doi: 10.17221/229/2015-CJFS

Rapid Detection of Total Nitrogen Content in Soy Sauce using NIR Spectroscopy

Jie XU[1] Furong HUANG[1], Yuanpeng LI[1], Zhe CHEN[1] and Yong WANG[2]

[1]Department of Opto-Electronic Engineering and [2]Department of Food Science and Engineering, Jinan University, Guangzhou, P.R. China

Abstract

Xu J., Huang F., Li Y., Chen Z., Wang Y. (2015): **Rapid detection of total nitrogen content in soy sauce using NIR spectroscopy**. Czech J. Food Sci., 33: 518–522.

A method for the rapid and nondestructive determination of total nitrogen content in soy sauce was explored. Prediction models were established using near-infrared spectroscopy combined with each of the following techniques: partial least squares (PLS), interval PLS, synergy interval PLS (siPLS), and backward interval PLS. Results showed that each improved forecast model was better than the PLS model. The siPLS method exhibited the best performance. The full spectrum of light soy sauce was divided into 20 subintervals. The combination of four subintervals, namely, 6, 11, 13, and 18, showed the best effect, with a correlation coefficient of 0.9977 and an RMSECV of 0.0198. The full spectrum of dark soy sauce was divided into 20 subintervals. The combination of three subintervals, namely, 14, 17, and 19, showed the best effect, with a correlation coefficient of 0.9818 and an RMSECV of 0.0640. Therefore, the siPLS method can realise a rapid and accurate detection of total nitrogen content in soy sauce.

Keywords: near-infrared spectroscopy; partial least squares

Soy sauce is a popular and widely used traditional Chinese condiment. This condiment improves the flavour and taste of food, and its rich content of selenium and other minerals helps prevent cancer. Soy sauce contains various vitamins and minerals that can reduce cholesterol and damage caused by free radicals to the human body as well as lower the incidence of cardiovascular diseases (Kataoka 2005). The total nitrogen content is an important indicator of soy sauce quality.

Traditional analytical methods for the determination of total nitrogen content in soy sauce are Kjeldahl nitrogen determination (Hiller *et al*. 1948) and sodium hypobromite (Kato *et al*. 2013) methods. However, these methods are time consuming and tedious. Therefore, a simple, rapid, and accurate method for the detection of total nitrogen content in soy sauce is urgently needed for food safety monitoring in the food industry and quality control agencies. In addition to its ability for online detection and simultaneous determination of multicomponents, near-infrared spectroscopy (NIRS) is fast, environment-friendly, easy to perform, economical, objective and nondestructive (Woo *et al*. 2005). Thus, this technique (NIRS) is usually preferred over traditional analytical methods.

With the rapid development of computer technology and stoichiometry, NIRS has been commonly used in the agricultural (Chang *et al*. 2001; Chen *et al*. 2008; Lu *et al*. 2011), biological (Throback *et al*. 2004), pharmaceutical (Luypaert *et al*. 2007), and food industries (Růžičková & Šustová 2006; Cen *et al*. 2007; Jirsa *et al*. 2007). Partial least squares (PLS) is currently the most frequently used technique for chemical metrology because of its higher prediction ability and simpler model compared with other conventional learning algorithms. However, the full spectrum of the PLS model can be easily disturbed by irrelevant components, which influence the quality and precision of the calibration model. Thus, selecting an appropriate spectral region is important. In recent years, improved PLS methods

doi: 10.17221/229/2015-CJFS

such as interval PLS (iPLS), forward iPLS (fiPLS), backward iPLS (biPLS), and synergy iPLS (siPLS) have shown excellent performance in eliminating irrelevant and nonlinear variables, simplifying models, and improving the robustness of models by screening the characteristic wavelength. Cen *et al*. (2007) detected citric acid and tartaric acid in orange juice using partial least squares (PLS) regression with cross-validation. NIR spectrometric determination of quality parameters in vegetable oils using iPLS and variable selection was done by Costa Pereira *et al*. (2008). The nondestructive detection of total acid content in Chinese vinegar using NIRS combined with synergy interval partial least square (Si-PLS) algorithm was performed by Chen *et al*. (2012). To improve and simplify the NIR prediction model of the soluble solid content (SSC) of strawberry, Shi combined backward interval partial least squares (BiPLS) and simulated annealing algorithm (SAA) to select the efficient wavelengths (Shi *et al*. 2011). However, only a few studies have determined the total nitrogen content in soy sauce using NIRS (Ouyang *et al*. 2012; Xu *et al*. 2013; Meng *et al*. 2014). What is more, their detection methods need some complex spectral preprocessing procedures thus cannot meet the "rapid" request in the practical application. The current study was designed to accurately determine the total nitrogen content of soy sauce using NIRS combined with different PLS methods and select the relatively best model and finally find a simple and rapid method.

MATERIAL AND METHODS

Five different batches of Lee Kum Kee soy sauce which were from different production times were bought from the local market. A total of 128 and 228 samples for light and dark soy sauces from 5 different batches were tested independently. 96 samples of light soy sauces were randomly selected as calibration set, the rest 32 were validation set. 171 samples of dark soy sauces were randomly selected as calibration set, the rest 57 were validation set. Each sample was carefully determined by both NIRS and Kjeldahl nitrogen determination method. The environmental temperature was kept around 25°C and the relative humidity was kept at a steady level in the laboratory. The statistics of total nitrogen content is listed in Table 1.

NIR equipment and software. The visible and NIR spectra were collected in the transmission mode using an XDS rapid liquid analyser with VISION software (Foss NIR Systems, Silver Spring, USA). The sample was held in a circular sample cuvette with plastic cap (optical path is 8 mm). Each spectrum was the average of 32 scans with a wavelength increment of 0.5 nm. The range of spectra was from 400 nm to 2500 nm. The spectral resolution was 8 nm. Each sample was analysed three times and the mean value of three spectra was used in the following analysis. Data analysis was performed using home-made routines programmed in MATLAB code (MATLAB; The Math Works, Massachusetts, USA). The toolbox used to select the most informative variables, called iPLS, biPLS and siPLS, was downloaded from www.models.kvl.dk/. The calculation of the accuracy profile was adopted by e•noval v 3.0 (Arlenda®, Belgium) software.

Relevant principles

iPLS. The iPLS (Norgaard *et al*. 2000; Ferrao *et al*. 2011) algorithm introduced by Norgaard was used for interval selection in this work. The principle of this algorithm is to split the spectra into smaller equidistant regions and, afterwards, develop PLS regression models for each of the sub-intervals, using the same number of latent variables. Thereafter, an average error is calculated for every sub-interval and for the full-spectrum model. The region with the lowest error is chosen. An optimised region can be found by reducing or increasing it by subtracting or adding new variables, symmetrically or asymmetrically. One of the main advantages of this method is the possibility to represent a local regression model in a graphical display, focusing on a choice of better

Table 1. Statistics of total nitrogen content

Species	Sample number	Maximum	Minimum	Mean	Standard deviation
		(g/100 ml)			
Light soy sauce	128	1.883	0.628	1.084	0.338
Dark soy sauce	228	1.937	0.636	1.547	0.338

doi: 10.17221/229/2015-CJFS

Table 2. Results of soy sauce iPLS calibration model with different interval number

Sauce	Interval	Factors	Wavelengths (nm)	RMSECV	*r*	Bias
Light soy	10	8	2080–2288	0.0281	0.9912	–0.0001
	15	9	1660–1798	0.0262	0.9972	–0.0002
	20	9	1668–1770	0.0250	0.9954	0.0003
	25	8	1660–1742	0.0282	0.9961	0.0000
	30	7	1730–1798	0.0347	0.9936	0.0004
Dark soy	10	8	2290–2498	0.0720	0.9742	0.0001
	15	14	2220–2358	0.0730	0.9785	0.0007
	20	5	1876–1978	0.0751	0.9624	–0.0005
	25	8	2164–2246	0.0789	0.9631	–0.0008
	30	8	2290–2358	0.0724	0.9810	–0.0004

intervals and allowing a comparison between interval models and the full-spectrum model.

siPLS. Synergy interval partial least squares algorithm (siPLS) (Lin *et al*. 2009; Wang *et al*. 2012) is the expansion of interval partial least squares algorithm (iPLS); first of all, it equally divides the whole spectrum into several subintervals, then does partial least squares regression through permutation and combination with different numbers and different subintervals, at last, using the synergy intervals spectrum which has the smallest root mean square error of cross validation (RMSECV) to establish the siPLS model and prediction.

siPLS can make up the limitation of iPLS modelling with only one single interval. Meanwhile, the run time of the algorithm is long, depending on the divided number of subintervals and the combined number of subintervals.

biPLS. Backward interval partial least squares (biPLS) (Li *et al*. 2009; Shi *et al*. 2011; Huang *et al*. 2013) is an improved method which was proposed by L. Norgaard to select a characteristic region based on the theory of iPLS. As in the interval PLS model the data set is split into a given number of intervals, now PLS models are calculated with each interval left out, i.e. if one chooses 40 intervals, then each model is based on 39 intervals leaving out one interval at a time. The first left out interval is the one that when left out gives the poorest performing model with respect to RMSECV. This procedure is continued until one interval remains.

Evaluation index. This study established cross-validation models by NIRS that is separately combined with PLS, iPLS, siPLS, and biPLS. With RMSECV as the main standard, the correlation coefficient (*r*) and bias as reference, the best modelling parameters in each PLS model and the most ideal one overall were selected. The spectral range of the PLS model was 400 nm to 2500 nm, with RMSECV, RMSECP values of 0.0637, 0.0630 for the light soy sauce samples and 0.1043, 0.1040 for the dark soy sauce samples, respectively. RMSECV(P) is principally used to assess the feasibility of modelling methods and the prediction ability of the corresponding models to present accurate results with low RMSECV(P).

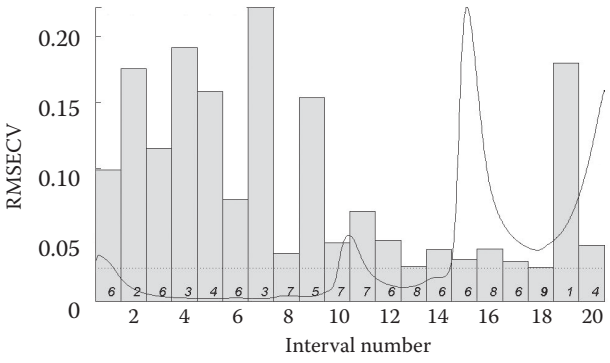

Figure 1. RMSECV of light soy sauce's iPLS calibration model (n = 20)

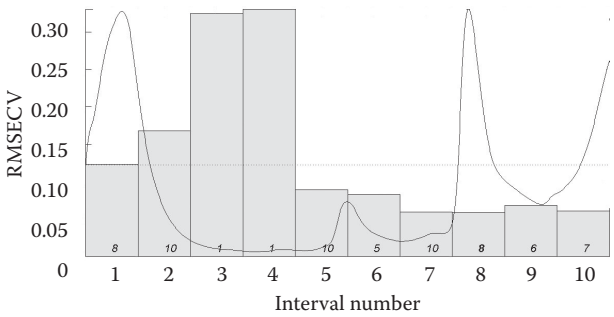

Figure 2. RMSECV of dark soy sauce's iPLS calibration model (n = 10)

doi: 10.17221/229/2015-CJFS

Table 3. Modelling results under different numbers of interval and combinations

Sauce	Interval	Combination	Interval number	RMSECV		*r*
Light soy	10	2	6, 7	0.0234	0.0232	0.9966
		3	4, 7, 9	0.0232		0.9967
		4	5, 6, 7, 9	0.0229		0.9968
	20	2	11, 18	0.0210	0.0204	0.9970
		3	3, 5, 18	0.0205		0.9973
		4	6, 11, 13, 18	0.0198		0.9977
	30	2	20, 27	0.0213	0.0205	0.9968
		3	2, 20, 27	0.0203		0.9979
		4	4, 15, 26, 27	0.0199		0.9966
Dark soy	10	2	5, 10	0.0733	0.0728	0.9781
		3	3, 4, 10	0.0727		0.9757
		4	3, 4, 6, 8	0.0724		0.9778
	20	2	14, 19	0.0671	0.0652	0.9806
		3	14, 17, 19	0.0640		0.9818
		4	8, 9, 17, 19	0.0644		0.9808
	30	2	17, 28	0.0685	0.0704	0.9796
		3	12, 24, 28	0.0670		0.9801
		4	5, 10, 12, 28	0.0756		0.9796

RESULTS AND DISCUSSION

Interval partial least squares (iPLS) modelling. In this experiment, the wavelength of the infrared spectrum is 400–2500 nm long and n = 10, 15, 20, 25, and 30, respectively. PLS regression models of the whole infrared spectrum and intervals were established. RMSECV and r of these models were compared.

It can be seen from Table 2 that when n = 20, the iPLS calibration model of the 18th interval achieves the minimum RMSECV (0.0250). The corresponding wavelength ranges from 2188 nm to 2292 nm and the optimal amount of factors is 9. RMSECV of light soy sauce iPLS calibration model when n = 20 is shown in Figure 1.

As for the dark soy sauce shown in Table 2, when n = 10, the iPLS calibration model of the 8th interval achieves the minimum RMSECV (0.0720). The corresponding wavelength ranges from 2290 to 2498 nm and the optimal amount of factors is 8. The RMSECV of dark soy sauce iPLS calibration model when n = 10 is shown in Figure 2.

Synergy interval partial least squares (siPLS) modelling. To improve the accuracy of the prediction model, SiPLS, an improved iPLS method, was applied for modelling treatment of two soy sauces. Numbers of the interval were determined 10, 20, and 30, respectively. Later, 2, 3, and 4 intervals were selected according to RMSECV as the optimal interval combination for final modelling.

According to Table 3, when n = 20, the light soy sauce SiPLS calibration model of four interval combination (6, 11, 13, and 18) presents the minimum RMSECV (0.0198), while the dark soy sauce SiPLS calibration

Table 4. Modelling results under different interval combinations

Sauce	Interval	Interval number	RMSECV	*r*
Light soy	10	4, 6, 9	0.0260	0.9960
	20	4, 7, 8	0.0483	0.9845
	30	12, 16, 25	0.0520	0.9851
Dark soy	10	8, 9, 10	0.1025	0.9741
	20	9, 10, 11	0.0901	0.9701
	30	12, 13, 14	0.1087	0.9631

doi: 10.17221/229/2015-CJFS

Table 5. Predictions of different optimal iPLS models

Sauce	Methods	Intervals	Interval range (nm)	RMSECV	RMSECP
Light soy	PLS	all	400–2500	0.0637	0.0630
	iPLS	20	1668–1770	0.0241	0.0239
	siPLS	20	925–1030, 1450–1555, 1660–1765, 2185–2290	0.0198	0.0203
	biPLS	10	1030–1240, 1450–1660, 2080–2290	0.0260	0.0266
Dark soy	PLS	all	400–2500	0.1043	0.1040
	iPLS	10	2290–2498	0.0801	0.0794
	siPLS	20	1765–1870, 2080–2185, 2290–2395	0.0640	0.0642
	biPLS	20	1240–1555	0.0901	0.0889

model of three interval combination (14, 17, and 19) has the minimum RMSECV (0.0640). Furthermore, the SiPLS calibration model has better result as the combining intervals increase. However, a continuous increase of interval combination will lead to excessive calculations, thus influencing the modelling time.

Backward interval partial least squares (biPLS) modelling. As another improved iPLS method, biPLS modelling is commonly applied to the processing of spectral data. In this experiment, the 400–2500 nm long near-infrared spectrum was divided into 10, 20, and 30 intervals, respectively. Optimal interval combinations were selected by using biPLS method (Table 4).

According to Table 4, the light soy sauce biPLS calibration model of three interval combination (4, 6, and 9) when n = 10 presents the minimum RMSECV (0.0260). The corresponding correlation coefficient (r) reaches 0.9960 and the wavelength ranges are 1030–1240, 1450–1660, and 2080–2290 nm. The dark soy sauce biPLS calibration model of three interval combination (9, 10, and 11) when n = 20 shows the minimum RMSECV (0.0901). The corresponding r is 0.9701 and the wavelength ranges from 1440 nm to 1555 nm.

Comprehensive comparison of different optimal iPLS models. Predictions of different optimal iPLS models are listed in Table 5. The RMSECP values were obtained from the validation set at the same interval ranges and combined interval number as the calibration set did. Compared to the common iPLS model, three improved iPLS models show higher prediction accuracy. Among them, the siPLS model has the highest prediction accuracy. The RMSECV and RMSECP of light soy sauce siPLS model are 0.0198 and 0.0203, while those of dark soy sauce are 0.0640 and 0.0642. Predicted values of both light and dark soy sauces show a good linear relation with measured values.

CONCLUSIONS

In this study we used near-infrared spectroscopy combined with different partial least squares methods to detect the total nitrogen content of two kinds of soy sauce samples nondestructively. The prediction accuracy of the different partial least squares methods and their different interval partition number were compared. The siPLS showed its superiority compared with the other PLS models; it is relatively simple, can effectively eliminate the interference spectrum area and reduce the modelling number, so as to reduce the processing time, and what is more, it has higher forecast efficiency and satisfactory forecast precision, with RMSECV = 0.0198, RMSECP = 0.0203 for light soy sauce, and RMSECV = 0.0640, RMSECP = 0.0642 for dark soy sauce. Therefore, it can realise a quick and nondestructive detection of total nitrogen content in soy sauce by near-infrared spectroscopy combined with synergy interval partial least squares. At the same time, we can apply this modelling idea to other detection fields widely. We can use NIR combined with partial least squares to detect the quality of liquid food or other liquid samples, for instance, the content of alcohol in wine, the content of blood sugar in the blood and the components and contents of edible oil etc. However, water is a strong absorber which will cover the near-infrared spectrum absorption signals of other components. So it is quite complex to balance the relationship between reducing the error of spectrum detection and obtaining a strong signal of the target components.

References

Cen H., Bao Y., He Y., Sun D. (2007): Visible and near infrared spectroscopy for rapid detection of citric and tartaric acids in orange juice. Journal of Food Engineering, 82: 253–260.

doi: 10.17221/229/2015-CJFS

Chang C.W., Laird D.A., Mausbach M.J., Hurburgh C.R. (2001): Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. Soil Science Society of America Journal, 65: 480–490.

Chen Q., Zhao J., Liu M., Cai J.(2008): Nondestructive identification of tea (*Camellia sinensis* L.) varieties using Ft-NIR spectroscopy and pattern recognition. Czech Journal of Food Sciences, 26: 360–367.

Chen Q., Ding J., Cai J., Sun Z., Zhao J. (2012): Simultaneous measurement of total acid content and soluble salt-free solids content in Chinese vinegar using near-infrared spectroscopy. Journal of Food Science, 77: 222–227

Ferrao M.F., Viera M.D.S., Panta Pazos R.E., Fachini D., Gerbase A.E., Marder L. (2011): Simultaneous determination of quality parameters of biodiesel/diesel blends using HATR-FTIR spectra and PLS, iPLS or siPLS regressions. Fuel, 90: 701–706.

Hiller A., Plazin J., Van Slyke D.D. (1948): A study of conditions for Kjeldahl determination of nitrogen in proteins; description of methods with mercury as catalyst, and titrimetric and gasometric measurements of the ammonia formed. The Journal of Biological Chemistry, 176: 1401–1420.

Huang W., Tian F., Liu Z., Yang Q., Zhao X., Yang J. (2013): Wavelength selection for FT-NIR spectroscopic analysis of protein in *Chimonobambusa quadrangularis* shoot based on iPLS and BiPLS models. Food Science, 34: 133–137.

Jirsa O., Hrušková M., Švec I. (2007): Bread features evaluation by NIR analysis. Czech Journal of Food Sciences, 25: 243–248.

Kataoka S. (2005): Functional effects of Japanese style fermented soy sauce (shoyu) and its components. Journal of Bioscience and Bioengineering, 100: 227–234.

Kato J., Koseki T., Aoki Y., Yamada A., Tanaka T. (2013): Coulometric titration of urea with electrogenerated hypobromite. Analytical Sciences, 29: 753–755.

Li P., Wang J., Cao N., Han D. (2009): Selection of variables for MLR in Vis/NIR spectroscopy based on BiPLS combined with Ga. Spectroscopy and Spectral Analysis, 29: 2637–2641.

Lin F., Chen Z., Wang K., Deng J., Xu H. (2009): Determination of nitrogen in rice leaf based on FTIR spectra and siPLS-GA-PLS algorithm. Journal of Infrared and Millimeter Waves, 28: 277–280.

Lu Q., Tang M., Cai J., Zhao J., Vittayapadung S. (2011): Vis/NIR hyperspectral imaging for detection of hidden bruises on kiwifruits. Czech Journal of Food Sciences, 29: 595–602.

Luypaert J., Massart D.L., Heyden Y.V. (2007): Near-infrared spectroscopy applications in pharmaceutical analysis. Talanta, 72: 865–883.

Meng Q., Hatakeyama M., Sugawara E. (2014): Formation by yeast of 2-furanmethanethiol and ethyl 2-mercaptopropionate aroma compounds in japanese soy sauce. Bioscience, Biotechnology, and Biochemistry, 78: 109–114.

Norgaard L., Saudland A., Wagner J., Nielsen J.P., Munck L., Engelsen S.B. (2000): Interval partial least-squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy. Applied Spectroscopy, 54: 413–419.

Pereira C.A.F., Pontes C.M.J., Neto G.F.F., Santos B.S.R., Galvao H.R.K., Araujo U.M.C. (2008): NIR spectrometric determination of quality parameters in vegetable oils using iPLS and variable selection. Food Research International, 41: 341–348.

Ouyang Q., Zhao J., Chen Q., Lin H., Sun Z. (2012): Rapid measurement of antioxidant activity in dark soy sauce by NIR spectroscopy combined with spectral intervals selection and nonlinear regression tools. Analytical Methods, 4: 940–946.

Růžičková J., Sustová K. (2006): Determination of selected parameters of quality of the dairy products by NIR spectroscopy. Czech Journal of Food Sciences, 24: 255–260.

Shi J., Zou X., Zhao J., Mao H. (2011): Selection of wavelength for strawberry NIR spectroscopy based on BiPLS combined with SAA. Journal of Infrared and Millimeter Waves, 30: 458–462.

Throback I.N., Enwall K., Jarvis A., Hallin S. (2004): Reassessing PCR primers targeting *nirS*, *nirk* and *nosZ* genes for community surveys of denitrifying bacteria with DGGE. FEMS Microbiology Ecology, 49: 401–417.

Wang X., Bao Y., Liu G., Li G., Lin L. (2012): Study on the best analysis spectral section of NIR to detect alcohol concentration based on SiPLS. International Workshop on Information and Electronics Engineering, 29: 2285–2290.

Woo Y.A., Kim H.J., Ze K.R., Chung H. (2005): Near-infrared (NIR) spectroscopy for the non-destructive and fast determination of geographical origin of Angelicae Gigantis Radix. Journal of Pharmaceutical and Biomedical Analysis, 36: 955–959.

Xu X., Wu H., He H., Huang B., Han J., Ren Y. (2013): Study of chloropropanols in soy sauce by gas chromatography-triple quadrupole mass spectrometry with coupled column separation without derivatisation. Food Additives & Contaminants: Part A, 30: 421–429.

Received: 2015–05–04
Accepted after corrections: 2015–09–13

Corresponding author:

Dr Furong Huang, Jinan University, Department of Opto-Electronic Engineering, Guangzhou, 510632, China;.
E-mail: furong_huang@163.com