

Prediction of quality traits in dry pepper powder using visible and near-infrared spectroscopy

^{1,2}Theanjumpol, P., ³Kaur, A., ^{1,2}Muenmanee, N., ^{1,2,4}Chanbang, Y. and ^{1,2*}Maniwara, P.

¹Postharvest Technology Research Center, Faculty of Agriculture, Chiang Mai University, Chiang Mai 50200, Thailand

²Postharvest Technology Innovation Center, Ministry of Higher Education, Science, Research and Innovation, Bangkok 10400, Thailand

³Department of Plant Sciences, University of California, Davis, California 95616, United States of America

⁴Department of Entomology and Plant Pathology, Faculty of Agriculture, Chiang Mai University, Chiang Mai 50200, Thailand

Article history

Received:

8 January 2022

Received in revised form:

23 June 2022

Accepted:

15 July 2022

Keywords

colour,

pungency,

rapid analysis,

near-infrared spectroscopy,

pepper powder quality

phenotyping

Abstract

Fruit quality phenotyping is a bottleneck in plant breeding. The present work aimed to investigate the applicability of visible (Vis) and near-infrared (NIR) spectroscopy for quality evaluation in dry red chili powder. We constructed prediction models for the American Spice Trade Association (ASTA)-colour and the Scoville Heat Unit (SHU)-pungency pepper traits using spectroscopy and multivariate statistical techniques. Predictive partial least squares (PLS) models were successfully achieved with high correlations (r) between the predicted and reference values for calibration and validation ($r = 0.955$ and 0.928 for ASTA-colour; $r = 0.941$ and 0.918 for SHU-pungency). Spectroscopy data from visible and short-wave radiation (Vis-SWNIR) provided the most robust (residual predictive deviation value) model for ASTA-colour (RPD = 2.84) and long-wave radiation (LWNIR) for SHU-pungency (RPD = 2.48). Spectral categories for wavelength range selection, variable importance for effective wavelength selection, and root mean press-statistic for factor selection were important criteria for PLS. Trait variance and distribution were also important criteria for the predictive capacity and power of the models. In conclusion, non-invasive spectroscopy was a promising tool in our study for dry red chili quality phenotyping.

DOI

<https://doi.org/10.47836/ifrj.30.1.16>

© All Rights Reserved

Introduction

Plant breeders must rapidly phenotype large numbers of plants to make selections. Plant phenotyping for fruit quality evaluation entails laborious post-harvest processing and biochemical laboratory analyses in addition to field phenotyping. Phenotyping, therefore, remains a bottleneck in the plant breeding pipeline for future breeding advances. Several high-throughput phenotyping tools have been investigated for plant growth and architecture phenotyping, including aerial- (*i.e.*, drones, helicopters, aerostats) and ground-driven (*i.e.*, tractors, vehicles, manual cameras) imaging-based remote sensing techniques (Furbank and Tester,

2011; White *et al.*, 2012; Araus and Cairns, 2014; Rebetzke *et al.*, 2019). For fruit quality phenotyping which involves physical and chemical evaluation of fruits, the suitability of spectroscopy has been investigated as a non-invasive technique for various fruit and vegetable commodities: Cheng *et al.* (2004) for cucumber; Sánchez *et al.* (2012) for strawberry; Sun *et al.* (2012) for orange; Pissard *et al.* (2013) for apple; Maniwara *et al.* (2019) for purple passion fruit; Donis-González *et al.* (2020) for grape and peach; and Kaur *et al.* (2020) for tomato and pepper.

Spectroscopic techniques study the absorption of energy from organic molecules in commodities in the electromagnetic spectrum regions of visible (380 - 720 nm) and near-infrared (780 - 2500 nm)

*Corresponding author.

Email: maniwara016@gmail.com

wavelengths. Signals of major structures and functional groups of organic compounds are detected within this wavelength range when irradiated by NIR (Wang *et al.*, 2015; Arendse *et al.*, 2018). The resulting spectrum from NIR irradiation reflects some physical attributes and/or the chemical constitution of the biological sample. However, the spectrum is heavily dominated by interfering factors such as absorbance by water, wavelength-dependent light scattering, instrumental noises, and low signal-to-noise ratio, thus obscuring the relevant information concerning fruit quality attributes (Nicolai *et al.*, 2007; Wang *et al.*, 2015). Multivariate statistical techniques (or chemometrics) are applied to extract useful spectral information using spectral pre-processing treatments and regression methods (Rinnan *et al.*, 2009) to build mathematical models that are predictive of product quality traits.

Our project was an inter-collaborative study between East-West Seed (EWS) International, India and Thailand, and the Postharvest Technology Research Center of Chiang Mai University, Thailand to investigate the feasibility of VIS and NIR spectroscopy for fruit quality evaluation in dry red chili (syn. dry pepper) of India. Dry pepper breeding in India for improved varieties is of economic importance. India is one of the major exporters of dry pepper and its derived products to different countries including Malaysia, Bangladesh, Sri Lanka, and the United States of America, with a net value reaching \$410 million (Reddy *et al.*, 2014; 2015). The current chili hybrid seed market in India is approximately 50 tons per year, with an estimated turnover worth \$16 million (Reddy *et al.*, 2015).

Pungency (or hotness) and colour (or redness) are two of the most desired red chili fruit quality traits in released cultivars of dry pepper. However, phenotyping (or quantification) of both traits requires laborious evaluations and biochemical analyses. Pepper pungency is typically quantified by the concentration of secondary metabolites known as capsaicinoids, which include compounds such as capsaicin, dihydrocapsaicin, and nordihydrocapsaicin (Srinivasan, 2016; Guzman and Bosland, 2017). Pepper pungency is measured in the Scoville Heat Unit (SHU) which indicates the number of times the substance must be diluted so that the pungency is not perceived. Pepper colour is quantified by the extractable colour or the colour compounds of exocarp (Ergüneş and Tarhan, 2006; Kim *et al.*, 2008). It is measured in the American Spice Trade

Association (ASTA) colour value, which signifies the intensity of redness or its saturation. We focused on these two fruit quality traits in our study, and incorporated germplasm with diverse backgrounds including commercial hybrids, advanced lines, and doubled-haploid lines of the EWS-India dry pepper breeding program.

Materials and methods

Plant material and field design

Samples ($n = 180$) representing biochemical variation in pungency and colour traits at the Aurangabad, India Research and Development (R&D) Station of EWS were selected. These samples encompassed F1 hybrids, accessions from doubled-haploid (DH), and F3-F8 populations. The dried and ground products of these pepper samples were kept in 4°C germplasm for the past six months to two years.

A novel set of genetic material ($n = 83$) consisting of DH lines and commercial F1 hybrids was also included in the project. Plants were greenhouse-grown in June 2019, and field-transplanted in July 2019 at the Mulani Farm of Aurangabad R&D Station into plots of five to ten plants per line. A few lines failed to germinate, resulting in a total of 79 lines. Plants were managed per farm guidelines. The phenotypic and yield data were collected by the breeding team as per routine work. The red fruits were harvested over two pickings at 111 and 147 days after transplant, respectively. Samples were dried in the sun for 10 - 15 days before further processing and analyses by the Aurangabad biochemistry laboratory.

Pepper processing and biochemical data collection

The dried pepper powder stored at 4°C for the 180 selected samples from the germplasm collection was used for biochemical analyses. The 79 sun-dried pepper samples were further dried in an air oven before ground into fine powder using routine spice/coffee blenders. For colour extraction, 0.1 g dry pepper powder was extracted in 100 mL absolute acetone with a 16-h dark incubation at room temperature using the ASTA-20.1 method (ASTA, 1986). The transparent extract of the samples was read against acetone reference at 460 nm using the UV-1800 Shimadzu Spectrophotometer with in-house UVProbe-Photometric software for the collection of optical density absorbance values. The absorbance data were then converted into American

Spice Trade Association (ASTA) colour values (ASTA, 1986; Kim *et al.*, 2008).

For capsaicinoid extraction, 5 g dry pepper powder was extracted in 80 mL absolute ethanol with a Soxhlet apparatus for 3 h. Samples were cooled to room temperature, filtered using filter paper, and incubated at 4°C until high-performance liquid chromatography (HPLC) injection. Each sample was filtered into a 2-mL HPLC vial with a 0.45- μ m filter and a 5-mL disposable syringe. Samples were read using the Prominence Shimadzu HPLC system equipped with a degasser, autosampler, column oven, LC Plus quaternary pump, and ultraviolet (UV) detector. The chromatographic conditions were Luna C₁₈ column (particle size 5 μ m, dimension 250 \times 4.6 mm) from Phenomenex (USA); column temperature: 40°C; sample temperature: 20°C; sample volume: 20 μ L; UV detection wavelength: 280 nm; mobile phase: binary mixture acetonitrile-water at 60:40, v/v; and flow rate: 1 mL/min. Capsaicin, dihydrocapsaicin, and nordihydrocapsaicin were determined by comparison to an external pure capsaicin reference standard, 10 ppm in ethanol, injected under the same conditions. The compounds were identified based on the retention times measured under identical HPLC conditions, and they were quantified using the peak areas of each sample. The in-house HPLC software was used to convert the peak area data into the Scoville Heat Unit (SHU)-pungency values using the coefficients of the heat value for each individual compound (Todd *et al.*, 1977; Sanatombi and Sharma, 2008; Othman *et al.*, 2011).

NIR data collection and pre-processing

A NIRSystem 6500 Multi-Mode Analyser BenchTop (Foss NIRSystems, Silver Spring, USA) with specifics of the spinning module, rotating cup, and tungsten halogen lamp was used to collect the absorbance data for ground dry pepper samples. The in-house VISION software was used to collect the reference spectrum in an open chamber of the NIR apparatus at 24°C. Sample (5 g) spectra were collected at every 2.0 nm over the 400 - 2500 nm wavelength region. Each sample was measured once by the apparatus. The reference calibration was completed at every ten pepper samples. The absorbance data was exported into several pre-processed spectral formats using the in-house software, including original, first and second derivative, standard normal variate (SNV), and

multiplicative scatter correction (MSC).

NIR model development, calibration, and validation

The partial least squares regression (PLSR) algorithm built into the JMP v.10.0 software (SAS Institute Inc., Cary, NC, USA) was used for model calibration and validation using the original and pre-processed spectral and trait data. PLSR predicts each trait vector Y from the spectral matrix X by modelling the shared structure between the two, and extracting a group of orthogonal latent variables (LV) (Nicolai *et al.*, 2007; Abdi, 2010).

The spectral and trait data were randomly divided into calibration and internal validation sets of 100 and 80 samples, respectively. The variable importance in projection (VIP) score, a weighted sum of squares of the PLSR-weights, was calculated using the JMP multivariate toolbox, which summarised the contribution a variable (wavelength) made to the calibration model (Eriksson *et al.*, 2001; Nordey *et al.*, 2017). Variables with a VIP score lower than 0.8 were removed from the PLS model (Eriksson *et al.*, 2001). The total number of variables (or effective wavelengths) was optimised into PLS factors (max fixed as 15) by the minimum root mean predicted residual sum of squares (PRESS)-statistic with a leave-one-out cross-validation method (Wold, 1994; Maniwaru *et al.*, 2014). These factors or LVs were employed by the regression algorithm to model the relationship between the X matrix and Y vectors.

The default PLSR statistics for evaluating model performance were calculated, including correlation coefficients of calibration (r_{cal}) and validation (r_{val}), and the root mean square errors of cross-validation (RMSECV) and prediction (RMSEP) (Wold *et al.*, 2001; Nicolai *et al.*, 2007). Models with the highest correlation coefficients and the lowest errors were selected as optimal for ASTA-colour and SHU-pungency traits. We used the optimal calibration model for each trait to predict the Y values of another 79 pepper samples (external validation set). Each model was assessed based on the correlation coefficient between the predicted values and measured values (r_{pre}), the root mean square error of external prediction (RMSEEP), and the ratio of performance to deviation, known as the residual predictive deviation value (RPD) (Magwaza *et al.*, 2012; Olarewaju *et al.*, 2019). Predictive and robust models were denoted by the highest prediction correlations and RPD values, and the lowest error values.

Results and discussion

Biochemical phenotypic data analysis

Pepper samples amongst initial calibration and validation datasets expressed phenotypic variation for ASTA-colour and SHU-pungency traits, thus indicating that these were good populations for NIR model development (Table 1). The mean biochemical data of each trait were significantly different among the three datasets (p -value < 0.0001). Additionally, the pooled data of each trait were normally distributed as illustrated in the distribution diagram. However, ASTA-colour data was more normal relative to SHU-pungency, as more accessions than expected from the lower pungency tail were selected. When the data were re-calibrated (see next section), a greater distribution was observed for both traits, thus suggesting statistically reliable data for the second calibration. Sánchez *et al.* (2013) and Li *et al.* (2017) advised quality variation or a great standard distribution (σ) of samples as important criteria to produce robust, reliable, and reproducible prediction models.

NIR spectral data characteristics and pruning windows

Original NIR spectra of samples for VIS and NIR wavelengths are presented in Figure 1. The spectral data exhibited minimal spectral noise, uniformity in the NIR absorption patterns, and clear sample-specific variation as indicated by colour differences. As the NIR equipment collected data from visible and short-wave radiation (400 - 1098 nm) to long-wave radiation wavelengths (1100 - 2498 nm) for each sample, the instrument detector transitioned from silicon to lead sulphide, thus causing an abrupt depression in the NIR spectra. For further model development, the spectral data was divided into three categories, visible and short-wave radiation (Vis-SWNIR, 400 - 1098 nm), short-wave radiation (SWNIR, 800 - 1098 nm), and long-wave radiation (LWNIR, 1100 - 2498 nm) wavelength regions.

Calibration, validation, and prediction of ASTA-colour and SHU-pungency

The ASTA-colour and SHU-pungency calibration models were independently developed using spectral data from each spectral category (Vis-SWNIR, SWNIR, and LWNIR), as well as each pre-processing treatment (original, first and second

derivatives, MSC, and SNV). The ASTA-colour model calibrations and predictions resulted in a low-high agreement between NIR predicted values and actual values (correlation coefficients (r_s) of 0.10 - 0.90). The most optimal model with the highest correlation and lowest error was achieved using Vis-SWNIR spectral data and first derivative pre-processing treatment with r_{cal} and r_{val} values of 0.909 and 0.924, as well as RMSECV and RMSEP values of 9.43 and 8.49, respectively (Table 2, bold). Although predictions of 79 external validation samples using this optimal model resulted in high prediction correlation (r_{pre} value of 0.890), a low number of extracted factors with the PRESS analysis (five PLS factors) and low prediction error (RMSEEP value of 16.26), the RPD value was only 2.18, thus indicating good approximate robustness and moderate reliability (Jiang *et al.*, 2016).

To further increase the model's robustness of the Vis-SWNIR first derivative optimal PLSR for future predictions, the regression was calibrated again by adding 40 samples from the external prediction dataset to the former 100 samples, and leaving 39 samples to validate the model together with the former 80 samples (Table 1). Consequently, the new calibration set consisted of 140 samples, and a new validation set consisted of 119 samples. Although the re-calibrated model elicited a higher PLS factor (nine factors) than the previous model, it provided high correlations of calibration and validation (r_{cal} and r_{val} values of 0.955 and 0.928, respectively) as well as low errors (RMSECV value of 10.08 and RMSEP value of 11.31), with a high RPD value of 2.84 (Figure 2A and 2B). This model emerged to be more predictive, robust, and reliable for ASTA-colour prediction, as now the calibration dataset was larger and the calibration phenotypic range was wider and more distributed. In the literature, colour and its parameters have also been successfully identified or predicted with high correlations in carrot and tomato ($r = 0.80$ and $0.86 - 0.91$) by Beghi *et al.* (2018), in Spanish paprika ($r = 0.92 - 0.95$) by Palacios-Morillo *et al.* (2016), and in rice ($r = 0.86 - 0.94$) by Saleh *et al.* (2008).

The SHU-pungency trait calibration models using the original calibration strategy of calibration with $n = 100$, internal validation with $n = 80$, and external prediction with $n = 79$, resulted in high r_{cal} values (0.934 - 0.958) and moderate r_{val} values (0.740 - 0.856). The most optimal model was achieved using LWNIR and first derivative pre-processed data with

Table 1. Descriptive statistics of traits, American Spice Trade Association (ASTA)-colour, and the total Scoville Heat Units (SHU)-pungency

Trait	Statistics	Calibration <i>n</i> = 100	Internal validation <i>n</i> = 80	External validation <i>n</i> = 79	Refined calibration <i>n</i> = 140	Internal validation <i>n</i> = 119	
	Mean	46.5	60.0	102.7	60.8	76.0	
	Max	100.2	103.8	221.1	221.1	171.2	
	Min	2.1	26.4	47.6	2.1	26.4	
	Range	98.1	77.4	173.5	219.0	144.8	
	SD	22.5	15.6	35.4	35.6	32.2	
ASTA colour	Distribution diagram						
	Mean	45549.0	61275.0	39057.0	45141.4	30833.7	
	Max	145300.0	134700.0	111000.0	145300.0	134700.0	
	Min	900.0	3800.0	400.0	900.0	400.0	
	Range	144400.0	130900.0	110600.0	144400.0	134300.0	
	SD	31667.0	29469.6	27859.4	31039.8	30833.7	
SHU pungency	Distribution diagram						

Min: minimum; Max: maximum; *n*: sample size; and SD: standard deviation.

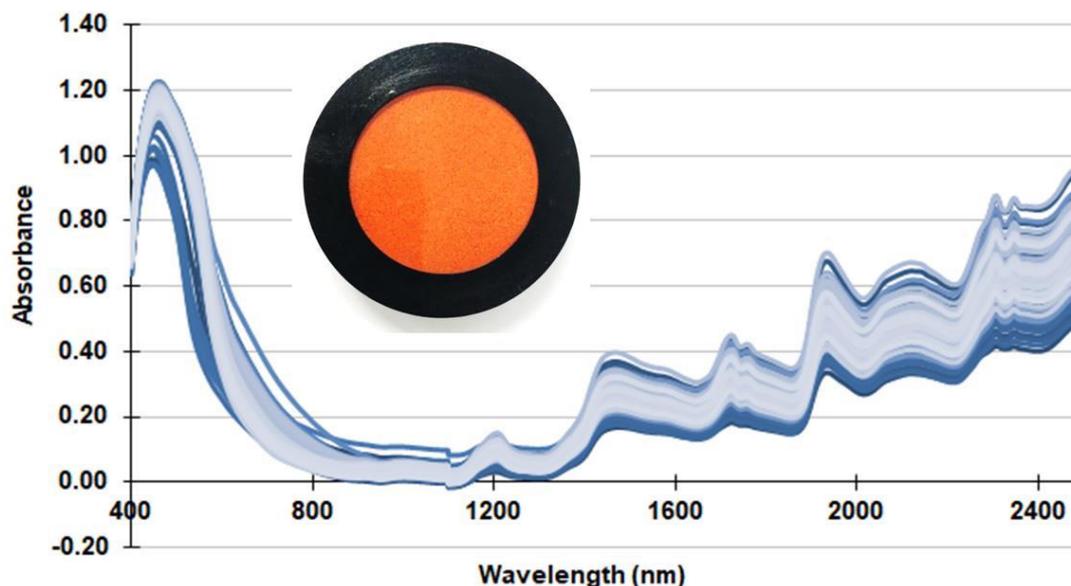


Figure 1. Original absorbance spectra of dry pepper powder with spectral data from visible and near infrared regions (400 - 2500 nm). Photo insert is the rotating cup used for scanning of dry pepper powder.

Table 2. Calibration, validation, and prediction results of American Spice Trade Association (ASTA)-colour trait in pepper with partial least squares (PLS) regression models using spectroscopy data from categories of visible and short-wave radiation (Vis-SWNIR), short-wave (SWNIR), and long-wave (LWNIR) radiation. Pre-processing treatments used were first and second derivatives, multiplicative scatter correction (MSC), and standard normal variate (SNV).

Trait	Spectral type/ Pre-processing	Calibration (<i>n</i> = 100)			Validation (<i>n</i> = 80)		Prediction (<i>n</i> = 79)		
		PLS factors (<i>f</i>)	Effective wavelengths (per total)	r_{cal}	RMSECV	r_{val}	RMSEP	r_{pre}	RMSEEP
Vis-SWNIR (400 - 1098 nm)									
	Original	7	166/350	0.912	8.46	0.915	9.47	0.898	15.66
	1st derivative	5	170/340	0.909	9.43	0.924	8.49	0.890	16.26
	2 nd derivative	2	141/340	0.897	9.99	0.893	9.05	0.875	17.26
	MSC	5	194/350	0.888	9.23	0.866	10.27	0.806	21.08
	SNV	5	211/350	0.889	9.20	0.862	10.39	0.784	22.08
SWNIR (800 - 1098 nm)									
ASTA colour	Original	7	99/140	0.706	11.31	0.259	11.99	0.291	13.27
	1 st derivative	6	87/140	0.666	11.24	0.223	11.76	0.315	12.02
	2 nd derivative	4	109/140	0.758	11.18	0.155	11.51	0.288	11.70
	MSC	4	63/150	0.529	10.16	0.357	9.93	0.261	34.38
	SNV	8	70/150	0.802	10.84	0.326	17.41	0.102	35.33
LWNIR (1100 - 2498 nm)									
	Original	15	556 / 700	0.897	8.96	0.512	11.76	0.444	18.33
	1 st derivative	14	485 / 690	0.928	7.81	0.690	9.94	0.647	13.90
	2 nd derivative	9	507 / 690	0.954	6.50	0.536	9.64	0.656	13.79
	MSC	15	603 / 700	0.922	8.09	0.627	10.26	0.255	388.90
	SNV	15	613 / 700	0.919	8.20	0.589	10.56	0.322	65.35

n: sample size; r_{cal} : correlation coefficient of calibration; RMSECV: root mean square error of cross-validation; r_{val} : correlation coefficient of validation; RMSEP: root mean square error of prediction; r_{pre} : correlation coefficient of prediction; and RMSEEP: root mean square error of external prediction.

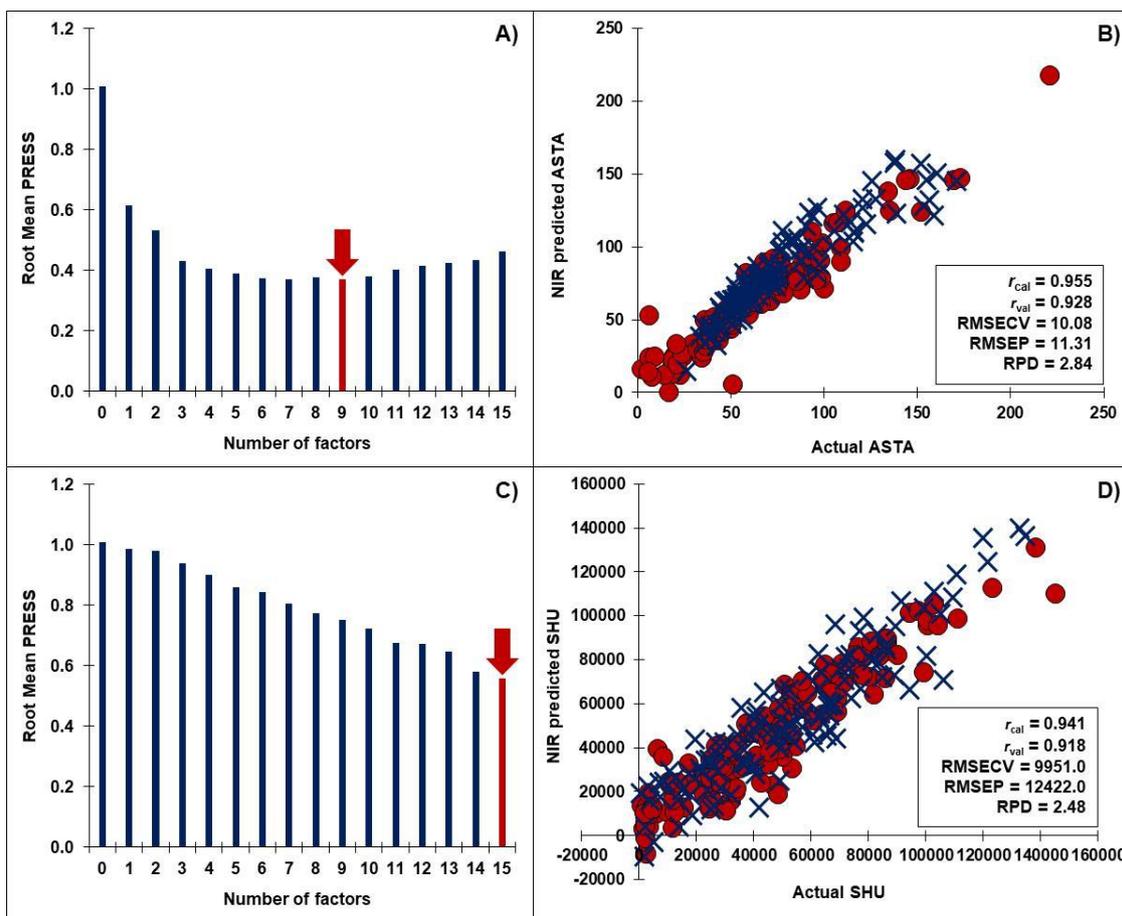


Figure 2. Root mean predicted residual sum of squares (PRESS)-statistic of refined calibrations for (A) American Spice Trade Association (ASTA)-colour, and (C) SHU-pungency traits together with their scatter plots of calibrations (marked as rounded symbol; $n = 140$) and predictions (marked as crossed symbol; $n = 119$) of (B) ASTA-colour and (D) SHU-pungency.

a high RPD value of 2.62 (Table 3, bold). However, low agreement (r of 0.487) between the NIR predicted and actual value was observed when the model was applied to the external prediction. The evident difference in the phenotypic ranges across the calibration and prediction datasets, as well as the negatively skewed distribution of the SHU-pungency trait, could have contributed to this pitfall (Table 1). The least pungent sample had a SHU-pungency of 400 in the external validation set, whereas 900 and 3,800 were the lowest pungency values in the calibration and internal validation datasets, respectively. The predicted range of SHU-pungency values was outside the range of values used for calibrating the models. For instance, the frequency of samples with high pungency was far greater in the external validation relative to the calibration set, making possible distinct NIR spectral characters that were not well fitted to the PLS regression.

The LWNIR first derivative SHU-pungency model was calibrated again similar to ASTA-colour by increasing the number of calibration samples to

140, and combining the internal and external validation samples to 119. The re-calibrated model provided high correlations of calibration and prediction (r_{cal} and r_{val} values of 0.941 and 0.918, respectively), and an RPD value of 2.48. Although this model used a high number of extracted factors (PLS factors of 15), it was highly predictive and robust (Figure 2C and 2D). The germplasm was broadly diverse for pungency with 145,300 and 900 as the maximum and minimum calibration pungency values, respectively, and 134,700 and 400 for validation, which could have resulted in requiring more PLS factors for the prediction model. In the future, this model can be further explored and assessed for robustness using an appropriate independent prediction set. In the literatures, predictive models for pungency related compounds have also been successfully developed in chili or pepper powder by Bonifazi *et al.* (2019) ($r = 0.98$), Jiang *et al.* (2018) ($r = 0.80 - 0.83$), Rahman *et al.* (2018) ($r = 0.68 - 0.88$), and Mo *et al.* (2013) ($r = 0.97 - 0.99$).

Table 3. Calibration, validation, and prediction results of Scoville Heat Unit (SHU)-pungency trait in pepper with partial least squares (PLS) regression models using spectroscopy data from categories of visible and short-wave radiation (Vis-SWNIR), short-wave (SWNIR), and long-wave (LWNIR) radiation. Pre-processing treatments used were first and second derivatives, multiplicative scatter correction (MSC), and standard normal variate (SNV).

Trait	Spectral type/ Pre-processing	Calibration (<i>n</i> = 100)			Validation (<i>n</i> = 80)		Prediction (<i>n</i> = 79)			
		PLS factors (<i>f</i>)	Effective wavelengths (per total)	<i>r</i> _{cal}	RMSECV	<i>r</i> _{val}	RMSEP	<i>R</i> _{pre}	RMSEEP	
Vis-SWNIR (400 - 1098 nm)										
SHU pungency	Original	8	204/350	0.626	15534	0.318	15211	0.132	12477	
	1 st derivative	5	211/340	0.605	15328	0.456	13856	0.120	11499	
	2 nd derivative	3	212/340	0.630	15575	0.282	14924	0.077	10609	
	MSC	6	219/350	0.514	14042	0.080	12790	0.187	75623	
	SNV	7	209/350	0.586	15113	0.121	15594	0.173	1e+5	
	SWNIR (800 - 1098 nm)									
	Original	9	118/150	0.759	15731	0.129	18413	0.112	20001	
	1 st derivative	6	91/140	0.643	15670	0.122	17360	0.077	16602	
	2 nd derivative	3	79/140	0.558	14734	0.292	13123	- 0.029	8503	
	MSC	4	62/150	0.377	11109	0.362	11084	0.107	5e+5	
	SNV	6	74/150	0.551	26568	0.041	18880	0.098	49812	
	LWNIR (1100 - 2498 nm)									
	Original	15	533/700	0.889	12937	0.663	17025	0.370	39783	
	1st derivative	15	422/690	0.958	9098	0.856	11263	0.487	33797	
	2 nd derivative	8	420/690	0.947	9679	0.779	14487	0.515	24919	
MSC	15	550/700	0.934	10593	0.740	14610	0.187	6e+5		
SNV	15	531/700	0.934	10590	0.747	14424	0.183	5e+5		

n: sample size; *r*_{cal}: correlation coefficient of calibration; RMSECV: root mean square error of cross-validation; *r*_{val}: correlation coefficient of validation; RMSEP: root mean square error of prediction; *r*_{pre}: correlation coefficient of prediction; and RMSEEP: root mean square error of external prediction.

Variable importance and effective wavelengths in re-calibrated PLS models

For NIR calibration model development, only the wavelengths with a VIP score greater than 0.8 (significance threshold) reflected the number of effective wavelengths, as practiced in chemometrics application (Wold, 1994; Eriksson *et al.*, 2001; Nordey *et al.*, 2017). The most predictive models for ASTA-colour relied on most of the Vis-SWNIR spectroscopy data, whereas the LWNIR spectroscopy data was important for the SHU-pungency trait. Figure 3 shows diagrams of the VIP scores for both traits using the wavelengths of their respective spectroscopy data category. The wavelength region between 400 and 670 nm was primarily important for the ASTA-colour trait, in addition to several other

wavelengths in the SWNIR region, totalling 172 effective wavelengths (Table 2, Figure 3A). For the optimal ASTA-colour trait prediction model, nine PLS factors were extracted based on PRESS analysis to prevent possible overfitting. The red colour of pepper is imparted by carotenoids (*i.e.*, pigments) with more than 50 identified structures, consistent with the signature wavebands observed in the VIS region for the ASTA-colour trait (Arimboor *et al.*, 2015; Kaur *et al.*, 2020). These wavelength regions were also central for cultivar discrimination in Cen *et al.* (2007) and Shao *et al.* (2009) for orange, and in Cao *et al.* (2010) for grape.

There was no particular wavelength region critical for SHU-pungency calibration as noted by the random spikes within the LWNIR wavelengths in the

VIP score plot (Figure 3B). Over 60% of the wavelengths were important for the optimal SHU-pungency trait prediction model. The complexity of variables required 15 extracted factors for producing reliable PLS regression (Figure 2C). Several wavebands signalled common organic compounds were observed in SHU-pungency VIP score plots. The scores showed great peaks due to water with first

overtone, and combination bands of the OH-bonds at 1450 and 1940 nm, respectively, as well as combination bands of C-H bonds at 1206 nm and between 2000 and 2500 nm (Nicolai *et al.*, 2007; Wang *et al.*, 2015; Currà *et al.*, 2019). The wavelength peaks at 1722, 1746 - 1758, and 2148 - 2200 nm could be associated with oil and protein, respectively (Manley, 2014).

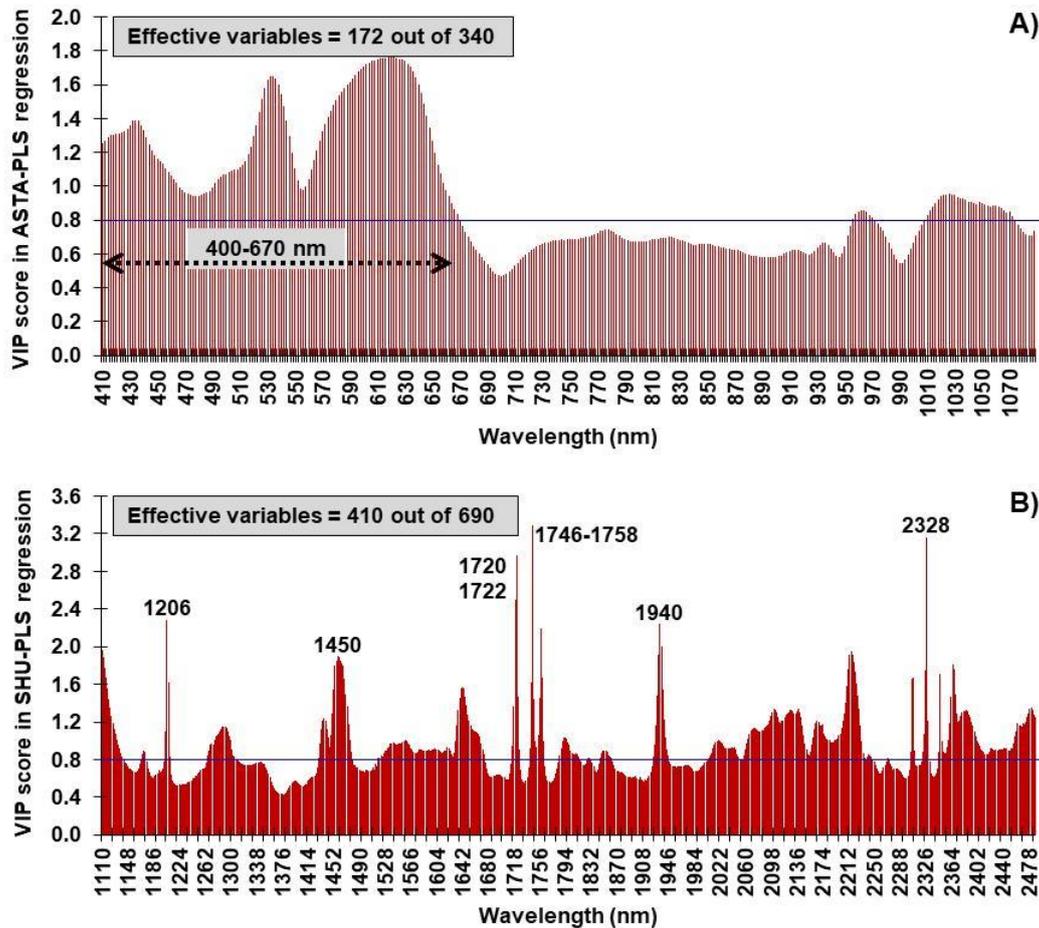


Figure 3. Variable importance in the projection (VIP) score for the refined PLS regressions of the (A) American Spice Trade Association (ASTA)-colour and (B) Scoville Heat Unit (SHU)-pungency traits. Wavelengths with VIP scores greater than 0.8 were effective for model development as indicated by the horizontal line. Wavelengths of visible and short-wave radiation (VIS-SWNIR) for ASTA-colour, and long-wave radiation (LWNIR) for SHU-pungency were important for the most optimal models.

Conclusion

The present work investigated the ability of NIR spectroscopy with multivariate statistics to predict the high-throughput phenotype of the biochemical quality of dry pepper of India. We successfully predicted ASTA-colour and SHU-pungency traits using robust and reliable trait-specific PLS prediction models. The visible and short-wave radiation spectroscopy with first derivative pre-processing treatment provided the most predictive

model for ASTA-colour. The long-wave radiation spectroscopy with first derivative pre-processing treatment provided the most predictive model for SHU-pungency. Variable importance in projection scores was useful for calibration of both traits as they significantly reduced the number of effective wavelengths. Quality distribution, the range of diversity within the germplasm, and trait phenotypic variation were influential factors for model performance. For future routine screening implementation in a seed company, ASTA-colour and

SHU-pungency PLS models should continue to be calibrated with larger calibrations ($n > 1000$) when tested across multiple locations, seasons, years, and other factors relevant to the breeding program.

Acknowledgement

The authors thank the Postharvest Technology Innovation Center, Ministry of Higher Education, Science, Research, and Innovation for providing the NIRS instrument and expertise on NIR data collection. The authors also thank Mr. Gopal Kulkarni and his field breeding team, and Dr. Gajanan Salunke and his biochemistry laboratory team for assistance with germplasm curation, field trials, pepper processing, and biochemical data collection. The authors also thank the EWS-Thailand administration and research teams, including Drs. Conrado Balatero, Darush Struss, Simon de Hoop, Yuenyad Teerawatsakul, and Ms. Youngin Kim for project inception and coordination.

References

- Abdi, H. 2010. Partial least squares regression and projection on latent structure regression (PLS regression). *Wiley Interdisciplinary Reviews - Computational Statistics* 2: 97-106.
- American Spice Trade Association (ASTA). 1986. Official analytical method of the ASTA - analytical method 20.1. United States: ASTA.
- Araus, J. L. and Cairns, J. E. 2014. Field high-throughput phenotyping: The new crop breeding frontier. *Trends in Plant Science* 19: 52-61.
- Arendse, E., Fawole, O. A., Magwaza, L. S. and Opara, U. L. 2018. Non-destructive prediction of internal and external quality attributes of fruit with thick rind: A review. *Journal of Food Engineering* 217: 11-23.
- Arimboor, R., Natarajan, R. B., Menon, K. R., Chandrasekhar, L. P. and Moorkoth, V. 2015. Red pepper (*Capsicum annuum*) carotenoids as a source of natural food colors: Analysis and stability - A review. *Journal of Food Science and Technology* 52: 1258-1271.
- Beghi, R., Giovenzana, V., Tugnolo, A. and Guidetti, R. 2018. Application of visible/near infrared spectroscopy to quality control of fresh fruits and vegetables in large-scale mass distribution channels: A preliminary test of carrots and tomatoes. *Journal of the Science of Food and Agriculture* 98: 2729-2734.
- Bonifazi, G., Gasbarrone, R. and Serranti, S. 2019. Dried red chili peppers pungency assessment by visible and near infrared spectroscopy. In *Proceedings of SPIE-Algorithms, Technologies and Applications for Multispectral and Hyperspectral Imagery*. Maryland, United States.
- Cao, F., Wu, D. and He, Y. 2010. Soluble solids content and pH prediction and varieties discrimination of grapes based on visible-near infrared spectroscopy. *Computers and Electronics in Agriculture* 71: S15-S18.
- Cen, H., He, Y. and Huang, M. 2007. Combination and comparison of multivariate analysis for the identification of orange varieties using visible and near infrared reflectance spectroscopy. *European Food Research and Technology* 225: 699-705.
- Cheng, X., Chen, Y. R., Tao, Y., Wang, C. Y., Kim, M. S. and Lefcourt, A. M. 2004. A novel integrated PCA and FLD method on hyperspectral image feature extraction for cucumber chilling damage inspection. *Transactions of the ASAE* 47: 1313-1320.
- Currà, A., Gasbarrone, R., Cardillo, A., Trompetto, C., Fattapposta, F., Pierelli, F., ... and Serranti, S. 2019. Near-infrared spectroscopy as a tool for *in vivo* analysis of human muscles. *Scientific Reports* 9: 8623.
- Donis-González, I. R., Valero, C., Momin, M. A., Kaur, A. and Slaughter, D. C. 2020. Performance evaluation of two commercially available portable spectrometers to non-invasively determine table grape and peach quality attributes. *Agronomy* 10: 1-16.
- Ergüneş, G. and Tarhan, S. 2006. Color retention of red peppers by chemical pretreatments during greenhouse and open sun drying. *Journal of Food Engineering* 76: 446-452.
- Eriksson, L., Johansson, E., Kettaneh-Wold, N. and Wold, S. 2001. Multi- and megavariate data analysis. Principles and applications. Sweden: Umetrics Academy.
- Furbank, R. T. and Tester, M. 2011. Phenomics - Technologies to relieve the phenotyping bottleneck. *Trends in Plant Science* 16: 635-644.
- Guzman, I. and Bosland, P. W. 2017. Sensory properties of chile pepper heat and its

- importance to food quality and cultural preference. *Appetite* 117: 186-190.
- Jiang, J., Cen, H., Zhang, C., Lyu, X., Weng, H., Xu, H. and He, Y. 2018. Nondestructive quality assessment of chili peppers using near-infrared hyperspectral imaging combined with multivariate analysis. *Postharvest Biology and Technology* 146: 147-154.
- Jiang, Q., Chen, Y., Guo, L., Fei, T. and Qi, K. 2016. Estimating soil organic carbon of cropland soil at different levels of soil moisture using VIS-NIR spectroscopy. *Remote Sensing* 8: 755.
- Kaur, A., Donis-González, I. R. and St. Clair, D. A. 2020. Evaluation of a visible and near-infrared hand-held spectrophotometer as an in-field high-throughput phenotyping tool for tomato and pepper fruit quality trait breeding. *The Plant Phenome Journal* 3: e20008
- Kim, S., Ha, T. Y. and Park, J. 2008. Characteristics of pigment composition and colour value by the difference of harvesting times in Korean red pepper varieties (*Capsicum annuum*, L.). *International Journal of Food Science* 43: 915-920.
- Li, M., Lv, W., Zhao, R., Guo, H., Liu, J. and Han, D. 2017. Non-destructive assessment of quality parameters in 'Friar' plums during low temperature storage using visible/near infrared spectroscopy. *Food Control* 73: 1334-1341.
- Magwaza, L. S., Opara, U. L., Nieuwoudt, H., Cronje, P. J. R., Saeys, W. and Nicolai, B. 2012. NIR spectroscopy applications for internal and external quality analysis of citrus fruit - A review. *Food and Bioprocess Technology* 5: 425-444.
- Maniwaru, P., Nakano, K., Boonyakiat, D., Ohashi, S., Hiroi, M. and Tohyama, T. 2014. The use of visible and near infrared spectroscopy for evaluating passion fruit postharvest quality. *Journal of Food Engineering* 143: 33-43.
- Maniwaru, P., Nakano, K., Ohashi, S., Boonyakiat, D., Sehanam, P., Theanjumol, P. and Poonlarp, P. 2019. Evaluation of NIRS as non-destructive test to evaluate quality traits of purple passion fruit. *Scientia Horticulturae* 257: 108712.
- Manley, M. 2014. Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. *Chemical Society Reviews* 43: 8200-8214.
- Mo, C., Hasegawa, M., Lee, K., Lim, J. G., Kim, M. S., Kang, S., ... and Cho, B. K. 2013. Development of a non-destructive on-line pungency measurement system for red pepper powder. *Journal of the Faculty of Agriculture, Kyushu University* 58: 137-144.
- Nicolai, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I. and Lammertyn, J. 2007. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology* 46: 99-118.
- Nordey, T., Joas, J., Davrieux, F., Chillet, M. and Léchaudel, M. 2017. Robust NIRS models for non-destructive prediction of mango internal quality. *Scientia Horticulturae* 216: 51-57.
- Olarewaju, O. O., Magwaza, L. S., Nieuwoudt, H., Poblete-Echeverría, C., Fawole, O. A., Tesfay, S. Z. and Opara, U. L. 2019. Model development for non-destructive determination of rind biochemical properties of 'Marsh' grapefruit using visible to near-infrared spectroscopy and chemometrics. *Spectrochimica Acta Part A* 209: 62-69.
- Othman, Z. A. A., Ahmed, Y. B. H., Habila, M. A. and Ghafar, A. A. 2011. Determination of capsaicin and dihydrocapsaicin in *Capsicum* fruit samples using high performance liquid chromatography. *Molecules* 16: 8919-8929.
- Palacios-Morillo, A., Jurado, J. M., Alcazar, A. and Pablos, F. 2016. Differentiation of Spanish paprika from protected designation of origin based on color measurements and pattern. *Food Control* 62: 243-249.
- Pissard, A., Fernández Pierna, J. A., Baeten, V., Sinnaeve, G., Lognay, G., Mouteau, A., ... and Lateur, M. 2013. Non-destructive measurement of vitamin C, total polyphenol and sugar content in apples using near-infrared spectroscopy. *Journal of the Science of Food and Agriculture* 93: 238-244.
- Rahman, A., Lee, H., Kim, M. S. and Cho, B. K. 2018. Mapping the pungency of green pepper using hyperspectral imaging. *Food Analytical Methods* 11: 3042-3052.
- Rebetzke, G. J., Berni, J. J., Fischer, R. A., Deery, D. M. and Smith, D. J. 2019. Review: High throughput phenotyping to enhance the use of crop genetic resources. *Journal of Plant Sciences* 282: 1-9.

- Reddy, M. K., Kumar, S., Kumar, R., Srivastava, A., Chawda, N., Ebert, A. W. and Vishwakarma, M. 2014. Chilli (*Capsicum annuum* L.) breeding in India: An overview. SABRAO Journal of Breeding and Genetics 46: 160-173.
- Reddy, M. K., Srivastava, A., Lin, S. W., Kumar, R., Shieh, H. C., Ebert, A. W., ... and Kumar, S. 2015. Exploitation of AVRDC's chili pepper (*Capsicum* spp.) germplasm in India. Journal of the Taiwan Society for Horticultural Science 61: 1-9.
- Rinnan, Å., Berg, F. V. D. and Engelsen, S. B. 2009. Review of the most common pre-processing techniques for near-infrared spectra. Trends in Analytical Chemistry 28: 1201-1222.
- Saleh, M. I., Meullenet, J. F. and Siebenmorgen, T. J. 2008. Development and validation of prediction models for rice surface lipid content and color parameters using near-infrared spectroscopy: A basis for predicting rice degree of milling. Cereal Chemistry 86: 787-791.
- Sanatombi, K. and Sharma, G. J. 2008. Capsaicin content and pungency of different *Capsicum* spp. cultivars. Notulae Botanicae Horti Agrobotanici Cluj-Napoca 36: 89-90.
- Sánchez, M. T., De la Haba, M. J., Benítez-López, M., Fernández-Novales, J., Garrido-Varo, A. and Pérez-Marín, D. 2012. Non-destructive characterization and quality control of intact strawberries based on NIR spectral data. Journal of Food Engineering 110: 102-108.
- Sánchez, M. T., De la Haba, M. J., Serrano, I. and Pérez-Marín, D. 2013. Application of NIRS for nondestructive measurement of quality parameters in intact oranges during on-tree ripening and at harvest. Food Analytical Methods 6: 826-837.
- Shao, Y., He, Y., Bao, Y. and Mao, J. 2009. Near-infrared spectroscopy for classification of orange and prediction of the sugar content. International Journal of Food Properties 12: 644-658.
- Srinivasan, K. 2016. Biological activities of red pepper (*Capsicum annuum*) and its pungent principle capsaicin: A review. Critical Reviews in Food Science and Nutrition 56: 1488-1500.
- Sun, T., Xu, W., Lin, J., Liu, M. and He, X. 2012. Determination of soluble solids content in navel oranges by Vis/NIR diffuse transmission spectra combined with CARS method. Spectroscopy and Spectral Analysis 32: 3229-3233.
- Todd, P., Bensinger, M. and Biftu, T. 1977. Determination of pungency due to *Capsicum* by gas-liquid chromatography. Journal of Food Science 42: 660-665.
- Wang, H., Peng, J., Xie, C., Bao, Y. and He, Y. 2015. Fruit quality evaluation using spectroscopy technology: A review. Sensors 15:11889-11927.
- White, J. W., Andrade-Sanchez, P., Gore, M. A., Bronson, K. F., Coffelt, T. A., Conley, M. M., ... and Wang, G. 2012. Field-based phenomics for plant genetics research. Field Crops Research 133: 101-112.
- Wold, S. 1994. PLS for multivariate linear modeling. In van de Waterbeemd, H. (ed). QSAR: Chemometric Methods in Molecular Design, Methods and Principles in Medicinal Chemistry, p. 195-218. Weinheim: Verlag Chemie.
- Wold, S., Sjostrom, M. and Eriksson, L. 2001. PLS-regression: A basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58: 109-130.