



## OPEN ACCESS

## EDITED BY

Xiaohui Lin,  
University College Dublin, Ireland

## REVIEWED BY

Qiaoli Ma,  
Jiangxi Agricultural University, China  
Wilson Castro,  
National University Frontera, Peru  
Rui Guerra,  
University of Algarve, Portugal

## \*CORRESPONDENCE

Xudong Sun  
2609@ecjtu.edu.cn;  
874916937@qq.com;  
sunxudong\_18@163.com

## SPECIALTY SECTION

This article was submitted to  
Nutrition and Food Science  
Technology,  
a section of the journal  
Frontiers in Nutrition

RECEIVED 23 June 2022

ACCEPTED 26 July 2022

PUBLISHED 24 August 2022

## CITATION

Sun X, Deng D, Liu J and 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.  
doi: 10.3389/fnut.2022.976178

## COPYRIGHT

© 2022 Sun, Deng, Liu and Feng. This  
is an open-access article distributed  
under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#).  
The use, distribution or reproduction  
in other forums is permitted, provided  
the original author(s) and the copyright  
owner(s) are credited and that the  
original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution  
or reproduction is permitted which  
does not comply with these terms.

# Model development and update of portable NIRS instrument for assessment of internal quality attributes of two navel orange varieties

Xudong Sun<sup>1,2\*</sup>, Di Deng<sup>1</sup>, Jiacheng Liu<sup>3</sup> and Shaoran Feng<sup>4</sup>

<sup>1</sup>School of Mechatronics and Vehicle Engineering, East China Jiaotong University, Nanchang, China, <sup>2</sup>Ministry of Education, Key Laboratory of Conveyance Equipment (East China Jiaotong University), Nanchang, China, <sup>3</sup>Cultivation Laboratory, Ganzhou Citrus Research Institute, Ganzhou, China, <sup>4</sup>Beijing Sunlight Yishida Technology Co., Ltd., Beijing, China

The variation of fruit among batches influences the performance of the portable near infrared spectroscopy (NIRS) instrument and then determines the success or failure for practical application in fruit industry. Model development and update methods were investigated for determining soluble solids contents (SSC) and titrable acidity (TA) of navel orange. The pretreatment and variable selection methods were explored for building partial least square regression (PLSR) models. The best models, developed by the combination of second derivative (2D) and variable sorting for normalization (VSN), could predict SSC but not TA. The root mean square error of prediction (RMSEP), coefficient of determination for prediction ( $R_p^2$ ) and ratio of prediction to deviation (RPD) for SSC were 0.66 °Brix, 0.66 and 1.73. Model maintain methods of model update (MU) and slope and bias correction (SBC) achieved the best results in predicting SSC for two external validation sets with  $R_p^2$ , RMSEP and RPD of 0.54, 0.83 °Brix, 1.60 and 0.52, 0.83 °Brix, 1.65, respectively. The results suggested model development and update with MU and SBC could improve the robustness of the portable NIRS instrument in predicting SSC of navel orange.

## KEYWORDS

navel orange, SSC, TA, near infrared spectroscopy, batch, portable instrument

## Introduction

The internal quality attributes of soluble solids content (SSC) and titrable acidity (TA) are main indexes for characterizing maturity level and taste of citrus fruit. The near infrared spectroscopy (NIRS) with the advantages of non-destructive, rapid and on field application was firstly integrated into package system for fruit sorting in Japan since the late 1980s (1). The portable NIRS instrument followed online sorting system was developed successfully in 2,000, which was suitable for determining fruit properties such as monitoring quality variation for supply chain or predicting optimal harvest date (OHD) (2). To some extent, model development and update determine the success or failure for practical application of NIRS technology.

Currently, partial least square regression (PLSR) algorithm is adopted widely to build a robust model combined with pretreatment methods (3). The first or second derivative spectra were frequently applied to develop PLSR models rather than raw absorption spectra because of baseline correction with derivative pretreatment. The combinations of derivative spectra and standard normal variate (SNV) or multiplicative scattering correction (MSC) were generally operated for removing light scatter and improving the accuracy of the model to the interest attributes of fruit (4). An extension algorithm of SNV named variable sorting for normalization (VSN) was proposed by Rabatel et al. (5) in 2019. VSN adjusted the weight to the wavelength for strengthening the target variables. The root mean square error of prediction (RMSEP) was reduced by 34% with VSN pretreatment for olive dry matter content (DMC) when the size of fruit did not match with the measurement window of the portable NIRS instrument (6). The normalized spectral ratio (NSR) was proposed by Li et al. (7) in 2020, which chose two wavelengths for replacing slope and offset parameters of MSC. The RMSEP of SSC for apple was decreased from 0.85 to 0.64% after NSR correction using an online sorting system.

Variable selection also is a critical step for extracting useful variables or eliminating variables containing mostly noise during model development (3). Recently, variable selection algorithms were explored such as variable combination population analysis (VCPA) (8), iteratively retaining informative variables (IRIV) (9), Monte Carlo uniformation variable elimination (MC-UVE) (10), successive projections algorithm (SPA) (11) and window search (6). For fruit quality assessment by the portable NIRS instrument, an optimal wavelength region often was recommended. For example, a good result with

determination coefficient ( $R^2$ ) of 0.90–0.96 and RMSEP of 0.29–0.33 °Brix was achieved in predicting SSC of pear, and the range 700–930 nm was recommended for pear sugar content using Vis-NIR instrumentation adopted in the current study (12). The region of 729–975 nm was recommended to build PLSR model for DMC of mango for Vis-NIR instrumentation (13). Therefore, variable selection methods should be considered in this work.

Mathematical models imbedded in the portable NIRS instrument performed poorly when the calibration model predicted new groups with different season, variety, population or environmental conditions (14). At present, the common solution is merging new samples into the original calibration set and recalibrated, here it is named as model updating (MU) (15). MU is necessary to support the portable NIRS instrument application success for fruit quality assessment (16). The variation between calibration and prediction sets causes the change of both slope and bias of PLSR model (17), slope and bias correction (SBC) is a simple and easy way to maintain PLSR model (18). The dynamic orthogonalization projection (DOP) algorithm adopts external parameter orthogonalisation (EPO) framework to correct the spectra between the original calibration and new prediction sets for updating the PLSR model. RMSEP was reduced by 66% in predicting DMC of olive fruit using DOP algorithm (19). The orthogonal signal correction (OSC) proposed by Wold et al. (20) provides another way to correct the spectra difference among fruit varieties (21). Mishra and Woltering (15) proposed a semi-supervised parameter-free calibration enhancement (PFCE) approach to update the model for predicting several new batches of pear and kiwi fruit. The performance of the model in predicting moisture

TABLE 1 Statistical results for 12 populations.

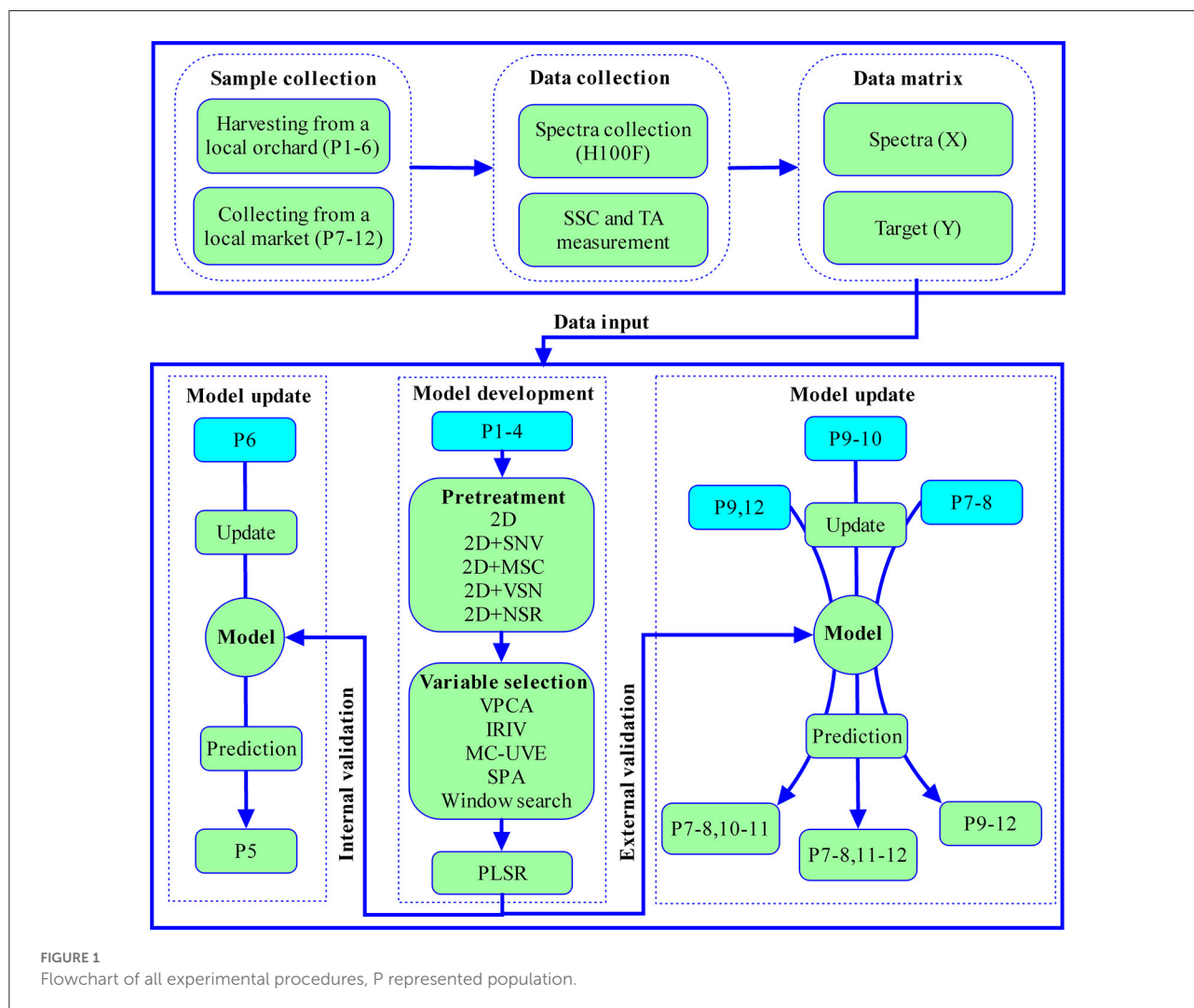
Population	Variety	Fruit #	Spectra #	SSC (°Brix)		TA (g/L)		Date
				Mean	SD	Mean	SD	
1	Newhall	22	66	11.3	0.77	101.1	22.64	14/10/2021
2	Newhall	29	87	11.2	0.92	88.8	14.56	24/10/2021
3	Newhall	29	87	12.1	1.02	18.4	6.00	09/11/2021
4	Newhall	29	87	11.9	1.07	17.4	6.65	23/11/2021
5	Newhall	30	87	12.6	1.14	17.8	7.21	10/12/2021
6	Newhall	40	120	12.5	0.87	16.0	4.38	31/12/2021
7	Lunwan	30	90	11.7	1.84			10/05/2022
8	Lunwan	30	90	11.8	1.27			20/05/2022
9	Lunwan	30	90	11.0	1.03			28/05/2022
10	Lunwan	30	90	11.3	1.22			10/06/2022
11	Lunwan	30	90	11.1	1.15			24/06/2022
12	Lunwan	30	90	11.2	1.10			10/07/2022
Total	359	1074						

SSC is soluble solids content; TA is titrable acidity; SD is standard deviation.

TABLE 2 The differences and similarities between two varieties derived from [www.baidu.com](http://www.baidu.com) with keywords of Newhall and Lunwan in Chinese.

Variety	Weight	Mature period	Edible rate	Skin color
Newhall	250–350 g	November to January	73–75%	Orange yellow
Lunwan	>200 g	March to April	74.1%	Shallow orange

Weight refers to average weight of single fruit.



content and total soluble solids of pear and kiwi fruit were greatly improved using PFCE approach. Therefore, Model updating also was considered for assessment the quality of navel orange.

The objective was to investigate an approach for model development and update in assessment the quality of navel orange using a portable NIRS instrument. The pretreatment and variable selection methods were attempted to build a robust PLSR model. The update methods were explored for improving the performance of PLSR model to new populations.

## Materials and methods

### Samples preparation

The populations of 1-6 (Newhall navel orange) were harvested from a local orchard from Oct. 14, 2021 to Dec. 31, 2021 located at Ganzhou Citrus Research Institute (114°51'2'' E, 25°46'36'' N), Ganzhou, China. The fruit tree was 6-year-old Newhall navel orange, a mainstream species of citrus. Thirty and forty samples were collected randomly from three trees at moderate height for the first five and sixth batches, respectively.

The samples were cleaned and transported to the laboratory at Nanchang, China. Removing damage samples, a total of 179 navel orange samples were obtained. To verify the applicability of the model to new batches, the populations of 7–12 (Lunwan navel orange) ( $n = 180$ ) were collected randomly from a local market from May 10, 2022 to Jul. 10, 2022. The statistical results for the samples were listed in Table 1. For understanding the differences and similarities well-between the two varieties, the main features for Newhall and Lunwan were listed in Table 2.

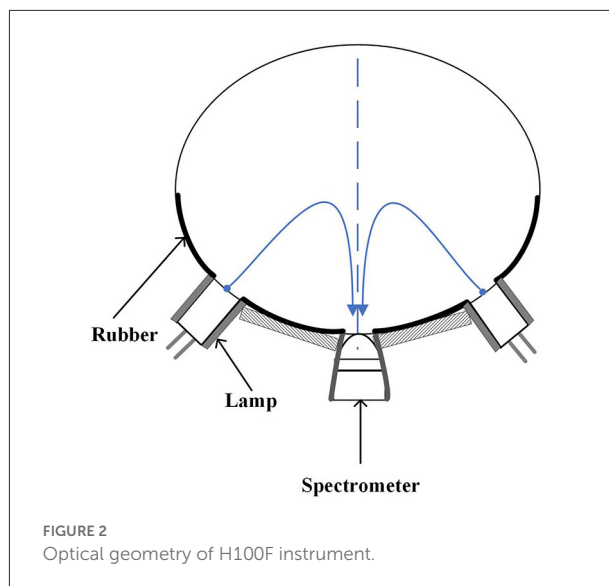
The procedures of the model development and update were shown in Figure 1. One internal and three external validation sets were applied to investigate influence of coverage of the update set to the prediction set (Figure 1).

## Near infrared spectra collection

The spectra of the samples were collected by hand-held non-destructive detector for fruit internal quality (H100F, Sunforest, Incheon, Korea). The full region from 650 to 950 nm composed of 151 variables with wavelength interval of 2 nm. The light source is two tungsten halogen lamps (10 W) arranged at an included angle of 100 degrees. In order to avoid the influence of external environment, the samples were placed at room temperature (about 20°C) for 24 h before spectra collection. After startup of the instrument, the reference and dark current spectrum were recorded with the integration time of 900 milliseconds. Then the spectrum of navel orange was recorded with the integration time of 900 milliseconds. During collection, the navel orange was tightly attached to the probe. The light emitted by two halogen tungsten lamps, after diffuse reflected by navel orange tissue, entered into the detector of the spectrometer and was converted into absorption spectrum (Figure 2). A spectrum was recorded every 120 degrees along the equator. Considering the variation of SSC distribution within a sample, each spectrum was regarded as a sample.

## Internal quality attributes measurement

After spectra collection, the flesh at the spectral measurement point was juiced, filtered and dripped into a beaker for measurement of SSC and TA. The SSC was measured by digital sugar meter (PAL-1, Atago Co., Tokyo, Japan). The measurement process is to suck juice drops onto the light window of the sugar meter until it covered the light window. The measurement was repeated for three times and the average value was obtained. TA was measured by potentiometric titration (PE28, Mettler-Toledo, Zurich, Switzerland), the end point of titration was reached when sodium hydroxide solution was used to titrate to pH 8.2, and TA was calculated according to the volume of sodium hydroxide solution consumed [formula (1)].



But it should be noted, for populations of 7–12 only SSC was measured.

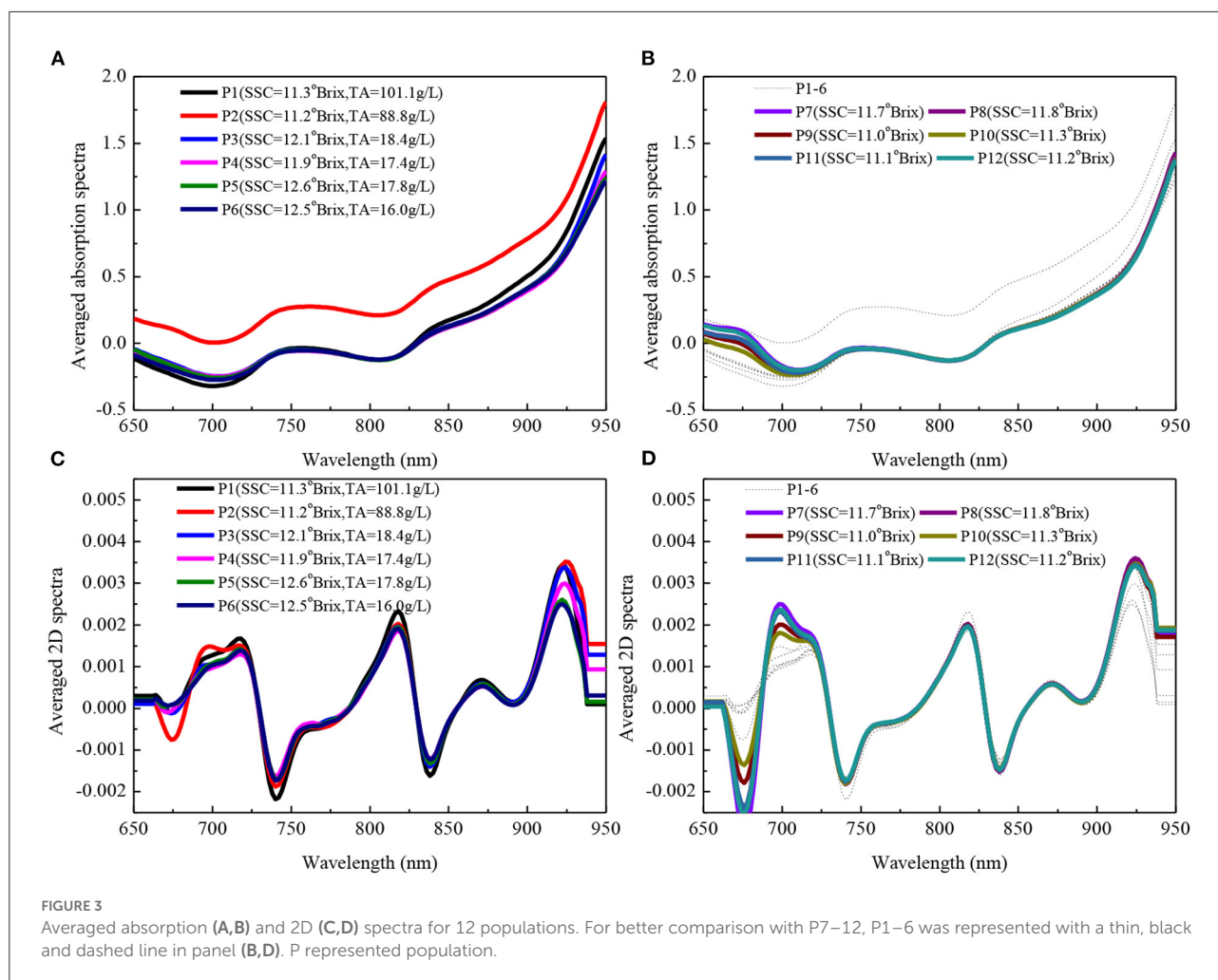
$$X = \frac{[c \times (V_1 - V_2)] \times k \times F}{m} \times 1000 \quad (1)$$

Where  $X$  is TA value,  $c$  is the concentration of sodium hydroxide solution,  $V_1$  is the volume of sodium hydroxide solution consumed when titrating fruit juice,  $V_2$  is the volume of sodium hydroxide solution consumed during blank experiment,  $k$  is the conversion factor of acid (citric acid is 0.064),  $F$  is the dilution ratio of fruit juice and  $m$  is volume of juice.

## Data analysis

The pretreatment algorithms of second derivative (2D), SNV, MSC, VSN and NSR, variable selection algorithms of VCPA, IRIV, MC-UVE, SPA and window search and model upgrade algorithms of SBC, DOP and OSC were carried out by use of Matlab 2020a (Mathworks Inc., Natick, MA). The VSN and DOP algorithms were provided by Prof. Roger (5, 19), the others were programmed by ourselves. All the algorithms were programmed based on PLSR.

The 5-fold cross validation was adopted to determine the number of latent variables (LVs). The determination coefficient of cross validation and prediction ( $R_{CV}^2$  and  $R_p^2$ ), root mean square error of cross validation and prediction (RMSECV and RMSEP) were calculated. A quick guideline of RPD defined as the ratio of standard deviation (SD) to the RMSECV or the RMSEP was applied for model assessment



(22, 23). The critical limit of RPD is one, above which some predictive power exists. Ideally, RPD should be  $>2$  for a good calibration.

## Results and discussion

### Analysis of near infrared spectral characteristic

Generally, the derivative spectra were applied to build models for removing baseline shift in fruit quality assessment by the portable NIRS instrument (3). The spectra of this study were transformed into second derivative (2D) spectra by second Savitzky-Golay derivative with window width of 13 points and fitting order of 2. The averaged absorption and 2D spectra were showed in Figure 3. For better comparison with the averaged spectra of populations 7–12, populations of 1–6 were represented with a thin, black and dashed line in panel

B and D of Figure 3. From Figure 3A, the spectral intensities for population 2 were higher than the others, however, the variation for six populations were consistent. The baseline shift was removed after 2D pretreatment, and the characteristic peaks of 680, 740, 815, 840, 880, and 920 nm could be seen from Figure 3B. The absorption peak around 680 nm is due to the decrease of chlorophyll content (24). The absorption peak near 740 nm is related to the third-overtone of O-H, and the absorption peak near 840 nm is assigned to an O-H combination (25, 26). The absorption peak near 815 nm is mostly related to the fourth-overtone N-H, and the absorption peak near 880 nm may be caused by the fourth overtone C-H (27). The absorption peak near 920 nm is interpreted as a combination of the third overtone CH-vibration and the third-overtone CH<sub>2</sub> vibration (28). In addition, the averaged SSC and TA for every population were also marked in Figure 3. But we could not observe pattern clearly, chemometrics methods should be adopted to interpret and develop the model.



## Pretreatment methods

Near infrared spectroscopy contains not only the target component information, but also other irrelevant information and noise. In the process of developing a robust mathematical model, pretreatment for the spectral data has become a common way to remove irrelevant information (29). The pretreatment algorithms of 2D, SNV, MSC, VSN, and NSR were adopted for finding out the suitable pretreatment method. Two wavelengths of 824 and 916 nm were chosen by NSR algorithm without any physical justification, respectively. The number of latent variables (LVs) was determined by the lowest RMSECV value

using 5-fold cross validation method. The optimal number of LVs for SSC and TA were 9 and 7, respectively. As shown in Table 3, the best model for SSC ( $R_p^2 = 0.66$ , RMSEP = 0.66°Brix) with RPD of 1.73 indicated that this model could screen SSC roughly. But this work also encouraged quality assessment of navel orange to forward. The RMSEP reduction of 17.5% for citrus was inferior to 34% reduction for olive DMC (6). Because citrus fruit size matched well with H100F test window, but single olive size was smaller than F750 test window. This result was superior to 10% reduction using SNV pretreatment method for citrus SSC by a portable device (30). However, several pretreatment methods did not achieve good results in

TABLE 3 Calibration (P1–4) and prediction (P5) results of PLSR models for SSC and TA with different pretreatment methods.

Attribute	Pretreatment	LVs	Calibration				Prediction			
			$R_{cv}^2$	RMSECV	Bias	RPD	$R_p^2$	RMSEP	Bias	RPD
SSC (°Brix)	2D	9	0.62	0.64	−0.0045	1.48	0.49	0.80	0.2548	1.43
	2D + SNV	9	0.58	0.67	−0.0024	1.42	0.52	0.78	0.3266	1.46
	2D + MSC	9	0.59	0.66	−0.0040	1.44	0.57	0.74	0.3737	1.54
	<b>2D + VSN</b>	<b>9</b>	<b>0.60</b>	<b>0.66</b>	<b>−0.0013</b>	<b>1.44</b>	<b>0.66</b>	<b>0.66</b>	<b>0.1382</b>	<b>1.73</b>
	2D + NSR	9	0.53	0.71	−0.0080	1.34	0.47	0.95	0.4764	1.20
TA (g/L)	<b>2D</b>	<b>7</b>	<b>0.83</b>	<b>16.16</b>	<b>−0.0079</b>	<b>0.77</b>	<b>0.66</b>	<b>21.82</b>	<b>0.6255</b>	<b>0.33</b>
	2D + SNV	7	0.79	17.86	−0.0154	0.70	0.51	26.18	5.2522	0.28
	2D + MSC	7	0.80	17.74	−0.0066	0.70	0.50	26.24	6.7637	0.27
	2D + VSN	7	0.79	17.92	−0.0087	0.71	0.53	25.53	7.1741	0.28
	2D + NSR	7	0.80	17.48	0.0129	0.71	0.54	25.29	−2.7920	0.29

The best calibration and prediction results for each parameter were shown in bold. P represented population.

TABLE 4 Calibration (P1–4) and prediction (P5) results of PLSR models for SSC with different variable selection methods based on 2D or 2D + VSN pretreatment.

Attribute	Pretreatment	Method	Variable	LVs	Calibration				Prediction			
					$R_{cv}^2$	RMSECV	Bias	RPD	$R_p^2$	RMSEP	Bias	RPD
SSC (°Brix)	2D	None	151	9	0.62	0.64	−0.0045	1.48	0.49	0.80	0.2548	1.43
	2D	VCPA	10	3	0.56	0.69	−0.0054	1.38	0.45	0.83	−0.1752	1.37
	2D	IRIV	30	5	0.62	0.63	−0.0022	1.51	0.54	0.76	0.1715	1.50
	2D	MC-UVE	45	4	0.59	0.66	−0.0018	1.44	0.45	0.84	0.1113	1.36
	2D	SPA	24	11	0.60	0.65	−0.0019	1.46	0.48	0.81	0.1481	1.41
	<b>2D</b>	<b>Window search</b>	<b>121</b>	<b>9</b>	<b>0.65</b>	<b>0.61</b>	<b>−0.0023</b>	<b>1.56</b>	<b>0.55</b>	<b>0.75</b>	<b>0.1016</b>	<b>1.52</b>
	2D	Previous report	116	7	0.62	0.64	−0.0051	1.48	0.51	0.79	0.1534	1.44
	<b>2D + VSN</b>	<b>None</b>	<b>151</b>	<b>9</b>	<b>0.60</b>	<b>0.66</b>	<b>−0.0013</b>	<b>1.44</b>	<b>0.66</b>	<b>0.66</b>	<b>0.1382</b>	<b>1.73</b>
	2D + VSN	VCPA	11	6	0.55	0.69	−0.0056	1.38	0.55	0.76	−0.0158	1.50
	2D + VSN	IRIV	41	6	0.60	0.65	−0.0019	1.46	0.63	0.68	0.1429	1.68
	2D + VSN	MC-UVE	35	10	0.55	0.69	−0.0038	1.38	0.40	0.87	−0.1349	1.31
	2D + VSN	SPA	27	11	0.59	0.65	−0.0040	1.46	0.56	0.75	−0.0714	1.52
	2D + VSN	Window search	129	9	0.61	0.64	−0.0020	1.48	0.56	0.75	−0.1605	1.52
	2D + VSN	Previous report	116	9	0.52	0.71	−0.0026	1.34	−24.99	5.75	1.4920	0.20

Previous report recommended wavelength band of 700–930 nm for H100F instrument in predicting SSC of pear by Choi et al. (12). The best result in calibration and prediction for each parameter was bolded. P represented population.

TA prediction. The best model ( $R_p^2 = 0.66$ , RMSEP = 21.82 g/L) with RPD of 0.33 for TA developed by 2D spectra did not have the predictive ability, this result reconsolidated the conclusion in the literatures (4, 31). It had been noted that the interactance

optical geometry for field-portable instrumentation was not appropriate for assessment of the acidity of intact low TA fruit of 1%. Therefore, only SSC model was developed and updated in the following section.

### Variable selection

Variable selection is a common method for extracting informative variables or removing irrelevant variables to target component during model development (32). Generally, two strategies involved wavelength combinations and wavelength bands were adopted to build a robust model. The wavelength range for most of the portable fruit selector instruments locates below 1,100 nm due to strong penetration ability and low-cost spectrometer. The variable number often did not exceed three hundred, for example, 151 variables for H100F and 306 for F750, respectively. Therefore, a reasonable window often is recommended, for example 700–930 nm for H100F and 729–975 nm for F750 in predicting SSC of pear and DMC of mango (12, 13). In this work, window search and compared methods of VCPA, IRIV, MC-UVE and SPA were adopted

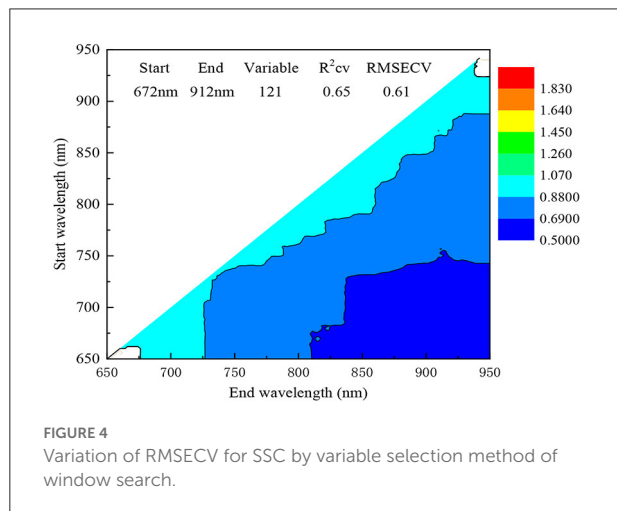


FIGURE 4 Variation of RMSECV for SSC by variable selection method of window search.

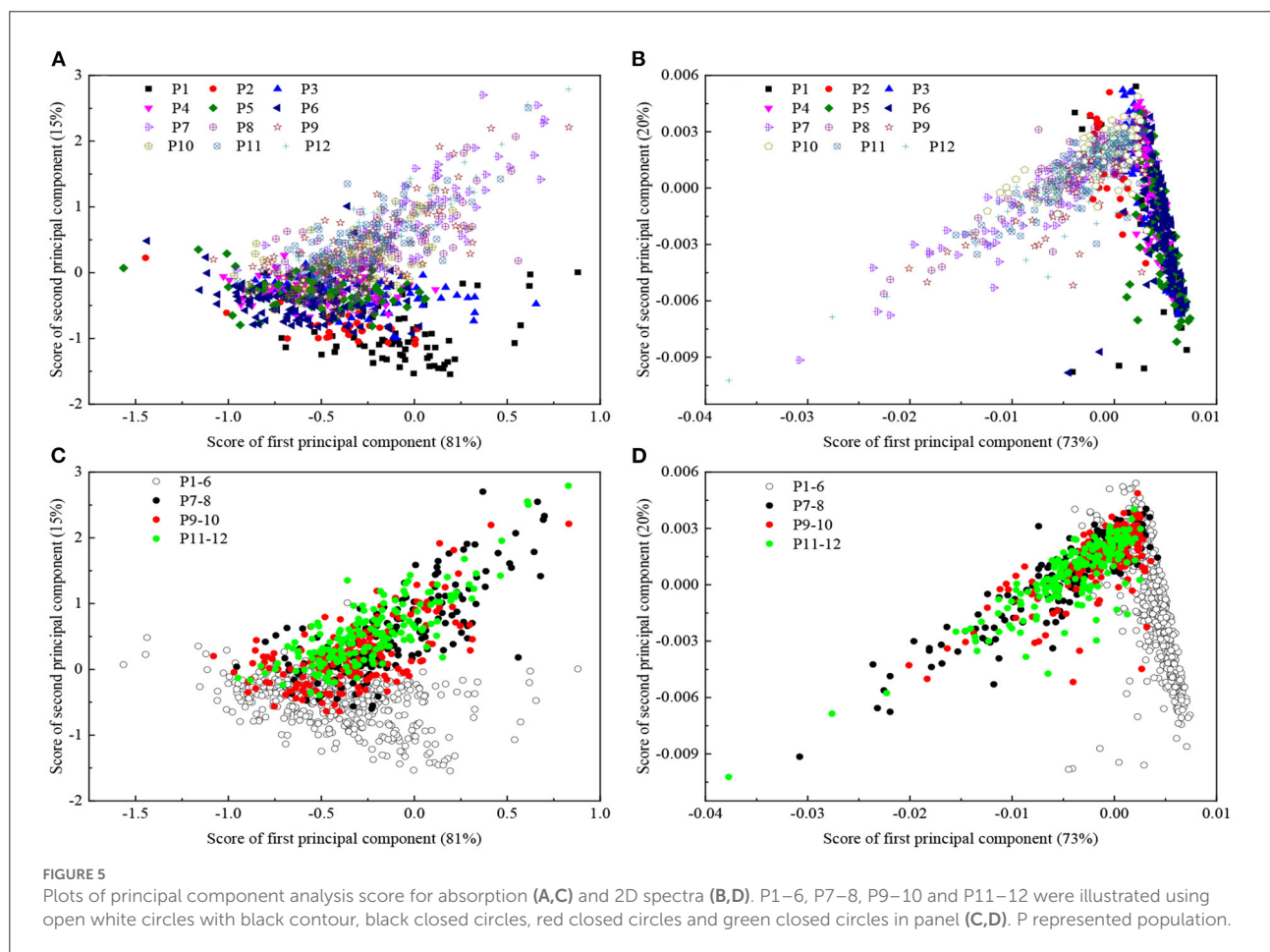


FIGURE 5 Plots of principal component analysis score for absorption (A,C) and 2D spectra (B,D). P1–6, P7–8, P9–10 and P11–12 were illustrated using open white circles with black contour, black closed circles, red closed circles and green closed circles in panel (C,D). P represented population.

TABLE 5 Model update results for H100F portable NIRS instrument.

Method	Parameters	Calibration					Prediction				
		Population	$R^2_{CV}$	RMSECV	Bias	RPD	Population	$R^2_p$	RMSEP	Bias	RPD
None	LVs = 9	1-4	0.60	0.66	-0.0013	1.44	5	0.66	0.66	0.1382	1.73
None	LVs = 9	1-4	0.60	0.66	-0.0013	1.44	9-12	-1.92	2.10	1.6978	0.62
None	LVs = 9	1-4	0.60	0.66	-0.0013	1.44	7-8, 10-11	-2.52	2.27	1.7183	0.60
None	LVs = 9	1-4	0.60	0.66	-0.0013	1.44	7-8, 11-12	-3.79	2.58	2.1336	0.52
OSC	LVs = 7	1-4	0.44	0.77	-0.0030	1.23	5	0.69	0.63	0.0369	1.81
OSC	LVs = 7	1-4	0.44	0.77	-0.0030	1.23	9-12	0.08	1.18	0.0839	1.13
OSC	LVs = 7	1-4	0.44	0.77	-0.0030	1.23	7-8, 10-11	0.14	1.12	0.1031	1.22
OSC	LVs = 7	1-4	0.44	0.77	-0.0030	1.23	7-8, 11-12	0.05	1.20	0.4515	1.12
MU	LVs = 9	1-4 + (6)**	0.61	0.66	-0.0036	1.41	5	0.64	0.67	0.3178	1.70
<b>MU</b>	<b>LVs = 9</b>	<b>1-4 + (7-8)**</b>	<b>0.55</b>	<b>0.74</b>	<b>0.0044</b>	<b>1.55</b>	<b>9-12</b>	<b>0.54</b>	<b>0.83</b>	<b>-0.0632</b>	<b>1.60</b>
MU	LVs = 9	1-4 + (9, 12)**	0.56	0.74	-0.0011	1.34	7-8, 10-11	0.52	0.84	-0.0300	1.63
MU	LVs = 9	1-4 + (9-10)**	0.59	0.73	-0.0016	1.37	7-8, 11-12	0.31	0.98	0.5340	1.37
SBC	LVs = 9	1-4 + (6)*#	0.60	0.66	-0.0013	1.41	5	0.71	0.60	0.0577	1.90
SBC	LVs = 9	1-4 + (7-8)*#	0.60	0.66	-0.0013	1.55	9-12	0.55	0.84	-0.0129	1.58
<b>SBC</b>	<b>LVs = 9</b>	<b>1-4 + (9, 12)*#</b>	<b>0.60</b>	<b>0.66</b>	<b>-0.0013</b>	<b>1.34</b>	<b>7-8, 10-11</b>	<b>0.52</b>	<b>0.83</b>	<b>0.0152</b>	<b>1.65</b>
<b>SBC</b>	<b>LVs = 9</b>	<b>1-4 + (9-10)*#</b>	<b>0.60</b>	<b>0.66</b>	<b>-0.0013</b>	<b>1.37</b>	<b>7-8, 11-12</b>	<b>0.33</b>	<b>0.96</b>	<b>0.5012</b>	<b>1.40</b>
<b>DOP</b>	<b>LVs = 9, K = 6</b>	<b>1-4 + (6)*</b>	<b>0.61</b>	<b>0.65</b>	<b>-0.0032</b>	<b>1.46</b>	<b>5</b>	<b>0.73</b>	<b>0.58</b>	<b>0.0261</b>	<b>1.97</b>
DOP	LVs = 9, K = 6	1-4 + (7-8)*	0.54	0.70	-0.0035	1.64	9-12	0.49	0.88	0.0114	1.51
DOP	LVs = 9, K = 6	1-4 + (9, 12)*	0.57	0.68	-0.0031	1.46	7-8, 10-11	0.47	0.88	0.0636	1.56
DOP	LVs = 9, K = 5	1-4 + (9-10)*	0.56	0.68	-0.0032	1.47	7-8, 11-12	0.20	1.05	0.6199	1.28

The PLSR model for SSC was developed based on 2D + VSN pretreatment. The best result in calibration set and prediction set for SSC was bolded. RPD value of MU was different with None, OSC, SBC and DOP. Because the number in predicting set was different. The symbol (\*) represents the update set for spectra correction between the original calibration set and the update set using DOP method. The symbol (#) represents the update set for model parameters correction between the original calibration set and the update set using SBC method. The symbol (\*\*) represents the update set adding to the original calibration set for recalibration.

to choose related variables to SSC (Table 4). For SSC, the highest RPD of 1.52 was obtained with the optimal wavelength band of 672–912 nm (121 variables). Figure 4 showed the result of variable selection by window search, and the minimum RMSECV can be obtained in the region of 672–912 nm. However, the best result ( $R^2_p = 0.55$ ,  $RMSEP = 0.75^\circ\text{Brix}$ ,  $RPD = 1.52$ ) with variable selection was still inferior to VSN pretreatment with RPD of 1.73. Because VSN attached different weights for adjusting the contribution of every variable to SSC. Further, all methods for variable selection failed based on 2D spectra after VSN pretreated. Therefore, 2D spectra combination of VSN pretreatment in the full region (650–950 nm) were applied for model update.

## Model update

The samples in calibration set may not cover with new samples well in prediction set, and the performance of the model will decrease (14). Therefore, model update is a crucial routine for practical application of the portable NIRS instrument. The principle of the model update is to figure out the pattern between calibration and prediction sets. So 2D spectra in the

region of 650–950 nm were applied for SSC model update in this section. The score plots of principal component analysis (PCA) for absorption and 2D spectra were shown in Figure 5. The update and prediction sets were designed as shown in Figure 1. The update set of populations 9–10 and prediction set (populations 7–8, 11–12) were used as a negative group during external validation because the region of populations 9–10 could not cover the populations 7–8 and 11–12 well. The methods of DOP, OSC, SBC and MU were employed to update the model, respectively.

DOP method is an extension of EPO (19), which removes the variation in spectra caused by the population change using EPO framework. OSC was compared because of good spectra correction ability was achieved in the literature (21). SBC was also considered because it provided a direct correction routine based on model parameters (18, 33). The number of the samples adding into the calibration set was determined via observing the change of the RMSEP (16). For internal and external validation sets, 40 and 30 fruit samples were used as the standard samples in the update sets, which were added into the original calibration set or corrected spectra. K, the parameter of DOP, was optimized from 1 to 20, was listed in Table 5.



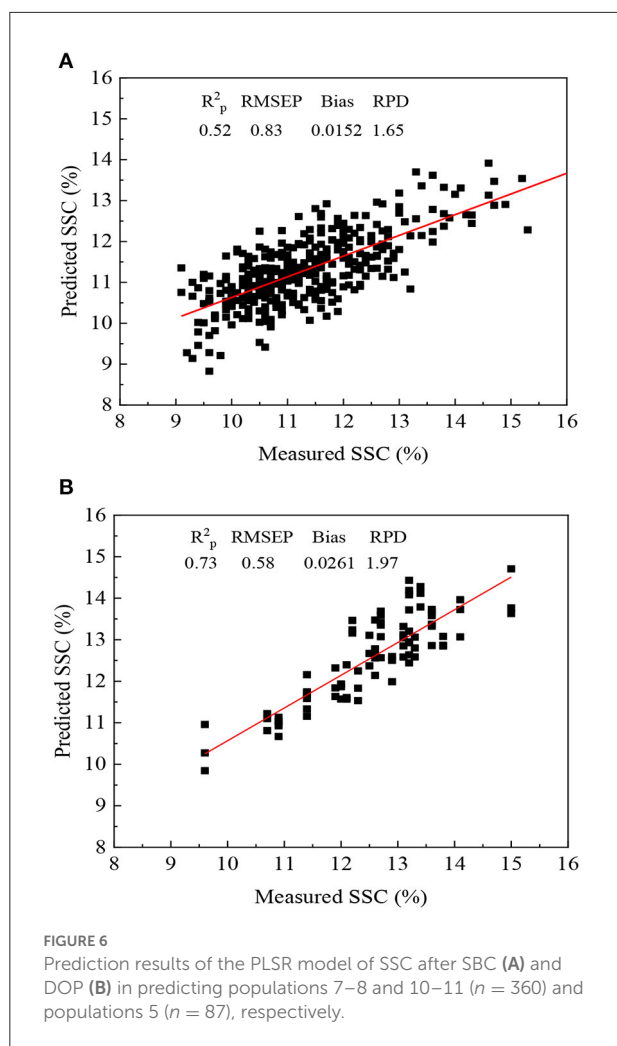


FIGURE 6 Prediction results of the PLSR model of SSC after SBC (A) and DOP (B) in predicting populations 7–8 and 10–11 ( $n = 360$ ) and populations 5 ( $n = 87$ ), respectively.

For the internal validation (population 5), the best result was obtained using DOP method followed by SBC, OSC and MU with  $R_p^2$  of 0.73, RMSEP of 0.58 °Brix and RPD of 1.97 (Table 5; Figure 6B). It could be found from rows 2 to 4 in Table 5 that direct application of the PLSR model failed completely in external validation ( $RPD < 1$ ). It should be noted that the PLSR model developed with Newhall navel orange only allowed to work on the new batch (Lunwan navel orange) after updating. For the three external validation sets, methods of MU, SBC and DOP for the two positive groups (populations 9–12 and 7–8, 10–11) achieved better results than the negative group (populations 7–8 and 11–12) except OSC. The negative group yielded always the worst results, because the performance for model update was determined by the representation of the update set to the prediction set (34). The prediction set could not be covered well by the update set in Figures 5C,D. The best results in predicting populations 9–12, populations 7–8 and 10–11 and populations 7–8 and 11–12 were achieved by MU and SBC, respectively (Table 5; Figure 6A).

TABLE 6 The previous reports on assessment of citrus quality using NIRS instrument.

Objective	Attribute	Method	Result	References
Citrus	SSC	SNV-SPA	$R_p = 0.92$ , RMSEP = 0.57	(30)
Orange	SSC	SpectraNet-53	$R_p = 0.40$ , RMSEP = 1.16	(35)
Mandarin	Sugar degree	BPNN	$R_p = 0.87$ , RMSEP = 0.74	(36)
Citrus	SSC	VABPLS	$R_p = 0.82$ , RMSEP = 0.60	(37)

$R_p$  is correlation coefficient of prediction; VABPLS is variable adaptive boosting partial least squares; SpectraNet-53 is a deep residual learning architecture; BPNN is backward propagation neural network.

The internal validation (population 5) and two positive validation (population 9–12, 7–8, and 10–11) sets with  $R_p^2$  of 0.52–0.73 and RMSEP of 0.58–0.83 were superior to the previous report in predicting citrus SSC for  $R_p$  of 0.4 and RMSEP of 1.16 Brix (30). However, our result was inferior to the reports in the literatures (35–37) (Table 6). Because an external prediction set with different variety and region was still a challenge for the portable NIRS instrument practical application. Methods of SBC and MU were also recommended for easy operation for the use.

## Conclusions

Model development of pretreatment and variable selection methods and model upgrade methods of MU, SBC, DOP, and OSC were investigated to ensure the predictive ability of portable NIRS instrument for SSC and TA of citrus fruit. The best models could predict SSC but not TA according to the critical limit of RPD of one. For model development, the combinations of VSN and second derivative pretreatment in the full region (650–950 nm) achieve the best results with  $R_p^2$  of 0.66, RMSEP of 0.66 °Brix and RPD of 1.73 in predicting SSC. For model update, MU and SBC achieved the best results in predicting SSC for two external validation sets with  $R_p^2$ , RMSEP and RPD of 0.54, 0.83 °Brix, 1.60 and 0.52, 0.83 °Brix, 1.65, respectively. In addition, this work provided a case for model development and upgrade for the portable NIRS instrument in practical application of fruit quality assessment.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

XS: conceptualization, methodology, review and editing, project administration, funding acquisition, supervision, and resources. DD: spectra collection, formal analysis, and original draft. JL: sample collection and analysis. SF: instrument support. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by Jiangxi Province Technology Innovation Guidance Project (No. 20212BDH 81008).

## Conflict of interest

Author SF is employed by Beijing Sunlight Yishida Technology Co., Ltd.

## References

- Kawano S. Present condition of nondestructive quality evaluation of fruits and vegetables in Japan. *JARQ*. (1994) 28:212–6.
- Zude M, Pflanz M, Kaprielian M, Aivazian BL. NIRS as a tool for precision horticulture in the citrus industry. *Biosyst Eng*. (2008) 99:455–9. doi: 10.1016/j.biosystemseng.2007.10.016
- Anderson NT, Walsh KB. Review: the evolution of chemometrics coupled with near infrared spectroscopy for fruit quality evaluation. *J Near Infrared Spec*. (2022) 30:3–17. doi: 10.1177/09670335211057235
- Walsh K B, Blasco J, Zude-Sasse M, Sun X. Visible-NIR 'point' spectroscopy in postharvest fruit and vegetable assessment: the science behind three decades of commercial use. *Postharvest Biol Technol*. (2020) 168:111246. doi: 10.1016/j.postharvbio.2020.111246
- Rabatel G, Marini F, Walczak B, Roger J. VSN: variable sorting for normalization. *J Chemom*. (2019) 34:e3164. doi: 10.1002/cem.3164
- Sun X, Subedi P, Walker R, Walsh KB. NIRS prediction of dry matter content of single olive fruit with consideration of variable sorting for normalisation pre-treatment. *Postharvest Biol Technol*. (2020) 163:11140. doi: 10.1016/j.postharvbio.2020.111140
- Li L, Peng Y, Yang C, Li Y. Optical sensing system for detection of the internal and external quality attributes of apples. *Postharvest Biol Technol*. (2020) 162:111101. doi: 10.1016/j.postharvbio.2019.111101
- Yun Y, Wang W, Deng B, Lai G, Liu X, Ren D, et al. Using variable combination population analysis for variable selection in multivariate calibration. *Anal Chim Acta*. (2015) 862:14–23. doi: 10.1016/j.aca.2014.12.048
- Yun Y, Wang W, Tan M, Liang Y, Li H, Cao D, et al. A strategy that iteratively retains informative variables for selecting optimal variable subset in multivariate calibration. *Anal Chim Acta*. (2014) 807:36–43. doi: 10.1016/j.aca.2013.11.032
- Cai W, Li Y, Shao X. A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra. *Chemom Intell Lab Syst*. (2008) 90:188–94. doi: 10.1016/j.chemolab.2007.10.001
- Soares SFCS, Gomes AA, Araujo MCU, Filho ARG, Galvão RKH. The successive projections algorithm. *TrAC, Trends Anal Chem*. (2013) 42:84–98. doi: 10.1016/j.trac.2012.09.006
- Choi JH, Chen PA, Lee BH, Yim SH, Kim MS, Bae YS, et al. Portable, non-destructive tester integrating VIS/NIR reflectance spectroscopy

The remaining authors declared that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnut.2022.976178/full#supplementary-material>

- for the detection of sugar content in Asian pears. *Sci Hortic*. (2017) 220:147–53. doi: 10.1016/j.scienta.2017.03.050
- Anderson NT, Subedi PP, Walsh KB. Manipulation of mango fruit by dry matter content to improve eating quality. *Sci Hortic*. (2017) 226:316–21. doi: 10.1016/j.scienta.2017.09.001
- Goisser S, Wittmann S, Mempel H. Food-scanner applications in the fruit and vegetable sector. *Landtechnik*. (2021) 76:52–67. doi: 10.15150/lt.2021.3264
- Mishra P, Woltering E. Handling batch-to-batch variability in portable spectroscopy of fresh fruit with minimal parameter adjustment. *Anal Chim Acta*. (2021) 1177:338771. doi: 10.1016/j.aca.2021.338771
- Hayes CJ, Walsh KB, Greensill CV. Improving calibration transfer between shortwave near infrared silicon photodiode array instruments. *J Near Infrared Spectrosc*. (2016) 24:59–68. doi: 10.1255/jnirs.1194
- Feng J, McGlone AV, Currie M, Clark CJ, Jordan BR. Assessment of yellow-fleshed kiwifruit (*Actinidia chinensis* 'Hort16A') quality in pre- and post-harvest conditions using a portable near-infrared spectrometer. *Hortscience*. (2011) 46:57–63. doi: 10.21273/HORTSCI.46.1.57
- Ji N, Li M, Li W, Liu R, Zhang Y, Han D. Study on NIR model transfer between similar kinds of fruits based on slope/bias algorithm. *Spectrosc Spect Anal*. (2017) 37:227–31. doi: 10.3964/j.issn.1000-0593(2017)01-0227-05
- Mishra P, Roger JM, Rutledge DN, Woltering E. Two standard-free approaches to correct for external influences on near-infrared spectra to make models widely applicable. *Postharvest Biol Technol*. (2020) 170:111326. doi: 10.1016/j.postharvbio.2020.111326
- Wold S, Antti H, Lindgren F, Ohman J. Orthogonal signal correction of near-infrared spectra. *Chemometr Intell Lab Syst*. (1998) 44:175–85. doi: 10.1016/S0169-7439(98)00109-9
- Wang J, Wang J, Chen Z, Han D. Development of multi-cultivar models for predicting the soluble solid content and firmness of European pear (*Pyrus communis* L.) using portable vis-NIR spectroscopy. *Postharvest Biol Technol*. (2017) 129:143–51. doi: 10.1016/j.postharvbio.2017.03.012
- Li J, Huang W, Chen L, Fan S, Zhang B, Guo Z, et al. Variable selection in visible and near-infrared spectral analysis for noninvasive determination of soluble solids content of 'Ya' pear. *Food Anal Methods*. (2014) 7:1891–902. doi: 10.1007/s12161-014-9832-8

23. Williams P. The RPD statistic: a tutorial note. *NIR News*. (2014) 25:22–6. doi: 10.1255/nirn.1419
24. Zude-Sasse M, Truppel I, Herold B. An approach to non-destructive apple fruit chlorophyll determination. *Postharvest Biol Technol*. (2002) 25:123–33. doi: 10.1016/S0925-5214(01)00173-9
25. Jamshidi B, Minaei S, Mohajerani E, Ghassemian H. Reflectance Vis/NIR spectroscopy for nondestructive taste characterization of Valencia oranges. *Comput Electron Agric*. (2012) 85:64–9. doi: 10.1016/j.compag.2012.03.008
26. Golic M, Walsh K, Lawson P. Short-wavelength near-infrared spectra of sucrose, glucose, and fructose with respect to sugar concentration and temperature. *Appl spectrosc*. (2003) 57:139–45. doi: 10.1366/000370203321535033
27. Yan Y. *Basic and Application of Near Infrared Spectroscopy Analysis*. Beijing: China Light Industry Press (2005).
28. Herold B, Kawano S, Sumpf B, Tillmann P, Walsh KB. *VIS/NIR Spectroscopy: Zude/Optical Monitoring of Fresh and Processed Agricultural Crop*. Boca Raton, FL: CRC Press (2008).
29. Shah SSA, Zeb A, Qureshi WS, Arslan M, Malik AU, Alasmay W, et al. Towards fruit maturity estimation using NIR spectroscopy. *Infrared Phys Technol*. (2020) 111:130479. doi: 10.1016/j.infrared.2020.103479
30. Tian X, Li J, Yi S, Jin G, Qiu X, Li Y. Nondestructive determining the soluble solids content of citrus using near infrared transmittance technology combined with the variable selection algorithm. *Artif Intell Agric*. (2020) 4:48–57. doi: 10.1016/j.aiaa.2020.05.001
31. Subedi PP, Walsh KB, Hopkins DW. Assessment of titratable acidity in fruit using short wave near infrared spectroscopy. Part B: intact fruit studies. *J Near Infrared Spectrosc*. (2012) 20:459–63. doi: 10.1255/jnirs.1011
32. Mishra P, Woltering E, Brouwer B, Hogeveen-van Echtelt E. Improving moisture and soluble solids content prediction in pear fruit using near-infrared spectroscopy with variable selection and model updating approach. *Postharvest Biol Technol*. (2021) 171:111348. doi: 10.1016/j.postharvbio.2020.111348
33. Wang A, Yang P, Chen J, Wu Z, Jia Y, Ma C, et al. A new calibration model transferring strategy maintaining the predictive abilities of NIR multivariate calibration model applied in different batches process of extraction. *Infrared Phys Techn*. (2019) 103:103046. doi: 10.1016/j.infrared.2019.103046
34. Cavaco AM, Pires R, Antunes MD, Panagopoulos T, Brázio A, Afonso AM, et al. Validation of short wave near infrared calibration models for the quality and ripen of 'Newhall' orange on tree across years and orchards. *Postharvest Biol Technol*. (2018) 141:86–97. doi: 10.1016/j.postharvbio.2018.03.013
35. Martins JA, Guerra R, Pires R, Antunes MD, Panagopoulos T, Brázio A, et al. SpectraNet-53: a deep residual learning architecture for predicting soluble solids content with VIS-NIR spectroscopy. *Comput Electron Agric*. (2022) 197:106945. doi: 10.1016/j.compag.2022.106945
36. Zhang K. Shatang mandarin sugar degree detection based on near infrared spectrum. In: *Proceedings of 2019 7th International Conference on Machinery, Materials and Computing Technology (ICMMCT 2019)*. Chongqing (2019).
37. Li P, Li S, Du G, Jiang L, Liu X, Ding S, et al. A simple and nondestructive approach for the analysis of soluble solid content in citrus by using portable visible to near-infrared spectroscopy. *Food Sci Nutr*. (2020) 8:2543–52. doi: 10.1002/fsn3.1550