Quality Assurance and Safety of Crops & Foods, 2023; 15(1): 139–151

Near-infrared spectroscopy of Chinese soy sauce for quality evaluation

Xiaoqian Chen[1], Chuanwei Li[1], Xiaofang Liu[2], Yu Su[1], Ziang Sun[1], Lei Yuan[1], Shuo Wang[1]*

[1]*College of Food Science and Engineering, Yangzhou University, Yangzhou, China;* [2]*School of Tourism and Cuisine, Yangzhou University, Yangzhou, China*

***Corresponding Author:** Shuo Wang, College of Food Science and Engineering, Yangzhou University, Yangzhou, Jiangsu 225127, China. Email: wangsh@yzu.edu.cn

Received: 20 August 2022; Accepted: 19 December 2022; Published: 17 January 2023

RESEARCH ARTICLE

Abstract

The feasibility of near-infrared (NIR) spectroscopy and partial least-squares regression (PLSR) was investigated for rapid prediction of the quality of Chinese soy sauce. Twenty-four soy sauce samples from eight common brands available in China were analyzed for the contents of various components that may affect the quality of soy sauce. Sensory evaluation was also conducted to determine the relationship between components and the sensory quality of soy sauce. Subsequently, NIR spectra (400–2500 nm) of the samples were obtained, and the raw spectra were subjected to different pretreatment methods. PLSR was performed on the raw and treated spectra to construct models using a calibration set. The performance of models was evaluated by comparing the determination coefficient of prediction (R^2_p) and root-mean-square error of prediction (RMSEP). The results showed that the models constructed using the moisture content (R^2_p of 0.825 and RMSEP of 1.73), amino acid nitrogen content (R^2_p of 0.785 and RMSEP of 0.071), and taste scores (R^2_p of 0.733 and RMSEP of 11.93) performed well, and the interactions between amino acid nitrogen content and taste of soy sauce were clarified. This study demonstrates that NIR spectroscopy can be used as a valid alternative method for rapid prediction of the sensory quality of soy sauce during processing.

Keywords: near-infrared (NIR) spectroscopy; partial least-squares regression (PLSR); rapid quality prediction; sensory evaluation; soy sauce

Introduction

Soy sauce is a condiment originating in China that is made by fermentation of soybeans. It is widely used in the East and Southeast Asian cuisines. China is a major producer and consumer of soy sauce. The current global annual production volume of soy sauce is approximately 8 million tons, and China produces approximately 5 million tons of this product (Shurtleff and Aoyagi, 2012). Compared with other countries, China has a wider variety of types of soy sauce, including braised soy sauce, steamed soy sauce, seafood soy sauce, and shrimp seed soy sauce. With improvements in quality of life, consumers are increasingly focusing on the sensory qualities of food (Jürkenbeck and Spiller, 2021). Consequently, the sensory characteristics of soy sauce have become important for determining its quality. Sensory evaluation is an important tool that can be used to explore consumer preferences and market demand (Stone *et al.*, 2020).

Sensory evaluation of food products involves the use of human sensory organs to assess their various quality characteristics and provides comparative descriptions (Sarkar *et al.*, 2022b; Steinhaus and Schieberle, 2007). However, this process is time-consuming and costly, and the results are highly subjective. Sensory evaluation can also be influenced by environmental factors (Xu *et al.* 2013), and it cannot be used for rapid testing of product sensory quality (Sarkar *et al.*, 2022a). Therefore, establishment of an objective, effective, and rapid quality evaluation model is currently considered to be the most critical aspect in quality testing of soy sauce.

Near-infrared (NIR) spectroscopy is a convenient, efficient, and low-cost analytical technique that has undergone rapid development in the recent years (Zareef *et al.*, 2020). It combines spectroscopic measurement analysis with chemometrics and is rapid, nonpolluting, nondestructive, and capable of determining multiple components simultaneously (Bázár *et al.*, 2016; Li *et al.*, 2007, 2020). Currently, NIR spectroscopy is widely used in food testing (Cortes *et al.*, 2017; Escribano *et al.*, 2017; Jamshidi *et al.*, 2016; Lan *et al.*, 2020; Lorenzo *et al.*, 2009), such as identification of chestnut varieties (Corona *et al.*, 2021), prediction of the moisture content in roasted pistachio kernels (Mohammadi-Moghaddam *et al.*, 2018), and evaluation of the sensory properties of wines (Cayuela *et al.*, 2017). Wang et al. (2021) investigated the feasibility of using NIR spectroscopy and partial least-squares regression (PLSR) to evaluate the quality of Japanese fermented soy sauce. They found that NIR spectroscopy could be used as an alternative to conventional methods for soy sauce quality control, and could quickly and economically grade soy sauce products. However, the brewing of Japanese soy sauce is different from that of Chinese soy sauce, and there is only one main variety (*Koikuchi* soy sauce) that differs significantly from Chinese soy sauce in its taste and texture (Diez-Simon *et al.*, 2020). Consequently, the feasibility of applying NIR spectroscopy to sensory evaluation of Chinese soy sauce is unclear.

The aim of this study was to explore the feasibility of using NIR spectroscopy for rapid and objective sensory evaluation of Chinese soy sauce. The relationship between the sensory characteristics of Chinese soy sauce and their quality was first investigated using 24 samples from common soy sauce brands. The results were believed to provide scientific data for identifying key components of soy sauce and to improve their appearance and taste. Subsequently, the correlation between the components, sensory scores, and NIR spectrum were clarified by performing different pretreatment methods on the spectral data. Overall, NIR spectroscopy was promising to rapidly assess the sensory evaluation of Chinese soy sauce.

Experimental

Sample preparation

Twenty-four bottles of eight commercially available soy sauce brands were obtained. The bottles were randomly grouped into a calibration set (n = 18) and a validation set (n = 6). The calibration set was used to construct a quantitative calibration model, and the validation set was used to predict the model accuracy. The soy sauce samples were divided into two groups, A and B, according to their sensory scores. The 12 samples with higher sensory scores were placed in Group A, and the 12 samples with lower sensory scores were placed in Group B. The groups were used to compare the characteristics between different soy sauces.

Analysis of physiochemical indicators

The color, salt content, Brix value, pH, moisture content, and amino acid nitrogen content were analyzed in triplicates using the previously described method (Wang *et al.*, 2018). Briefly, the color of each soy sauce sample was measured using a spectrophotometer (CM-5, Konica Minolta, Tokyo, Japan); the analyses of salt content and Brix values were carried out by using a conductivity salinometer (PAL-SALT, Atago Corporation, Tokyo, Japan) and a digital saccharimeter (PAL-1, Atago Co), respectively; the pH of the soy sauce was measured using a pH meter (Remag PHS-2F, Yidian Scientific Instruments Corporation, Shanghai, China); and the measurement of moisture content was performed by thermophysical drying. The determination of the amino acid nitrogen content in soy sauce was carried out by absorbance measurements at 400 nm, according to the national standards method reported in GB 5009.235-2016 *Determination of amino acid nitrogen in food*.

Analysis of organic acid and sugar

The contents of organic acids and sugars were determined by using a high-performance liquid chromatography (HPLC) as previously reported (Sarkar *et al.*, 2020). The HPLC measurement was carried out by using the Shimadzu LC-20A system (Shimadzu Corporation, Kyoto, Japan). For the separation of organic acid, a Shodex KC-811 (i.d. 8 mm × 300 mm, Showa Denko Corporation, Tokyo, Japan) column was used, and the temperature was maintained at 50°C. The eluent (3.0 mmol/L $HClO_4$ solution) flow rate was 1.0 mL/min, and the chromatogram of samples was recorded with UV detection at 210 nm. As for sugar analysis, a Shodex KS-801 column (i.d. 8 mm × 300 mm, Showa Denko Co.) was used for separation, the column temperature was set at 80°C, ultrapure water used as eluent with a flow rate of 0.7 mL/min, and the detection was performed on a differential refractive index detector. Both analyses were measured in triplicates.

Sensory evaluation

The sensory evaluation of soy sauce was carried out by employing 37 assessors (19–30 years old; 12 men and 25 women) with professional training and experience. For evaluation process, the 24 soy sauce samples were first poured onto white plastic plates to observe the color and texture. The assessors also put their noses 5 cm above the soy sauce to evaluate the aroma of soy sauce. Next, fresh cucumbers were cut into small pieces and dipped in the soy sauce for taste tests. Finally, each sample was scored according to its appearance, texture, taste, and aroma on a scale of 1–5 for each component. A higher score indicated higher satisfaction. The best possible score of a sample was 20 for each evaluation. The ratings given by all 37 individuals were added together to obtain a final total sensory evaluation score for each sample.

NIR spectroscopy

The acquisition of NIR spectroscopy for assessing the quality of soy sauces was performed according to a previous method (Wang *et al*., 2021). Briefly, the NIR spectra were collected using diffuse reflectance mode (Cary 5000, Varian Corporation, California, USA) with the wavelength ranging from 400 to 2500 nm. All samples were scanned on the same day to ensure that all measurement conditions were consistent, including ambient temperature and humidity. The pretreatment methods of spectra could be used to eliminate errors caused by disturbances, such as high-frequency random noise, baseline drift, and stray light, and to improve the reliability of the NIR model (Chen *et al*., 2013). In this work, nine pretreatments including first derivative (1st derivative), second derivative (2nd derivative), multiple scattering correction (MSC), and standard normal variate (SNV) methods, and combinations of these methods were employed into the spectrum processing, and the Savitsky–Golay algorithm with 10 points of smoothing was used to optimize the raw spectroscopic data.

Modelling and validation

The NIR predictive model was established by correlating the spectra and measured data with the aim to predicting the values of unknown samples (Xie *et al*., 2009). In this study, all soy sauce samples were randomly separated into two subsets of 16 and 8 samples. The 16 samples were classified into the calibration set and used for model development and cross validation, while the other 8 samples were classified into the validation set and used to test the practical performance of the established models.

The measured indicators, sensory scores, and spectral data of calibration set were first used to establish the predictive model. The modelling process was performed by PLSR using the Unscrambler data processing software (version 10.4, CAMO Software, Oslo, Norway). Subsequently, the established calibration models were validated by both internal full cross-validation and external validation of the validation set (Cámara-Martos *et al*., 2012). The correction coefficient of determination (R^2_c), root-mean-square error of correction (RMSEC), cross-validation coefficient of determination (R^2_{cv}), and root-mean-square error of cross-validation (RMSECV) were calculated to clarify the performance. The validation set was further employed to evaluate the feasibility of calibration model according to prediction coefficient of determination (R^2_p), root-mean-square error of prediction (RMSEP), and deviation rate (bias). For these indexes, R^2 indicates the degree of linear correlation between the predicted value from the model and the reference value. While RMSEC, RMSECV, and RMSEP represent the standard deviations between model predictions and reference values during model calibration, cross-validation, and independent validation, respectively. For the same batch of samples, the smaller values of RMSEC, RMSECV, and RMSEP indicate better model performance. The calculation of these indexes was as following:

$$R^2 = 1 - \frac{\sum (y_{\text{i, actual}} - y_{\text{i, predicted}})^2}{\sum (y_{\text{i, actual}} - \hat{y}_{\text{i, actual}})^2} \tag{1}$$

$$\text{RMSEC} = \sqrt{\frac{\sum (y_{\text{i, actual}} - y_{\text{i, predicted}})^2}{m-1}} \tag{2}$$

Where, m is the number of samples in the model.

$$\text{RMSECV} = \sqrt{\frac{\sum (y_{\text{i, actual}} - y_{\text{i, predicted}})^2}{n-1}} \tag{3}$$

Where, n is the number of cycles of cross-validation.

$$\text{RMSEP} = \sqrt{\frac{\sum (y_{\text{i, actual}} - y_{\text{i, predicted}})^2}{k-1}} \tag{4}$$

Where, k is the number of samples in the validation set used for model testing.

$$\text{Bias} = \frac{1}{k} \sum (y_{\text{i, predicted}} - y_{\text{i, actual}}) \tag{5}$$

Where, k is the number of samples in the prediction set.

Among these equations, $y_{\text{i,actual}}$ denotes the measured value of the i-th sample, $\hat{y}_{\text{i,actual}}$ denotes the average values in the calibration set (or validation set), and $y_{\text{i,predicted}}$ denotes the predicted value of the i-th sample.

Statistical analysis

The means and standard deviation were calculated to analyze the sensory attributes in soy sauce samples tested. All the results were reported as mean ± standard deviation of at least three measurements. The analysis of variance (ANOVA) was applied by using DPS software (Data Processing System, Hangzhou Rui feng Information Technology Co., Ltd., Zhejiang, China) to determine significant differences between soy sauces. The statistical significance level was set at $P < 0.05$.

Results and Discussion

Sensory evaluation results

In the sensory evaluation, the highest score was 512 and the lowest score was 332 (Table 1). Samples with high taste and appearance scores tended to have high overall scores. According to the results of the sensory evaluation, the 24 samples were divided into two groups of 12 with the higher-ranking samples in Group A and the lower-ranking samples in Group B.

Table 1. Reference data on the sensory scores of soy sauce.

Ranking	Name	Total score	Taste	Appearance
1	Factory A	512	121	140
2	Factory B	510	124	138
3	Factory C	509	128	130
4	Factory D	507	128	136
5	Factory E	507	127	129
6	Factory F	496	115	134
7	Factory G	494	116	135
8	Factory H	472	124	125
9	Factory I	467	116	128
10	Factory J	464	118	117
11	Factory K	462	122	115
12	Factory L	461	109	126
13	Factory M	458	119	116
14	Factory N	450	110	122
15	Factory O	448	107	113
16	Factory P	441	118	111
17	Factory Q	437	104	119
18	Factory R	423	113	107
19	Factory S	416	105	107
20	Factory T	410	83	99
21	Factory U	408	80	101
22	Factory V	382	100	87
23	Factory W	382	70	106
24	Factory X	332	78	109

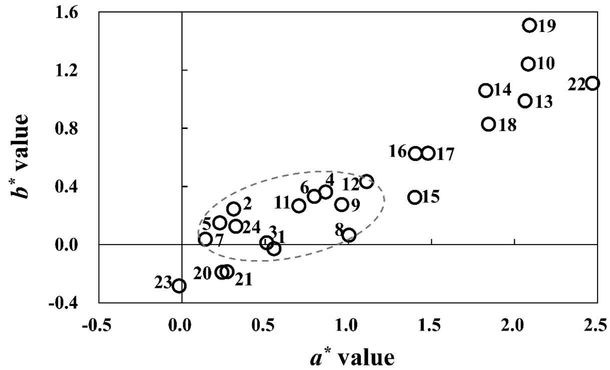

Figure 1. Measurements of a^* and b^* values of samples by using a CM-5 spectrophotometer. The numbering in figure is the same as the ranking in Table 1.

Soy sauce composition

Color analysis

The results of color measurements are shown in Figure 1. L^*, a^*, b^* are the three elements of the lab color space. L^* indicates the brightness, with an L^* value of 0 indicating pure black and an L^* value of 100 indicating pure white. a^* denotes the red index and indicates a spectral change from red to green. Larger positive values of a^* indicate a reddish color and smaller negative values indicate a greenish color. The value of b^* indicates a spectral change from yellow to blue, with larger positive values indicating a yellowish color and smaller negative values indicating a bluish color (Aliakbarian *et al.*, 2016).

The a^* values of soy sauce samples with higher rankings ranged from 0.2 to 1.0, and the b^* values ranged from –0.1 to 0.4 (Figure 1). These results were within an appropriate range for the desired appearance of soy sauce. It is generally believed that as the Japanese soy sauce becomes redder (i.e., a larger a^* value), the quality improves (Wang *et al.*, 2021). However, the results in this study showed that the a^* value of Chinese soy sauce could vary within a certain range. Very high values of a^* and b^* will not improve the quality of soy sauce. The L^* values of the soy sauce ranged from 2.6 to 3.2 (Table 2), and did not greatly contribute to its appearance.

Taste analysis

A certain amount of brine is usually added to soy sauce during the fermentation process to inhibit the growth of unwanted bacteria (Syifaa *et al.*, 2016). The salt contents and Brix values of the 24 soy sauce samples are shown in Figure 2. The high-quality soy sauce samples had salt contents between 12 and 17% and Brix values between 41 and 47%. If the salt content is too high, it will adversely affect the taste and health benefits of the soy sauce (Bibbins-Domingo *et al.*, 2010). By contrast, if the salt content is

Table 2. Average physicochemical compositions in soy sauce samples.

Rank	a*	b*	L*	Salt (%)	Brix (%)
1	0.55 ± 0.06	−0.03 ± 0.02	2.64 ± 0.01	12.78 ± 0.15	46.66 ± 0.55
2	0.31 ± 0.02	0.25 ± 0.06	3.12 ± 0.04	15.57 ± 0.62	41.78 ± 1.61
3	0.51 ± 0.02	0.01 ± 0.03	2.84 ± 0.02	13.40 ± 0.06	45.79 ± 1.16
4	0.86 ± 0.04	0.36 ± 0.04	2.86 ± 0.01	15.45 ± 0.75	51.18 ± 0.97
5	0.23 ± 0.06	0.15 ± 0.02	2.71 ± 0.04	14.42 ± 0.17	43.29 ± 0.55
6	0.80 ± 0.03	0.33 ± 0.02	2.92 ± 0.02	13.97 ± 0.15	52.42 ± 0.57
7	0.14 ± 0.05	0.04 ± 0.04	2.83 ± 0.09	17.89 ± 0.06	41.45 ± 1.12
8	1.00 ± 0.06	0.07 ± 0.05	3.07 ± 0.01	15.78 ± 0.34	42.45 ± 1.45
9	0.96 ± 0.04	0.28 ± 0.04	3.54 ± 0.03	12.96 ± 0.32	41.33 ± 1.52
10	2.08 ± 0.03	1.24 ± 0.04	3.28 ± 0.01	14.45 ± 0.46	45.38 ± 0.97
11	0.70 ± 0.04	0.27 ± 0.01	2.99 ± 0.02	10.56 ± 0.16	32.86 ± 1.21
12	1.11 ± 0.03	0.43 ± 0.03	3.19 ± 0.02	13.46 ± 0.30	41.19 ± 0.54
13	2.06 ± 0.07	0.99 ± 0.05	3.69 ± 0.02	7.18 ± 0.15	37.66 ± 0.55
14	1.83 ± 0.03	1.06 ± 0.07	3.38 ± 0.01	11.90 ± 0.30	35.56 ± 1.15
15	1.40 ± 0.07	0.33 ± 0.01	3.42 ± 0.04	16.45 ± 0.86	40.41 ± 1.09
16	1.40 ± 0.02	0.63 ± 0.02	3.08 ± 0.01	15.56 ± 0.19	41.87 ± 0.53
17	1.48 ± 0.05	0.63 ± 0.02	3.03 ± 0.01	14.31 ± 0.11	52.31 ± 1.62
18	1.84 ± 0.03	0.83 ± 0.04	3.45 ± 0.06	13.24 ± 0.28	46.37 ± 0.53
19	2.09 ± 0.01	1.51 ± 0.02	4.62 ± 0.04	16.38 ± 0.00	32.36 ± 0.98
20	0.24 ± 0.02	−0.19 ± 0.03	1.34 ± 0.03	13.77 ± 0.26	56.24 ± 1.77
21	0.27 ± 0.07	−0.19 ± 0.05	2.19 ± 0.01	14.92 ± 0.72	54.19 ± 0.57
22	2.47 ± 0.03	1.11 ± 0.04	3.61 ± 0.01	15.55 ± 0.31	50.44 ± 0.56
23	−0.02 ± 0.02	−0.28 ± 0.04	2.12 ± 0.02	15.43 ± 0.21	40.63 ± 1.72
24	0.32 ± 0.04	0.13 ± 0.05	2.50 ± 0.01	16.27 ± 0.49	49.21 ± 0.01
Rank	**Moisture (%)**	**Glucose***	**Galactose**	**Oxalic acid**	**Citric acid**
1	59.11 ± 0.05	12.95 ± 0.56	ND	1.61 ± 0.05	5.12 ± 0.01
2	64.07 ± 0.03	5.76 ± 0.09	ND	0.54 ± 0.02	6.42 ± 0.15
3	63.52 ± 0.01	7.48 ± 0.05	ND	0.82 ± 0.05	17.21 ± 0.19
4	60.84 ± 0.01	11.01 ± 0.10	6.07 ± 0.15	0.59 ± 0.01	8.57 ± 0.03
5	61.46 ± 0.11	7.66 ± 0.24	ND	0.76 ± 0.04	11.59 ± 0.12
6	59.42 ± 0.04	6.54 ± 0.06	7.96 ± 0.29	0.61 ± 0.00	14.06 ± 0.18
7	60.52 ± 0.58	8.62 ± 0.08	ND	0.65 ± 0.02	9.74 ± 0.39
8	64.89 ± 0.09	3.25 ± 0.02	7.94 ± 0.43	1.69 ± 0.00	5.11 ± 0.19
9	63.31 ± 0.10	11.94 ± 0.21	6.85 ± 0.14	0.91 ± 0.04	15.15 ± 0.47
10	63.15 ± 0.08	10.85 ± 0.30	8.10 ± 0.11	0.94 ± 0.02	7.83 ± 0.16
11	72.76 ± 0.13	5.68 ± 0.16	4.34 ± 0.18	0.66 ± 0.02	5.29 ± 0.18
12	63.87 ± 0.57	12.51 ± 0.12	10.32 ± 0.14	2.44 ± 0.05	7.10 ± 0.33
13	66.52 ± 0.29	3.25 ± 0.03	4.46 ± 0.07	0.29 ± 0.01	6.33 ± 0.16
14	61.69 ± 0.11	13.22 ± 0.35	9.85 ± 0.37	1.05 ± 0.02	8.66 ± 0.45
15	67.33 ± 0.20	5.29 ± 0.05	ND	0.86 ± 0.01	6.40 ± 0.16
16	63.73 ± 0.10	10.86 ± 0.12	6.80 ± 0.37	0.91 ± 0.01	5.41 ± 0.13
17	60.82 ± 0.34	9.65 ± 0.36	8.35 ± 0.03	0.60 ± 0.00	14.03 ± 0.07
18	60.24 ± 0.10	10.55 ± 0.05	ND	0.97 ± 0.03	7.02 ± 0.12
19	71.03 ± 0.30	7.57 ± 0.10	3.75 ± 0.16	0.58 ± 0.02	5.05 ± 0.14
20	47.25 ± 0.39	18.57 ± 0.13	ND	30.08 ± 0.93	ND
21	51.30 ± 0.13	11.62 ± 0.13	ND	10.76 ± 0.92	6.83 ± 0.21
22	64.21 ± 0.01	11.28 ± 0.14	8.17 ± 0.14	0.85 ± 0.04	5.61 ± 0.43
23	61.32 ± 0.64	21.95 ± 0.34	ND	8.39 ± 0.14	ND
24	62.98 ± 0.11	5.81 ± 0.10	ND	1.82 ± 0.03	7.57 ± 0.32

(Continues)

Table 2. Continued.

Rank	Tartaric acid	Lactic acid	Pyroglutamic acid	Amino acid nitrogen
1	35.96 ± 0.97	3.97 ± 0.11	2.64 ± 0.05	1.06 ± 0.00
2	60.56 ± 1.95	16.87 ± 0.16	4.33 ± 0.05	0.93 ± 0.00
3	82.52 ± 1.19	12.05 ± 0.39	1.76 ± 0.08	1.01 ± 0.00
4	10.12 ± 0.04	ND	4.19 ± 0.14	1.05 ± 0.00
5	27.11 ± 0.37	ND	5.19 ± 0.03	1.08 ± 0.00
6	9.86 ± 0.13	ND	4.43 ± 0.05	1.05 ± 0.00
7	98.03 ± 3.60	ND	5.04 ± 0.07	1.01 ± 0.00
8	3.01 ± 0.12	9.66 ± 0.41	2.18 ± 0.06	0.88 ± 0.00
9	54.18 ± 2.31	ND	5.47 ± 0.20	1.07 ± 0.00
10	20.99 ± 0.59	ND	4.19 ± 0.14	0.98 ± 0.00
11	8.68 ± 0.36	17.78 ± 0.38	3.89 ± 0.08	0.72 ± 0.00
12	12.40 ± 0.28	22.73 ± 0.78	4.67 ± 0.10	0.90 ± 0.00
13	3.78 ± 0.15	8.67 ± 0.29	2.16 ± 0.04	0.85 ± 0.00
14	25.48 ± 0.92	ND	5.16 ± 0.19	0.91 ± 0.00
15	5.20 ± 0.13	6.09 ± 0.04	2.94 ± 0.06	0.79 ± 0.00
16	11.96 ± 0.34	ND	4.70 ± 0.12	1.10 ± 0.00
17	12.27 ± 0.05	ND	4.81 ± 0.20	0.93 ± 0.00
18	14.89 ± 0.36	15.21 ± 0.20	2.90 ± 0.04	0.52 ± 0.00
19	9.82 ± 0.33	4.18 ± 0.15	1.33 ± 0.05	0.48 ± 0.00
20	11.71 ± 0.30	ND	2.58 ± 0.02	1.00 ± 0.00
21	54.27 ± 2.09	ND	4.30 ± 0.13	0.98 ± 0.00
22	14.47 ± 0.59	9.42 ± 0.14	4.73 ± 0.13	1.09 ± 0.00
23	9.46 ± 0.46	ND	3.88 ± 0.09	0.66 ± 0.00
24	20.37 ± 0.89	ND	4.04 ± 0.09	0.78 ± 0.00

*The units of organic acids, sugars, and amino acid nitrogen are g/kg; ND: not detected.

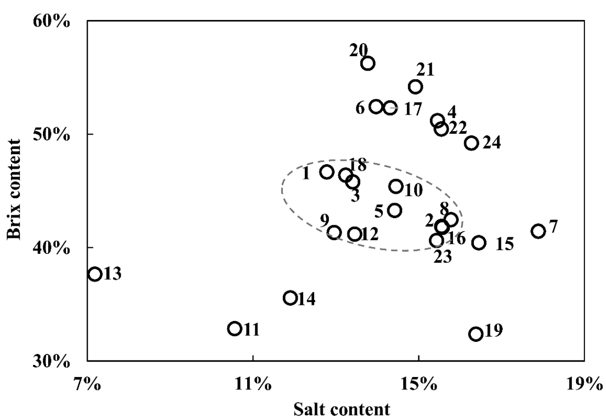

Figure 2. Plots of salt content versus Brix content for soy sauces. The numbering in figure is the same as the ranking in Table 1.

too low, the growth of unwanted bacteria during the fermentation process will not be inhibited and this affects the quality of the soy sauce (Taormina *et al.*, 2010).

The moisture content affects both the fermentation process of the soy sauce and the texture of the finished product. The moisture contents of the soy sauce samples are shown in Table 2. The results showed that moisture contents did not adversely affect the sensory quality of the soy sauce.

The pH values and amino acid nitrogen contents of the samples are shown in Figure 3. The pH of soy sauce affects its flavor, with a lower pH resulting in a more prominent sour taste. The pH values of the higher-quality soy sauces were all greater than 5.1, and the pH values of the lower-quality soy sauces were all less than 4.9. These results showed that higher acidity might be detrimental to the taste of soy sauce. Amino acid nitrogen represents the nitrogen content of free amino acids in soy sauce, and amino acids are closely related to the umami taste of soy sauce (Yanfang and Wenyi, 2009). The average levels of amino acid nitrogen in Groups A and B were 0.978 and 0.841 g/kg, respectively. The difference in the amino acid nitrogen between Groups A and B was significant ($P < 0.05$), and this might be related to the taste results. Consequently, the samples with high sensory scores had high amino acid nitrogen contents, which supported the positive effect of the amino acid nitrogen content on the taste of soy sauce.

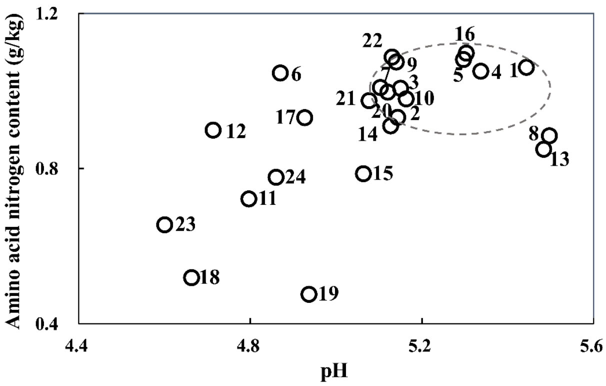

Figure 3. Plots of pH and amino acid nitrogen content for soy sauces. The numbering in figure is the same as the ranking in Table 1.

Five organic acids including oxalic, citric, tartaric, lactic, and pyroglutamic acids were detected in this experiment (Table 2). Among these acids, tartaric acid had the highest content. The average contents of oxalic, citric, tartaric, lactic, and pyroglutamic acids were 1.018, 9.433, 35.285, 6.922, and 3.998 g/kg, respectively, in Group A, and 4.763, 6.076, 16.14, 3.631, and 3.628 g/kg, respectively, in Group B. These results showed that amino acid nitrogen, citric acid, tartaric acid, and lactic acid are beneficial to the taste of soy sauce, while oxalic acid is detrimental to the taste of soy sauce and pyroglutamic acid has little effect on the taste of soy sauce.

The sugars in soy sauce provide sweetness and act as a source of carbon for other chemical reactions during fermentation (Chiou *et al.*, 1999; Kwak and Lim, 2004). Both glucose and galactose were detected in the samples (Table 2), with glucose present at higher levels. Glucose was detected in all of the samples at 3.25–21.95 g/kg. The average level in Group A (8.688 g/kg) was lower than that in Group B (10.802 g/kg). The lower level in Group A might be because glucose is an important carbon source that is consumed during the later stages of fermentation to produce some taste and flavor substances. This reduces the level of glucose and makes the overall flavor of the soy sauce richer. Galactose was detected in only 13 of the samples (Table 2) at 3.75–10.32 g/kg, and there was no significant correlation between the taste quality of the soy sauce and galactose. Overall, the physical and chemical properties could be used to objectively evaluate the quality of soy sauce.

Spectral analysis

In the NIR spectra (Figure 4), the samples showed multiple absorption peaks. Large differences were observed in the 400–800 nm region, which was probably because of the different colors of the soy sauce samples. All spectra showed large absorption peaks at 996–1134 nm, 1134–1325 nm, 1800–1950 nm, and 2140–2380 nm. The first two absorption peaks corresponded to the C-O and O-H functional groups in alcohols and phenols, and the latter two were attributed to O-H, N-H, and C-H groups.

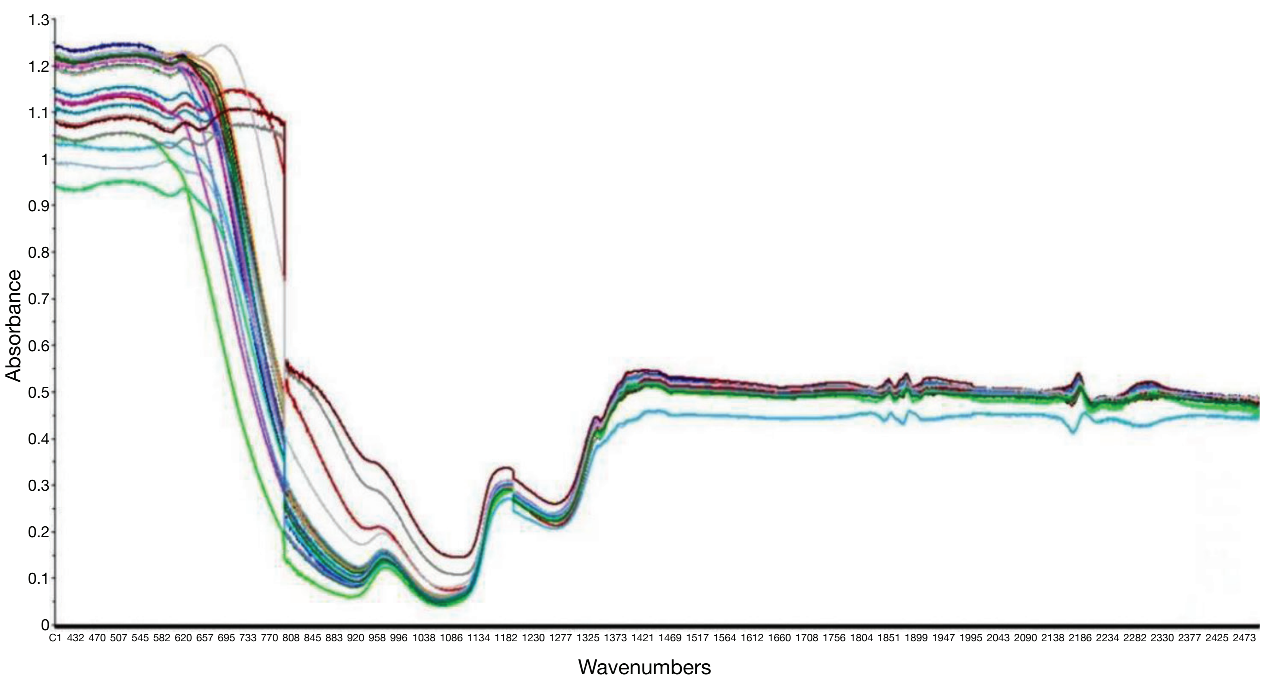

Figure 4. Original near-infrared spectra of soy sauces tested.

Establishment and validation of the quantitative analysis models

PLSR was used to build the model, and nine preprocessing methods (no treatment, 1st derivative, 2nd derivative, MSC, MSC + 1st derivative, MSC + 2nd derivative, SNV, SNV + 1st derivative, SNV + 2nd derivative) were used to improve the model accuracy. The data obtained by modelling and prediction using PLSR are shown in Table S1 *Accuracy parameters for judging the quality of the PLSR models based on the calibration and validation sets of different components and sensory scores of soy sauce samples*. When the NIR spectra were modelled using the total sensory evaluation score, the model had poor predictive performance as shown by the low R^2_p and high RMSEP (Table 3). The four scores were then

Table 3. Accuracy parameters for judging the quality of the PLSR models based on the calibration and validation sets of sensory scores of soy sauce samples using the full spectrum.

			Calibration				Validation		
Parameters	Pretreatment	LVs	R^2_c	RMSEC	R^2_{cv}	RMSECV	R^2_p	RMSEP	Bias
Total sensory evaluation score	None	3	0.313	30.122	0.152	35.434	0.222	59.621	–1.310
	1st derivative	2	0.339	29.531	0.228	33.805	0.228	59.420	3.733
	2nd derivative	3	0.439	27.208	0.204	34.329	0.252	58.507	2.146
	MSC	2	0.312	30.146	0.183	34.775	0.201	60.448	5.495
	MSC + 1st derivative	2	0.345	29.400	0.236	33.623	0.218	59.823	2.104
	MSC + 2nd derivative	3	0.434	27.333	0.196	34.498	0.234	59.197	1.189
	SNV	2	0.315	30.065	0.188	34.666	0.228	59.444	4.525
	SNV + 1st derivative	2	0.348	29.347	0.241	33.518	0.227	59.465	2.108
	SNV + 2nd derivative	2	0.440	27.202	0.208	34.241	0.236	59.106	1.357
Taste score	None	3	0.675	7.392	0.550	9.210	0.716	12.303	–1.281
	1st derivative	2	0.677	7.367	0.590	8.789	0.713	12.381	1.589
	2nd derivative	3	0.762	6.325	0.598	8.670	0.711	12.410	0.322
	MSC	2	0.653	7.637	0.513	9.583	0.680	13.074	1.571
	MSC + 1st derivative	1	0.564	8.560	0.524	9.476	0.733	11.930	1.486
	MSC + 2nd derivative	1	0.563	8.572	0.509	9.615	0.727	12.058	1.016
	SNV	2	0.656	7.609	0.523	9.486	0.702	12.603	1.348
	SNV + 1st derivative	1	0.658	7.584	0.539	9.318	0.687	12.930	1.671
	SNV + 2nd derivative	1	0.555	8.646	0.507	9.642	0.730	11.991	1.149
Amino acid nitrogen	None	5	0.78	0.081	0.51	0.129	NA	0.177	0.085
	1st derivative	2	0.84	0.069	0.64	0.11	0.128	0.144	0.051
	2nd derivative	4	0.96	0.034	0.604	0.1149	0.259	0.132	0.052
	MSC	4	0.95	0.038	0.72	0.096	0.732	0.0796	–0.027
	MSC + 1st derivative	2	0.75	0.087	0.610	0.114	NA	0.186	0.084
	MSC + 2nd derivative	3	0.98	0.023	0.627	0.1116	0.451	0.114	0.049
	SNV	4	0.97	0.030	0.730	0.095	0.785	0.071	–0.026
	SNV + 1st derivative	2	0.74	0.087	0.607	0.1146	NA	0.186	0.083
	SNV + 2nd derivative	3	0.99	0.021	0.626	0.1117	0.33	0.126	0.052
Moisture	None	2	0.975	0.845	0.787	2.592	0.784	1.931	1.159
	1st derivative	1	0.953	1.153	0.716	2.998	0.525	2.862	–0.361
	2nd derivative	1	0.701	2.903	0.659	3.282	NA	7.084	–3.962
	MSC	2	0.962	1.033	0.619	3.47	0.825	1.738	0.152
	MSC + 1st derivative	1	0.955	1.132	0.831	2.312	0.703	2.261	–0.169
	MSC + 2nd derivative	1	0.714	2.838	0.697	3.097	NA	7.428	–3.822
	SNV	2	0.743	2.691	0.559	3.736	NA	6.698	–3.246
	SNV + 1st derivative	1	0.956	1.119	0.793	2.559	0.689	2.314	–0.269
	SNV + 2nd derivative	1	0.709	2.864	0.686	3.149	NA	7.286	–3.765

LVs, number of latent variables; MSC, multiple scattering correction; NA, not available; R^2_c, the correction coefficient of determination; R^2_{cv}, the coefficient of determination in full cross-validation; R^2_p, the coefficient of determination for prediction; RMSEC, root mean square error of correction; RMSECV, root mean square error of cross-validation; RMSEP, root mean square error of prediction; SNV: standard normal variate.

modelled separately, and the model with the taste score was found to have the best predictive performance. The best taste score model was obtained using MSC + 1st derivative, which had a R^2_p of 0.733 and RMSEP of 11.93 (Table 3). The appearance score model was similar to the total sensory evaluation score model and had poor predictive performance. The aroma score model and the texture score model gave no valid prediction set data, which indicated that they had poor feasibility. Each parameter that may affect the quality of soy sauce was modelled with the spectra, and the moisture and amino acid nitrogen contents had the best modelling results and predictive ability. The best moisture content model was developed using MSC and had a R^2_p of 0.825 and RMSEP of 1.73. The best amino acid nitrogen content model was developed using SNV correction and had a R^2_p of 0.785 and RMSEP of 0.071. Both models had good predictive performance. The predictive performances of the models constructed using the remaining constituents were poor because either there was no valid prediction set data or the models had low R^2_p and high RMSEP. In summary, the models for the moisture content, amino acid nitrogen content, and taste score had good predictive performances and could be used to rapidly predict the soy sauce quality.

Amino acid nitrogen is the main umami substance in soy sauce. The taste analysis showed that soy sauce samples with high sensory scores and good taste also had high amino acid nitrogen contents. The good predictive ability of NIR spectroscopy for both the amino acid nitrogen content and taste score was likely related to the close relationship between these two components. Generally, high-quality soy sauce contains more amino acid nitrogen than low-quality soy sauce. This means that amino acid nitrogen is important for evaluating the quality of soy sauce.

Conclusions

The rich substances in soy sauce that are related to sensory quality have brought new challenges on how to quantitatively evaluate its final quality. This study performed a comprehensive investigation of physicochemical and sensory experiments on 24 commercial soy sauce samples in China. The contents of amino acid nitrogen, citric acid, tartaric acid, and lactic acid, beneficial for improving the quality of soy sauce, and the interaction between them were clarified. Glucose showed a negative effect on the soy sauce quality. Furthermore, with the aim to developing a rapid and objective method for predicting the measured components and sensory scores of soy sauces, the feasibility of NIR spectroscopy was tested and validated by modelling with PLSR. The predictive models showed good performances on predicting the moisture content, amino acid nitrogen content, and the taste score of soy sauce, which were considered useful for classifying the sensory quality of soy sauce quickly and economically. Finally, this work has some limitations, such as ignoring the aroma-producing compounds. Therefore, further studies should focus on analyzing the factors that may affect the quality of soy sauce, and improve the predictive performance for routine use.

Conflict of Interest

Authors have no conflicts of interest to declare for this article.

References

Aliakbarian, B., Bagnasco, L., Perego, P., Leardi, R. and Casale, M., 2016. UV-VIS spectroscopy for monitoring yogurt stability during storage time. Analytical Methods 8(30): 5962–5969. https://doi.org/10.1039/C6AY00607H

Bázár, G., Romvári, R., Szabó, A., Somogyi, T., Éles, V. and Tsenkova, R., 2016. NIR detection of honey adulteration reveals differences in water spectral pattern. Food Chemistry 194: 873–880. https://doi.org/10.1016/j.foodchem.2015.08.092

Bibbins-Domingo, K., Chertow, G.M., Coxson, P.G., Moran, A., Lightwood, J.M., Pletcher, M.J., et al., 2010. Projected effect of dietary salt reductions on future cardiovascular disease. New England Journal of Medicine 362(7): 590–599. https://doi.org/10.1056/NEJMoa0907355

Cámara-Martos, F., Zurera-Cosano, G., Moreno-Rojas, R., García-Gimeno, R.M. and Pérez-Rodríguez, F., 2012. Identification and quantification of lactic acid bacteria in a water-based matrix with near-infrared spectroscopy and multivariate regression modeling. Food Analytical Methods 5(1): 19–28. https://doi.org/10.1007/s12161-011-9221-5

Cayuela, J.A., Puertas, B. and Cantos-Villar, E., 2017. Assessing wine sensory attributes using Vis/NIR. European Food Research and Technology 243(6): 941–953. https://doi.org/10.1007/s00217-016-2807-9

Chen, H., Song, Q., Tang, G., Feng, Q. and Liang, L., 2013. The combined optimization of savitzky-golay smoothing and multiplicative scatter correction for FT-NIR PLS models. ISRN Spectroscopy 2013: 1–9. https://doi.org/10.1155/2013/642190

Chiou, R.Y., Ferng, S. and Beuchat, L.R., 1999. Fermentation of low-salt miso as affected by supplementation with ethanol. International Journal of Food Microbiology 48(1): 11–20. https://doi.org/10.1016/S0168-1605(99)00033-1

Corona, P., Frangipane, M.T., Moscetti, R., Lo Feudo, G., Castellotti, T. and Massantini, R., 2021. Chestnut cultivar identification through the data fusion of sensory quality and FT-NIR spectral data. Foods 10(11): 2575. https://doi.org/10.3390/foods10112575

Cortes, V., Rodriguez, A., Blasco, J., Rey, B., Besada, C., Cubero, S., et al., 2017. Prediction of the level of astringency in persimmon using visible and near-infrared spectroscopy. Journal of Food Engineering 204: 27–37. https://doi.org/10.1016/j.jfoodeng.2017.02.017

Diez-Simon, C., Eichelsheim, C., Mumm, R. and Hall, R.D., 2020. Chemical and sensory characteristics of soy sauce: a review. Journal of Agricultural and Food Chemistry 68(42): 11612–11630. https://doi.org/10.1021/acs.jafc.0c04274

Escribano, S., Biasi, W.V., Lerud, R., Slaughter, D.C. and Mitcham, E.J., 2017. Non-destructive prediction of soluble solids and dry matter content using NIR spectroscopy and its relationship with sensory quality in sweet cherries. Postharvest Biology and Technology 128: 112–120. https://doi.org/10.1016/j.postharvbio.2017.01.016

Jamshidi, B., Mohajerani, E. and Jamshidi, J., 2016. Developing a Vis/NIR spectroscopic system for fast and non-destructive pesticide residue monitoring in agricultural product. Measurement 89: 1–6. https://doi.org/10.1016/j.measurement.2016.03.069

Jürkenbeck, K. and Spiller, A., 2021. Importance of sensory quality signals in consumers' food choice. Food Quality and Preference 90: 104155. https://doi.org/10.1016/j.foodqual.2020.104155

Kwak, E.J. and Lim, S.I., 2004. The effect of sugar, amino acid, metal ion, and NaCl on model Maillard reaction under pH control. Amino Acids 27(1): 85–90. https://doi.org/10.1007/s00726-004-0067-7

Lan, W., Jaillais, B., Leca, A., Renard, C.M. and Bureau, S., 2020. A new application of NIR spectroscopy to describe and predict purees quality from the non-destructive apple measurements. Food Chemistry 310: 125944. https://doi.org/10.1016/j.foodchem.2019.125944

Li, X., He, Y. and Fang, H., 2007. Non-destructive discrimination of Chinese bayberry varieties using Vis/NIR spectroscopy. Journal of Food Engineering 81(2): 357–363. https://doi.org/10.1016/j.jfoodeng.2006.10.033

Li, X., Zhang, L., Zhang, Y., Wang, D., Wang, X., Yu, L., et al., 2020. Review of NIR spectroscopy methods for nondestructive quality analysis of oilseeds and edible oils. Trends in Food Science & Technology 101: 172–181. https://doi.org/10.1016/j.tifs.2020.05.002

Lorenzo, C., Garde-Cerdán, T., Pedroza, M.A., Alonso, G.L. and Salinas, M.R., 2009. Determination of fermentative volatile compounds in aged red wines by near infrared spectroscopy. Food Research International 42(9): 1281–1286. https://doi.org/10.1016/j.foodres.2009.03.021

Mohammadi-Moghaddam, T., Razavi, S., Sazgarnia, A. and Taghizadeh, M., 2018. Predicting the moisture content and textural characteristics of roasted pistachio kernels using Vis/NIR reflectance spectroscopy and PLSR analysis. Journal of Food Measurement and Characterization 12(1): 346–355. https://doi.org/10.1007/s11694-017-9646-7

Sarkar, T., Mukherjee, A., Chatterjee, K., Shariati, M.A., Rebezov, M., Rodionova, S., et al., 2022a. Comparative analysis of statistical and supervised learning models for freshness assessment of oyster mushrooms. Food Analytical Methods 15(4): 917–939. https://doi.org/10.1007/s12161-021-02161-7

Sarkar, T., Salauddin, M., Hazra, S.K. and Chakraborty, R., 2020. The impact of raw and differently dried pineapple (*Ananas comosus*) fortification on the vitamins, organic acid and carotene profile of dairy rasgulla (sweetened cheese ball). Heliyon 6(10): e05233. https://doi.org/10.1016/j.heliyon.2020.e05233

Sarkar, T., Salauddin, M., Pati, S., Chakraborty, R., Shariati, M.A., Rebezov, M., et al., 2022b. The fuzzy cognitive map–based shelf-life modelling for food storage. Food Analytical Methods 15(3): 579–597. https://doi.org/10.1007/s12161-021-02147-5

Shurtleff, W. and Aoyagi, A., 2012. History of soy sauce (160 CE to 2012): extensively annotated bibliography and sourcebook. Soyinfo Center, Lafayette, CA, USA.

Steinhaus, P. and Schieberle, P., 2007. Characterization of the key aroma compounds in soy sauce using approaches of molecular sensory science. Journal of Agricultural and Food Chemistry 55(15): 6262–6269. https://doi.org/10.1021/jf0709092

Stone, H., Bleibaum, R. and Thomas, H.A., 2020. Sensory evaluation practices. 5th Edition, Elsevier Academic Press, Boston, MA, USA.

Syifaa, A.S., Jinap, S., Sanny, M. and Khatib, A., 2016. Chemical profiling of different types of soy sauce and the relationship with its sensory attributes. Journal of Food Quality 39(6): 714–725. https://doi.org/10.1111/jfq.12240

Taormina, P.J., 2010. Implications of salt and sodium reduction on microbial food safety. Critical Reviews in Food Science and Nutrition 50(3): 209–227. https://doi.org/10.1080/10408391003626207

Wang, S., Liu, X., Tamura, T., Kyouno, N., Zhang, H. and Chen, J.Y., 2021. Rapid detection of the quality of miso (Japanese fermented soybean paste) using visible/near-infrared spectroscopy. Analytical Letters 54(14): 2304–2314. https://doi.org/10.1080/00032719.2020.1858092

Wang, S., Zhang, H., Liu, X., Tamura, T., Kyouno, N. and Chen, J.Y., 2018. Relationship between chemical characteristics and sensory evaluation of koikuchi soy sauce. Analytical Letters 51(14): 1813–1827. https://doi.org/10.1080/00032719.2017.1419252

Xie, L., Ye, X., Liu, D. and Ying, Y., 2009. Quantification of glucose, fructose and sucrose in bayberry juice by NIR and PLS. Food Chemistry 114(3): 1135–1140. https://doi.org/10.1016/j.foodchem.2008.10.076

Xu, D., Li, C., Zhao, M., Fengc, Y., Sun, L. and Wang, Y., 2013. Assessment on the improvement of soy sauce fermentation by *Aspergillus oryzae* HG76. Biocatalysis and Agricultural Biotechnology 2(4): 344–351. https://doi.org/10.1016/j.bcab.2013.07.002

Yanfang, Z. and Wenyi, T., 2009. Biochemical changes in low-salt fermentation of solidstate soy sauce. African Journal of Biotechnology 8(24): 7028–7034

Zareef, M., Chen, Q., Hassan, M.M., Arslan, M., Hashim, M.M., Ahmad, W., et al., 2020. An overview on the applications of typical non-linear algorithms coupled with NIR spectroscopy in food analysis. Food Engineering Reviews 12(2): 173–190. https://doi.org/10.1007/s12393-020-09210-7

Supplementary

Table S1. Accuracy parameters for judging the quality of the PLSR models based on the calibration and validation sets of different components and sensory scores of soy sauce samples.

			Calibration				Validation		
Parameters	Pretreatment	LVs	R^2_c	RMSEC	R^2_{cv}	RMSECV	R^2_p	RMSEP	Bias
Amino acid nitrogen	None	5	0.777	0.081	0.510	0.129	NA	0.177	0.085
	1st derivative	2	0.840	0.069	0.640	0.110	0.128	0.144	0.051
	2nd derivative	4	0.961	0.034	0.604	0.149	0.259	0.132	0.052
	MSC	4	0.951	0.038	0.720	0.096	0.732	0.080	–0.027
	MSC + 1st derivative	2	0.747	0.087	0.610	0.114	NA	0.186	0.084
	MSC + 2nd derivative	3	0.982	0.023	0.627	0.116	0.451	0.114	0.049
	SNV	4	0.969	0.030	0.730	0.095	0.785	0.071	–0.026
	SNV + 1st derivative	2	0.745	0.087	0.607	0.146	NA	0.186	0.083
	SNV + 2nd derivative	3	0.985	0.021	0.626	0.117	0.330	0.126	0.052
D-Galactose	None	3	0.321	3.130	0.106	3.802	NA	4.781	3.466
	1st derivative	2	0.353	3.054	0.220	3.551	NA	5.131	3.986
	2nd derivative	2	0.274	3.235	0.155	3.696	NA	4.102	3.637
	MSC	2	0.328	3.310	0.187	3.625	NA	5.158	4.248
	MSC + 1st derivative	2	0.366	3.024	0.253	3.476	NA	5.012	3.826
	MSC + 2nd derivative	2	0.271	3.243	0.155	3.697	NA	4.285	3.694
	SNV	2	0.331	3.106	0.207	3.582	NA	5.088	4.136
	SNV + 1st derivative	2	0.367	3.022	0.254	3.475	NA	4.997	3.825
	SNV + 2nd derivative	2	0.262	3.263	0.142	3.728	NA	4.253	3.708
Glucose	None	1	0.267	3.173	NA	3.946	0.352	4.657	1.690
	1st derivative	1	0.262	3.183	0.040	3.844	0.445	4.310	1.316
	2nd derivative	1	0.279	3.146	0.097	3.729	0.484	4.157	1.502
	MSC	1	0.247	3.216	0.050	3.823	0.424	4.391	1.398
	MSC + 1st derivative	1	0.276	3.153	0.088	3.747	0.531	3.966	1.337
	MSC + 2nd derivative	1	0.287	3.128	0.136	3.647	0.585	3.728	1.443
	SNV	1	0.248	3.213	0.053	3.818	0.418	4.415	1.391
	SNV + 1st derivative	1	0.271	3.163	0.077	3.769	0.513	4.039	1.310
	SNV + 2nd derivative	1	0.282	3.140	0.125	3.669	0.568	3.806	1.414
Oxalic acid	None	4	0.990	0.691	0.611	4.570	NA	9.592	3.689
	1st derivative	2	0.997	0.406	0.684	4.123	0.382	4.853	0.899
	2nd derivative	3	0.928	1.862	0.424	5.563	NA	7.617	3.355
	MSC	5	0.990	0.698	0.667	4.227	NA	7.955	3.086
	MSC + 1st derivative	2	0.997	0.379	0.789	3.363	NA	9.785	3.831
	MSC + 2nd derivative	3	0.959	1.404	0.604	4.611	NA	8.600	3.691
	SNV	6	0.994	0.542	0.698	4.029	NA	9.000	3.897
	SNV + 1st derivative	2	0.997	0.384	0.774	3.482	NA	9.602	3.713
	SNV + 2nd derivative	3	0.952	1.509	0.576	4.773	NA	8.328	3.594
Lactic acid	None	1	0.093	6.149	NA**	7.064	0.217	7.231	–1.553
	1st derivative	1	0.080	6.193	0.018	6.778	0.202	7.301	–1.283
	2nd derivative	1	0.066	6.242	0.034	6.721	0.173	7.431	–1.737
	MSC	1	0.099	6.131	0.038	6.706	0.221	7.212	–1.418
	MSC + 1st derivative	1	0.091	6.159	0.040	6.699	0.192	7.347	–1.330
	MSC + 2nd derivative	1	0.074	6.215	0.047	6.676	0.156	7.506	–1.353
	SNV	1	0.096	6.141	0.030	6.734	0.223	7.203	–1.384

(Continues)

Table S1. Continued.

Parameters	Pretreatment	LVs	Calibration R^2_c	RMSEC	R^2_{cv}	RMSECV	Validation R^2_p	RMSEP	Bias
	SNV + 1st derivative	1	0.090	6.161	0.036	6.712	0.196	7.326	–1.304
	SNV + 2nd derivative	1	0.074	6.214	0.046	6.679	0.161	7.484	–1.335
Citric acid	None	1	0.163	3.856	NA	4.835	0.350	2.959	1.910
	1st derivative	2	0.687	2.357	0.275	3.801	0.535	2.504	1.359
	2nd derivative	3	0.986	0.500	2.706	3.813	0.404	2.833	1.135
	MSC	1	0.097	4.007	NA	5.140	0.474	2.663	0.855
	MSC + 1st derivative	4	0.673	2.410	0.261	3.838	0.624	2.250	0.947
	MSC + 2nd derivative	5	0.950	0.941	0.402	3.453	NA	4.073	0.200
	SNV	1	0.126	3.941	NA	5.163	0.542	2.485	0.810
	SNV + 1st derivative	4	0.680	2.385	0.304	3.727	NA	5.088	0.487
	SNV + 2nd derivative	5	0.950	0.943	0.406	3.441	NA	4.099	0.293
Moisture	None	2	0.975	0.845	0.787	2.592	0.784	1.931	1.159
	1st derivative	1	0.953	1.153	0.716	2.998	0.525	2.862	–0.361
	2nd derivative	1	0.701	2.903	0.659	3.282	NA	7.084	–3.962
	MSC	2	0.962	1.033	0.619	3.470	0.825	1.738	0.152
	MSC + 1st derivative	1	0.955	1.132	0.831	2.312	0.703	2.261	–0.169
	MSC + 2nd derivative	1	0.714	2.838	0.697	3.097	NA	7.428	–3.822
	SNV	2	0.743	2.691	0.559	3.736	NA	6.698	–3.246
	SNV + 1st derivative	1	0.956	1.119	0.793	2.559	0.689	2.314	–0.269
	SNV + 2nd derivative	1	0.709	2.864	0.686	3.149	NA	7.286	–3.765
Brix	None	1	0.459	4.777	0.341	5.585	NA	7.738	5.723
	1st derivative	1	0.463	4.761	0.405	5.307	NA	7.714	5.181
	2nd derivative	1	0.420	4.947	0.387	5.388	NA	8.234	5.475
	MSC	1	0.446	4.824	0.345	5.569	NA	7.603	5.391
	MSC + 1st derivative	1	0.443	4.849	0.397	5.343	NA	8.114	5.165
	MSC + 2nd derivative	1	0.403	5.019	0.356	5.521	NA	8.658	5.294
	SNV	1	0.456	4.792	0.358	5.514	NA	7.558	5.379
	SNV + 1st derivative	1	0.451	4.812	0.403	5.315	NA	7.993	5.129
	SNV + 2nd derivative	1	0.412	4.982	0.368	5.470	NA	8.547	5.264
*L**	None	2	0.690	0.365	0.559	0.462	NA	0.519	0.237
	1st derivative	1	0.665	0.380	0.546	0.468	NA	0.486	0.227
	2nd derivative	1	0.648	0.389	0.574	0.453	NA	0.467	0.182
	MSC	1	0.674	0.374	0.569	0.456	NA	0.430	0.195
	MSC + 1st derivative	1	0.684	0.369	0.614	0.432	NA	0.510	0.220
	MSC + 2nd derivative	1	0.668	0.378	0.639	0.417	NA	0.537	0.197
	SNV	1	0.677	0.373	0.572	0.455	NA	0.434	0.198
	SNV + 1st derivative	1	0.685	0.369	0.606	0.436	NA	0.505	0.227
	SNV + 2nd derivative	1	0.671	0.376	0.637	0.419	NA	0.529	0.203
*b**	None	1	0.436	0.372	0.354	0.421	NA	0.379	0.26
	1st derivative	1	0.412	0.380	0.369	0.417	NA	0.429	0.318
	2nd derivative	1	0.368	0.394	0.333	0.428	NA	0.402	0.300
	MSC	1	0.439	0.371	0.400	0.406	NA	0.395	0.297
	MSC + 1st derivative	1	0.392	0.386	0.346	0.424	NA	0.423	0.322
	MSC + 2nd derivative	1	0.338	0.403	0.272	0.448	NA	0.414	0.316
	SNV	1	0.444	0.369	0.406	0.404	NA	0.399	0.299
	SNV + 1st derivative	1	0.399	0.384	0.354	0.421	NA	0.426	0.324
	SNV + 2nd derivative	1	0.347	0.400	0.286	0.443	NA	0.416	0.318

(Continues)

Table S1. Continued.

Parameters	Pretreatment	LVs	Calibration R^2_c	Calibration RMSEC	Calibration R^2_{cv}	Calibration RMSECV	Validation R^2_p	Validation RMSEP	Validation Bias
*a**	None	1	0.474	0.479	0.425	0.530	NA	0.883	0.764
	1st derivative	1	0.448	0.490	0.421	0.532	NA	0.965	0.842
	2nd derivative	1	0.394	0.514	0.374	0.553	NA	0.939	0.818
	MSC	1	0.468	0.482	0.452	0.518	NA	0.921	0.816
	MSC + 1st derivative	1	0.433	0.497	0.403	0.540	NA	0.969	0.845
	MSC + 2nd derivative	1	0.588	0.424	0.379	0.551	NA	0.929	0.833
	SNV	1	0.472	0.480	0.457	0.515	NA	0.925	0.819
	SNV + 1st derivative	1	0.440	0.494	0.412	0.536	NA	0.972	0.849
	SNV + 2nd derivative	1	0.592	0.422	0.378	0.551	NA	0.934	0.835
Appearance score	None	3	0.359	10.948	0.223	12.770	0.189	11.320	–6.205
	1st derivative	2	0.389	10.693	0.309	12.038	0.410	9.652	–4.902
	2nd derivative	3	0.470	9.963	0.275	12.334	0.424	9.542	–5.610
	MSC	2	0.338	11.128	0.229	12.715	0.418	9.590	–3.950
	MSC + 1st derivative	2	0.374	10.821	0.294	12.168	0.334	10.256	–5.802
	MSC + 2nd derivative	3	0.469	9.969	0.265	12.414	0.261	10.808	–6.382
	SNV	2	0.340	11.113	0.230	12.712	0.425	9.531	–4.352
	SNV + 1st derivative	2	0.376	10.808	0.298	12.134	0.362	10.037	–5.790
	SNV + 2nd derivative	3	0.475	9.914	0.277	12.314	0.300	10.520	–6.303
Taste score	None	3	0.675	7.392	0.550	9.210	0.716	12.303	–1.281
	1st derivative	2	0.677	7.367	0.590	8.789	0.713	12.381	1.589
	2nd derivative	3	0.762	6.325	0.598	8.670	0.711	12.410	0.322
	MSC	2	0.653	7.637	0.513	9.583	0.680	13.074	1.571
	MSC + 1st derivative	1	0.564	8.560	0.524	9.476	0.733	11.930	1.486
	MSC + 2nd derivative	1	0.563	8.572	0.509	9.615	0.727	12.058	1.016
	SNV	2	0.656	7.609	0.523	9.486	0.702	12.603	1.348
	SNV + 1st derivative	1	0.658	7.584	0.539	9.318	0.687	12.930	1.671
	SNV + 2nd derivative	1	0.555	8.646	0.507	9.642	0.730	11.991	1.149
Total sensory evaluation score	None	3	0.313	30.122	0.152	35.434	0.222	59.621	–1.310
	1st derivative	2	0.339	29.531	0.228	33.805	0.228	59.420	3.733
	2nd derivative	3	0.439	27.208	0.204	34.329	0.252	58.507	2.146
	MSC	2	0.312	30.146	0.183	34.775	0.201	60.448	5.495
	MSC + 1st derivative	2	0.345	29.400	0.236	33.623	0.218	59.823	2.104
	MSC + 2nd derivative	3	0.434	27.333	0.196	34.498	0.234	59.197	1.189
	SNV	2	0.315	30.065	0.188	34.666	0.228	59.444	4.525
	SNV + 1st derivative	2	0.348	29.347	0.241	33.518	0.227	59.465	2.108
	SNV + 2nd derivative	2	0.440	27.202	0.208	34.241	0.236	59.106	1.357
Texture score	None	1	0.268	9.944	0.208	10.948	NA	25.978	5.527
	1st derivative	1	0.321	9.578	0.187	11.095	NA	25.737	4.179
	2nd derivative	2	0.997	0.616	0.130	11.475	NA	25.388	3.988
	MSC	1	0.232	10.184	0.141	11.400	NA	25.102	4.235
	MSC + 1st derivative	2	0.303	9.703	0.163	11.259	NA	24.821	3.354
	MSC + 2nd derivative	3	0.429	8.778	0.181	11.136	NA	24.610	3.231
	SNV	1	0.237	10.147	0.151	11.337	NA	25.213	4.210
	SNV + 1st derivative	2	0.300	9.719	0.157	11.298	NA	24.824	3.296
	SNV + 2nd derivative	3	0.429	8.778	0.149	11.351	NA	24.657	3.267

LVs, number of latent variables; MSC, multiple scattering correction; NA, not available; R^2_c, the correction coefficient of determination; R^2_{cv}, the coefficient of determination in full cross-validation; R^2_p, the coefficient of determination for prediction; RMSEC, root mean square error of correction; RMSECV, root mean square error of cross-validation; RMSEP, root mean square error of prediction; SNV: standard normal variate.