RESULTS AND DISCUSSION

1. Development of Quantitative Models for Predicting the Chemical and Physical Properties of Thai Commercial Fish Sauces

Partial least square (PLS) regression with the aids of wavelength interval selection methods were used to develop the quantitative models for chemical and physical of 100 Thai fish sauce samples. The following four steps describe the procedure for quantitative modeling. First, the chemical and physical properties of fish sauce samples were measured (section 1.1) and used as dependent variables (Y-data). In the second step, the NIR spectra of all samples were measured (section 1.2) and used as independent variables (X-data). Then, in the third step, the input wavelength variables were selected by using the wavelength interval selection methods (section 1.3). In the final step, the chemical and physical data (Y-data) and selected input wavelength variables (X-data) were applied to develop the quantitative models (section 1.4). In this step, the performance of developed models was calculated and compared. The best prediction model with the lowest error was selected.

1.1 Chemical and Physical Concentrations

The total nitrogen, sodium chloride, pH, reducing sugar, density, baume, total soluble solid, refractive index and color (L*, a*, b*) in Thai fish sauce samples were analyzed. In order to develop the quantitative models, the concentrations of measured chemical and physical parameters (Y-data) were used to develop the PLS models as shown in section 1.4. Among the samples analyzed for the different parameters, seventy samples were randomly selected for the calibration set and the remaining samples were used for the prediction set. The statistical characteristics of calibration and prediction sets are summarized in Table 7.

Table 7 Chemical and physical mean, range and standard deviation (SD) values of Thai fish sauce samples in the NIR calibration and prediction sets.

Analytes	Calibration set (n=70)			Prediction set (n=30)		
	Mean	Range	SD	Mean	Range	SD
Total nitrogen *(%w/v)*	1.46	0.23-2.83	0.79	1.55	0.29-2.77	0.79
Sodium chloride *(%w/v)*	20.17	12.96-25.46	3.04	20.52	14.50-25.23	2.77
pH (-)	5.04	4.12-5.92	0.39	5.10	4.29-5.84	0.36
Reducing sugar (mg/mL)	1.00	0.13-2.38	0.57	1.04	0.23-2.41	0.59
Density *(g/cm^3)*	1.21	1.10-1.80	0.08	1.21	1.13-1.33	0.04
Baume *(°Baume)*	23.98	13.60-28.80	4.62	24.57	15.80-28.75	4.24
Total soluble solid *(°Brix)*	35.32	20.00-44.80	7.23	36.19	22.00-44.50	6.77
Refractive index (-)	1.39	1.36-1.41	0.01	1.39	1.37-1.41	0.01
Color: CIE Lab scale						
L*	12.93	6.71-24.39	4.57	12.62	7.33-23.84	3.98
a*	2.51	-0.46-10.10	2.39	2.46	-0.20-7.92	2.06
b*	16.64	8.77-37.20	6.24	15.87	9.36-27.49	4.78

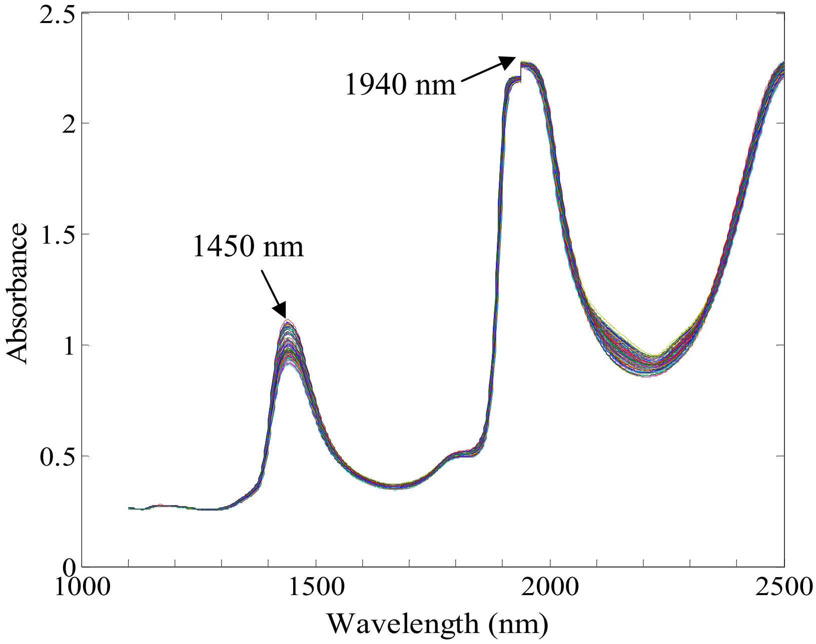
As shown in Table 7, Thai commercial fish sauce samples showed differences in the chemical and physical characteristics. These data, especially for total nitrogen content (TN), indicate that the fish sauce samples used in this study consisted of three typical grades as following the Ministry of Public Health (2000); standard pure fish sauce (TN≥0.9%w/v), standard mixed fish sauce (TN≥0.4%w/v), and out of standard (TN< 0.4%w/v). Therefore, these samples containing difference levels of total nitrogen content would be suitable for developing the quantitative and qualitative models.

1.2 NIR Spectra

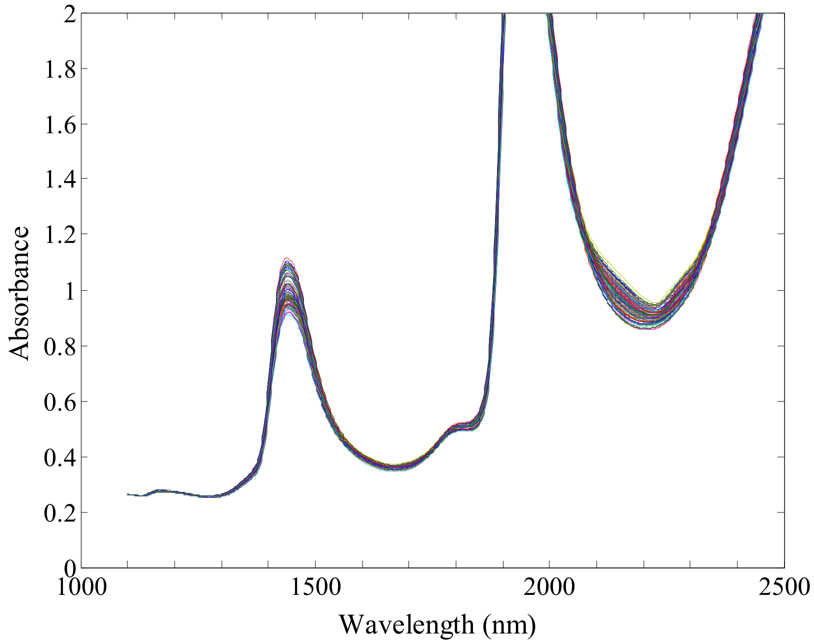
The NIR transflectance spectra in the region of 1100-2500 nm of 100 Thai fish sauce samples are shown in Figure 21(a). This figure shows the major absorption bands of water at 1450 nm and 1940. The position of these bands can be shifted by temperature changes or hydrogen boding interactions with sample components. The band at 1450 nm corresponds to the first, second and third O-H stretching overtone; the band at 1940 nm are O-H stretching and bending combination band (Maeda *et al*., 1995; Blanco *et al*., 2000). NIR absorbance bands rely on anharmonicity in the overtones of fundamental vibrations with energies in the infrared region. It can be difficult to assign wavelengths to structures because of matrix-dependent wavelength shifts. Therefore, chemometrics are used to correlate spectral features (X-data) with analyte concentrations (Y-data) (Osborne *et al*., 1993).

For chemometric analysis, the 1900-2000 nm region was not employed in order to avoid heavily overlapping absorption bands. William and Norris (1990) explained the phenomenon of overlapping is caused by a problem with gratings that does not exist with prism in the NIR instrument, that is light with several wavelengths leaves the grating at the same angle of dispersion. Figure 21(b) shows the NIR transflectance spectra in the regions of 1100-1900 and 2000-2440 nm of 100 Thai fish sauce samples used for quantitative modeling.

Both data in section 1.1 (Y-data) and section 1.2 (X-data) were applied to develop the PLS models as shown in section 1.4. Before developing the PLS models, the suitable input wavelength variables were investigated by the two wavelength interval selection methods named i) moving window partial least squares regression (MWPLSR) and ii) searching combination moving window partial least squares (SCMWPLS) as shown in section 1.3.1 and 1.3.2, respectively.



(a)



(b)

Figure 21 NIR transflectance spectra of 100 Thai fish sauce samples. (a) the NIR region of 1100-2500 nm and (b) the NIR region of 1100-1900 and 2000-2440 was used for chemometric analysis.

1.3 Wavelength Interval Selection Methods

In the quantitative determination of the chemical and physical parameters of Thai fish sauces; two wavelength interval selection methods were applied. Moving window partial least squares regression (MWPLSR) was applied for selecting the informative regions and subsequently searching combination moving window partial least squares (SCMWPLS) was used to combine and optimize the informative regions.

1.3.1 Moving Window Partial Least Squares Regression (MWPLSR)

MWPLSR was performed to search informative regions from the spectral regions of 1100-1900 and 2000-2440 nm. Informative regions mean that they contain useful information for a PLS model building and are helpful to improve the performance of the model (Jiang *et al.*, 2002). Figure 22 shows residue lines (Log(SSR)) for total nitrogen content obtained by MWPLSR.

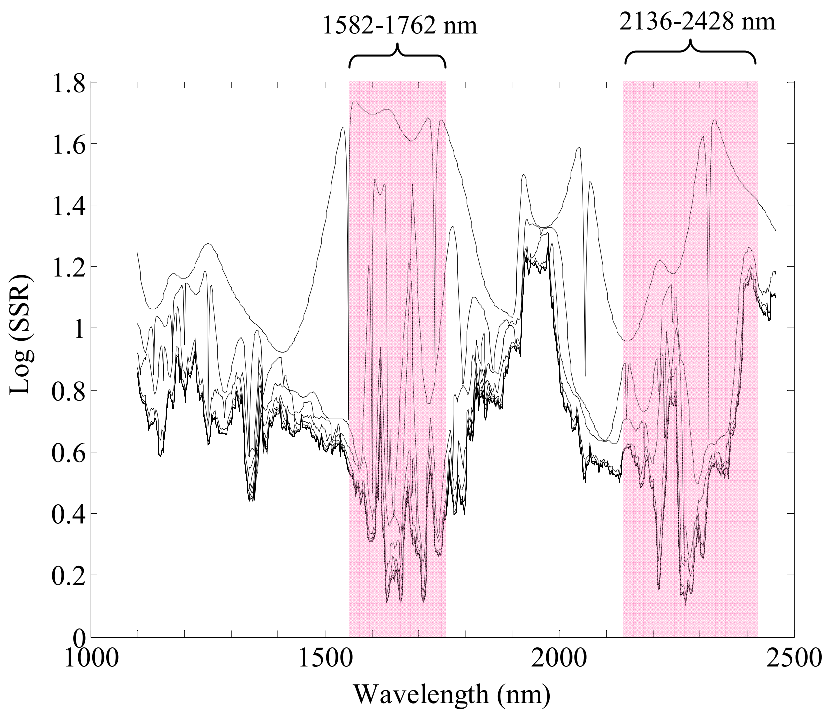
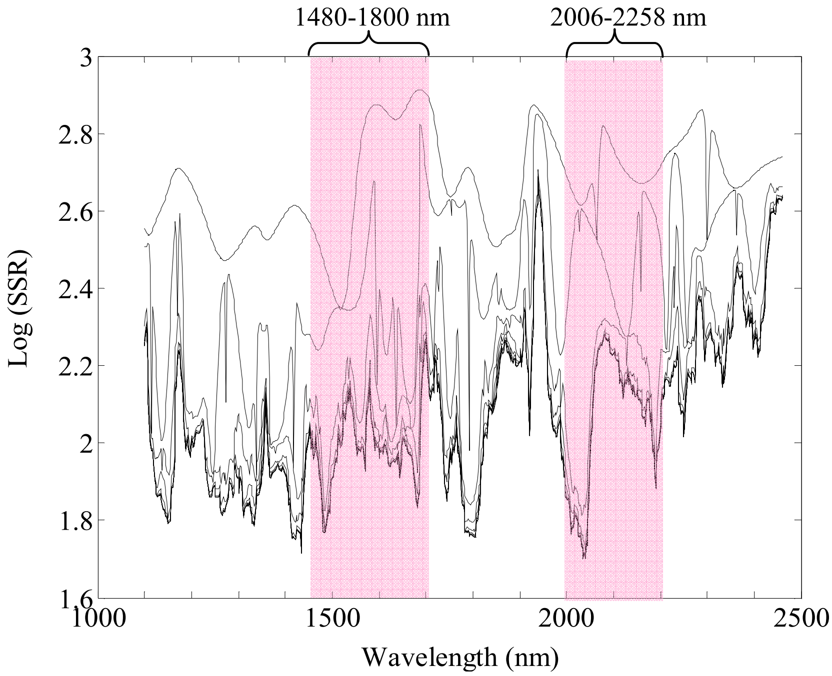


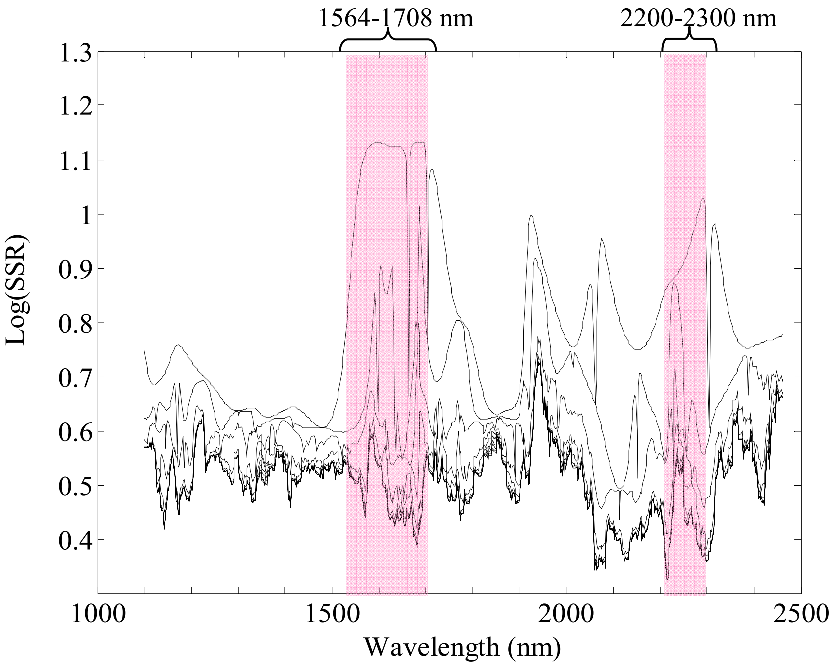
Figure 22 Residue lines for total nitrogen content of Thai fish sauces obtained by moving window partial least squares regression (MWPLSR). The shade areas are final informative regions.

It is noted in Figure 22 that there are two informative spectra regions of 1582-1762 and 2136-2428 nm, which give small values of SSR (the sum of squared residues) or the minimum error level. These informative regions contain useful information for PLS model building of TN content. The first informative region contains the 1590-1650 nm region, where bands due to the first overtones of NH stretching modes of proteins and amino acids appear (Williams and Norris, 1990). The second informative region contains the 2140-2170 and 2200-2250 nm regions, where several bands arising from the combinations of amide modes are located (Williams and Norris, 1990; Siesler *et al*., 2002).

The residue lines for other ten components i.e. sodium chloride, pH, reducing sugar, density, baume, total soluble solid, refractive index, color L*, color a*, and color b*, obtained by MWPLSR are shown in Figure 23 (a), (b), (c), (d), (e), (f), (g), (h), (i), and (j) respectively. Like total nitrogen content, these components have a few informative regions, which give small values of SSR. The informative spectra regions of total nitrogen content and those parameters are presented in Table 8. All regions correspond to the regions for the combinations of stretching and deformation modes of amino acids, which are one of the most important effective parameters for the qualities of fish sauces. Amino acids show common bands in the 1500-1540 nm and 1980-2010 nm regions due to the first overtone of NH_2 stretching mode and the combination of NH_2 stretching and NH_2 deformation modes, respectively (Siesler *et al*., 2002). These results indicated that the chemical and physical properties of fish sauces are related to the amino acids. As mentioned previously, fish proteins are degraded by fermentation to amino acids that producing a wide variety of tastes, flavors, colors and compositions of fish sauces (Gildberg, 2001; Shih *et al*., 2003).

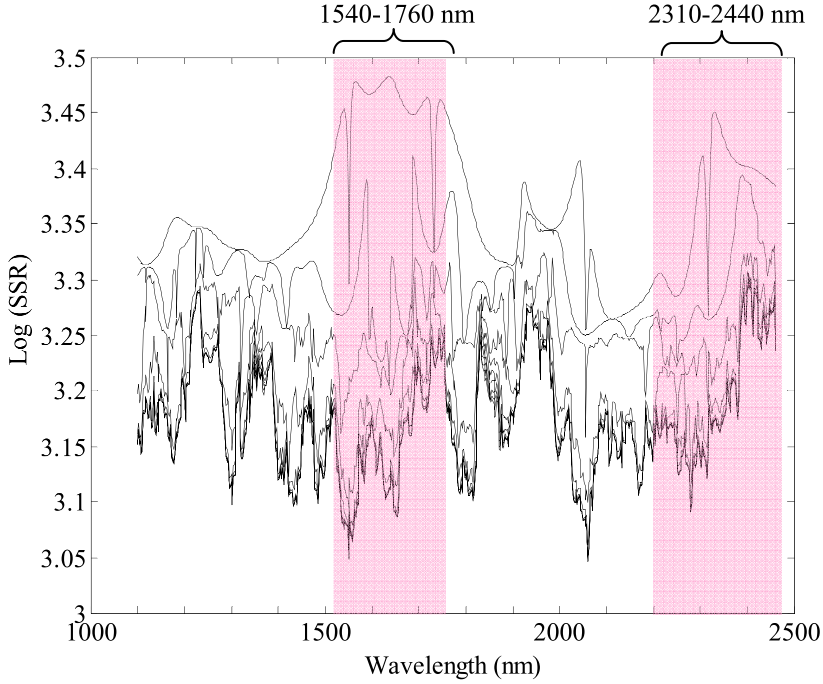


(a) Sodium chloride

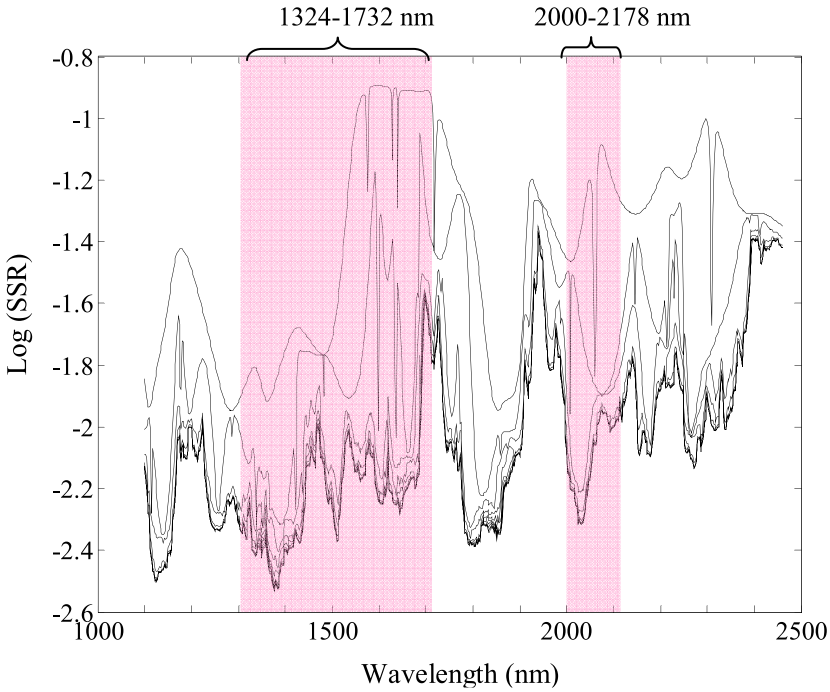


(b) pH

Figure 23 Residue lines for chemical and physical parameters of Thai fish sauces obtained by moving window partial least squares regression (MWPLSR). The shade areas are final informative regions.

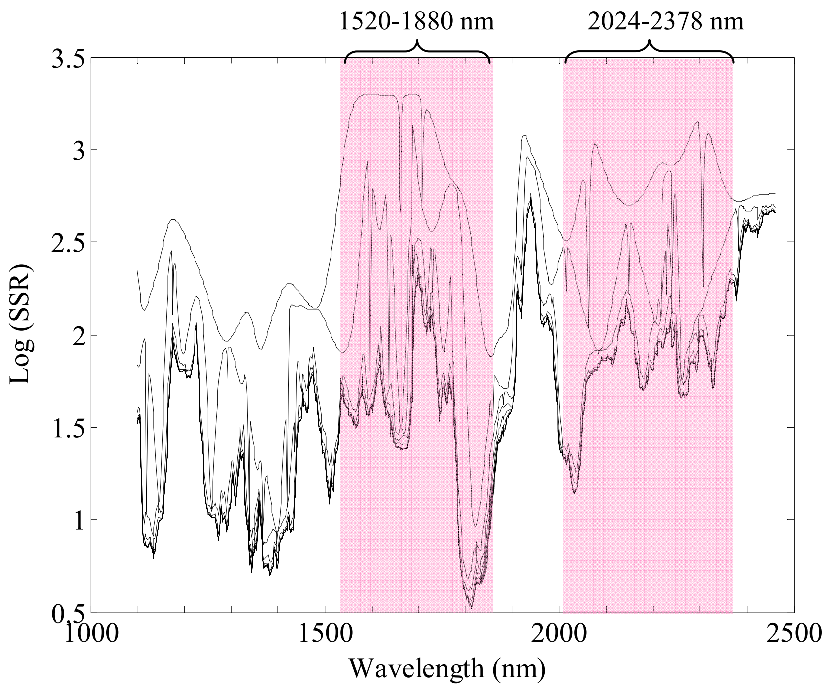


(c) Reducing sugar

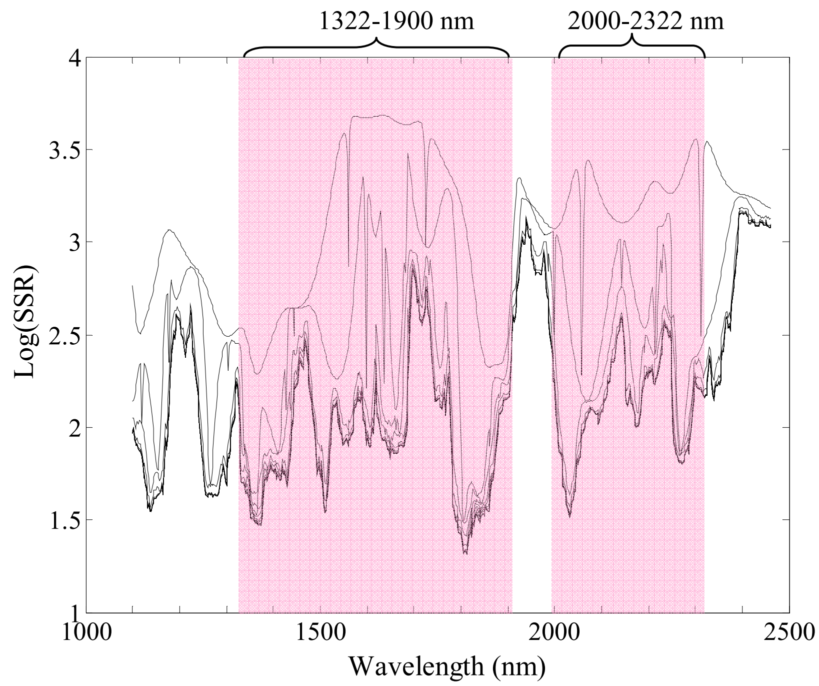


(d) Density

Figure 23 (Continued)

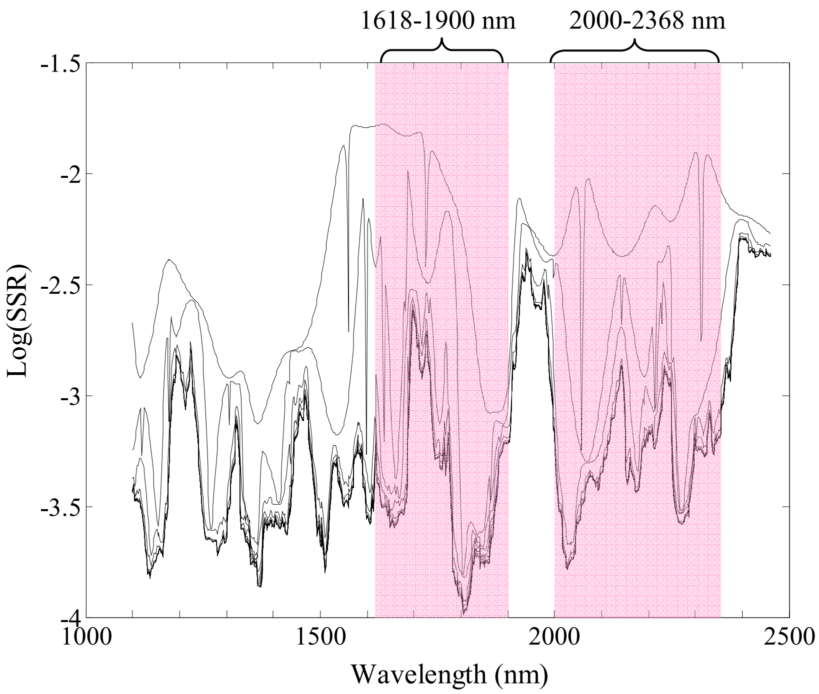


(e) Baume

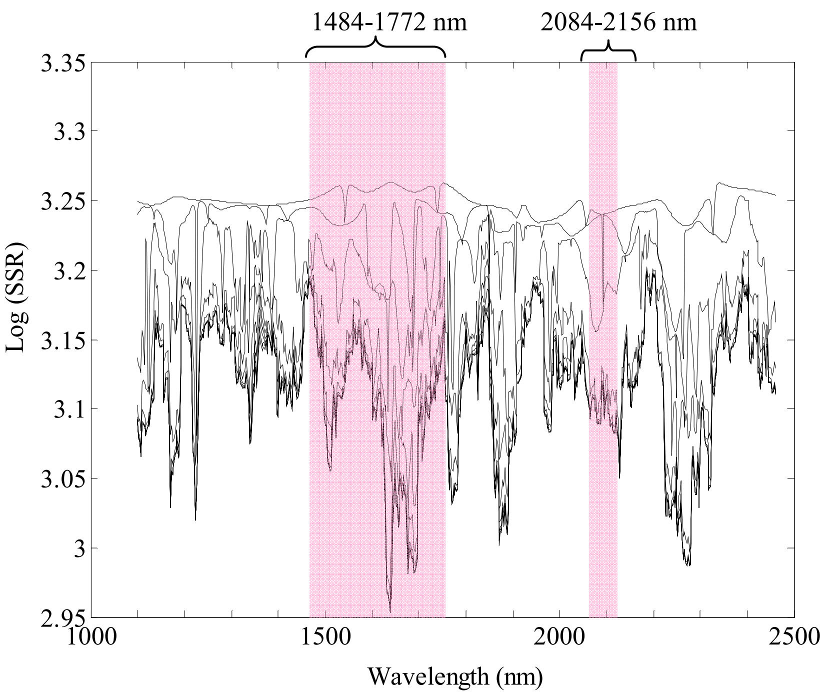


(f) Total soluble solid

Figure 23 (Continued)

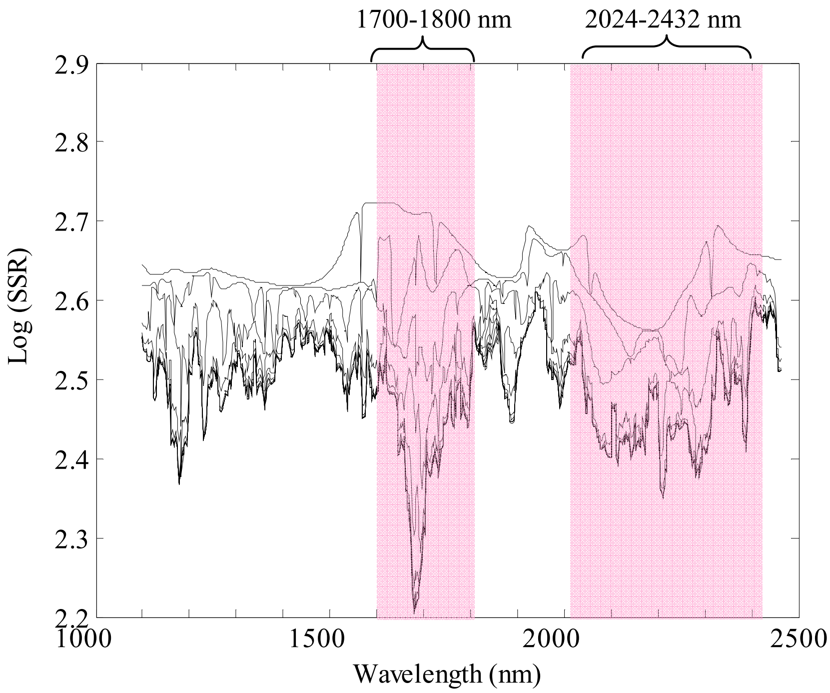


(g) Refractive index

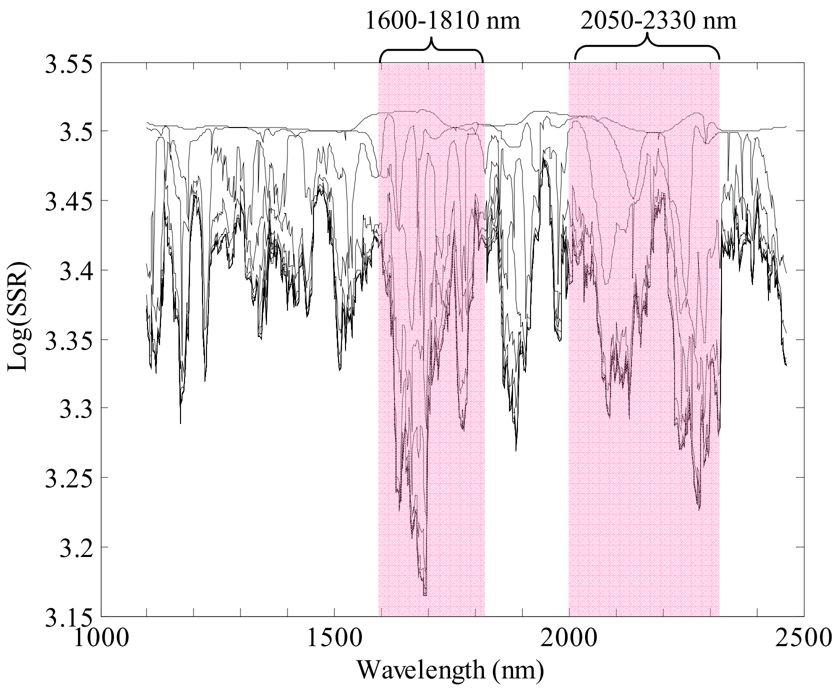


(h) Color L*

Figure 23 (Continued)



(i) Color a*



(j) Color b*

Figure 23 (Continued)

Table 8 Chemical assignments of spectral regions suggested by moving window partial least squares regression (MWPLSR) for the chemical and physical parameters of Thai fish sauces.

Analytes	Informative regions (nm)	Functional groups and molecules[1]
Total nitrogen	1582-1762,	CH_3, CH_2, -SH, $CONH_2$
	2136-2428	Amino acids, HC=CH, $CONH_2$, CONHR, Protein, -CHO, CH_3
Sodium chloride	1480-1800,	$CONH_2$, CONHR, NH, Protein, ROH, RNH_2, C=H, -CONH-, $=CH_2$, CH_2, CH_3, -SH,
	2006-2258	Amino acids, $CONH_2$, CONHR, Protein,
pH	1564-1708,	-CONH-, CH_2, CH_3
	2200-2300	-CHO, Amino acids, CH_2
Reducing sugar	1540-1760,	-CONH-, CH_2, CH_3
	2310-2440	-CHO, Amino acids, CH_2
Density	1324-1732,	CH_2, CH_3, $CONH_2$, CONHR, NH, Protein, RNH_2
	2000-2178	Amino acids, HC=CH, $CONH_2$, CONHR, Protein,
Baume	1520-1880,	ROH, RNH_2, C=H, -CONH-, $=CH_2$, CH_2, CH_3, -SH,
	2024-2378	Amino acids, $CONH_2$, CONHR, Protein, -CHO, CH_2
Total soluble solid	1322-1900,	CH_2, CH_3, $CONH_2$, CONHR, NH, Protein, RNH_2 , $-CO_2H$
	2000-2322	Amino acids, $CONH_2$, CONHR, Protein, -CHO, CH_2
Refractive index	1618-1900,	$=CH_2$, CH_2, CH_3, -SH, $-CO_2H$
	2000-2368	Amino acids, $CONH_2$, CONHR, Protein, -CHO, CH_2

Table 8 (Continued)

Analytes	Informative regions (nm)	Functional groups and molecules[1]
Color		
L*	1484-1772,	$CONH_2$, CONHR, NH, Protein, ROH, RNH_2, C=H, -CONH-, $=CH_2$, CH_2, CH_3, -SH,
	2084-2156	ROH, $CONH_2$, CONHR, Amino acids, HC=CH,
a*	1700-1800,	$=CH_2$, CH_2, CH_3, -SH, $-CO_2H$
	2024-2432	$CONH_2$, CONHR, Amino acids, HC=CH, Protein, $HC=CHCH_2$,
b*	1600-1810,	$=CH_2$, CH_2, CH_3, -SH, $-CO_2H$
	2050-2330	Amino acids, $CONH_2$, CONHR, Protein, -CHO, CH_2

[1]Williams and Norris (1990) and Siesler *et al*. (2002)

1.3.2 Searching Combination Moving Window Partial Least Squares (SCMWPLS)

SCMWPLS was performed to search for optimized combinations of the informative regions obtained by MWPLSR. In order to eliminate uninformative regions, SCMWPLS optimizes the informative regions and combines them altogether (Kasemsumran *et al*., 2003; Du *et al*., 2004; Kasemsumran *et al*., 2004). The optimized combination of informative regions obtained by SCMWPLS are presented in Table 9.

Table 9 Selected wavelength variables obtained by moving window partial least square regression (MWPLSR) and searching combination moving window partial least squares (SCMWPLS) for the chemical and physical parameters of Thai fish sauces.

Analytes	Selected wavelength variables (nm)	
	MWPLSR method	SCMWPLS method
Total nitrogen	1582-1762, 2136-2428	2264-2428
Sodium chloride	1480-1800, 2006-2258	1480-1798, 2252-2258
pH	1564-1708, 2200-2300	1676-1708, 2208-2260
Reducing sugar	1540-1760, 2310-2440	1608-1760
Density	1324-1732, 2000-2178	1358-1438
Baume	1520-1880, 2024-2378	1580-1670, 2224-2354
Total soluble solid	1322-1900, 2000-2322	1322-1442, 2000-2076
Refractive index	1618-1900, 2000-2368	1774-1846, 2078-2114
Color L*	1484-1772, 2084-2156	2122-2272
Color a*	1700-1800, 2024-2432	1164-1288
Color b*	1600-1810, 2050-2330	1696-1726, 2086-2158

To demonstrate the performance of these two wavelength selection methods, four kinds of input wavelength variables (X-data) were applied to develop the PLS calibration models as shown in section 1.4. These four input variables were i) the whole spectra region (section 1.2), ii) the informative regions obtained by MWPLSR (section 1.3.1), iii) the direct combination of informative regions and iv) the optimized combination of informative regions obtained by SCMWPLS (section 1.3.2). In order to compare the performance of developed models, the root mean square errors of prediction (RMSEP) were calculated. The best prediction model was selected by considering the model yielded the lowest RMSEP value.

1.4 Model analysis

There are four main differences between input wavelength variables (X-data) for developing the PLS models. A first input wavelength variable is the whole spectra region (1100-1900 nm and 2000-2440 nm) as shown in section 1.2. The second difference input variable is the selected wavelengths obtained by MWPLSR method as shown in section 1.3.1. The third difference input variable is the direct combination of selected wavelengths obtained by MWPLSR. The fourth difference input variable is the selected wavelengths obtained by SCMWPLS method as shown in section 1.3.2. These four differences input wavelength variables were used to build PLS models with the calibration set (n=70). The performance of these models was evaluated by using the prediction set (n=30). The calculated root-mean-square error of prediction (RMSEP), correlation coefficient and the selected PLS factor numbers are summarized in Table 10.

Table 10 Prediction results of PLS calibration models for total nitrogen content, sodium chloride, pH, reducing sugar, density, baume, total soluble solid, refractive index, color (L*, a*, b*) of Thai fish sauces.

Analytes	Wavelength selection methods	Spectra regions (nm)	PLS factors	R	RMSEP
1. Total nitrogen (%w/v)	Full spectra	1100-1900, 2000-2440	8	0.987	0.131
	MWPLSR[1]	1582-1762	5	0.989	0.120
	MWPLSR[2]	2136-2428	4	0.991	0.106
	MWPLSR[3]	1582-1762, 2136-2428	5	0.991	0.107
	SCMWPLS[4]	2264-2428	5	0.992	0.100
2. Sodium chloride (%w/v)	Full spectra	1100-1900, 2000-2440	5	0.966	0.699
	MWPLSR[1]	1480-1800	5	0.965	0.719
	MWPLSR[2]	2006-2258	4	0.964	0.729
	MWPLSR[3]	1480-1800, 2006-2258	5	0.970	0.664
	SCMWPLS[4]	1480-1798, 2252-2258	5	0.972	0.647
3. pH (-)	Full spectra	1100-1900, 2000-2440	7	0.919	0.158
	MWPLSR[1]	1564-1708	5	0.868	0.184
	MWPLSR[2]	2200-2300	4	0.903	0.159
	MWPLSR[3]	1564-1708, 2200-2300	6	0.896	0.161
	SCMWPLS[4]	1676-1708, 2208-2260	5	0.909	0.155
4. Reducing Sugar (mg/mL)	Full spectra	1100-1900, 2000-2440	9	0.662	0.440
	MWPLSR[1]	1540-1760	7	0.690	0.438
	MWPLSR[2]	2310-2440	3	0.644	0.475
	MWPLSR[3]	1540-1760, 2310-2440	4	0.618	0.482
	SCMWPLS[4]	1608-1760	6	0.745	0.407

R: Correlation coefficient, RMSEP: Root mean square error of prediction.
MWPLSR: Moving window partial least squares regression
SCMWPLS: Searching combination moving window partial least squares
[1]The method was used to search for the informative regions in the region of 1100-1900 nm.
[2]The method was used to search for the informative regions in the region of 2000-2440 nm.
[3]The method was used to combine the informative regions obtained by MWPLSR.
[4]The method was used to optimize the combination of the informative regions obtained by MWPLSR.

Table 10 (Continued)

Analytes	Wavelength selection methods	Spectra regions (nm)	PLS factors	R	RMSEP
5. Density (g/cm^3)	Full spectra	1100-1900, 2000-2440	5	0.962	0.009
	MWPLSR[1]	1324-1732	4	0.969	0.008
	MWPLSR[2]	2000-2178	4	0.954	0.010
	MWPLSR[3]	1324-1732, 2000-2178	5	0.964	0.009
	SCMWPLS[4]	1358-1438	2	0.977	0.007
6. Baume (°Baume)	Full spectra	1100-1900, 2000-2440	8	0.999	0.120
	MWPLSR[1]	1520-1880	6	0.999	0.133
	MWPLSR[2]	2024-2378	8	0.998	0.251
	MWPLSR[3]	1520-1880, 2024-2378	8	0.999	0.152
	SCMWPLS[4]	1580-1670, 2224-2354	8	0.999	0.118
7. Total soluble solid (°Brix)	Full spectra	1100-1900, 2000-2440	9	0.997	0.507
	MWPLSR[1]	1322-1900	6	0.998	0.470
	MWPLSR[2]	2000-2322	6	0.998	0.458
	MWPLSR[3]	1322-1900, 2000-2322	4	0.998	0.481
	SCMWPLS[4]	1322-1442, 2000-2076	5	0.998	0.435
8. Refractive Index (-)	Full spectra	1100-1900, 2000-2440	7	0.997	0.00092
	MWPLSR[1]	1618-1900	5	0.998	0.00084
	MWPLSR[2]	2000-2368	7	0.998	0.00088
	MWPLSR[3]	1618-1900, 2000-2368	6	0.997	0.00091
	SCMWPLS[4]	1774-1846, 2078-2114	5	0.998	0.00079

R: Correlation coefficient, RMSEP: Root mean square error of prediction.
MWPLSR: Moving window partial least squares regression
SCMWPLS: Searching combination moving window partial least squares
[1]The method was used to search for the informative regions in the region of 1100-1900 nm.
[2]The method was used to search for the informative regions in the region of 2000-2440 nm.
[3]The method was used to combine the informative regions obtained by MWPLSR.
[4]The method was used to optimize the combination of the informative regions obtained by MWPLSR.

Table 10 (Continued)

Analytes	Wavelength selection methods	Spectra regions (nm)	PLS factors	R	RMSEP
9. Color L*	Full spectra	1100-1900, 2000-2440	12	0.510	3.940
	MWPLSR[1]	1484-1772	3	0.505	3.349
	MWPLSR[2]	2032-2272	6	0.585	3.160
	MWPLSR[3]	1484-1772, 2032-2272	4	0.490	3.401
	SCMWPLS[4]	2122-2272	6	0.672	2.914
10. Color a*	Full spectra	1100-1900, 2000-2440	11	0.615	2.082
	MWPLSR[1]	1700-1800	6	0.760	1.516
	MWPLSR[2]	2024-2432	5	0.737	1.528
	MWPLSR[3]	1700-1800, 2024-2432	8	0.510	2.085
	SCMWPLS[4]	1164-1288	5	0.878	1.023
11. Color b*	Full spectra	1100-1900, 2000-2440	4	0.443	5.197
	MWPLSR[1]	1600-1810	6	0.483	4.947
	MWPLSR[2]	2050-2330	6	0.381	5.360
	MWPLSR[3]	1600-1810, 2050-2330	7	0.437	5.120
	SCMWPLS[4]	1696-1726, 2086-2158	6	0.604	4.803

R: Correlation coefficient, RMSEP: Root mean square error of prediction.
MWPLSR: Moving window partial least squares regression
SCMWPLS: Searching combination moving window partial least squares
[1]The method was used to search for the informative regions in the region of 1100-1900 nm.
[2]The method was used to search for the informative regions in the region of 2000-2440 nm.
[3]The method was used to combine the informative regions obtained by MWPLSR.
[4]The method was used to optimize the combination of the informative regions obtained by MWPLSR.

It can be seen from Table 10 that the PLS models based on the informative regions selected by MWPLSR show better performance than that based on the whole spectra region. With the optimized combination of informative regions obtained by SCMWPLS, the developed model improves the prediction ability significantly and gives the best prediction results. The lowest RMSEP for total nitrogen content, sodium chloride, pH, reducing sugar, density, baume, total soluble solid, refractive index, color L*, color a*, and color b* were 0.100%w/v, 0.647%w/v, 0.155, 0.407 mg/mL, 0.007 g/cm^3, 0.118 °Baume, 0.435 °Brix, 0.00079, 2.914, 1.023, and 4.803, respectively. Of particular note in Table 10 is that the optimization of the informative regions by SCMWPLS is more robust to optimize the informative regions than the direct combination of informative regions.

From all the results above, we can conclude that i) NIR-PLS method can be used to develop the quantitative models for the chemical and physical of Thai fish sauces, ii) the combination of MWPLSR and SCMWPLS is powerful to find out the suitable input wavelength variable, which can be used to improve the performance of a PLS model with low RMSEP, small number of PLS factors, high correlation coefficients and small number of data points, iii) the wavelength interval selection methods have the advantages of being able to handle collinear X-variables in NIR data which are many redundant variables and highly correlated and iv) the small number of data obtained by wavelength interval selection methods can be used to avoid causing the over-fitting problem due to selecting too many wavelengths.

2. **Development of Qualitative Models for Classifying Thai Fish Sauces Based on their Total Nitrogen Content**

For the classification of Thai fish sauces based on their total nitrogen content, five supervised pattern recognition methods were used to develop the classification models. These methods were i) Linear discriminant analysis (LDA), ii) Factor analysis-Linear discriminant analysis (FALDA), iii) Soft independent modeling of class analog (SIMCA), iv) K nearest neighbors (KNN) and v) Artificial neural networks (ANNs). The selected wavelengths for the total nitrogen content obtained by SCMWPLS (2264-2428 nm) shown in section 1.3.2 were used as input variables (X-data). The purposes of using the selected wavelengths for developing the classification models were to reduce the number of variables (wavelengths) and avoid the ill-posed problem. The output variables (Y-data) were the category memberships of samples for standard pure fish sauce (SPF), standard mixed fish sauce (SMF), and out of standard fish sauce (OF). The three steps for developing the classification models were i) categorization the memberships of Thai fish sauce samples, ii) Selecting the input wavelength variables and iii) developing the model by using the classification rules.

2.1 Categorization the Memberships of Thai Fish Sauce Samples

According to the principle of supervised pattern recognition methods, the known category memberships of samples are required. In this study, the classification models were used to classify fish sauce based on their total nitrogen following the Notification of the Ministry of Public Health (2000). Therefore, the total nitrogen content of a hundred fish sauce samples determined in section 1.1 was used. All samples were classified into three groups based on their total nitrogen content; i) standard pure fish sauce (TN $\geq$ 0.9 %w/v), ii) standard mixed fish sauce (TN $\geq$ 0.4 %w/v), and iii) out of standard fish sauce (TN $<$ 0.4 %w/v). Table 11 shows the number of samples belonging to three groups of Thai fish sauces and their total nitrogen contents.

Table 11 Minimum (Min), maximum (Max), mean, and standard deviation (SD) values of total nitrogen content in each group of Thai fish sauce.

Groups of fish sauce	Number of samples	Total nitrogen content (%w/v)			
		Min	Max	Mean	SD
1. Standard pure fish sauce (SPF) ($TN \geq 0.9$ %w/v)	71	1.07	2.83	1.90	0.52
2. Standard mixed fish sauce (SMF) ($TN \geq 0.4$ %w/v)	16	0.41	0.76	0.59	0.09
3. Out of standard fish sauce (OF) ($TN < 0.4$ %w/v)	13	0.23	0.39	0.32	0.05

The total nitrogen content was determined by the Kjeldahl method (AOAC, 2000).

For classification, sample of each three groups were split randomly into two sets, the training and test sets. The training set was composed of 80 fish sauce samples having the number of SPF, SMF and OF samples 55, 14, and 11, respectively. The test set consisted of the remaining samples having the number of SPF, SMF and OF samples 16, 2, and 2, respectively. The training set was used to establish the model, and the test set was used to validate the SIMCA, KNN and ANNs models. The LDA and FALDA models were validated by using the cross-validation methods because these methods can be optimally applied when the data sets have more objects than variables.

2.2 NIR Spectra

The selected wavelength region for the total nitrogen content (2264-2428 nm) obtained by SCMWPLS method in section 1.3.2 were used to develop the classification models as input variables (X-data). The 83 input wavelength variables from 2264 to 2428 nm at 2 nm intervals and the groups of fish sauces (section 2.1) were applied to the five supervised pattern recognition methods as shown in section 2.3. The purposes of using the selected wavelength regions were i) to select the input wavelength variables that are meaningful for the classification and elimination of those that have no discriminating and ii) to reduce the redundant variables in NIR data that can be solved the ill-problem in some classification methods. Figure 24 shows the NIR transflectance in the spectra region of 2264-2428 of three groups of Thai fish sauces.

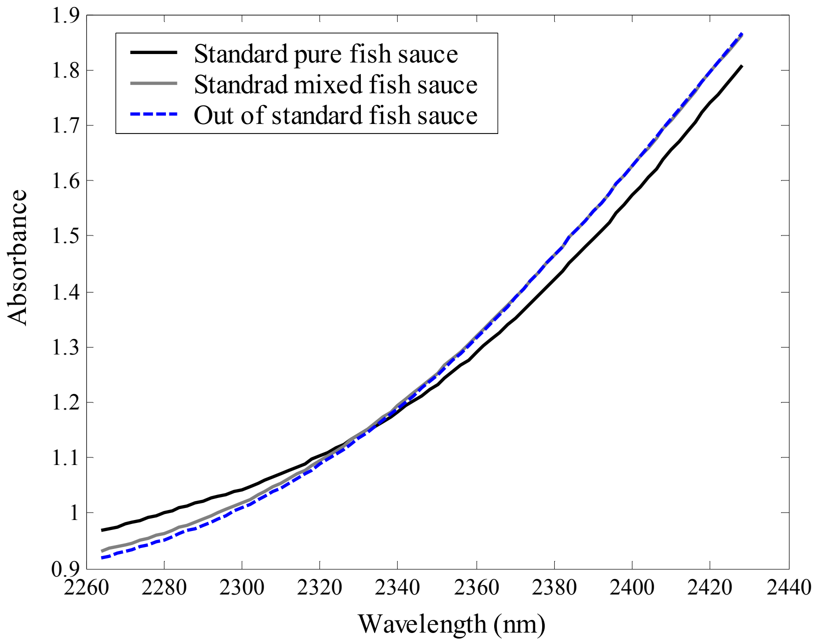


Figure 24 The average NIR transflectance in the spectra region of 2264-2428 nm corresponding to the total nitrogen content for three groups of Thai fish sauces.

2.3 Classification Modeling

Five supervised pattern recognition methods were used to develop the five classification models. Both data in section 2.1 (Y-data) and section 2.2 (X-data) were used to develop the models. The accuracy of developed model was calculated as the corrective classification rate (%). The classification results for differences five models developed by five supervised pattern recognition methods are followings:

2.3.1 Linear Discriminant Analysis (LDA)

The original 83 wavelengths from 2264 to 2428 nm at 2 nm intervals were used as input variables (X-data) to develop the LDA model. The stepwise method was applied to select the predictors. An important reason for eliminating predictors is that many predictors are often useless and undermine the predictive ability of the classification model. LDA can be optimally applied when the data sets have more objects than variables. Therefore, to solve this disadvantage, the cross validation method was used to validate the developed model. Two discriminant functions were developed to classify fish sauces into three groups. The stepwise LDA continued up to step 4 and the wavelengths of 2268, 2412, 2424 and 2428 nm were selected to build the discriminant function and their coefficients are:

$$\text{Function 1} = 1.00A_{2268} - 5.77A_{2412} + 0.65A_{2424} + 4.29A_{2428}$$
$$\text{Function 2} = 0.24A_{2268} - 4.72A_{2412} - 23.90A_{2424} + 19.38A_{2428}$$

where A_x is the absorbance of fish sauce at wavelength x.

With these two functions, the centroid for each group was computed. For an unknown sample *u* with variables values A_x, the same functions were calculated. Subsequently, the Euclidean distance between an unknown sample and the centroid for each group was calculated. An unknown *u* is classified with the class which has the lowest Euclidance distance. Figure 25 illustrates three groups scatter plots obtained by the LDA classified model. This plot shows the distribution of the centroid groups within a main overlap corresponding to SMF- and OF- groups. Table 12 shows the values of cross-validation of LDA where the discriminant functions were determined using the four important spectral differences at 2268, 2412, 2424 and 2428 nm. In total, this classification model attained 82.00% of cross-validated grouped cases correctly classified. The SPF-, SMF- and OF-groups were classified correctly 85.92%, 75.00% and 69.23%, respectively. Figure 25 illustrates three groups scatter plots obtained by the LDA classified model. This plot shows the distribution of the centroid groups within a main overlap corresponding to SMF and OF groups.

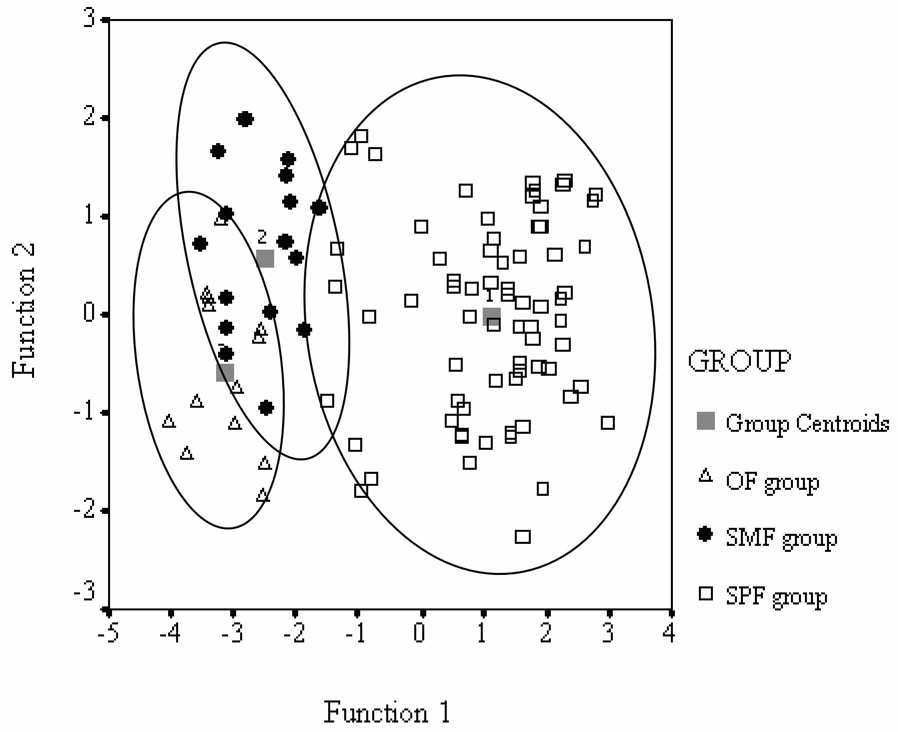


Figure 25 Three groups scatter plots obtained by the Linear Discriminant Analysis (LDA) classified model (□: standard pure fish sauces (SPF), •: standard mixed fish sauces (SMF), Δ: out of standard fish sauces (OF)).

Table 12 The cross-validation results of linear discriminant analysis (LDA) model for 100 Thai fish sauce samples.

Fish sauce groups	No. of samples	Predicted group memberships			Corrective classification rate (%)
		SPF	SMF	OF	
1) Standard pure fish sauce (SPF)	71	61	6	4	85.92
2) Standard mixed fish sauce (SMF)	16	0	12	4	75.00
3) Out of standard fish sauce (OF)	13	0	4	9	69.23
Total corrective classification rate (%)					82.00

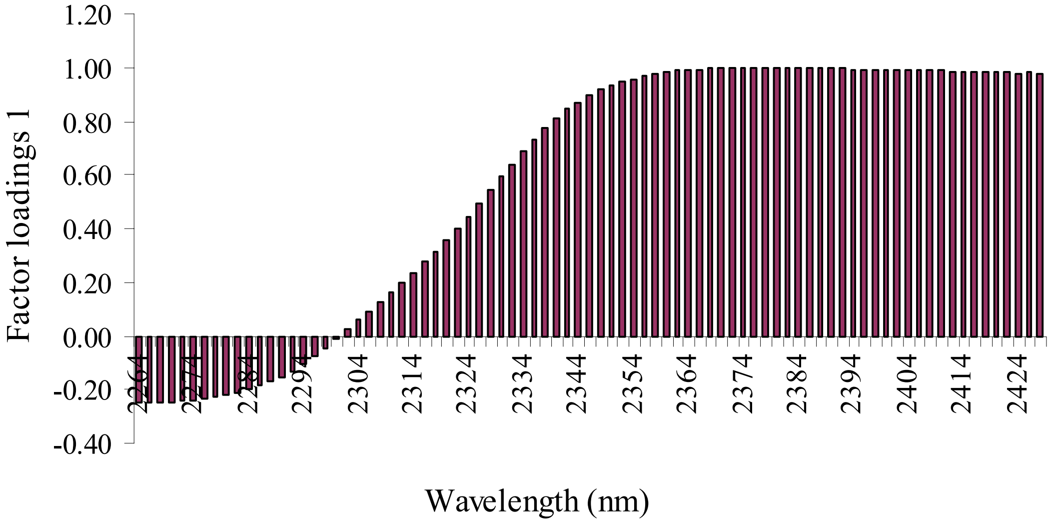
In cross-validation, each case is classified by the functions derived from all cases other than that case. The discriminant functions were developed using the wavelength variables of 2268, 2412, 2424 and 2428 nm selected by stepwise method.

2.3.2 Factor Analysis- Linear Discriminant Analysis (FALDA)

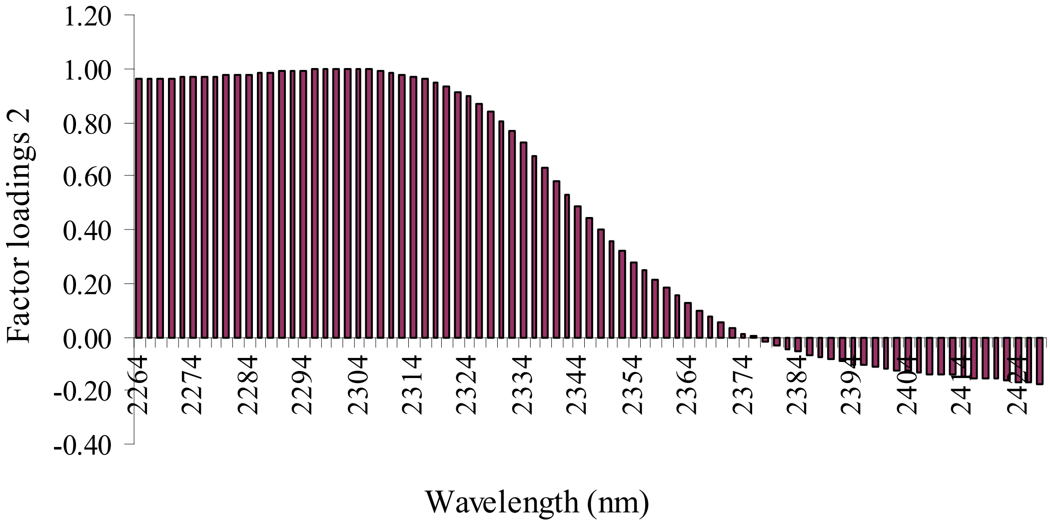
According to the disadvantage of LDA technique that can be optimally applied when the data sets have more objects, a multivariate technique for reducing the data is required to solve this problem. Factor analysis (FA), a multivariate technique for reducing the original variables by creating new variables called factors, was used in this study. For developing the FALDA model, the FA was applied in order to reduce the number of 83 wavelength variables from 2264-2428 nm at 2 nm intervals as factors (section a). Subsequently, the extracted factor scores were used as input variables (X-data) to develop the LDA model (section b).

(a) Factor Analysis of NIR Spectra

In FALDA, factor scores were extracted from a correlation matrix in FA prior LDA to include as many factors as possible. Major factors calculated from NIR data sets and eigenvalues of factors often reflect background effects and contain significant information about chemical components. From the FA applied to NIR spectra, the dimensionality of the data was reduced from 83 wavelength variables to two uncorrelated factors which explained a total of 99.7% of the variation. The factor 1 and factor 2 accounted 58.0% and 41.7% of variance, respectively. The factor loadings for the first two factors were calculated as shown in Figure 26. Factor loadings are the correlation coefficients between the wavelength variables and factors. Factor loadings are the basis for imputing a label to the different factors. Loadings above 0.6 are usually considered "high" and those below 0.4 are "low." It can be seen from the Figure 26 that the wavelength region of 2264-2334 was highly related to the factor 1, whereas the wavelength region of 2336-2428 was highly related to the factor 2. To obtain the factor scores, the factor score coefficient matrix for the first two factors were calculated as shown in Figure 27. Factor scores are the scores of each case on each factor. The factor score coefficients are used to calculate the factor scores of each case for each of the two factors. Subsequently, the calculated factor scores were used as input variable (X-data) to develop the LDA model.

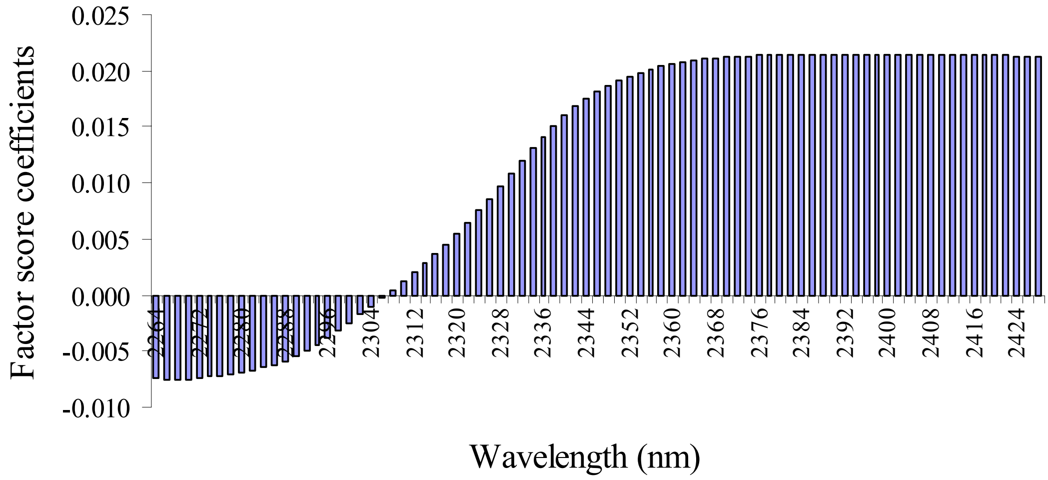


(a) Factor 1

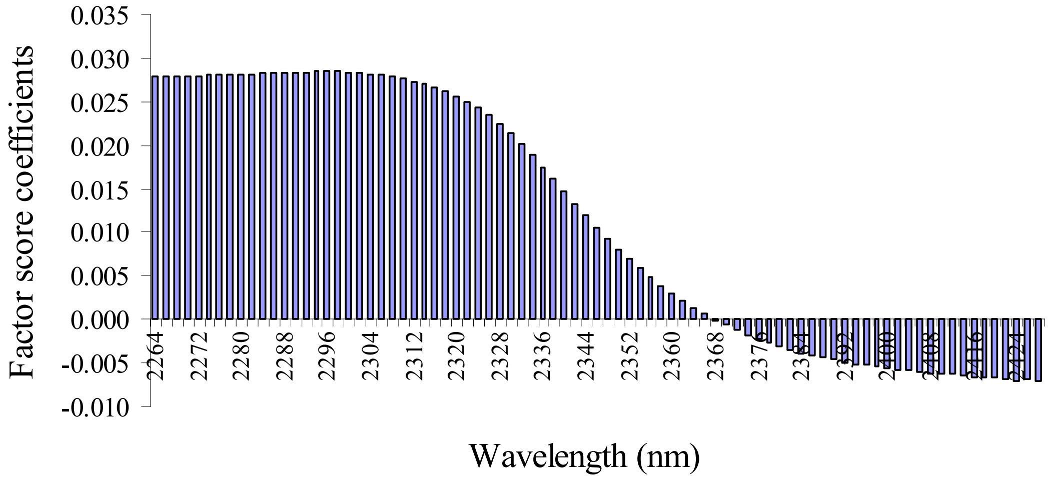


(b) Factor 2

Figure 26 Factor loadings for the first two factors extracted from the NIR region of 2264-2428 nm using factor analysis; (a) Factor 1 and (b) Factor 2. The extraction and rotation methods used in factor analysis were principal component analysis and Varimax with Kaiser Normalization methods, respectively.



(a) Factor 1



(b) Factor 2

Figure 27 Factor score coefficients based on the first two factors extracted from the NIR region of 2264-2428 nm using factor analysis; (a) Factor 1 and (b) Factor 2. The extraction and rotation methods used in factor analysis were principal component analysis and Varimax with Kaiser Normalization methods, respectively.

Figure 28 shows a distribution of 100 fish sauce samples based on factor scores for factor 1 and factor 2. This figure shows the tendency to form three groups of fish sauce samples. Factor 1 and 2 contain 58.0% and 41.7% of variance, respectively. Although more than 99% of variances were accumulated in the first two factors extracted from NIR spectra, samples were not clearly separated. Factor score plots shown in Figure 28 indicated great difficulty in discovering apparent difference between whole spectral patterns (2264-2428 nm) in three fish sauce groups. Therefore, the LDA techniques, one of supervised pattern recognition techniques, using a classification criterion were applied to find specific factors or regions where significant difference among SPF-, SMF- and OF- groups could be expected.

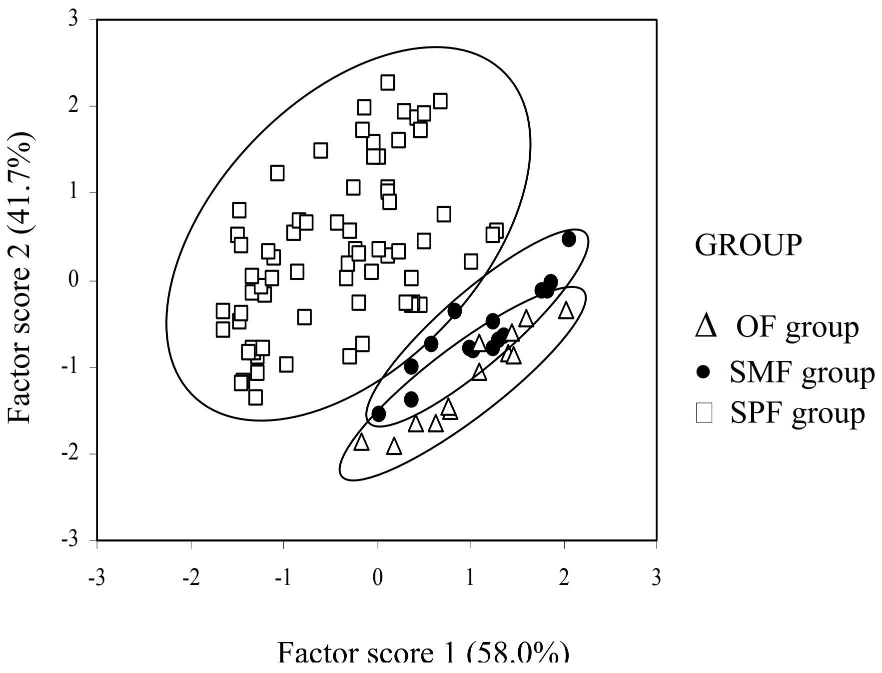


Figure 28 Distribution of 100 fish sauce sample based on factor scores of the first two factors calculated from the spectra region of 2264-2428 nm (□: standard pure fish sauces, •: standard mixed fish sauces, Δ: out of standard fish sauces).

(b) Factor Analysis- Linear Discriminant Analysis (FALDA) model

From Factor analysis result as shown in section (a), the factor scores based on the first two factors which accounted for 99.7% of variation were applied to LDA as predictors (X-data). The stepwise method was used to select the suitable predictors for developing the model. For comparing the performance of LDA and FALDA models, the cross validation test was used to validate the classified model. Like LDA model, two discriminant functions were developed. The stepwise LDA continued up to step 2 and both of factors were selected to build the discriminant function and their coefficients are:

$$\text{Function 1} = 1.13\,(\textit{factor scores}\_\mathrm{F}_1) - 1.07\,(\textit{factor scores}\_\mathrm{F}_2)$$
$$\text{Function 2} = 0.50\,(\textit{factor scores}\_\mathrm{F}_1) + 0.62\,(\textit{factor scores}\_\mathrm{F}_2)$$

where *factor scores*_F_x are the factor scores based on the factor X extracted from the NIR region of 2264-2428 nm using factor analysis. The extraction and rotation methods used in factor analysis were principal component analysis and Varimax with Kaiser Normalization methods, respectively.

With these two functions, the centroid for each group was computed. For an unknown sample u with variables values A_x, the same functions were calculated. Subsequently, the Euclidean distance between an unknown sample and the centroid for each group was calculated. An unknown u is classified with the class which has the lowest Euclidance distance. Figure 29 shows three groups scatter plots obtained by the FALDA classified model. This plot shows the distribution of the centroid groups within a main overlap corresponding to SMF- and OF- groups. Table 13 shows the values of cross-validation of LDA where the discriminant function were determined using the two factors extracted from the 2264-2428 nm region of the NIR spectra. In total, the FALDA classified model attained 85.0% of cross-validated grouped cases correctly classified. The SPF-, SMF- and OF-groups are classified correctly 87.32%, 81.25% and 76.92%, respectively.

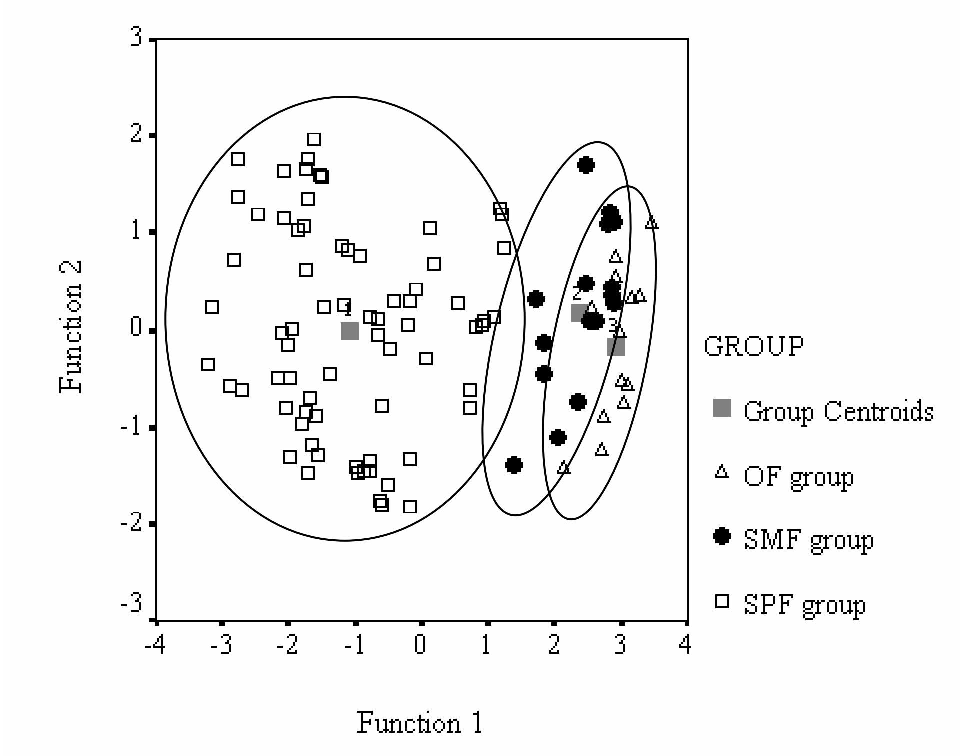


Figure 29 Three groups scatter plots obtained by the Factor Analysis- Linear Discriminant Analysis (FALDA) classified model (□: standard pure fish sauces (SPF), ●: standard mixed fish sauces (SMF), Δ: out of standard fish sauces (OF)).

Table 13 The cross-validation results of Factor analysis linear discriminant analysis (FALDA) model for 100 Thai fish sauce samples.

Fish sauce groups	No. of samples	Predicted group memberships			Corrective classification rate (%)
		SPF	SMF	OF	
1) Standard pure fish sauce (SPF)	71	62	9	0	87.32
2) Standard mixed fish sauce (SMF)	16	0	13	3	81.25
3) Out of standard fish sauce (OF)	13	0	3	10	76.92
Total corrective classification rate (%)					85.00

In cross-validation, each case is classified by the functions derived from all cases other than that case. The discriminant functions were developed using the first two factors extracted from the NIR region of 2264-2428 nm by factor analysis.

It can be obviously seen that LDA performed by using factor scores as predictor variables (FALDA) yielded the higher correct discrimination percentage than conventional LDA using the original spectral data. This may be explained from the basic principle of Factor analysis. Factor analysis, one of data reduction techniques, can separate information about chemical properties from noise terms into different factors (Martens and Naes, 1989). Note that FALDA can solve the main disadvantage of LDA, which cannot be applied to data sets having more variables than objects.

2.3.3 Soft Independent Modeling of Class Analog (SIMCA)

For SIMCA method, one of parametric methods, all the NIR data of 100 samples were split randomly into the training and test sets. The training set was composed of 80 fish sauce samples including 55, 14, and 11 of SPF-, SMF- and OF- samples, respectively. The test set consisted of the remaining samples including 16, 2, and 2 of SPF-, SMF- and OF- samples, respectively. Principal component (PC) numbers were determined for each class with the SIMCA method from the training set. Random cross-validation was used to develop the classified models. The number of principal components employed to each class and their total explained variance are summarized in Table 14.

Table 14 Number of principal components used to develop the best SIMCA model for standard pure fish sauce (SPF), standard mixed fish sauce (SMF), and out of standard fish sauce (OF) groups.

SIMCA models	Number of principal components (PCs)	Total explained variance (%)
1) Standard pure fish sauce (SPF)	2	99.68
2) Standard mixed fish sauce (SMF)	3	99.97
3) Out of standard fish sauce (OF)	3	99.98

Numbers of principal components for each class were determined by the random cross-validation method.

From Table 14, the numbers of principal component used to develop the best SIMCA models for SPF-, SMF-, and OF- groups were 2, 3, and 3, which accounted for 99.68%, 99.97%, and 99.98% of the total variance. These three models were used to classify fish sauce in the test set (n=20). The SIMCA errors can be of two types: i) Type-I errors: object not included in its own class and ii) Type-II errors: object included in a wrong class (Duda *et al.*, 2001). Table 15 summarizes the classification error resulting in terms of those two error types in the test set.

Table 15 Number of classification error of the SIMCA models in the fish sauce test set (n=20) at 5% significance levels for the F-test.

Fish sauce groups	No. of samples	Number of misclassified samples		% Corrective classification rate
		Type-I errors	Type-II errors	
1) Standard pure fish sauce (SPF)	16	0	0	100.00
2) Standard mixed fish sauce (SMF)	2	0	1	50.00
3) Out of standard fish sauce (OF)	2	0	2	0.00
Total	20	0	3	85.00

Type-I error: The object not included in its own class.
Type-II error: The object included in a wrong class.
The numbers of principal components employed in each SIMCA class of SPF, SMF and OF were 2, 3, and 3, respectively.

. The Cooman's plot shown in Figure 30 shows the classification of samples by SIMCA analysis of NIR spectra at 5% significant level that attained 85% of correct discrimination percentage. The Cooman's plot shows distances between samples and the center of each group. It is easy to visualize how certain a classification is. All of SPF and one of SMF were correctly assigned but all of OF were falsely assigned to the SMF-OF group (overlap of SMF- and OF- group zones). Therefore, SIMCA was not able to differentiate between the SMF and OF groups probably because their spectral data similarities (Figure 24) make multidimensional boxes construct quit close. However, the SIMCA model yielded the good classification results for classifying the SPF groups from the others.

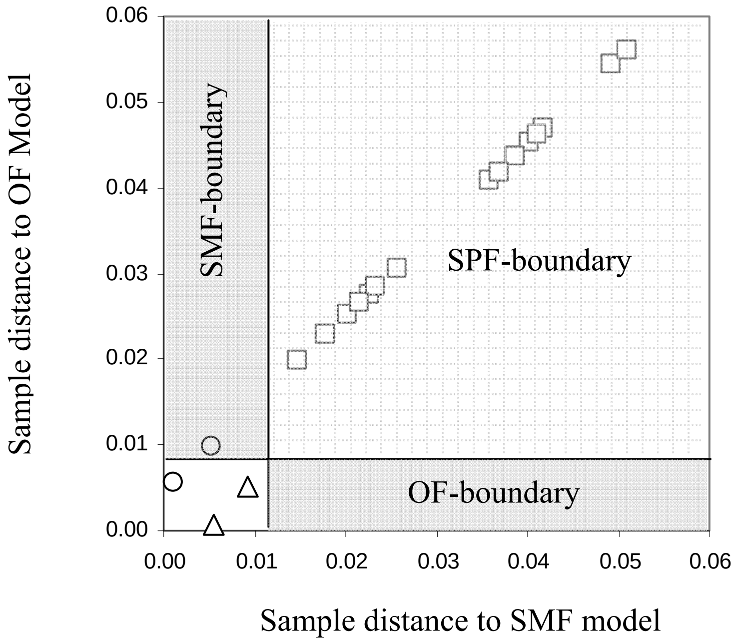


Figure 30 Cooman's plot of the classification models performed on the region of 2264- 2428 nm using the test set (n=20); □: standard pure fish sauces,○: standard mixed fish sauces, Δ : out of standard fish sauces.

To compare the classification ability between the hard modeling and soft modeling methods the classification results of LDA and SIMCA models were considered. According to the differences of basic idea, in LDA, the combinations of descriptors are selected as to place line or plane boundaries between classes. On the other hand, SIMCA is constructed for each class according to distribution of samples belonging to a class in the problem. The LDA and SIMCA results indicated that SIMCA, a soft modeling method, was powerful to differentiate SPF-sample from the SMF- and OF-samples. The SIMCA model yielded a corrective classification rate of 100% for SPF group. SIMCA method has the advantage of being able to handle collinear X-variable, missing data and noisy variables and can deal with overlapped classes (Vandeginste *et al*., 1998; Roggo *et al*., 2003a).

2.3.4 K Nearest Neighbors (KNN)

KNN is one of non-parametric methods. The principle of the KNN classification method is that the test object is assigned to a class according to the so-called majority vote procedure i.e. to the class which is most represented in the set of K nearest training objects (Wu and Massart, 1997). For comparing the performance of non-parametric and parametric methods, the fish sauce samples were again split into two sets as in the case of SIMCA, named the training set and the test set. Twenty unknown samples of the test set were classified according to the majority of its K - nearest neighbors in the training set. Several values (1-15) of the number of neighbors K were tested by cross-validation procedure. Figure 31 shows that K = 1 with the Euclidean distance yields the best values which attained 95% of corrective classification rate. Table 16 show the classification results of KNN model with K =1.

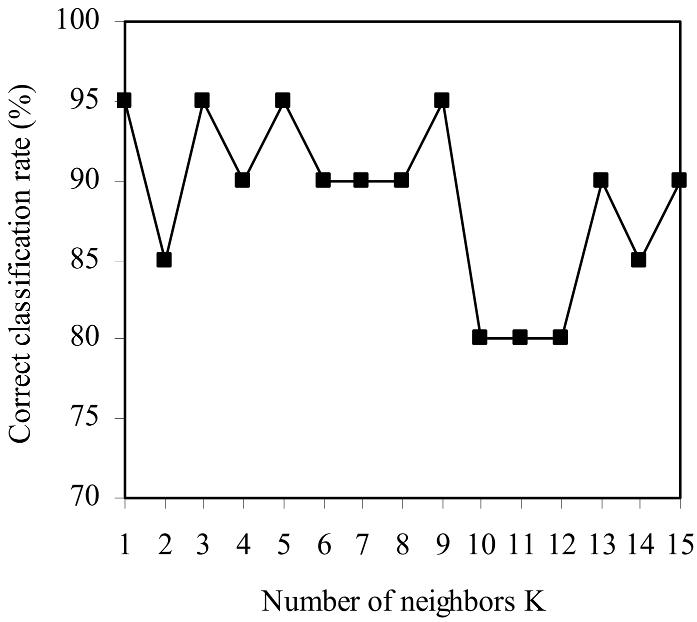


Figure 31 Corrective classification rate versus K (from K = 1 to K = 15).

Table 16 The classification results of KNN model for the fish sauce test set.

Fish sauce groups	No. of samples	Predicted group memberships			Corrective classification rate (%)
		SPF	SMF	OF	
1) Standard pure fish sauce (SPF)	16	16	0	0	100.00
2) Standard mixed fish sauce (SMF)	2	0	2	0	100.00
3) Out of standard fish sauce (OF)	2	0	1	1	50.00
Total corrective classification rate (%)					95.00

According to Figure 31, the KNN classifier was implemented with K=1.

Comparing the non-parametric and parametric methods with the same test set, the KNN classifier performed better (95.00%) than the SIMCA classifier (85.00%). However, the KNN model performed relatively poorly with OF groups. This is because of the unbalance of training data. The disadvantage of KNN is itself often introduces inductive biases for a data record takes the assumption that training data are equally distributed among all categories. Therefore, if the supposition is violated by unbalance of training data, the KNN classifier often delivers quit poorer performance.

2.3.5 Artificial Neural Networks (ANNs)

The 83 wavelengths in the NIR region of 2264-2428 nm were used as input variables (X-data) to build the ANNs model. The output vectors (Y-data) of fish sauce samples were set to the values of 1, 2 and 3 for corresponding to SPF-, SMF- and OF- groups, respectively. The NIR spectra of 100 samples were divided into the training set (n=80) and the test set (n=20). The network weights were trained by the back-error propagation algorithm. The transfer functions of hidden layer (f_1) and output layer (f_2) were logsig and purelin functions, respectively. In order to acquire the optimal architecture of neural network, several ANNs systems with different number of hidden nodes varied from 1 to 4 were tested as shown in Table 17.

Table 17 Classification results of ANNs models for the fish sauce test set (n=20).

Fish sauce groups	No. of samples	No. of misclassified samples			
		1 hidden node	2 hidden nodes	3 hidden nodes	4 hidden nodes
1) Standard pure fish sauce (SPF)	16	0	0	0	0
2) Standard mixed fish sauce (SMF)	2	2	2	2	0
3) Out of standard fish sauce (OF)	2	0	0	0	0
Total	20	2	2	2	0
Corrective classification rate (%)		90.00	90.00	90.00	100.00

The ANNs model showed an excellent data classification performance. The corrective classification rate of 100% was achieved. The number of hidden nodes had effect on the classification result. Table 17 showed that increasing the number of nodes in the hidden layers from 1 to 4 could significantly improve the corrective classification rate from 90% to 100%. Therefore, the optimal ANNs architecture consisted of the 83 wavelengths as input neurons with connected with 4 hidden neurons and a single output neuron as shown Figure 32. The selected wavelengths obtained by SCMWPLS method can be used to reduce the number of input nodes. With the SCMWPLS method, the building and training of ANNs model became feasible because the training time of the network is shortened.

ANNs have been proved as a good tool to extract useful information and reveals inherent relationship from mass and complicated data. The advantage of ANNs is in their inherent ability to in corporate nonlinear and cross-product terms into the model. It makes no assumptions about normality of data or noncollinearity among variables. With the same validation methods, a comparison of ANNs, SIMCA and KNN results revealed that the ANNs provides a far greater number of correct classifications but it suffers from the perception of being a “black box” heuristic tool. In spite of ANNs lead to better results than SIMCA and KNN, it has to be taken into account that ANNs need many samples to get a good model. However, for this study, it was not a drawback because a high number of samples (n=100) was used.

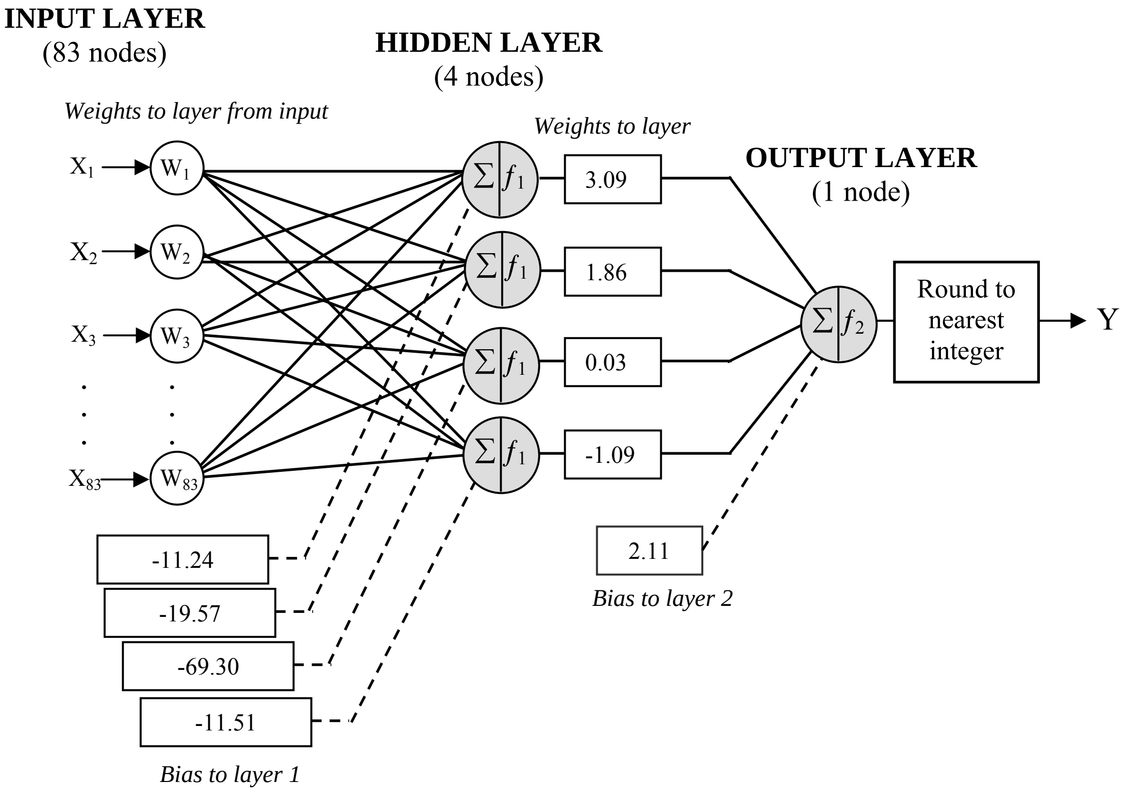


Figure 32 Diagram showing the network structure with 83 wavelength input variables. This is a network comprising four hidden neurons and a single output neuron. Transfer functions (f_i) for hidden and output layers are LOGSIG (f_1) and PURELIN (f_2), respectively. The weights to layer from input (W_i) are given in an Appendix C.

2.4 Comparison of Five Classification Models

From all the results above, we conclude that an optimized combination of informative regions for the total nitrogen content obtained by SCMWPLS can be used to develop classified models with the correct classification rate of more than 82%, and ANNs classified model had the highest correct classification rate (100%). The corrective classification rates of the five supervised pattern recognition methods are summarized in Table 18.

Table 18 Corrective classification rate of the classification models developed by five supervised pattern recognition methods. The developed models performed on the region of 2264-2428 nm.

Methods	Validation methods	Corrective classification rate (%)			
		SPF-group	SMF-group	OF-group	Total
LDA	Cross validation	85.92	75.00	69.23	82.00
FALDA	Cross validation	87.32	81.25	76.92	85.00
SIMCA	Test set	100.00	50.00	0.00	85.00
KNN	Test set	100.00	100.00	50.00	95.00
ANNs	Test set	100.00	100.00	100.00	100.00

LDA: Linear discriminant analysis
FALDA: Factor analysis-Linear discriminant analysis
SIMCA: Soft independent modeling of class analog
KNN: K nearest neighbors
ANNs: Artificial neural networks

The non-parametric method (ANNs and KNN) was found to be more satisfactory than others parametric method (LDA, FALDA, and SIMCA). The most important advantage of non-parametric method is that it is free from statistical assumptions such as normality of the distribution of the variables, while the parametric methods such as LDA and SIMCA are explicitly based on distribution statistics (Vandeginste *et al.*, 1998). LDA can perform well in situations where the sample size is large enough compared to the number of variables. A comparison was also performed between LDA and FALDA. The classification performance of LDA and FALDA was similar. However, the FALDA model was found to be considerably more robust to reduce the number of input dimensions which can be solved the disadvantage of traditional LDA.

In this study, it can be seen that there are unbalanced data in the training set for each groups. The training sets for SPF-, SMF- and OF- groups were 16, 2, and 2, respectively. The classification performance for classifying SPF- group was higher than SMF-and OF- groups. This can be explained that in an unbalanced data set, the majority class is represented by a large portion of all the samples, while the other, the minority class has only a small percentage of all samples. When a sample classifier encounters an unbalanced sample, the classification performance often decreases.

3. <u>Categorization of Thai Fish Sauces Based on Sensory Characteristics and Principal Component Analysis</u>

The sensory characteristics of twenty fish sauces consisted of pure fish sauces and mixed fish sauces were investigated using the generic descriptive analysis method. Subsequently, all samples were categorized based on their sensory properties using the combination of cluster analysis and principal component analysis (PCA).

3.1 Sensory Evaluation

The sensory characteristics of the Thai fish sauces were profiled by a trained descriptive panel (12 judged) and the literature review (Meilgarrd *et al*., 1999). The final list was composed of 15 attributes. They were brown color, five aromatic descriptors (sweet aromatic, caramelized aromatic, fermented aromatic, fishy aromatic, and musty aromatic), four taste descriptors (sweet taste, salty taste, bitter taste, and umami taste), three aftertaste descriptors (sweet aftertaste, salty aftertaste, and bitter aftertaste) and two flavor descriptors (caramelized flavor, and fishy flavor). The final list of those sensory attributes, their definitions and references are given in Table 19.

Mean values of sensory attributes of the 12 pure fish sauces and 8 mixed fish sauces are shown in Table 20 and 21, respectively. Figure 33 shows the sensory profiles of Thai pure fish sauce and mixed fish sauce based on generic descriptive analysis test.

Table19 Descriptive attributes and their definitions used in the descriptive analysis of Thai fish sauces.

Attributes	Definitions	References and Intensities	Sources
Color			
Brown intensity	The intensity or strength of the brown color from light to dark.	1 Tea bag/ 1min soak (5.5) 1 Tea bag/ 5min soak (7.5) 1 Tea bag/ 60 min soak (9.5)	Lipton Yellow Label Tea (PT Unilever Indonesia)
Aromatics			
Sweet aromatic	The aromatic associated with sugar perceived by smell.	2% Natural mineral sugar solution (0.5) 5% Natural mineral sugar solution (2.0) 10% Natural mineral sugar solution (4.0)	Mitr Phol Gold Brand (Mitr-Kalasin Sugar Co., Ltd., Thailand)
Caramelized aromatic	The aromatic associated with brown sugar perceived by smell.	2% Natural brown sugar solution (2.0) 5% Natural brown sugar solution (6.0) 10% Natural brown sugar solution (11.0)	Wang Kanai Brand (T.N. Sugar Industry Co., Ltd. Thailand)
Fermented aromatic	The aromatic associated with fermented fish perceived by smell.	Shrimp paste (8.5)	Trachang Brand (Tang Thai Chiang Fish-Sauce Manufacturing Co., Ltd., Thailand)
Fishy aromatic	The aromatics or volatiles which are derived from fish products perceived by smell.	Tuna sandwich in brine (3.0) Tuna sandwich in brine (7.5)	Carrefour Brand, Carrefour TCB Brand (Tropical Canning Thailand Public Co.,Ltd)

Table 19 (Continued)

Attributes	Definitions	References and Intensities	Sources
Musty aromatic	The aromatic associated with closed air spaces that are perceived by smell such as attic and closets.	Thai mixed fish sauce(6.0) Thai mixed fish sauce (11.5)	Kum Ka brand, Tesco Lotus Red bow brand, Phairod Co.,Ltd., Thailand
Basic tastes			
Sweet taste	The fundamental taste factor associated with a sucrose solution perceived by tongue.	2% Pure refined sugar solution (2.0) 5% Pure refined sugar solution (5.0) 10% Pure refined sugar solution (10.0)	Mitr Phol Gold Brand (Mitr-Kalasin Sugar Co., Ltd., Thailand)
Salty taste	The fundamental taste factor associated with a sodium chloride solution perceived by tongue.	15% Sodium chloride solution (6.0) 20% Sodium chloride solution (9.0) 25% Sodium chloride solution (12.0)	Prung Thip Brand, Thai Refined Salt Co., Ltd.
Bitter taste	The fundamental taste factor associated with a caffeine solution perceived by tongue.	0.05% Caffeine solution (2.0) 0.08% Caffeine solution (5.0)	MERCK (Darmstadt, Germany)
Umami taste	The fundamental taste factor associated with a monosodium glutamate (MSG) solution perceived by tongue.	0.5% MSG solution (5.0) 1.0% MSG solution (10.0)	Ajinomoto Brand, Ajinomoto Co., (Thailand) Ltd.

Table 19 (Continued)

Attributes	Definitions	References and Intensities	Sources
Aftertastes			
Sweet taste	The fundamental taste factor associated with a pure refined sugar in solution remaining after swallowing the sample.	2% Pure refined sugar solution (0.5) 5% Pure refined sugar solution (3.0) 10% Pure refined sugar solution (5.5)	Mitr Phol Gold Brand (Mitr-Kalasin Sugar Co., Ltd., Thailand)
Salty taste	The fundamental taste factor associated with a sodium chloride solution remaining after swallowing the sample.	15% Sodium chloride solution (4.0) 20% Sodium chloride solution (7.0) 25% Sodium chloride solution (8.0)	Prung Thip Brand, Thai Refined Salt Co., Ltd.
Bitter taste	The fundamental taste factor associated with a caffeine solution remaining after swallowing the sample.	0.05% Caffeine solution (1.5) 0.08% Caffeine solution (4.0)	MERCK (Darmstadt, Germany)
Flavors			
Caramelized flavor	The characteristic aroma of natural brown sugar in solution perceived by tasting and smelling during swallowing.	2% Natural brown sugar solution (1.5) 5% Natural brown sugar solution (4.0) 10% Natural brown sugar solution (7.0)	Wang Kanai Brand (T.N. Sugar Industry Co., Thailand)
Fishy flavor	The characteristic aroma of fish products perceived by tasting and smelling during swallowing..	Tuna sandwich in brine (2.5) Tuna sandwich in brine (5.0)	Carrefour Brand, Carrefour TCB Brand (Tropical Canning Thailand Public Co.,Ltd)

Intensity is given on a 15-point scale; 0 = none, 15 = extremely high.

Table 20 Mean values[1] on descriptive analysis of the twelve pure fish sauce samples.

Sensory attributes	Pure fish sauce sample code											
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
Brown intensity	5.37	4.95	7.50	7.75	6.20	6.92	6.00	7.87	7.00	6.50	6.58	6.00
Sweet aromatic	1.67	0.92	2.75	4.58	0.92	2.42	2.33	1.00	1.33	1.00	1.42	1.25
Caramelized aromatic	2.67	2.58	3.00	1.75	1.92	2.67	2.08	1.83	1.67	1.25	1.67	1.33
Fermented aromatic	5.33	4.58	2.75	5.33	4.92	4.00	4.67	5.75	4.83	4.75	5.83	5.58
Fishy aromatic	5.67	4.67	4.42	3.92	4.75	4.25	3.33	4.00	3.92	2.92	3.25	3.25
Musty aromatic	3.50	4.17	5.67	2.33	4.58	2.00	3.58	4.42	4.83	4.25	5.42	4.67
Sweet taste	0.58	0.00	2.58	1.42	0.42	2.75	1.67	0.00	1.42	0.75	1.17	1.00
Salty taste	7.75	9.42	7.08	7.75	9.83	7.17	5.75	6.00	8.08	7.25	7.00	9.50
Bitter taste	0.83	0.67	1.17	2.83	2.58	2.25	2.00	2.08	1.25	1.17	1.92	1.75
Umami taste	6.42	6.58	5.83	6.42	7.00	5.83	5.17	6.08	6.50	5.42	6.00	5.33
Sweet aftertaste	4.25	3.25	3.58	3.33	1.58	4.08	3.08	2.92	2.17	2.17	2.92	1.42
Salty aftertaste	6.42	7.33	4.50	4.92	7.75	4.25	3.67	5.08	5.33	4.25	4.83	6.42
Bitter aftertaste	0.42	1.25	0.67	1.67	1.83	1.58	0.92	1.25	0.83	0.58	1.33	0.92
Caramelized flavor	4.17	3.17	2.50	1.75	1.50	2.33	1.58	2.25	1.50	1.33	2.00	1.00
Fishy flavor	3.50	3.42	3.42	3.50	3.92	3.17	2.50	3.42	2.75	2.42	2.58	2.17

[1]Intensities scored on a 15-point intensity scale where 0 = none and 15 = extremely high. For fish sauce descriptions, see Table 19.

Table 21 Mean values[1] on descriptive analysis of the eight mixed fish sauce samples.

Sensory attributes	Mixed fish sauce samples code							
	M1	M2	M3	M4	M5	M6	M7	M8
Brown intensity	4.92	5.75	3.83	5.08	6.58	4.42	5.33	5.17
Sweet aromatic	2.67	1.42	1.75	1.17	1.08	1.75	1.08	1.08
Caramelized aromatic	2.67	0.83	1.83	0.83	1.00	1.00	1.83	1.08
Fermented aromatic	1.50	2.42	4.33	4.17	3.75	3.17	4.33	4.33
Fishy aromatic	0.75	3.25	2.75	2.83	2.00	1.50	2.58	2.50
Musty aromatic	1.33	3.00	5.00	5.92	4.58	3.67	4.17	4.50
Sweet taste	1.83	1.00	0.75	0.75	1.67	1.83	1.67	1.25
Salty taste	4.33	8.58	8.67	6.58	5.67	7.42	5.92	6.50
Bitter taste	1.67	2.50	1.25	2.00	1.25	1.83	2.00	2.42
Umami taste	7.17	5.67	5.00	5.00	5.42	4.83	5.42	5.83
Sweet aftertaste	3.92	1.25	2.83	1.25	2.08	1.75	3.08	1.58
Salty aftertaste	2.58	6.17	4.75	3.75	3.75	4.67	3.83	4.25
Bitter aftertaste	0.92	1.17	0.75	1.08	0.83	1.33	1.42	1.67
Caramelized flavor	2.50	1.08	1.08	0.92	1.08	1.25	1.67	0.83
Fishy flavor	0.58	2.58	1.92	2.50	1.92	1.75	2.33	1.92

[1]Intensities scored on a 15-point intensity scale where 0 = none and 15 = extremely high. For fish sauce descriptions, see Table 19.

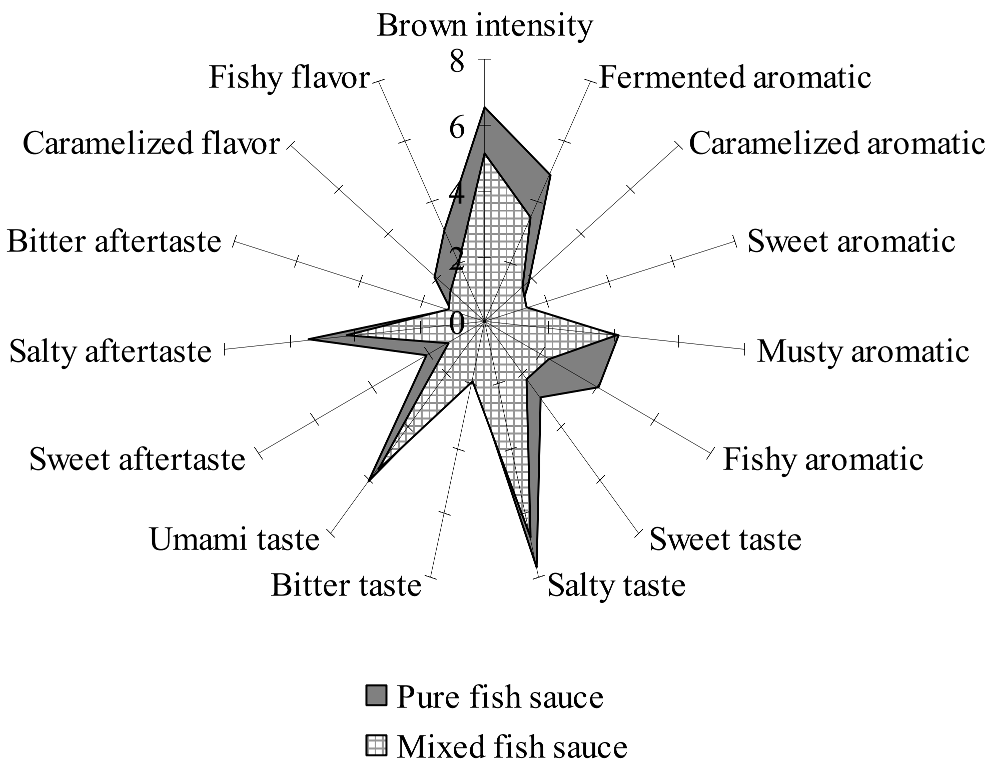


Figure 33 The sensory profiles of Thai pure fish sauce and mixed fish sauce based on generic descriptive analysis test. The distance from the center is the mean value for the attributes.

3.2 Analysis of Variance (ANOVA)

ANOVA is used to determine the significance of sample effects and other relevant model effects (O'Mahony, 1986). The two factors, panelist and sample, were studied in this study. The analyses of variance for each attributes were assessed independently at a significance level of 5%. Table 22 shows the sensory attributes and results of ANOVA. At the 5% level, the sample effect was significant for all attributes while the panelist effect was not significant. This can be indicated that the intensity of sensory attributes differs among samples. Therefore, all the fifteen sensory attributes were subjected to cluster analysis and principal component analysis (PCA).

Table 22 Analysis of variance on the 15 sensory attributes rated for fish sauce samples.

Sensory attributes	Panelist effect		Sample effect	
	F-ratio	*p-value*	*F-ratio*	*p-value*
Brown intensity	0.05	> 0.05	64.37	< 0.05
Sweet aromatic	0.28	> 0.05	15.73	< 0.05
Caramelized aromatic	1.03	> 0.05	14.28	< 0.05
Fermented aromatic	0.30	> 0.05	16.96	< 0.05
Fishy aromatic	0.23	> 0.05	19.48	< 0.05
Musty aromatic	0.28	> 0.05	16.38	< 0.05
Sweet taste	0.36	> 0.05	17.82	< 0.05
Salty taste	0.07	> 0.05	33.68	< 0.05
Bitter taste	0.25	> 0.05	7.02	< 0.05
Umami taste	0.58	> 0.05	7.47	< 0.05
Sweet aftertaste	0.29	> 0.05	14.26	< 0.05
Salty aftertaste	0.13	> 0.05	30.31	< 0.05
Bitter aftertaste	0.72	> 0.05	4.38	< 0.05
Caramelized flavor	0.57	> 0.05	8.57	< 0.05
Fishy flavor	0.56	> 0.05	11.66	< 0.05

3.3 Cluster Analysis

Cluster analysis was performed using the sensory scores of fifteen sensory attributes as predictor variables. Prior to cluster analysis, all sensory attributes were converted to *Z*-scores to standardize and normalize the sensory scores. The dendrogram obtained by the Euclidian distance partly classified samples according to types (Figure 34). Four clusters each mainly composed of P, P+M, P, M were observed. It can be seen that the sensory characteristics of some samples in the pure fish sauces (P) group similar to the mixed fish sauces (M). Although the branches of the dendrogram represent the distance between the samples or clusters, no information is provided by the dendrogram on the specific attribute differences among the samples. The final output does not provide the reason why two types of fish sauces were grouped together. Therefore, other test results, such as principal component analysis, need to be inspected to explain the clustering (Muñoz *et al*., 1996).

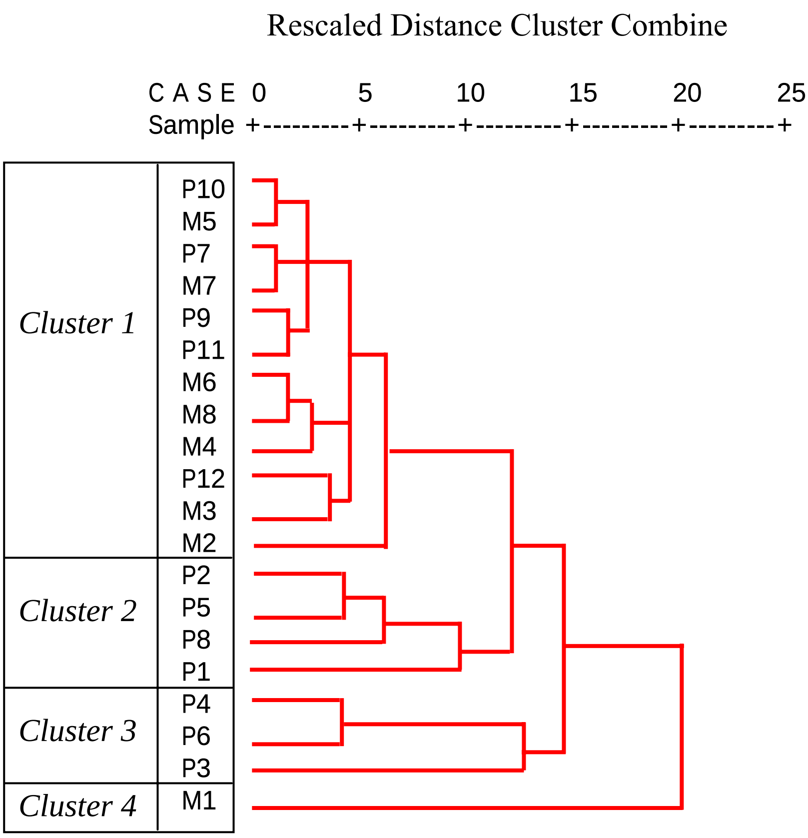


Figure 34 Clustering of 20 fish sauces based on their sensory properties (P, pure fish sauce; M, mixed fish sauce) using cluster analysis.

3.4 Principal Component Analysis (PCA)

Principal component analysis (PCA) applied to the fifteen sensory attributes of twenty fish sauce samples indicated that the first four principal components explained a total of 79.88% of the variation (30.44%, 24.70%, 15.32% and 9.42%, respectively). The validity of these four principal components was confirmed by full cross-validation. Therefore, the dimensionality of the data was reduced from fifteen variables to four uncorrelated components with 20.12% loss of variation.

The correlations between the sensory attributes and corresponding factor loadings of the first four principal components were investigated as shown in Table 23. Using guidelines provided by Stevens (1992) to inspect for significance of attribute loadings, an attribute was considered to load heavily on a given component if the factor loadings was greater than 0.72. Table 23 shows that a total of seven attributes loaded heavily on the first three principal components. Four attributes of fishy aromatic, sweet aftertaste, caramelized flavor and fishy flavor loaded heavily on the first PC, indicating strong correlations of these attributes with PC1. These attributes were positively loaded, which will be referenced as the *fishy flavor* component or PC1. Caramelized aromatic was loaded heavily on the second PC, which will be referenced as the *caramelized aromatic* component or PC2. Finally, bitter taste and bitter aftertaste were loaded heavily on the third PC, which will be referenced as the *bitterness* component or PC3.

Table 23 Factor loadings and percentage variance for the first four principal components (PCs) performed by principal component analysis (PCA).

Factor loading of parameters	Principal Components			
	PC1	PC2	PC3	PC4
Loadings*				
Brown intensity	0.392	0.098	0.262	0.673
Sweet aromatic	-0.234	0.696	0.074	0.314
Caramelized aromatic	0.218	**0.738**	0.325	0.155
Fermented aromatic	0.407	-0.492	0.144	0.464
Fishy aromatic	**0.896**	-0.228	-0.049	0.179
Musty aromatic	-0.109	-0.642	-0.338	0.426
Sweet taste	0.589	0.710	-0.239	0.065
Salty taste	0.436	-0.695	0.142	-0.234
Bitter taste	0.025	0.055	**0.942**	-0.202
Umami taste	0.706	0.176	-0.011	-0.333
Sweet aftertaste	**0.793**	0.368	-0.280	-0.212
Salty aftertaste	0.605	-0.670	0.081	-0.296
Bitter aftertaste	0.030	0.031	**0.940**	-0.016
Caramelized flavor	**0.726**	0.540	-0.170	-0.106
Fishy flavor	**0.879**	-0.253	0.139	0.230
Percentage variance	30.44	24.70	15.32	9.42

* Factor loadings marked in bold indicate situations where the attribute loadings meet the criteria of significance (0.72p = 0.01).

Besides the demonstration of the associations among the attributes, PCA can be used to display the relative "locations" of the samples with respect to each other and their characterizing attributes (Meilgaard *et al.*, 1999). The sensory data were analyzed to identify the relationships and differences between the samples and the sensory attributes using biplot analysis as shown in Figure 35 and 36, respectively.

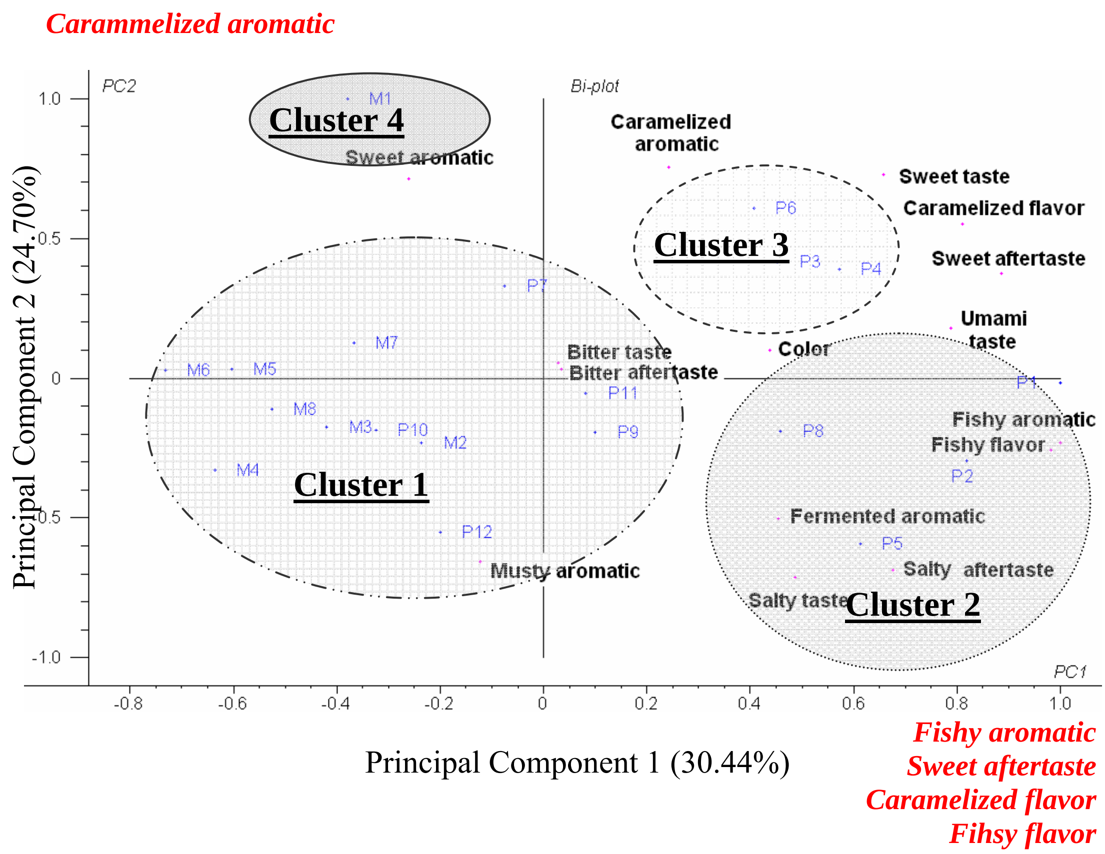


Figure 35 Scores for the twenty fish sauce samples and correlation coefficients of the fifteen sensory attributes with principal component 1 and 2. (P, pure fish sauce; M, mixed fish sauce)

Based on cluster analysis, fish sauce samples could be categorized into four clusters and were grouped together in the sensory map. Figure 35 shows the relationships and difference between sensory attributes and clusters obtained from cluster analysis results (Figure 34). According to *fishy flavor component* (PC1) which explained 30.44% of the total variance, the samples in cluster 2 and 3 were identified by higher degrees of *fishy aromatic, sweet aftertaste, caramelized flavor and fishy* *flavor* than the samples in cluster 1 and 4. It can be seen that pure fish sauce samples (P) and mixed fish sauce samples (M) were distinguished well on this PC. There is a clear distinction of the scores from the two types of fish sauce on PC1. The pure fish sauce samples (P) is somewhat more shifted towards the higher values of PC1 than the mixed fish sauce samples (M). These results indicated that pure fish sauces had higher degree of fishy aromatic, caramelized flavor, sweet aftertaste and fishy flavor than mixed fish sauce samples. Considering the *caramelized aromatic component* (PC2) which explained an additional of 24.70% of the variation, the samples in cluster 3 and 4 had higher degree of caramelized aromatic than the samples in cluster 1 and 2.

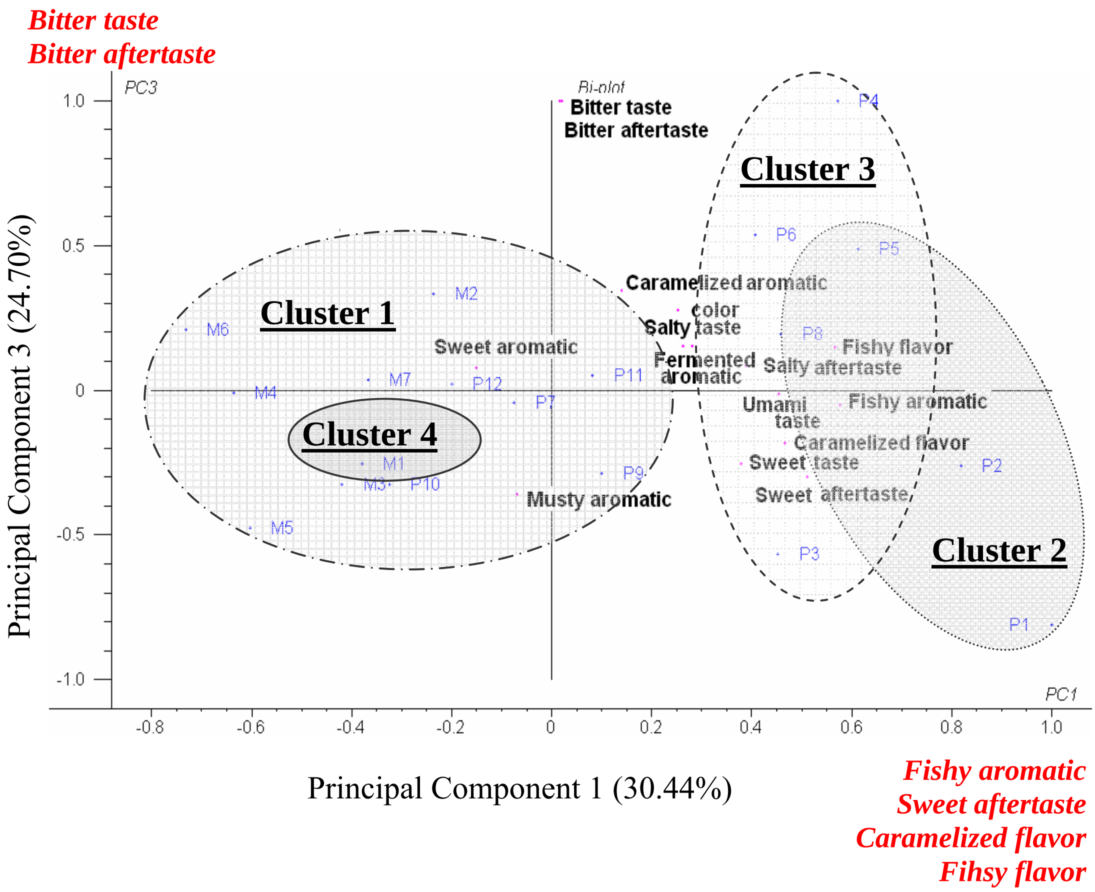


Figure 36 Scores for the twenty fish sauce samples and correlation coefficients of the fifteen sensory attributes with principal component 1 and 3. (P, pure fish sauce; M, mixed fish sauce)

To describe the taste and aftertaste of bitter of those four clusters, the *bitterness component* (PC3) which explained an additional of 15.32% of the variation was considered. Figure 36 shows the biplot of principal component 1 and 3. The samples in four groups from cluster analysis could not distinguish well on this plot. There were two overlapping zones between cluster 1 and 4, and cluster 2 and 3, respectively. These results indicated that the commercial pure fish sauces and the mixed fish sauces could be classified by the first two principal components which highly correlated with the sensory attributes of fishy aromatic, sweet aftertaste, caramelized flavor, fishy flavor, and caramelized aromatic. A summary of the characteristic of fish sauces classified by principal component 1 and 2 is shown in Table 24.

Table 24 Sensory characteristics of four clusters of Thai fish sauce samples classified by principle component analysis (PCA).

Clusters	Sensory characteristics				Sample codes***
	Fishy aromatic, sweet aftertaste, caramelized flavor and fishy flavor*		Caramelized aromatic**		
	Strong	Low	Strong	Low	
1		√		√	P7, P9,P10, P11,P12, M2, M3, M4, M5, M6, M7, M8
2	√			√	P1, P2, P5, P8
3	√		√		P4, P3, P6,
4		√	√		M1

* Explained by principal component 1, which accounted for 30.44% of the variation.

** Explained by principal component 2, which accounted for 24.70% of the variation.

*** P: pure fish sauce, M: mixed fish sauce

4. Correlations between Sensory, Chemical and Physical Properties of Thai Fish Sauces Studied by NIR Spectroscopy Combined with Chemometrics

The twenty fish sauce samples and their sensory properties studied in Part 3 were used to investigate the correlations between sensory, chemical and physical properties. With the same samples, the NIR transflectance of the samples were measured and subsequently used as input variables (X-data) to predict the chemical and physical properties by using the PLS models developed from the study in Part 1. The correlations between the sensory properties and the predicted chemical and physical properties of the fish sauce samples were investigated by principal component analysis (PCA).

4.1 NIR Spectra

NIR transflectance spectra of twenty investigated Thai fish sauce samples are shown in Figure 37. For chemometrics analyses the 1900-2000 nm region was not employed in order to avoid heavily overlapping absorption bands.

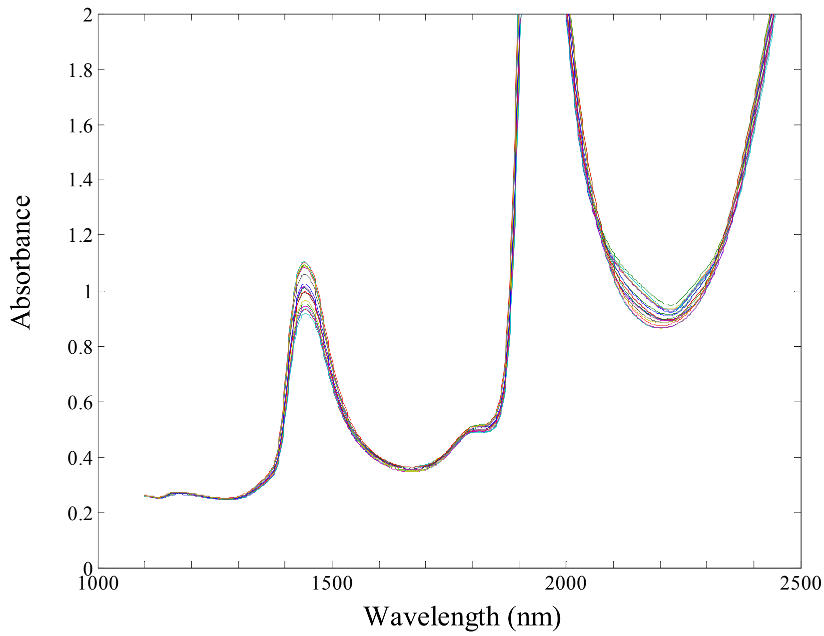


Figure 37 NIR transflectance spectra of 20 Thai commercial fish sauce samples.

4.2 Correlations between the Sensory Attributes and the NIR Spectra

4.2.1 PCA Loadings for the NIR Spectra

In order to reduce the number of spectral data, PCA was performed on the wavelength regions of 1100-1900 and 2000-2440 nm, which composed of 622 individual wavelengths. The first two principal components explained for 92 and 7% of the total variation, respectively. Therefore, the dimensionality of the data was reduced from 622 wavelengths to two uncorrelated components (PC1 and PC2). The PCA rely on the correlation matrix for all wavelengths as shown in Figure 38.

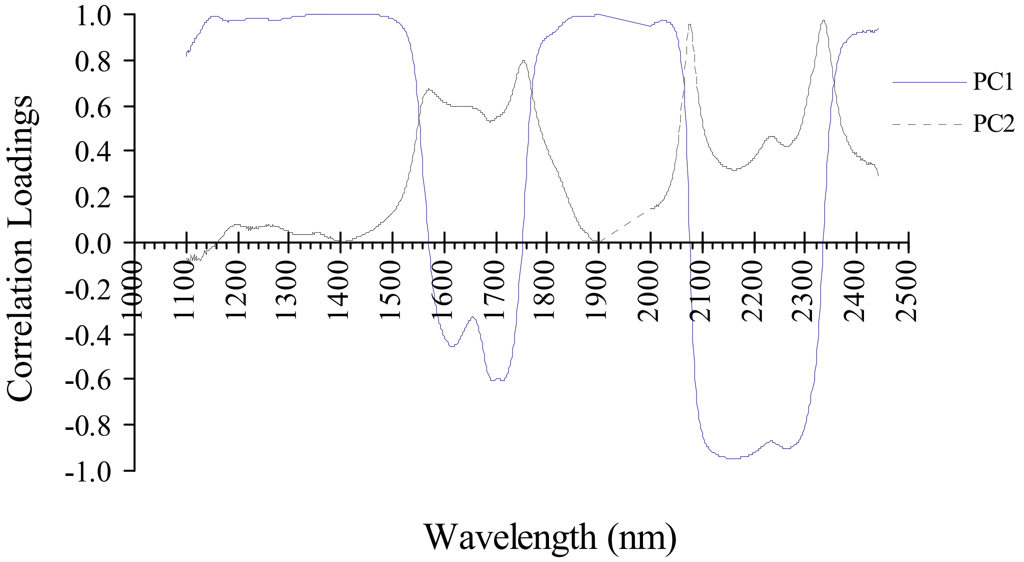


Figure 38 NIR spectra loadings for the first two principal components performed by principal component analysis (PCA).

Using the guidelines provided by Stevens (1992) to inspect for significance of wavelength loadings, a wavelength was considered to load heavily on a given component if the factor loadings was greater than 0.72. From Figure 38, the first PC was highly correlated with the wavelength regions of 1100-1544, 1774-2062, 2092-2308, and 2358-2440 nm. The wavelength regions of 1100-1544, 1774-2062 and 2358-2440 were positively correlated with PC1, while the other region of 2092-2308 nm was negatively correlated. The second PC was positively correlated with the wavelength regions of 1742-1764, 2066-2088, and 2312-2354. These two uncorrelated components and the fifteen sensory scores from the study in Part 3 were subjected principal component analysis to investigate their relationships.

4.2.2 Correlations between the Sensory Attributes and the NIR Spectra Studied by PCA

As the results in section 4.2.3, the NIR spectral data were reduced from 622 wavelengths to two uncorrelated components (PC1 and PC2). These components were used to investigate a possible relationship between NIR spectra and sensory data using PCA. The first three principal components explained a total of 66.86% of the variation. The correlations between the NIR spectra and sensory attributes were investigated using the PCA factor loadings as shown in Table 24.

Using the guidelines provided by Stevens (1992) to inspect for significance of attribute loadings, an attribute was considered to load heavily on a given component if the factor loadings was greater than 0.72. Table 24 shows that fishy aromatic, sweet aftertaste and fishy flavor were highly correlated with NIR1 on the first PC, which will be referenced as *fishy flavor component* or PC1. As mentioned above, NIR1 was highly correlated with the wavelength regions of 1100-1544, 1774-2062, 2092-2308, and 2358-2440 nm. These regions contain the 2140-2170 and 2200-2250 nm regions, where several bands arising from the combinations of proteins and amide modes are located (Williams and Norris, 1990; Siesler *et al*., 2002). It can be interpreted that the degree of fishy aromatic, sweet aftertaste and fishy flavor in the Thai fish sauce samples were strongly correlated with their proteins and amino acids contents. Caramelized aromatic and sweet taste were loaded heavily on the second PC, which will be referenced as *sweetness component* or PC2. These attributes were negatively correlated with PC2. Finally, bitter taste and bitter aftertaste were loaded heavily on the third PC, which will be referenced as the *bitterness component* or PC3.

Table 24 NIR spectra and sensory attributes loadings for the first three principal components performed by principal component analysis.

Parameters	Principal components		
	PC1	PC2	PC3
Loadings*			
- NIR spectra**			
NIR1	**-0.900**	0.017	0.004
NIR2	0.042	-0.176	-0.030
- Sensory attributes			
Brown intensity	0.409	0.123	0.267
Sweet aromatic	-0.217	0.684	0.070
Caramelized aromatic	0.223	**0.744**	0.323
Fermented aromatic	0.445	-0.469	0.150
Fishy aromatic	**0.905**	-0.210	-0.047
Musty aromatic	-0.059	-0.650	-0.338
Sweet taste	0.575	**0.724**	-0.239
Salty taste	0.462	-0.687	0.143
Bitter taste	0.009	0.056	**0.941**
Umami taste	0.674	0.191	-0.010
Sweet aftertaste	**0.764**	0.385	-0.280
Salty aftertaste	0.608	-0.652	0.085
Bitter aftertaste	0.026	0.028	**0.938**
Caramelized flavor	0.711	0.557	-0.171
Fishy flavor	**0.886**	-0.234	0.141
Percentage variance	31.41	21.93	13.52

* Factor loadings marked in bold indicate situations where the attribute loadings meet the criteria of significance (0.72p = 0.01).

** NIR 1 and NIR 2; The NIR spectra scores for the first two principal components (PC1 and PC2), respectively, performed on the wavelength regions of 1100-1900 and 2000-2440 nm using Principal component analysis.

Besides the demonstration of the associations among the attributes, PCA can be used to display the relative "locations" of the samples with respect to each other and their characterizing attributes (Meilgaard *et al.*, 1999). The sensory and NIR data were analyzed to identify the relationships and difference between the samples and the attributes using biplot analysis. According to the cluster analysis and PCA results from the study in Part 3, fish sauce samples were categorized into four clusters based on their sensory characteristics. Those four clusters were grouped together in the biplot of the first two PCs (53.34% of variability) as shown in Figure 39.

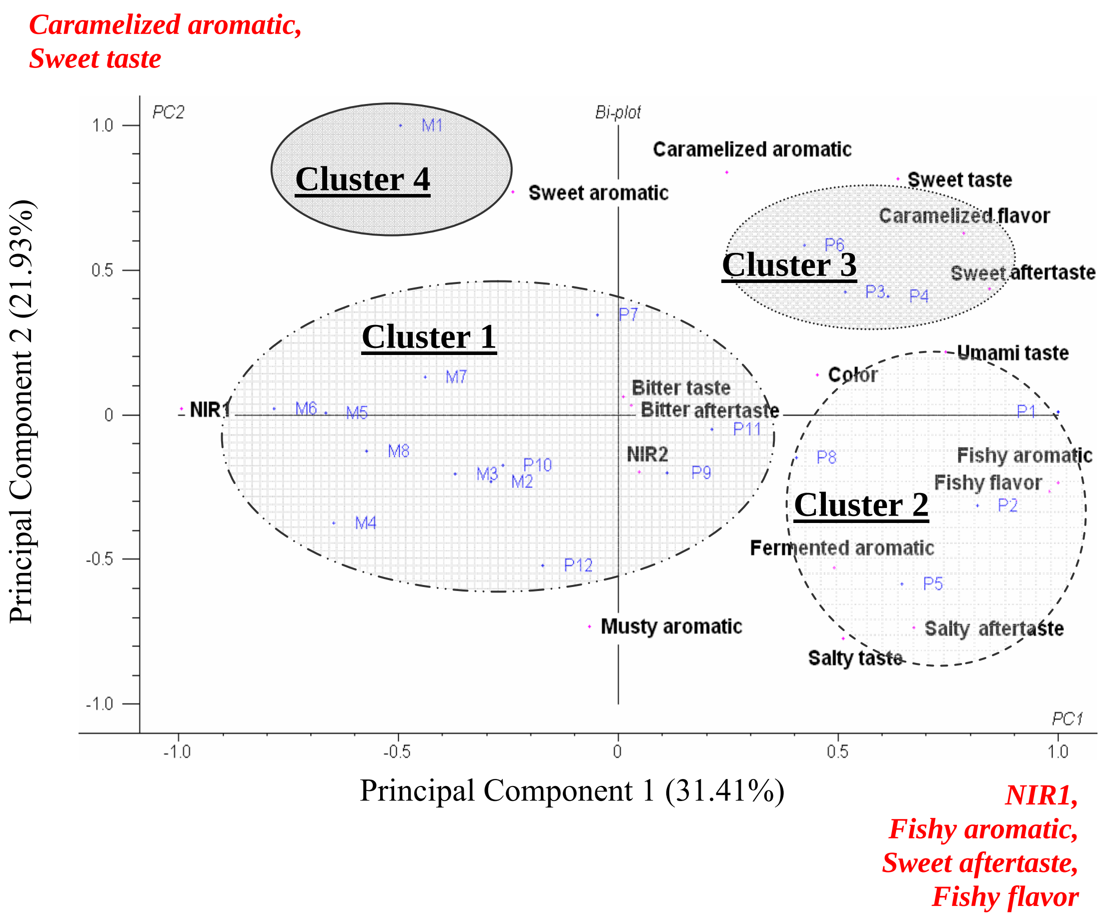


Figure 39 Scores for the twenty fish sauce samples and correlation coefficients of the sensory attributes and NIR spectra with principal component 1 and 2. (P, pure fish sauce; M, mixed fish sauce)

Figure 39 shows the relationships and differences between sensory attributes and NIR data for four clusters. According to *fishy flavor component* (PC1) which explained 31.41% of the total variance, the samples in cluster 2 and 3 were identified by higher degrees of *fishy aromatic, sweet aftertaste and fishy flavor* than the samples in cluster 1 and 4. In addition, the differences observed in the *sweetness component* (PC2) which explained an additional of 21.93% of the variation, the samples in cluster 3 and 4 had higher degree of *caramelized aromatic and sweet taste* than the samples in cluster 1 and 2.

It can be seen in Figure 39 that the pure fish sauce samples (P) and the mixed fish sauce samples (M) were distinguished well on this PC. There is a clear distinction of the scores from the two types of fish sauce on PC1. The pure fish sauce samples (P) is somewhat more shifted towards the higher values of PC1 than the mixed fish sauce samples (M). These results indicated that the pure fish sauces exhibited the higher degree of fishy aromatic, sweet aftertaste and fishy flavor than mixed fish sauce samples. The main sensory characteristics and NIR data describing each cluster are summarized in Table 25.

Table 25 Sensory characteristics and corresponding NIR spectra of four clusters of the Thai fish sauce samples classified by principal component analysis (PCA).

Clusters	Wavelength region (nm)		Sensory characteristics				Types of fish sauce	Sample codes***
	NIR1*		Fishy aromatic , Sweet aftertaste and Fishy flavor*		Caramelized aromatic and sweet taste**			
	High	Low	Strong	Low	Strong	Low		
1	√			√		√	Pure and mixed	P7, P9,P10, P11,P12, M2, M3, M4, M5, M6, M7,M8
2		√	√			√	Pure	P1, P2, P5, P8
3		√	√		√		Pure	P4, P3, P6,
4	√			√	√		Mixed	M1

* Explained by principal component 1, which accounted for 35.86% of the variation.

** Explained by principal component 2, which accounted for 20.03% of the variation.

*** P: pure fish sauce, M: mixed fish sauce

NIR 1; The NIR spectra scores for the first principal component (PC1) performed on the wavelength regions of 1100-1900 and 2000-2440 nm using Principal component analysis (PCA). NIR1 was highly correlated to the wavelength regions of 1100-1544, 1774-2062, 2092-2308, and 2358-2440 nm.

4.3 Prediction of the Chemical and Physical Properties using the NIR Spectra

To investigate the relationships between sensory, chemical and physical data, the chemical and physical properties of twenty fish sauce samples were predicted using the NIR spectra. According to the PLS prediction results for the chemical and physical properties of fish sauces (Table 10), the developed SCMWPLS models which yielded the correlation coefficient (R) more than 0.70 were used to predict the chemical and physical properties of samples. The total nitrogen content, sodium chloride, pH, reducing sugar, density, baume, total soluble solid, refractive index and color a* were predicted with their selected model using the NIR spectra as input variables. The statistical characteristics of the predicted properties and their optimized NIR spectra are summarized in Table 26.

Table 26 The optimized input NIR wavelength regions and predicted chemical and physical values of Thai fish sauces (n=20).

Analytes	Input variables/ Wavelength regions (nm)[1]	Predicted values			
		Min	Max	Mean	SD
Total nitrogen (%w/v)	2264-2428	0.00	2.17	0.80	0.77
Sodium Chloride (%w/v)	1480-1798, 2252-2258	14.19	22.47	19.90	2.24
pH (-)	1676-1708, 2208-2260	4.50	5.49	5.00	0.31
Reducing sugar (mg/mL)	1608-1760	0.00	1.80	1.01	0.48
Density (g/cm^3)	1358-1438	1.12	1.25	1.20	0.04
Baume (°Baume)	1580-1670, 2224-2354	15.49	28.97	23.99	4.02
Total soluble solid (°Brix)	1322-1442, 2000-2076	15.43	40.46	29.73	8.52
Refractive index (-)	1774-1846, 2078-2114	1.36	1.40	1.38	0.02
Color a*	1696-1726, 2086-2158	7.87	23.62	15.72	5.42

[1]Input variables/Wavelength regions means the optimized informative regions obtained from searching combination moving window partial least squares (SCMWPLS) from the study in Part 1 (Table 10).

[2]Min: Minimum, Max: Maximum, SD: Standard Deviation

4.4 Correlations between the Sensory Attributes and the Predicted Chemical Properties Studied by PCA

The four predicted chemical parameters of total nitrogen content, sodium chloride, pH and reducing sugar using SCMWPLS models and all fifteen sensory attributes from descriptive analysis were subjected to PCA as input variables. The first three principal components explained a total of 68.03% of the variation. The correlations between the sensory attributes and predicted chemical parameters were investigated by using the PCA factor loadings as shown in Table 27.

Using the guidelines provided by Stevens (1992) to inspect for significance of attribute loadings, an attribute was considered to load heavily on a given component if the factor loadings was greater than 0.72. Table 27 shows that total nitrogen, sodium chloride, pH, fishy aromatic and fishy flavor were loaded heavily on the first PC, which will be referenced as *fishy flavor component* or PC1. This indicated that the aromatic and flavor of fishy were highly correlated with the total nitrogen content, sodium chloride, and pH. Caramelized aromatic and sweet taste were loaded heavily on the second PC, which will be referenced as *sweetness component* or PC2. These attributes were negatively correlated with PC2. Samples with the highest degree of caramelized aromatic and sweet taste had larger negative scores, while samples with the lower degree of caramelized aromatic and sweet taste had positive scores. Finally, bitter taste and bitter aftertaste were loaded heavily on the third PC, which will be referenced as the *bitterness component* or PC3.

Table 27 Chemical parameters and sensory attributes loadings for the first three principal components performed by principal component analysis.

Parameters	Principal components		
	PC1	PC2	PC3
Loadings*			
Chemical parameters			
Total nitrogen	**0.922**	-0.103	0.060
Sodium Chloride	**0.726**	0.268	0.037
pH	**0.938**	0.053	0.039
Reducing sugar	0.460	-0.147	0.035
Sensory attributes			
Brown intensity	0.412	-0.133	0.256
Sweet aromatic	-0.204	-0.673	0.114
Caramelized aromatic	0.225	**-0.750**	0.334
Fermented aromatic	0.490	0.455	0.134
Fishy aromatic	**0.900**	0.142	-0.091
Musty aromatic	0.024	0.654	-0.314
Sweet taste	0.506	**-0.758**	-0.261
Salty taste	0.522	0.649	0.118
Bitter taste	0.028	-0.052	**0.929**
Umami taste	0.620	-0.239	-0.062
Sweet aftertaste	0.701	-0.435	-0.326
Salty aftertaste	0.631	0.611	0.037
Bitter aftertaste	0.072	-0.028	**0.937**
Caramelized flavor	0.657	-0.602	-0.203
Fishy flavor	**0.888**	0.168	0.097
Percentage variance	35.86	20.03	12.15

* Factor loadings marked in bold indicate situations where the attribute loadings meet the criteria of significance (0.72p = 0.01).

Besides the demonstration of the associations among the attributes, PCA can be used to display the relative "locations" of the samples with respect to each other and their characterizing attributes (Meilgaard *et al.,* 1999). The sensory and predicted chemical data were analyzed to identify the relationships and differences between the samples and the attributes using biplot analysis. According to the cluster analysis and PCA results from the study in Part 3, fish sauce samples were categorized into four clusters based on their sensory characteristics. Those four clusters were grouped together in the biplot of the first two PCs (55.89% of variability) as shown in Figure 40.

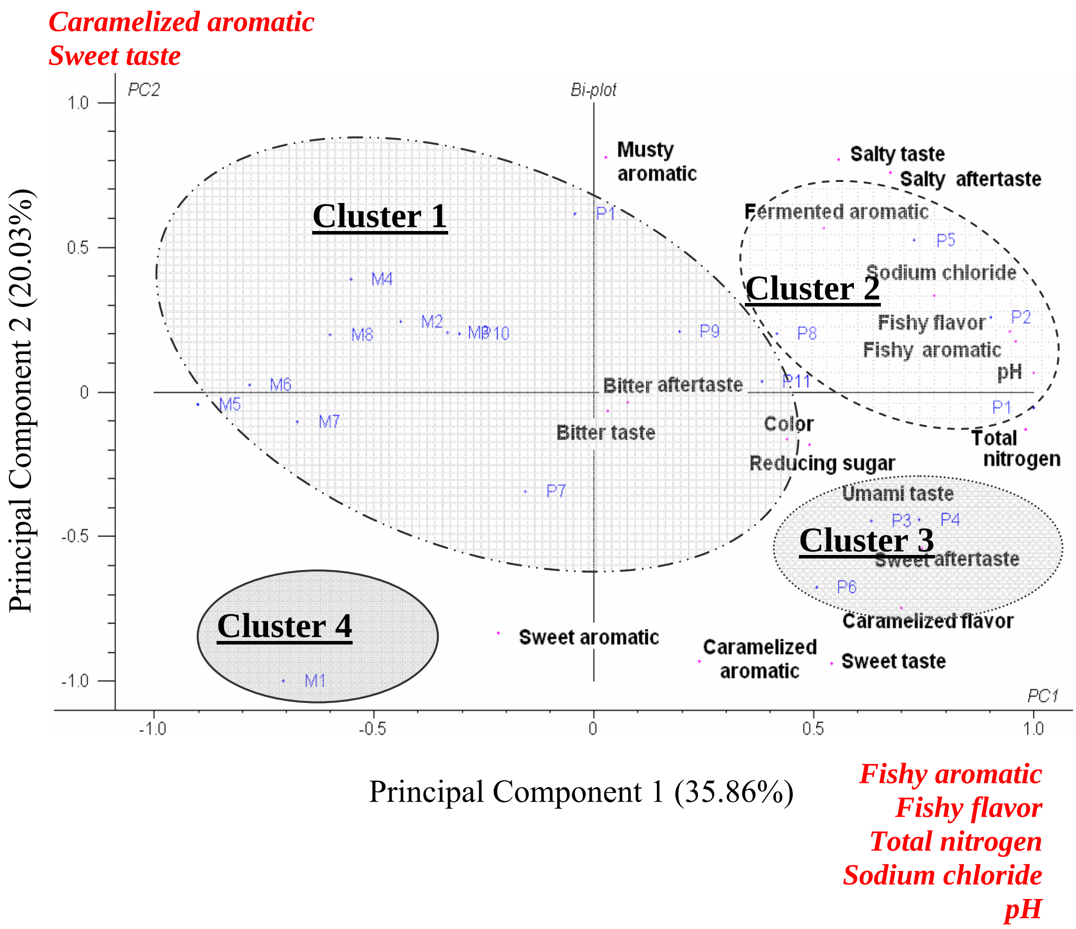


Figure 40 Scores for the twenty fish sauce samples and correlation coefficients of the sensory and chemical attributes with principal component 1 and 2. (P, pure fish sauce; M, mixed fish sauce)

Figure 40 shows the relationships and differences between the sensory attributes and the predicted chemical data for four clusters. According to *fishy component* (PC1) which explained 35.86% of the total variance, the samples in cluster 2 and 3 were identified by higher degrees of fishy aromatic, fishy flavor, total nitrogen, sodium chloride, and pH than the samples in cluster 1 and 4. In addition, the differences observed in the *sweetness component* (PC2) which explained an additional of 20.03% of the variation, the samples in cluster 3 and 4 had higher degree of *caramelized aromatic and sweet taste* than the samples in cluster 1 and 2.

It can be seen in Figure 40 that the pure fish sauce samples (P) and the mixed fish sauce samples (M) were distinguished well on this PC. There is a clear distinction of the scores from the two types of fish sauce on PC1. The pure fish sauce samples (P) is somewhat more shifted towards the higher values of PC1 than the mixed fish sauce samples (M). These results indicated that the pure fish sauces with higher contents of total nitrogen, sodium chloride and pH exhibited the higher degree of fishy aromatic and fishy flavor than the mixed fish sauce samples. The main chemical and sensory characteristics describing each cluster are summarized in Table 28.

Table 28 Sensory characteristics and chemical properties of four clusters of Thai fish sauce samples classified by principal component analysis (PCA).

Clusters	Chemical properties		Sensory characteristics				Types of fish sauce	Sample codes***
	Total nitrogen content, sodium chloride and pH*		Fishy aromatic and Fishy flavor*		Caramelized aromatic and sweet taste**			
	High	Low	Strong	Low	Strong	Low		
1		√		√		√	Pure and mixed	P7, P9,P10, P11,P12, M2, M3, M4, M5, M6, M7,M8
2	√		√			√	Pure	P1, P2, P5, P8
3	√		√		√		Pure	P4, P3, P6,
4		√		√	√		Mixed	M1

* Explained by principal component 1, which accounted for 35.86% of the variation.

** Explained by principal component 2, which accounted for 20.03% of the variation.

*** P: pure fish sauce, M: mixed fish sauce

4.5 Correlation between the Sensory Attributes and the Predicted Physical Properties Studied by PCA

The five predicted physical properties of density, baume, total soluble solid, refractive index and color a* using SCMWPLS models and all fifteen sensory attributes from descriptive analysis were subjected to PCA as input variables. The first three principal components explained for 43.03%, 18.63% and 11.66% of the total variation. The correlations between the sensory attributes and the predicted physical parameters were investigated by using the PCA factor loadings as shown in Table 29.

Using the guidelines provided by Stevens (1992) to inspect for significance of attribute loadings, an attribute was considered to load heavily on a given component if the factor loadings was greater than 0.72. Table 29 shows that all predicted physical properties, fishy aromatic and fishy flavor were loaded heavily on the first PC, which will be referenced as *fishy flavor component* or PC1. This indicated that the aromatic and flavor of fishy were highly correlated with the density, baume, total soluble solid, refractive index and color a* of fish sauce samples. Caramelized aromatic and sweet taste were loaded heavily on the second PC, which will be referenced as *sweetness component* or PC2.

Table 29 Physical parameters and sensory attributes loadings for the first three principal components performed by principal component analysis.

Parameters	Principal components		
	PC1	PC2	PC3
Loadings*			
- Physical parameters			
Density	**0.943**	-0.071	0.004
Baume	**0.943**	-0.031	-0.012
Total soluble solid (TSS)	**0.957**	0.001	-0.032
Refractive index (N_D)	**0.957**	-0.006	-0.015
Color a*	**0.897**	-0.032	0.115
- Sensory attributes			
Brown intensity	0.412	0.142	0.170
Sweet aromatic	-0.200	0.649	0.068
Caramelized aromatic	0.208	**0.751**	0.213
Fermented aromatic	0.526	-0.441	0.075
Fishy aromatic	**0.898**	-0.109	-0.054
Musty aromatic	0.068	-0.657	-0.216
Sweet taste	0.484	**0.781**	-0.167
Salty taste	0.547	-0.634	0.085
Bitter taste	0.006	0.056	0.614
Umami taste	0.578	0.274	-0.017
Sweet aftertaste	0.649	0.475	-0.200
Salty aftertaste	0.625	-0.581	0.040
Bitter aftertaste	0.055	0.029	0.615
Caramelized flavor	0.636	0.631	-0.120
Fishy flavor	**0.873**	-0.137	0.069
Percentage variance	43.03	18.63	11.66

* Factor loadings marked in bold indicate situations where the attribute loadings meet the criteria of significance (0.72p = 0.01).

Besides the demonstration of the associations among the attributes, PCA can be used to display the relative "locations" of the samples with respect to each other and their characterizing attributes (Meilgaard *et al.*, 1999). The sensory and the predicted physical data were analyzed to identify the relationships and differences between the samples and the attributes using biplot analysis. According to the cluster analysis and PCA results, fish sauce samples were categorized into four clusters based on their sensory characteristics. Those four clusters were grouped together in the biplot of the first two PCs (61.66% of variability) as shown in Figure 41.

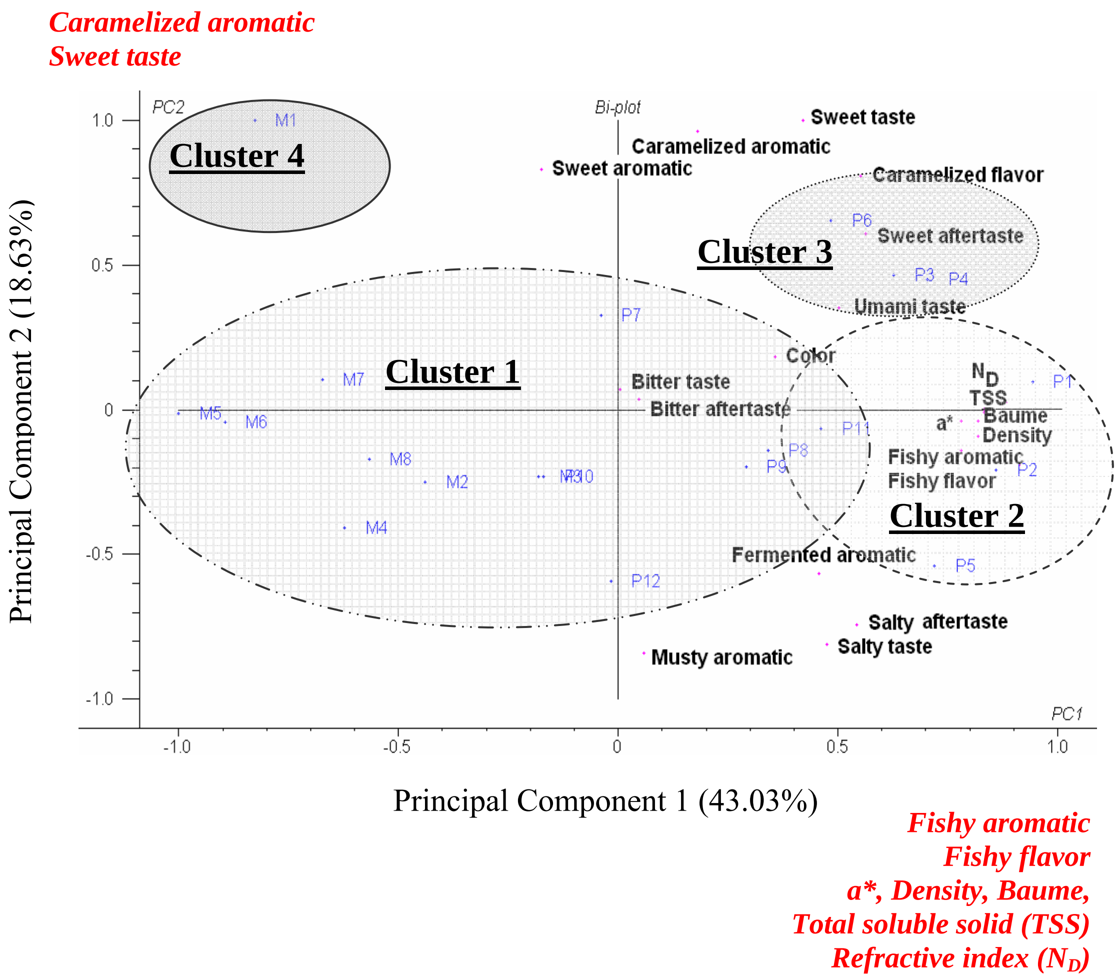


Figure 41 Scores for the twenty fish sauce samples and correlation coefficients of the sensory and physical attributes with principal component 1 and 2. (P, pure fish sauce; M, mixed fish sauce)

Figure 41 shows the relationships and differences between the sensory attributes and the predicted physical data for four clusters. According to *fishy component* (PC1) which explained 43.03% of the total variance, the samples in cluster 2 and 3 were identified by higher degrees of *fishy aromatic, fishy flavor, density, baume, total soluble solid, refractive index and color a** than the samples in cluster 1 and 4. In addition, the differences observed in the *sweetness component* (PC2) which explained an additional of 18.63% of the variation, the samples in cluster 3 and 4 had higher degree of *caramelized aromatic and sweet taste* than the samples in cluster 1 and 2.

It can be seen in Figure 41 that the pure fish sauce samples (P) and the mixed fish sauce samples (M) were distinguished well on this PC. There is a clear distinction of the scores from the two types of fish sauce on PC1. The pure fish sauce samples (P) is somewhat more shifted towards the higher values of PC1 than the mixed fish sauce samples (M). These results indicated that the pure fish sauces with higher degree of density, baume, total soluble solid, refractive index and color a* exhibited the higher degree of fishy aromatic and fishy flavor than the mixed fish sauce samples. The main physical and sensory characteristics describing each cluster are summarized in Table 30.

Table 30 Sensory characteristics and physical properties of four clusters of Thai fish sauce samples classified by principal component analysis (PCA).

Clusters	Physical properties: Density, Baume, TSS, N_D, Color[1]*		Sensory characteristics: Fishy aromatic and Fishy flavor*		Sensory characteristics: Caramelized aromatic and Sweet taste**		Types of fish sauce	Sample codes***
	High	Low	Strong	Low	Strong	Low		
1		√		√		√	Pure and mixed	P7, P9,P10, P11,P12, M2, M3, M4, M5, M6, M7,M8
2	√		√			√	Pure	P1, P2, P5, P8
3	√		√		√		Pure	P4, P3, P6,
4		√		√	√		Mixed	M1

[1]TSS; Total soluble solid, ND; Refractive index; Color; a*

* Explained by principal component 1, which accounted for 43.03% of the variation.

** Explained by principal component 2, which accounted for 18.63% of the variation.

*** P: pure fish sauce, M: mixed fish sauce

5. <u>Development of Quantitative Models for Predicting the Sensory Properties of Thai Fish Sauces</u>

The sensory properties and NIR spectral data of twenty fish sauce from the studies in Part 3 and 4 were applied to develop the quantitative models using PLS regression.

5.1 Sensory Evaluation

To develop the PLS models, the fifteen sensory attribute scores of Thai fish sauces from the study in Part 3 were used as dependent variables (Y-data). These attributes were brown color, five aromatic descriptors (sweet aromatic, caramelized aromatic, fermented aromatic, fishy aromatic, and musty aromatic), four taste descriptors (sweet taste, salty taste, bitter taste, and umami taste), three aftertaste descriptors (sweet aftertaste, salty aftertaste, and bitter aftertaste) and two flavor descriptors (caramelized flavor and fishy flavor). Their mean sensory attribute values of the 12 pure fish sauces and 8 mixed fish sauces are shown in Table 20 and 21.

5.2 NIR Spectra

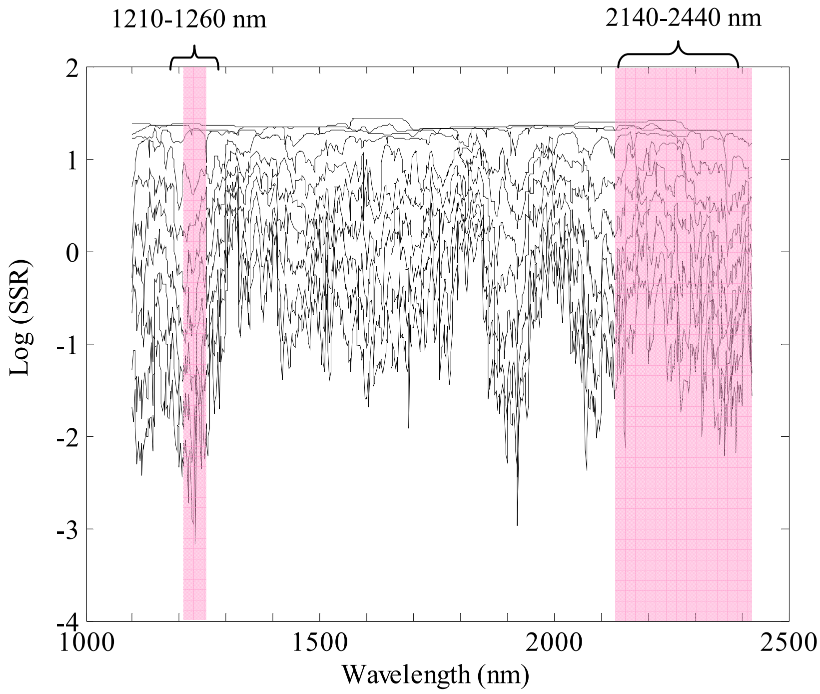
NIR transflectance spectra of 20 investigated Thai fish sauce samples are shown in Figure 37. Their absorbances of each wavelength in each sample were used as independent variables (X-data) to develop the models. For chemometric analysis, the 1900-2000 nm region was not employed in order to avoid heavily overlapping absorption bands. William and Norris (1990) explained the phenomenon of overlapping is caused by a problem with gratings that does not exist with prism in the NIR instrument, that is light with several wavelengths leaves the grating at the same angle of dispersion. Therefore, the wavelength regions of 1100-1900 and 2000-2440 nm will be referenced as the whole spectra regions and used for chemometric analysis.

5.3 Model Analysis

To find out the suitable input wavelength variables (X-data) for developing the PLS model, two kinds of input wavelength variables were compared. They were i) the whole spectra regions of 1100-1900 and 2000-2440 nm and ii) the selected wavelength regions. In this study, the selected wavelength regions were identified by the two wavelength interval selection methods named i) moving window partial least squares regression (MWPLSR) and ii) searching combination moving window partial least squares (SCMWPLS). The MWPLSR method was applied for selecting informative regions as shown in section 5.3.1. Subsequently the SCMWPLS method was used to combine and optimize the informative regions as shown in section 5.3.2. Finally, the whole spectra regions and the selected wavelength regions were applied to develop the PLS model. The performance of developed models was calculated and compared. The best prediction model with the lowest error was selected.

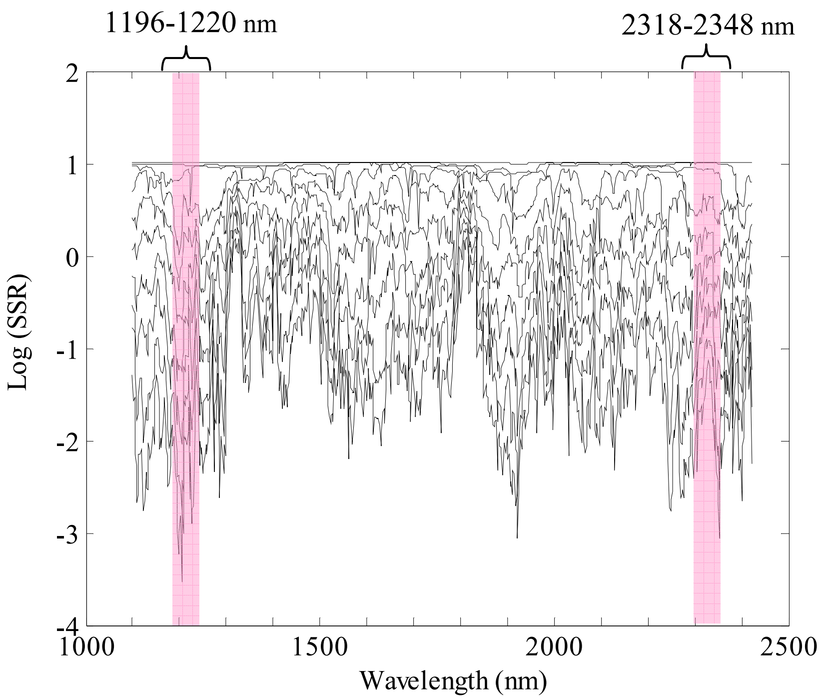
5.3.1 Moving Window Partial Least Squares Regression (MWPLSR)

MWPLSR was performed to search informative regions from the spectral regions of 1100-1900 and 2000-2440 nm for all fifteen sensory attributes. Informative regions mean that they contain useful information for a PLS model building and are helpful to improve the performance of the model (Jiang *et al*., 2002). The residue lines for brown color, five aromatic descriptors, four taste descriptors, three aftertaste descriptors and two flavor descriptors are shown in Figure 42, 43, 44, 45 and 46, respectively.



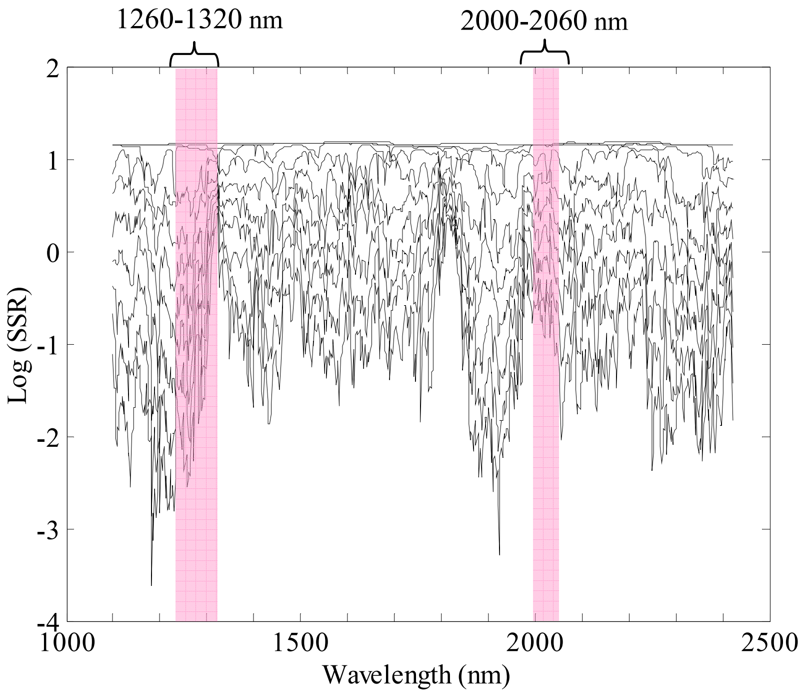
Brown intensity

Figure 42 Residue lines for brown color of Thai fish sauces obtained by moving window partial least squares regression (MWPLSR). The shade areas are final informative regions.

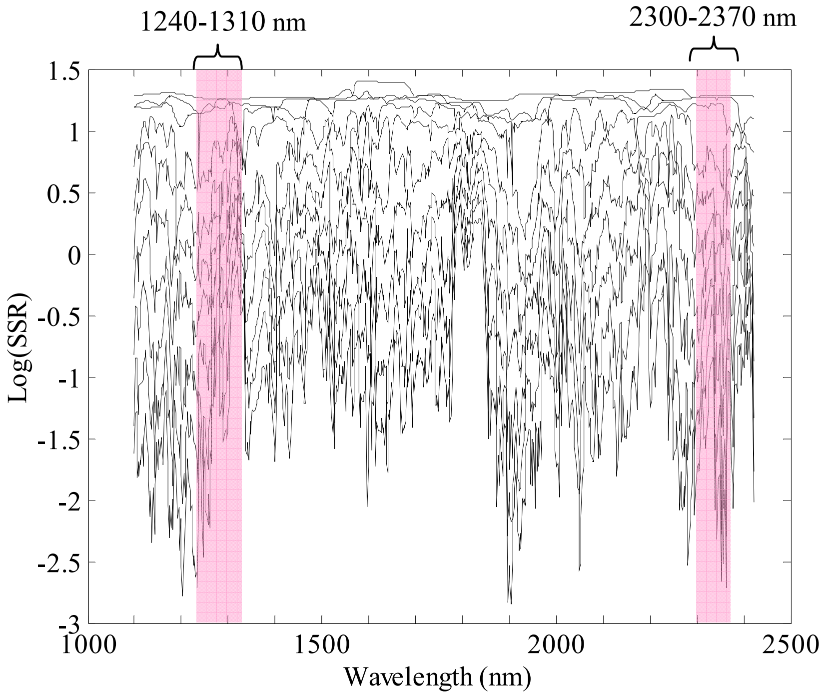


(a) Sweet aromatic

Figure 43 Residue lines for five aromatic descriptors of Thai fish sauces obtained by moving window partial least squares regression (MWPLSR). The shade areas are final informative regions. (a) sweet aromatic, (b) caramelized aromatic, (c) fermented aromatic, (d) fishy aromatic, and (e) musty aromatic.

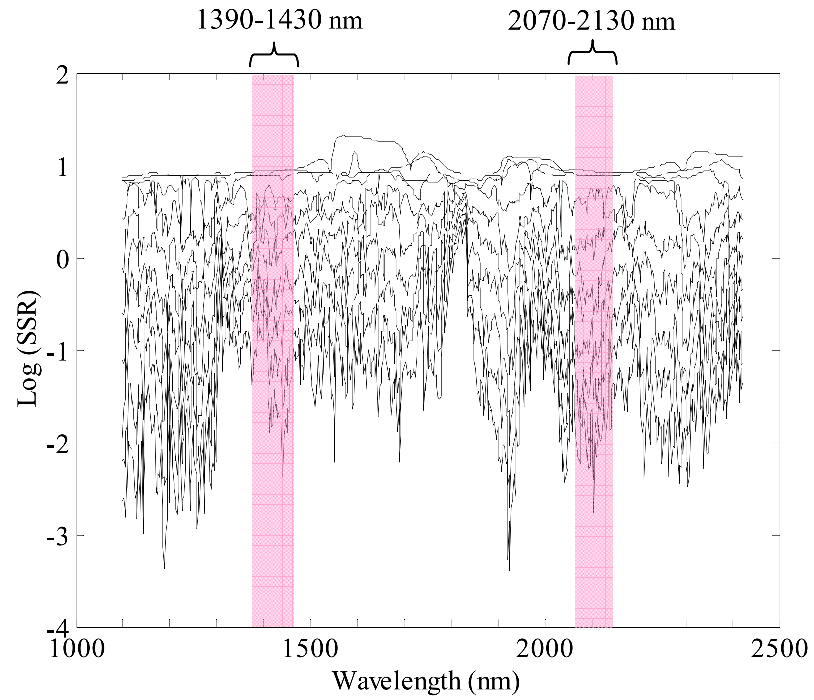


(b) Caramelized aromatic

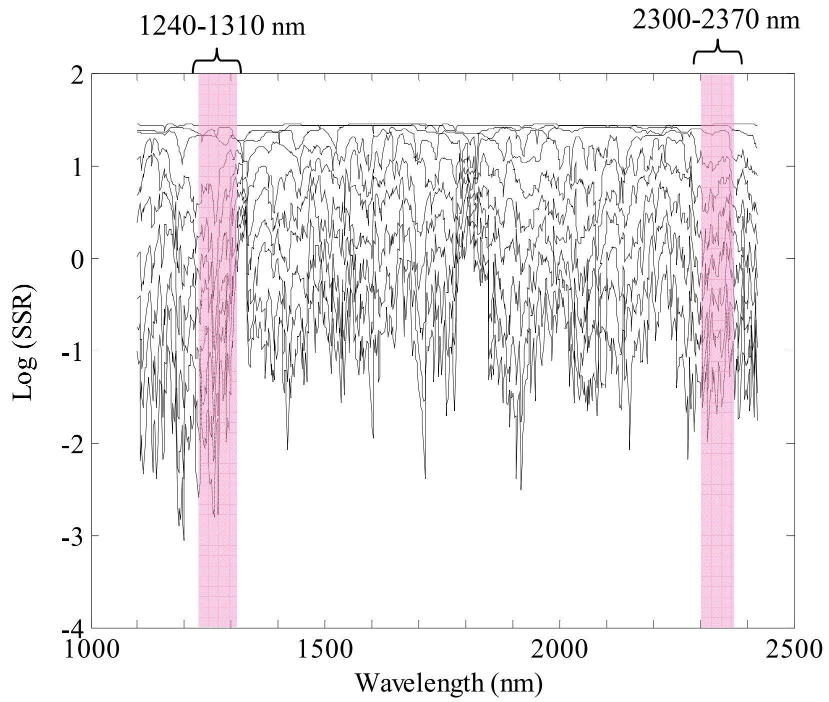


(c) Fermented aromatic

Figure 43 (Continued)

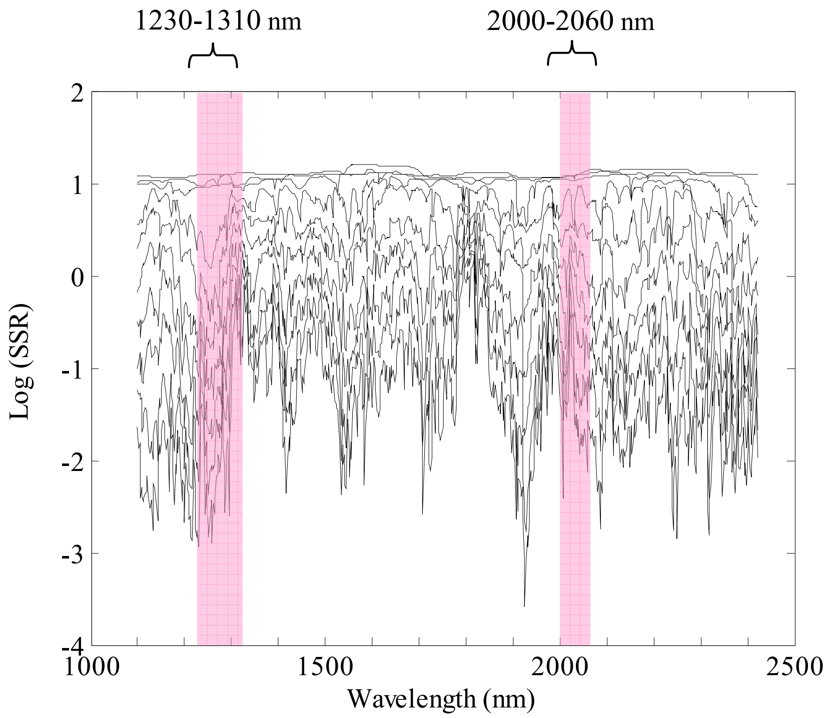


(d) Fishy aromatic



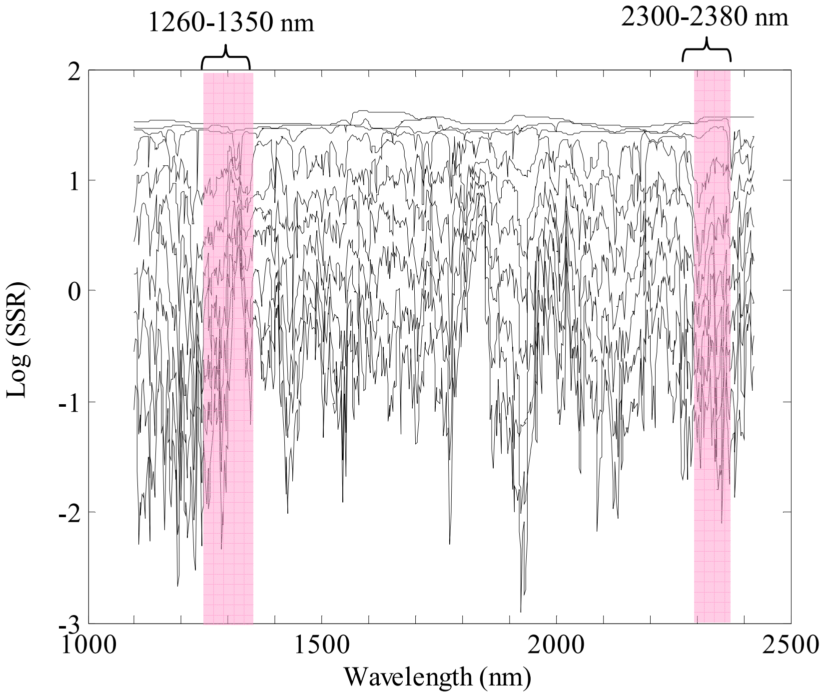
(e) Musty aromatic

Figure 43 (Continued)

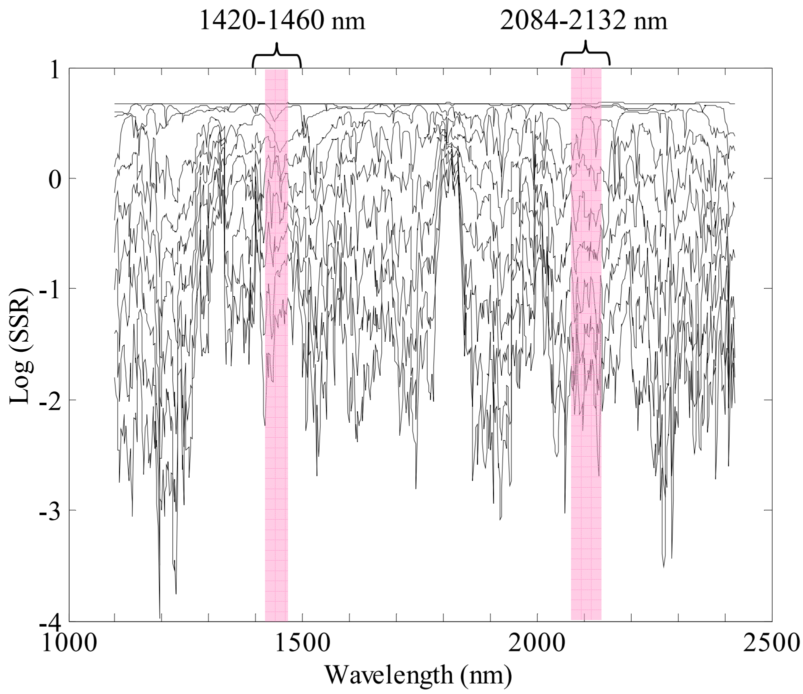


(a) Sweet taste

Figure 44 Residue lines for four taste descriptors of Thai fish sauces obtained by moving window partial least squares regression (MWPLSR). The shade areas are final informative regions. (a) sweet taste, (b) salty taste, (c) bitter taste, and (d) umami taste.

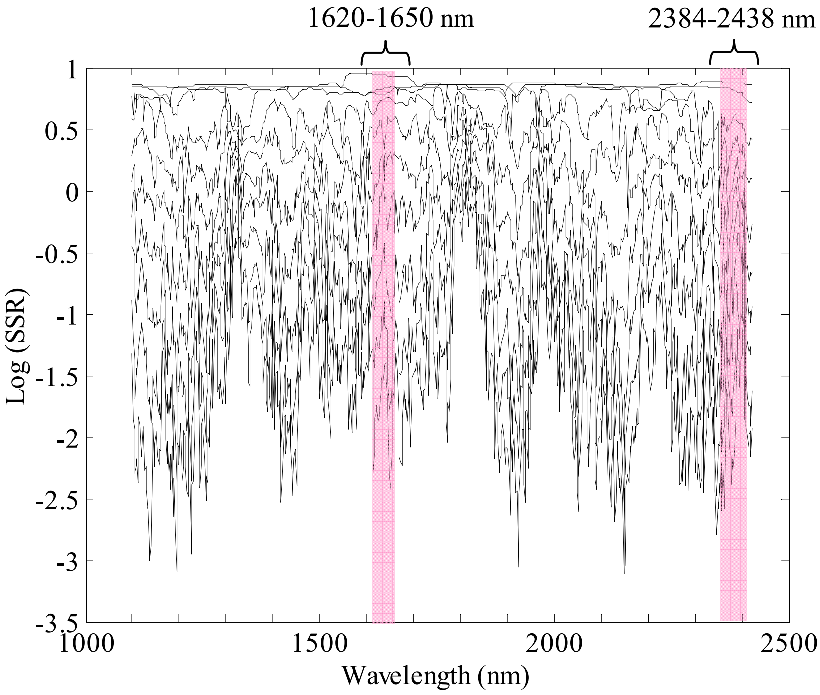


(b) Salty taste



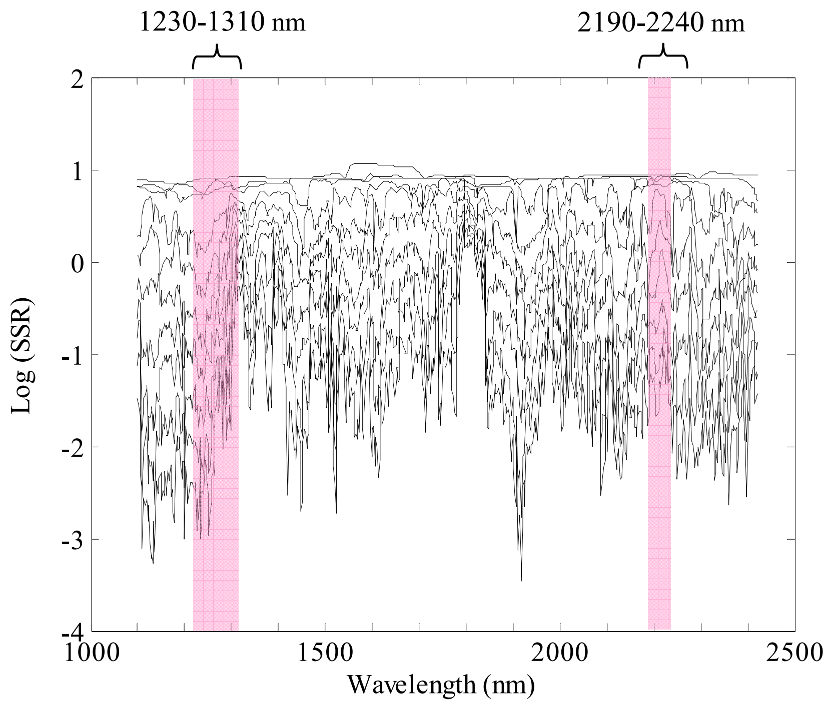
(c) Bitter taste

Figure 44 (Continued)



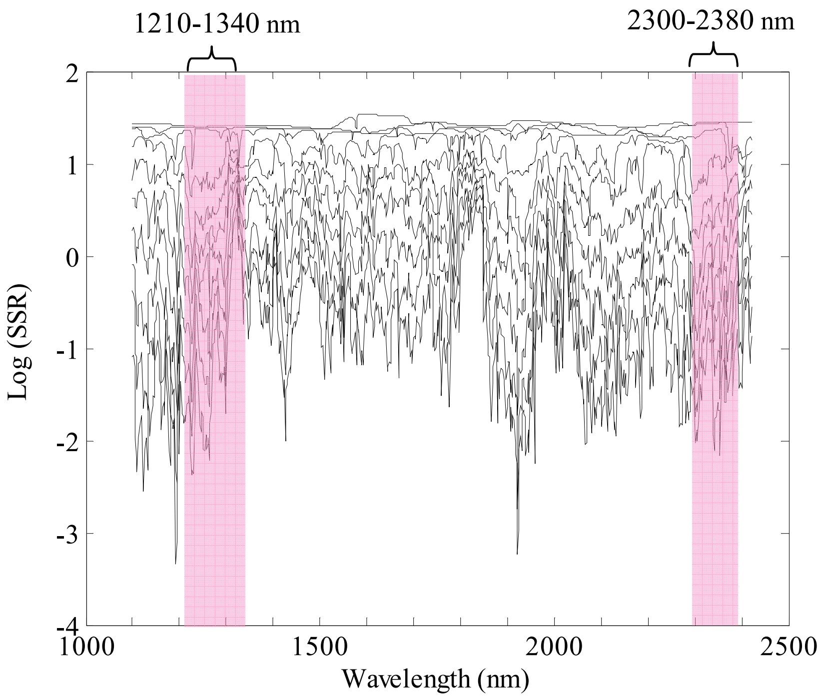
(j) Umami taste

Figure 44 (Continued)

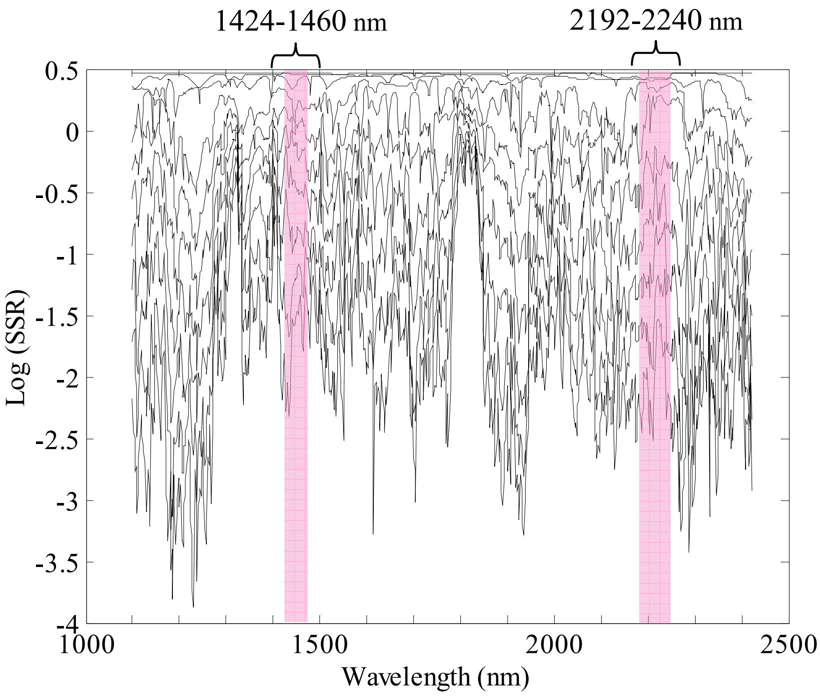


(a) Sweet aftertaste

Figure 45 Residue lines for three aftertaste descriptors of Thai fish sauces obtained by moving window partial least squares regression (MWPLSR). The shade areas are final informative regions. (a) sweet aftertaste, (b) salty aftertaste, and (c) bitter aftertaste.

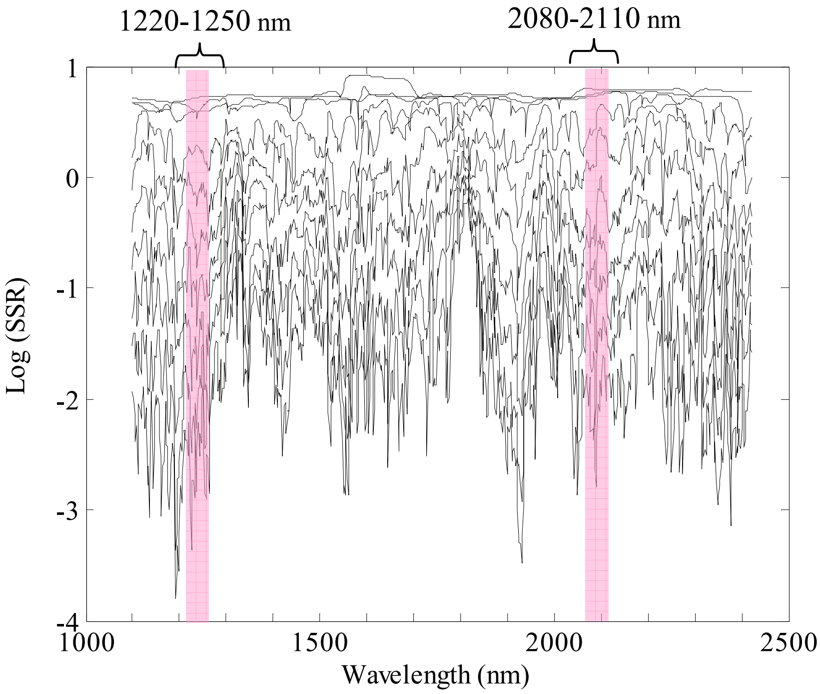


(b) Salty aftertaste



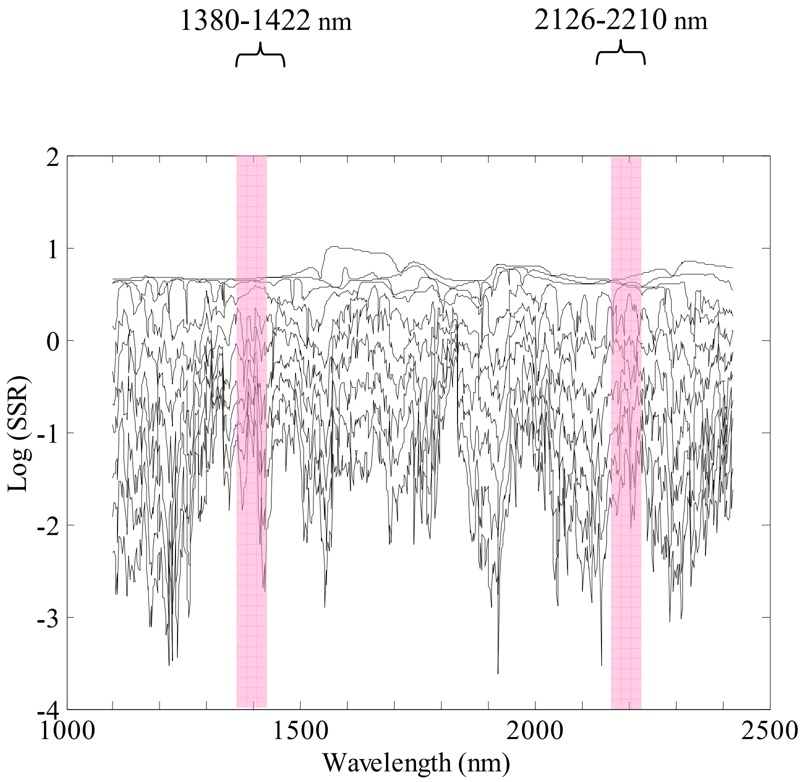
(c) Bitter aftertaste

Figure 45 (Continued)



(n) Caramelized flavor

Figure 46 Residue lines for two flavor descriptors of Thai fish sauces obtained by moving window partial least squares regression (MWPLSR). The shade areas are final informative regions. (a) caramelized flavor and (b)fishy flavor.



(o) Fishy flavor

Figure 46 (Continued)

Like chemical and physical parameters, the sensory attributes have a few informative regions, which give small values of the sum of squared residues (SSR). The informative spectra regions of those sensory attributes are presented in Table 30. All regions correspond to the regions for the combinations of stretching and deformation modes of amino acids, which are one of the most important effective parameters for the qualities of fish sauces. The fish proteins are degraded by fermentation to amino acids that were produced a wide varieties of tastes, flavors, colors and compositions of fish sauces (Gildberg, 2001).

Table 30 Chemical assignments of spectral regions suggested by moving window partial least squares regression (MWPLSR) for the sensory attributes of Thai fish sauces.

Sensory attributes	Informative regions (nm)	Functional groups and molecules[1]
Brown intensity	1210-1260	CH_2
	2140-2440	HC=CH, $CONH_2$, CONHR, protein, amino acids,CH_3, -CHO, CH_2, HC=$CHCH_2$
Sweet aromatic	1196-1220	CH_3, CH_2
	2318-2348	CH_2, HC=$CHCH_2$
Caramelized aromatic	1260-1320	CH_2
	2000-2060	$CONH_2$, CONHR, protein,
Fermented aromatic	1240-1310	CH_2
	2300-2370	CH_2, HC=$CHCH_2$
Fishy aromatic	1390-1430	CH_2, $CONH_2$
	2070-2130	$CONH_2$, CONHR, amino acids
Musty aromatic	1240-1310	CH_2
	2300-2370	CH_2, HC=$CHCH_2$
Sweet taste	1230-1310	CH_2
	2000-2060	$CONH_2$, CONHR, protein,
Salty taste	1260-1350	CH_3, CH_2
	2300-2380	CH_2, HC=$CHCH_2$, ROH
Bitter taste	1420-1460	$CONH_2$, CH
	2084-2132	ROH, $CONH_2$, CONHR, amino acids
Umami taste	1620-1650	=CH_2
	2384-2438	ROH
Sweet aftertaste	1230-1310	CH_2
	2190-2240	HC=CH, -CHO, amino acids
Salty aftertaste	1210-1340	CH_2
	2300-2380	CH_2, HC=$CHCH_2$, ROH

[1]Williams and Norris (1990) and Siesler *et al*. (2002)

Table 30 (Continued)

Sensory attributes	Informative regions (nm)	Functional groups and molecules[1]
Bitter aftertaste	1424-1460	$CONH_2$
	2192-2240	HC=CH, -CHO, amino acids
Caramelized flavor	1220-1250	CH_2
	2080-2110	ROH, $CONH_2$, CONHR,
Fishy flavor	1380-1422	CH_2
	2126-2210	amino acids, $CONH_2$, CONHR,

[1]Williams and Norris (1990) and Siesler *et al*. (2002)

5.3.2 Searching Combination Moving Window Partial Least Squares (SCMWPLS)

SCMWPLS was performed to search for optimized combinations of the informative regions obtained by MWPLSR. In order to eliminate uninformative regions, SCMWPLS optimizes the informative regions and combines them altogether (Kasemsumran et al., 2003; Du et al., 2004; Kasemsumran et al., 2004). The optimized combinations of informative regions obtained by SCMWPLS are presented in Table 31.

Table 31 Selected wavelength variables obtained by moving window partial least square regression (MWPLSR) and searching combination moving window partial least squares (SCMWPLS) for sensory attributes of Thai fish sauces.

Analytes	Selected wavelength variables (nm)	
	MWPLSR method	SCMWPLS method
Brown intensity	1210-1260, 2140-2440	2340-2440
Sweet aromatic	1196-1220, 2318-2348	2318-2348
Caramelized aromatic	1260-1320, 2000-2060	1260-1274, 2000-2052
Fermented aromatic	1240-1310, 2300-2370	1228-1308
Fishy aromatic	1390-2430, 2070-2130	2070-2130
Musty aromatic	1240-1310, 2300-2370	1228-2308
Sweet taste	1230-1310, 2000-2060	1274-1306, 2000-2058
Salty taste	1260-1350, 2300-2380	1260-1306, 2324-2374
Bitter taste	1420-1460, 2084-2132	2086-2130
Umami taste	1620-1650, 2384-2438	2384-2438
Sweet aftertaste	1230-1310, 2190-2240	1230-1310
Salty aftertaste	1210-1340, 2300-2380	2302-2380
Bitter aftertaste	1424-1460, 2192- 2240	2194-2240
Caramelized flavor	1220-1250, 2080-2110	2080-2180
Fishy flavor	1380-1422, 2126-2210	2154-2206

To demonstrate the performance of these two wavelength selection methods, four kinds of input wavelength variables (X-data) were applied to develop the PLS calibration models as shown in section 5.3.3. These four input variables were i) the whole spectra region (section 5.2), ii) the informative regions obtained by MWPLSR (section 5.3.1), iii) the direct combination of informative regions and iv) the optimized combination of informative regions obtained by SCMWPLS (section 5.3.2).

5.3.3 PLS Model Analysis

There are four main difference among input wavelength variables (X-data) for developing the PLS models. A first input wavelength variable is the whole spectra region (1100-1900 nm and 2000-2440 nm) as shown in section 5.2. The second difference input variable is the selected wavelengths obtained by MWPLSR method as shown in section 5.3.1. The third difference input variable is the direct combination of selected wavelengths obtained by MWPLSR. The fourth difference input variable is the selected wavelengths obtained by SCMWPLS method as shown in section 5.3.2. These four differences input wavelength variables were used to build PLS models. To compare the performance of developed models, the coefficient of determination (R^2), the root mean square errors of cross-validation (RMSECV), the ratio of prediction to deviation (RPD), the range error ratio (RER) were calculated as shown in Table 32.

Table 32 Prediction results of PLS calibration models for 15 sensory attributes of Thai fish sauces.

Sensory attributes	Wavelength selection methods	Spectra regions (nm)	PLS factors	R^2	RMSECV	RPD	RER
Brown intensity	Full spectra	1100-1900, 2000-2440	1	0.048	1.171	1.03	3.73
	MWPLSR[1]	1210-1260	1	0.039	1.166	1.03	3.75
	MWPLSR[2]	2140-2440	1	0.081	1.164	1.03	3.75
	MWPLSR[3]	1210-1260, 2140-2440	1	0.140	1.099	1.09	3.98
	SCMWPLS[4]	2340-2440	1	0.176	1.072	1.12	4.08
Sweet aromatic	Full spectra	1100-1900, 2000-2440	1	0.453	0.793	0.92	2.92
	MWPLSR[1]	1196-1220	1	0.302	0.779	0.94	2.97
	MWPLSR[2]	2318-2348	4	0.144	0.658	1.11	3.52
	MWPLSR[3]	1196-1220, 2318-2348	5	0.199	0.666	1.10	3.47
	SCMWPLS[4]	2318-2348	4	0.144	0.658	1.11	3.52
Caramelized aromatic	Full spectra	1100-1900, 2000-2440	1	0.019	0.971	0.94	3.79
	MWPLSR[1]	1260-1320	1	0.002	0.920	0.99	4.00
	MWPLSR[2]	2000-2060	1	0.006	0.920	0.99	4.00
	MWPLSR[3]	1260-1320, 2000-2060	1	0.016	0.981	0.93	3.75
	SCMWPLS[4]	1260-1274, 2000-2052	1	0.032	0.883	1.03	4.17

R^2: Coefficient of determination , RMSECV: Root mean square error of cross-validation
RPD: Ratio of prediction deviation (Standard deviation/ RMSECV)
RER: Range error ratio (maximum – minimum/RMSECV)
MWPLSR: Moving window partial least squares regression
SCMWPLS: Searching combination moving window partial least squares
[1]The method was used to identify the informative regions in the region of 1100-1900 nm.
[2]The method was used to identify the informative regions in the region of 2000-2440 nm.
[3]The method was used to directly combine the informative regions obtained by MWPLSR.
[4]The method was used to optimize the combination of the informative regions obtained by MWPLSR.

Table 32 (Continued)

Sensory attributes	Wavelength selection methods	Spectra regions (nm)	PLS factors	R^2	RMSECV	RPD	RER
Fermented aromatic	Full spectra	1100-1900, 2000-2440	1	0.048	1.171	0.98	3.68
	MWPLSR[1]	1240-1310	2	0.211	1.042	1.10	4.14
	MWPLSR[2]	2300-2370	3	0.121	1.093	1.05	3.95
	MWPLSR[3]	1240-1310, 2300-2370	1	0.023	1.177	0.98	3.66
	SCMWPLS[4]	1228-1308	2	0.238	0.998	1.15	4.32
Fishy aromatic	Full spectra	1100-1900, 2000-2440	2	0.623	0.668	1.67	7.08
	MWPLSR[1]	1390-1430	1	0.507	0.776	1.43	6.10
	MWPLSR[2]	2070-2130	5	0.717	0.594	1.87	7.96
	MWPLSR[3]	1390-1430, 2070-2130	1	0.559	0.733	1.52	6.45
	SCMWPLS[4]	2070-2130	5	0.717	0.594	1.87	7.96
Musty aromatic	Full spectra	1100-1900, 2000-2440	1	0.048	1.171	1.03	3.72
	MWPLSR[1]	1240-1310	2	0.211	1.042	1.16	4.19
	MWPLSR[2]	2300-2370	3	0.121	1.093	1.10	3.99
	MWPLSR[3]	1240-1310, 2300-2370	1	0.023	1.177	1.02	3.71
	SCMWPLS[4]	1228-1308	2	0.238	0.998	1.21	4.37

R^2: Coefficient of determination , RMSECV: Root mean square error of cross-validation
RPD: Ratio of prediction deviation (Standard deviation/ RMSECV)
RER: Range error ratio (maximum – minimum/RMSECV)
MWPLSR: Moving window partial least squares regression
SCMWPLS: Searching combination moving window partial least squares
[1]The method was used to identify the informative regions in the region of 1100-1900 nm.
[2]The method was used to identify the informative regions in the region of 2000-2440 nm.
[3]The method was used to directly combine the informative regions obtained by MWPLSR.
[4]The method was used to optimize the combination of the informative regions obtained by MWPLSR.

Table 32 (Continued)

Sensory attributes	Wavelength selection methods	Spectra regions (nm)	PLS factors	R^2	RMSECV	RPD	RER
Sweet taste	Full spectra	1100-1900, 2000-2440	1	0.155	0.859	1.10	3.61
	MWPLSR[1]	1230-1310	2	0.250	0.839	1.13	3.69
	MWPLSR[2]	2000-2060	4	0.282	0.842	1.12	3.68
	MWPLSR[3]	1230-1310, 2000-2060	1	0.090	0.924	1.02	3.35
	SCMWPLS[4]	1274-1306, 2000-2058	1	0.187	0.811	1.17	3.82
Salty taste	Full spectra	1100-1900, 2000-2440	1	0.056	1.457	1.02	3.91
	MWPLSR[1]	1260-1350	3	0.162	1.42	1.05	4.01
	MWPLSR[2]	2300-2380	7	0.222	1.439	1.03	3.96
	MWPLSR[3]	1260-1350, 2300-2380	6	0.106	1.468	1.01	3.88
	SCMWPLS[4]	1260-1306, 2324-2374	1	0.101	1.398	1.06	4.07
Bitter taste	Full spectra	1100-1900, 2000-2440	1	0.511	0.567	0.88	3.15
	MWPLSR[1]	1420-1460	1	0.315	0.544	0.92	3.28
	MWPLSR[2]	2084-2132	1	0.388	0.562	0.89	3.18
	MWPLSR[3]	1420-1460, 2084-2132	1	0.510	0.550	0.91	3.25
	SCMWPLS[4]	2086-2130	1	0.361	0.544	0.92	3.28

R^2: Coefficient of determination , RMSECV: Root mean square error of cross-validation
RPD: Ratio of prediction deviation (Standard deviation/ RMSECV)
RER: Range error ratio (maximum – minimum/RMSECV)
MWPLSR: Moving window partial least squares regression
SCMWPLS: Searching combination moving window partial least squares
[1]The method was used to identify the informative regions in the region of 1100-1900 nm.
[2]The method was used to identify the informative regions in the region of 2000-2440 nm.
[3]The method was used to directly combine the informative regions obtained by MWPLSR.
[4]The method was used to optimize the combination of the informative regions obtained by MWPLSR.

Table 32 (Continued)

Sensory attributes	Wavelength selection methods	Spectra regions (nm)	PLS factors	R^2	RMSECV	RPD	RER
Umami taste	Full spectra	1100-1900, 2000-2440	1	0.078	0.671	1.04	3.45
	MWPLSR[1]	1620-1650	1	0.100	0.654	1.07	3.54
	MWPLSR[2]	2384-2438	2	0.131	0.651	1.08	3.55
	MWPLSR[3]	1620-1650, 2384-2438	1	0.106	0.655	1.07	3.53
	SCMWPLS[4]	2384-2438	2	0.131	0.651	1.08	3.55
Sweet aftertaste	Full spectra	1100-1900, 2000-2440	1	0.243	0.693	1.17	4.35
	MWPLSR[1]	1230-1310	3	0.428	0.609	1.33	4.96
	MWPLSR[2]	2190-2240	5	0.292	0.680	1.19	4.44
	MWPLSR[3]	1230-1310, 2190-2240	1	0.217	0.711	1.14	4.24
	SCMWPLS[4]	1230-1310	3	0.428	0.609	1.33	4.96
Salty aftertaste	Full spectra	1100-1900, 2000-2440	1	0.060	1.323	1.01	3.81
	MWPLSR[1]	1210-1340	1	0.176	1.191	1.13	4.24
	MWPLSR[2]	2300-2380	9	0.454	0.969	1.38	5.21
	MWPLSR[3]	1210-1340, 2300-2380	1	0.094	1.268	1.06	3.98
	SCMWPLS[4]	2302-2380	9	0.674	0.750	1.79	6.73

R^2: Coefficient of determination , RMSECV: Root mean square error of cross-validation
RPD: Ratio of prediction deviation (Standard deviation/ RMSECV)
RER: Range error ratio (maximum – minimum/RMSECV)
MWPLSR: Moving window partial least squares regression
SCMWPLS: Searching combination moving window partial least squares
[1]The method was used to identify the informative regions in the region of 1100-1900 nm.
[2]The method was used to identify the informative regions in the region of 2000-2440 nm.
[3]The method was used to directly combine the informative regions obtained by MWPLSR.
[4]The method was used to optimize the combination of the informative regions obtained by MWPLSR.

Table 32 (Continued)

Sensory attributes	Wavelength selection methods	Spectra regions (nm)	PLS factors	R^2	RMSECV	RPD	RER
Bitter aftertaste	Full spectra	1100-1900, 2000-2440	2	0.013	0.415	0.95	3.27
	MWPLSR[1]	1424-1460	1	0.148	0.451	0.87	3.01
	MWPLSR[2]	2192-2240	4	0.112	0.374	1.05	3.63
	MWPLSR[3]	1424-1460, 2192-2240	1	0.279	0.412	0.95	3.29
	SCMWPLS[4]	2194-2240	1	0.114	0.374	1.05	3.63
Caramelized flavor	Full spectra	1100-1900, 2000-2440	1	0.234	0.589	1.15	3.76
	MWPLSR[1]	1220-1250	1	0.321	0.549	1.24	4.03
	MWPLSR[2]	2080-2110	5	0.460	0.511	1.33	4.33
	MWPLSR[3]	1220-1250, 2080-2110	2	0.296	0.566	1.20	3.91
	SCMWPLS[4]	2080-2108	6	0.582	0.457	1.49	4.84
Fishy flavor	Full spectra	1100-1900, 2000-2440	1	0.513	0.530	1.45	5.62
	MWPLSR[1]	1380-1422	1	0.508	0.532	1.45	5.60
	MWPLSR[2]	2126-2210	1	0.543	0.511	1.51	5.83
	MWPLSR[3]	1380-1422, 2126-2210	1	0.548	0.541	1.42	5.51
	SCMWPLS[4]	2154-2206	1	0.546	0.509	1.51	5.85

R^2: Coefficient of determination , RMSECV: Root mean square error of cross-validation
RPD: Ratio of prediction deviation (Standard deviation/ RMSECV)
RER: Range error ratio (maximum – minimum/RMSECV)
MWPLSR: Moving window partial least squares regression
SCMWPLS: Searching combination moving window partial least squares
[1]The method was used to identify the informative regions in the region of 1100-1900 nm.
[2]The method was used to identify the informative regions in the region of 2000-2440 nm.
[3]The method was used to directly combine the informative regions obtained by MWPLSR.
[4]The method was used to optimize the combination of the informative regions obtained by MWPLSR.

Using the full spectral region, the optimal number of the PLS factors for the prediction of the attributes fishy aromatic and bitter flavor was 2. For all other attributes, the optimal number was 1. The best prediction of the investigated variables was obtained with fishy aromatic and fishy flavor. The RMSECV for the attributes fishy aromatic and fishy flavor was 0.668 and 0.530, respectively. These RMSECV values resulted in RER values of 7.08 and 5.62, respectively. The fishy aromatic and fishy flavor attributes might also have practical utility since the RER was higher than 5 (Karoui *et al*., 2006). Considering the other sensory attributes, the RPD ratios were less than 1.50. These indicated that the developed models yielded poor predictions and cannot be used for further prediction.

It can be seen from Table 32 that the PLS models based on the informative regions selected by MWPLSR and SCMWPLS show better performance than that based on the whole spectra region. With the aids of wavelength interval selection methods, the models were improved the prediction ability. The sensory attributes might also have practical utility were fishy aromatic (RPD = 1.87, RER = 7.96), salty flavor (RPD = 1.79, RER = 6.73), and fishy flavor (RPD = 1.51, RER = 5.85).

It is noted in Table 32 that predictive information related to sensory properties and NIR spectra was not clear. It has been reported that the correlation between NIR spectroscopy and sensory properties did not seem to be related to a specific chemical moiety in the sample. It might be caused by collinearity between compositional variables such as between wavelengths or between other sensory properties (Hildrum *et al.*, 1995; Byrne *et al.*, 1998; Martens and Martens, 2001).