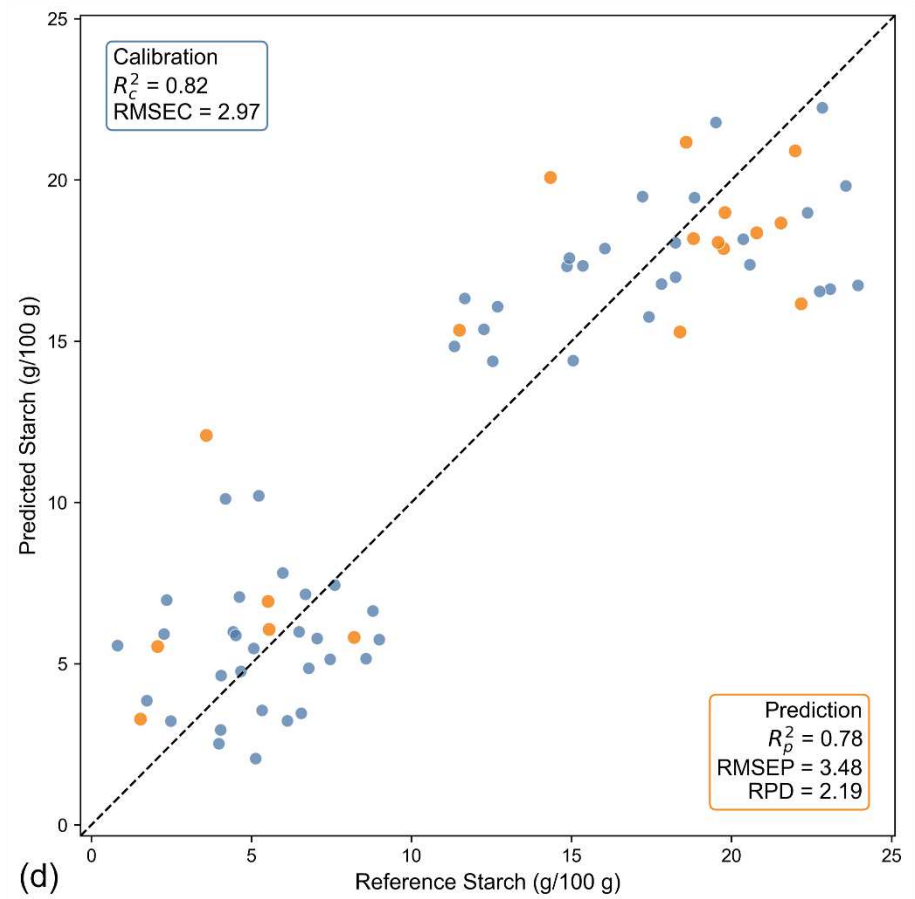
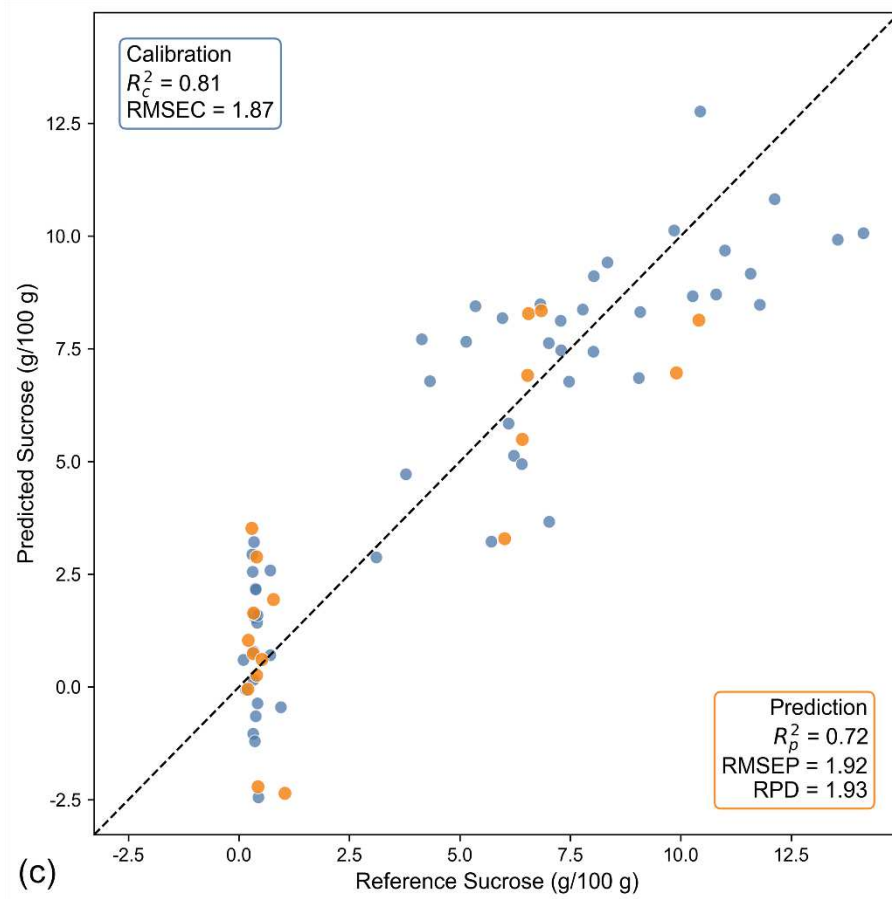


Fig. 6 VIP scores and PLS regression coefficients for wavelength selection and interpretation. The left panels show VIP scores, and the right panels show PLS regression coefficients. (a) Fructose; (b) Glucose; (c) Sucrose; (d) Starch; (e) TSS; (f) Firmness. The dashed line in the PLS-regression-coefficient panels denotes a coefficient value of 0





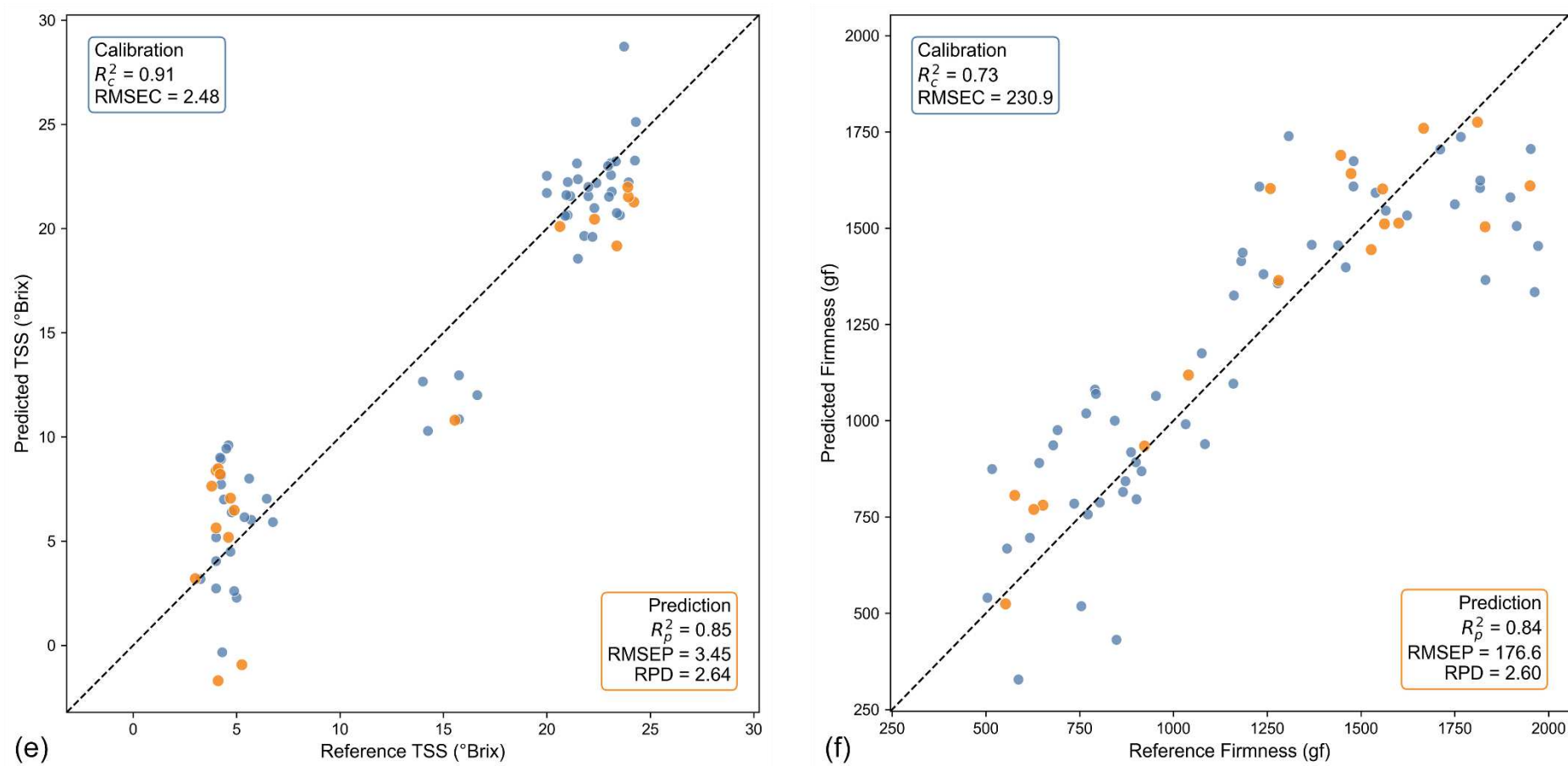


Fig. 7 Scatter plots of reference (x) versus predicted (y) values for eight physicochemical attributes obtained from the informative-wavelength PLSR models with optimal prediction

performance. (a) Fructose; (b) Glucose; (c) Sucrose; (d) Starch; (e) TSS; (f) Firmness. Blue symbols denote calibration-set samples; orange symbols denote prediction-set samples