16

Nondestructive Estimation of the Contents of the Functional Elements in Soybean by Near Infrared Reflectance Spectroscopy

Tetsuo Sato
National Agricultural Research Center for Kyushu Okinawa Region (KONARC), National Agriculture and Food Research Organization (NARO) Japan

1. Introduction

In recent days, the consumers' demands for the agricultural products are highly upgraded and widely diversified. Their attentions are focused not only on the major constituents, the nutritional constituents, or the palatability, but also on the physiologically functional activities. Soybean (*Glycine max* L.) is a major (oilseed) crop, and a good source of nutrition such as protein and oil. In Japan, soybean is used for producing the excellent traditional foods such as tofu, miso, soy-sauce, and boiled beans. It is also used for inventing the new industrial foods such as snack foods. In order to make the consumption of the soybeans increase, it is indispensable to make the high-values added to them. For example, the proper contents of physiologically functional elements give highly added values to soybeans. Soybean contains some functional elements such as isoflavones, thiamine (Vitamin B_1), riboflavin (Vitamin B_2) and tocopherols (Vitamin E). Isoflavones have function for preventing osteoporosis (Yamori, 2001). Thiamine (Vitamin B_1), riboflavin (Vitamin B_2) and tocopherol have relations to carbohydrate metabolism and anti-oxidant function, respectively (Tsujimura, 2004). However, the conventional methods for the determination of these physiologically functional elements are much labor intensive. A simple and rapid method for the estimation of them is necessary for screening soybean varieties for the plant breeding. Further, a nondestructive analytical method is needed so that the sample seeds can be used for breeding and sowing after the analyses and selections.

Near infrared spectroscopy (NIRS) has been understood as one of the most powerful analytical tools in the agro-food sector (Shenk et al., 2007; Williams, 2006). Hymowitz et al. (1974), Choung et al. (2001), and Tajuddin et al. (2002) reported the oil and protein analysis of soybean using NIRS. Now, NIRS is one of the official methods for determining major constituents of soybean in the trade based on the quality (Osoborne & Fearn, 1986; USDA FIGS, 1996). Furter, Pazderniket al. (1997), and Kovalenko et al. (2006) tried the determination of amino acid composition of soybeans by NIRS. Hollung et al. (2005) reported the evaluation of nonstarch polysaccharides and oligosaccharide content by NIRS. Li et al. (2009) analyzed lecithin and by-products in the soybean oil processing by using NIRS as a quality control tool. Also, Sato et al. (1998, 2002) reported the estimation of the contents of major constituents, the level of the deterioration indices, and the fatty acid

composition in soybean by NIRS. Then, if other criteria can be estimated by NIRS method, this method will gain greater position in the soybean analysis. In this chapter, the feasibility of NIRS for the estimation of the contents of some functional elements, i.e., isoflavones, thiamine, riboflavin and tocopherol in soybean seeds was examined.

2. Materials and methods

2.1 Soybean samples

Forty-eight samples were cultivated in various areas from northern to southern part of Japan in 2003. The varieties of samples used were as follows: Toyokomachi, and Toyomusume (produced in Hokkaido Prefectural Tokachi Agricultural Experiment Station); Ohsuzu, Suzuyutaka, and Ryuhou (produced in National Agricultural Research Center for Tohoku Region (Akita)); Ayakogane, Enrei, Koganedaizu, Sakukei-4-gou, Suzuyutaka, Tachinagaha, Tachiyutaka, Tamaurara, Tamahomare, Harosoy, Fukuyataka, Houen, Miyagi'oojiro, Yumeminori, Ohsodenomai, Kiyomidori, Akikogane, Enrei, Sachiyutaka, Suzuotome, NattoShouryuu, and Miyagi'ohshiro, (produced in National Institute of Crop Science (Tsukuba)); Enrei, Sachiyutaka, Tamahomare, Fukuyutaka, and Shintanbaguro (produced in National Agricultural Research Center for Western Region (Kagawa)); Fukuyutaka, Kurodamaru, Kiyomidori, and Shinanoguro (produced in our National Agricultural Research Center for Kyushu Okinawa Region (Kumamoto)). One Chinese (Baimei_Baishan) and two USA varieties (from Harrowvinton and from Ohio) were also included. These samples were collected and sent to our research center and were milled by a ultra-centrifugal mill ZM1000 (Retsch Co., Germany) through a screen (φ=1.0mm). All the powdered samples and all the whole grain samples were packed in sealed polystyrene containers (LABORAN Pack, AS ONE Co., Osaka) and were stored at 5°C until being analyzed.

2.2 Chemical measurements

Table 1 described the contents of the isoflavones, thiamine, riboflavin and tocopherols determined by HPLC method (Nishiba et al., 2007). The respective components such as glycosides (daidzin, glycitin, and genistin), malonyl glycosides, acetyl glycosides and aglycons (daidzein, glycitein, and genistein) were determined in this process, and the total isoflavone content was calculated as the summation. Also, the respective tocopherol content was determined, and the total content was calculated.

2.3 Near infrared spectroscopic measurements and statistical analysis.

2.3.1 NIR analysis with an InfraAlyzer 500 (IA500)

An InfraAlyzer 500 (Bran + Luebbe (B+L) GmbH, Norderstedt, Germany) (Photo 1) was used to measure the NIR reflectance spectra in the wavelength range from 1100 to 2500 nm at 2-nm intervals. Samples were packed in a standard cell on a standard drawer for soybean powder (about 3 g), or packed in a whole grain cell on a moving drawer for intact plural soybean seeds (about 60 g). Also, NIR spectra were measured for a single seed in a single grain cup on a standard drawer. See Chapter 28 for the sample presentation method. The samples were divided into two sets: a calibration set (n=36) and a prediction set (n=12) as in Table 1, where their fundamental statistics were described. The unit is mg $(100\ g\ DW)^{-1}$. By the way, the amounts of acetyl glycitin and glycitein were almost none as described in Table 1, and then, their statistical analyses were not carried out in the following. In the tocopherol

analysis case, samples were divided into two sets: a calibration set (n=16) and a prediction set (n=7) as in Table 1.

		calibration set (n = 36 or 16)				prediction set (n = 12 or 7)			
		range		mean	std	range		mean	std
glycoside	daidzin	8.41 -	56.71	24.55	11.68	11.57 -	43.59	21.34	10.63
	glycitin	2.31 -	11.81	6.37	2.50	2.27 -	12.81	6.10	3.36
	genistin	14.40 -	87.49	36.24	16.26	17.72 -	57.65	32.70	13.56
	total	26.28 -	154.01	67.16	28.77	34.31 -	109.93	60.14	25.25
malonyl glycoside	malonyl daidzin	31.82 -	192.27	95.77	40.54	39.36 -	162.15	84.90	39.08
	malonyl glycitin	4.77 -	22.22	12.60	4.49	4.70 -	25.98	11.81	6.02
	malonyl genistin	51.79 -	264.06	133.72	50.15	65.12 -	200.18	125.77	42.95
	total	102.96 -	473.30	242.08	89.84	119.91 -	368.42	222.48	82.47
acetyl glycoside	acetyl daidzin	0.04 -	1.64	0.69	0.35	0.06 -	1.53	0.71	0.47
	acetyl glycitin	0.00 -	0.00	0.00	0.00	0.00 -	0.00	0.00	0.00
	acetyl genistin	0.21 -	1.82	0.88	0.40	0.00 -	1.76	0.87	0.48
	total	0.26 -	3.46	1.56	0.73	0.10 -	3.25	1.59	0.92
aglycone	daidzein	0.25 -	2.66	1.07	0.59	0.26 -	1.72	0.78	0.45
	glycitein	0.00 -	0.23	0.01	0.04	0.00 -	0.00	0.00	0.00
	genistein	0.35 -	2.82	1.37	0.65	0.32 -	1.87	1.08	0.50
	total	0.60 -	5.48	2.45	1.20	0.58 -	3.58	1.86	0.93
total isoflavone	total	133.44 -	633.42	313.26	116.83	156.96 -	482.24	286.07	107.79
Vitamin B	thiamine	0.57 -	0.90	0.70	0.09	0.56 -	0.81	0.69	0.08
	riboflavin	0.20 -	0.28	0.23	0.02	0.21 -	0.26	0.23	0.01
tocopherol (toc)	α-toc	0.77 -	10.83	3.23	2.44	2.26 -	7.08	3.49	1.60
	β-toc	0.41 -	4.21	1.33	1.00	0.58 -	1.70	1.07	0.36
	γ-toc	11.19 -	21.69	17.09	2.80	9.10 -	19.16	16.08	3.13
	δ-toc	5.59 -	15.46	9.85	3.02	5.32 -	12.45	7.90	2.33
	total	25.27 -	37.59	31.50	3.69	20.67 -	35.93	28.53	4.60
	α-toc equivalence	2.76 -	13.29	5.57	2.60	4.09 -	9.41	5.60	1.72

The unit is [mg (100 g DW)$^{-1}$]. std: standard deviation

Table 1. The fundamental statistics of the samples to be analyzed.

Photo 1. An instrument : An InfraAlyzer 500

Multiple linear regression (MLR) analysis of the NIRS data with the HPLC data was carried out using IDAS software (B+L), an accessory software of IA500, on the calibration set. When the first- and second-derivative NIR spectra were calculated, the default parameters were used. The validations of the calibration equations obtained, or the prediction process, were

carried out using the prediction set. The Unscrambler (version 9.6; Camo Co., Norway), which was a software for the data-analysis and is sold separately, was also used on the IA500 data for partial least square regression (PLSR) or principal component regression (PCR) analysis. The authors analyzed the data not only on the original spectra, but also on the derivative spectral data, i.e., pretreated spectral data. In this case, the conditions to calculate the derivatives were as follows: gap 11, segment 10 for the first derivative (abbreviated as d1); and gap 10, segment 11 for the second derivative (d2). The gap and segment are the parameters in the Gap-Segment derivatives. Gap is the length of the interval that separates the two segments that are being averaged, and segment is an interval over which data values are averaged.

2.3.2 NIR analysis with a SpectraStar 2400 (SS2400)

A SpectraStar 2400 (Unity Scientific, USA) (Photo 2) was also used to measure the NIR reflectance spectra in the wavelength range from 1200 to 2400 nm at 1-nm interval. The sample presentation methods for powder and plural whole seeds were in the same manner as IA500 case.

Photo 2. An instrument: A SpectraStar 2400

The SensoLogic (Sensologic GmbH, Germany) was used on SpectraStar 2400 data with chemical data for MLR analysis, and PLSR/PCR analysis. The conditions to obtain the derivatives for the pretreatment of NIR spectra were as follows: gap 10, segment 10 for the first derivative (abbreviated as d1), gap 10, segment 10 for the second derivative (d2), and gap 5, segment 5 for the other second derivative (d22). First, the data analyses were carried out on the calibration set. Then, the validations of the calibration equations obtained were checked using the prediction set.

2.3.3 NIR analysis with an MPA, Multi Purpose FT-NIR Analyzer

An MPA (Multi Purpose FT-NIR Analyzer, Bruker Optics, Germany) was also used for the NIR measurements: wavenumber = 4000-12000cm^{-1}, resolution = 16cm^{-1} (Photo 3). The sample types analyzed were powdered soybeans, plural whole soybean seeds, and a single soybean seed. They were measured using the specified cells on the specified modules depending upon their types (Photo 4). The Opus (Bruker Optics, Germany) was used for the statistical analysis of PLSR and PCR with an automatic analysis. By the way, the calibration and prediction sample sets were different from above because of the automatic analysis.

Photo 3. An instrument: The FT-NIR spectrometry MPA

Photo 4. The sample presentation methods for an FT-NIR spectrometry MPA
(powder, plural seeds, and a single seed)

2.3.4 NIR analysis with an NIRFlex N-500

An NIRFlex N-500 (Buchi, Swiss) was also used to measure the FT-NIR reflectance spectra of the powder samples and whole kernel soybean seeds in the wavenumber range from 4000 to 12000 cm^{-1} with resolution = 8 cm^{-1} (Photo 5). Each sample was packed in a test tube or a petri dish for the measurement with using the specified module (Photo 6). NIRCal5 (Buchi, Swiss) was used for the automatic statistical analysis.

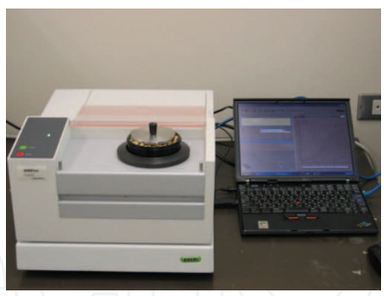


Photo 5. An instrument: An NIRFlex N-500

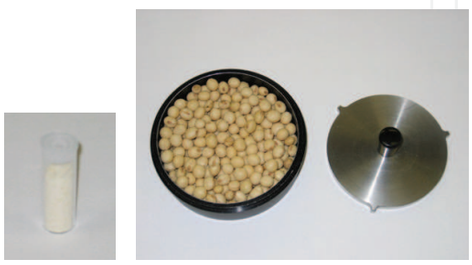


Photo 6. The sample presentation methods for an NIRFlex N-500.
(a test tube for a powder, and a dish and a lid for plural seeds)

3. Results and discussion

3.1 An InfraAlyzer 500

3.1.1 MLR analysis

Table 2 describes the calibration process (left side) and the prediction results (right side) developed for powdered soybean analysis with IDAS: the selected wavelengths in the calibration equations, the correlation coefficient (r), the standard error of calibration (SEC), the standard error of prediction (SEP), mean-corrected SEP (MC-SEP), and bias. These calibrations provided the best prediction in the validation process. In the table, when the contribution ratio, r^2, exceeded 0.5 in the prediction, the results are described in bold letters. For the total isoflavone content described at the last column of the isoflavone section, SEP was adequate for the estimation. The selected wavelengths were mainly due to C-H bonds (Osborne et al. 1993). The counts of wavelengths selected were also adequate, i.e., not so many wavelengths. Furthermore, especially as for powdered soybean, the respective components, such as glycosides and malonyl glycosides, also could be estimated separately, as described in Table 2, where the contribution ratios of the respective components, glycosides and malonyl glycosides, exceeded 0.5. On the other hand, the contents of acetyl glycosides and aglycons were poorly estimated because of their small range fluctuations.

		Calibration				Validation			
		treatment	selectred wavelengths	r	SEC	r	SEP	MC-SEP	Bias
glycoside	**daidzin**	**raw**	**1680, 2236**	**0.74**	**8.17**	**0.83**	**6.00**	**6.26**	**-0.24**
	glycitin	**d1**	**1747, 1943, 2147, 2383, 2391, 2471**	**0.90**	**1.22**	**0.80**	**2.22**	**2.10**	**0.94**
	genistin	**raw**	**1188, 1692, 2184, 2236**	**0.89**	**8.10**	**0.93**	**6.63**	**6.92**	**-0.09**
	total	**raw**	**1188, 1692, 2184, 2236**	**0.88**	**14.68**	**0.90**	**12.99**	**13.56**	**0.10**
malonyl glycoside	**malonyl daidzin**	**d1**	**1115, 1131, 2195, 2275, 2311,2347, 2383**	**0.96**	**12.30**	**0.95**	**15.43**	**14.98**	**5.69**
	malonyl glycitin	**d1**	**1851, 1939, 2307, 2419, 2463, 2475**	**0.91**	**2.03**	**0.74**	**4.30**	**4.33**	**1.12**
	malonyl genistin	**raw**	**1700, 1724, 2220**	**0.85**	**27.84**	**0.89**	**21.03**	**20.61**	**-7.26**
	total	**raw**	**1680, 2244**	**0.79**	**57.01**	**0.84**	**159.45**	**51.81**	**151.53**
acetyl glycoside	**acetyl daidzin**	**d1**	**1115, 1651, 1743, 2171, 2235**	**0.83**	**0.22**	**0.79**	**0.33**	**0.32**	**-0.11**
	acetyl genistin	**raw**	**1912, 1936, 2396, 2400**	**0.75**	**0.29**	**0.75**	**0.34**	**0.35**	**-0.04**
	total	d1	1271, 1751, 1755, 2107, 2191, 2347, 2363	0.93	0.30	0.66	0.69	0.71	0.05
aglycone	daidzein	d2	1354, 1746	0.59	0.50	0.57	0.53	0.39	0.38
	genistein	raw	2188, 2192	0.54	0.57	0.39	0.53	0.51	0.20
	total	raw	1556, 1668, 2296, 2364, 2376, 2380	0.82	0.76	0.02	1.36	1.26	0.63
total isoflavone	**total**	**raw**	**1188, 1688, 2184, 2236**	**0.92**	**48.88**	**0.95**	**38.51**	**39.94**	**-4.54**
Vitamin B	thiamine	raw	1852 2296 2320	0.57	0.08	0.38	0.08	0.08	0.02
	riboflavin	d2	1386 1482 1926 2338	0.71	0.01	0.62	0.01	0.01	0.00
tocopherol (toc)	α-toc	raw	1596	0.28	2.50	0.55	1.38	1.45	-0.32
	β-toc	**raw**	**1100**	**0.19**	**1.04**	**0.78**	**0.45**	**0.26**	**0.38**
	γ-toc	**d2**	**2142 2442**	**0.85**	**1.62**	**0.75**	**2.19**	**2.26**	**-0.63**
	δ-toc	raw	1176 1244 1260 1264 1412	0.98	0.80	0.49	3.06	2.61	1.87
	total	d2	1190 1322 1694 2282 2370 2434	0.90	2.11	0.40	6.04	4.55	4.33
	α-toc equivalence	raw	1596	0.26	2.69	0.57	1.42	1.53	-0.08

r : Correlation coefficient between chemical method and NIR method.
SEC: Standard error of calibration.; SEP: Standard error of prediction.; MC-SEP: Mean-corrected SEP.
(The bold letters mean that the contribution ratio, r^2, exceeded 0.5 in the validation.)

Table 2. The calibration and the prediction results (powdered soybean) for IA500 data.

Table 3 describes the calibration process and the prediction results developed for intact plural seeds analysis. SEP for the total isoflavone content described at the last column of the isoflavone section, was also small enough for the estimation. The selected wavelengths were mainly due to C-H bonds. Some of the respective components in intact plural soybean seeds still could be estimated. However, the contents of acetyl glycosides and aglycons were poorly estimated.

		Calibration				Validation			
		treatment	selected wavelengths	r	SEC	r	SEP	MC-SEP	Bias
glycoside	**daidzin**	**d2**	**1630, 1746, 2294**	**0.79**	**7.62**	**0.80**	**6.50**	**6.73**	**0.88**
	glycitin	d1	1203, 1947, 2335	0.81	1.55	0.49	2.92	3.05	-0.08
	genistin	raw	2236, 2280, 2336, 2360	0.80	10.50	0.50	12.84	13.41	0.18
	total	raw	1712, 1732, 2236, 2296, 2320, 2368	0.89	14.39	0.59	22.37	22.42	-6.30
malonyl glycoside	**malonyl daidzin**	**d2**	**1186, 1230, 1322, 1370, 1742, 2290, 2378**	**0.95**	**14.03**	**0.82**	**25.68**	**26.68**	**-2.65**
	malonyl glycitin	d1	1199, 1443, 1739, 1959, 2335, 2415	0.89	2.25	0.58	5.14	5.30	0.84
	malonyl genistin	**d1**	**1395, 1723, 2099, 2255, 2291, 2327, 2351**	**0.94**	**19.54**	**0.87**	**21.75**	**22.62**	**-6.30**
	total	**raw**	2192, 2208, 2232, 2236, 2272, 2336, 2356	0.92	38.97	0.44	151.29	73.54	133.91
acetyl glycoside	acetyl daidzin	raw	1112, 1124, 1132, 1136, 1144, 1304	0.79	0.24	0.66	0.35	0.36	-0.07
	acetyl genistin	raw	1648, 1660, 2268, 2280, 2288	0.83	0.24	0.57	0.39	0.41	-0.03
	total	d1	1395, 1655, 2255	0.68	0.56	0.69	0.70	0.73	-0.08
aglycone	**daidzein**	**raw**	**1256, 1260, 1268**	**0.66**	**0.47**	**0.85**	**0.39**	**0.24**	**0.31**
	genistein	d1	1263, 1643, 1927, 2287	0.66	0.52	0.40	0.49	0.49	0.16
	total	raw	2140, 2148, 2184, 2212, 2236, 2256, 2264	0.76	0.88	0.56	1.00	0.84	0.60
total isoflavone	**total**	**d2**	**1630, 1746, 2294**	**0.85**	**65.89**	**0.82**	**63.43**	**65.63**	**8.67**
Vitamin B	thiamine	d1	1527 1931 2095 2371	0.63	0.08	0.33	0.08	0.08	-0.02
	riboflavin	**raw**	**1120 1132 1256 1268 1296**	**0.77**	**0.01**	**0.80**	**0.01**	**0.01**	**0.00**
tocopherol (toc)	α-toc	d2	1554	0.41	2.37	0.05	1.78	1.92	0.09
	β-toc	raw	1160 1256 1456 2040	0.93	0.44	0.46	0.48	0.50	0.12
	γ-toc	d2	2314	0.55	2.50	0.57	2.58	2.79	0.03
	δ-toc	d2	1202 1246	0.88	1.57	0.14	2.96	2.94	1.14
	total	**raw**	**1176 1228 2428 2432**	**0.91**	**1.88**	**0.78**	**5.03**	**3.57**	**3.79**
	α-toc equivalence	d2	1442 1550	0.57	2.38	0.30	1.69	1.82	-0.15

see footnotes in Table2.

Table 3. The calibration and the prediction results (intact plural soybean seeds) for IA500 data.

The counts of wavelengths selected were proper, i.e., not too many. Generally, one wavelength can be selected for each 5 to 15 samples in MLR analysis (Hruschka, 2001), i.e., three to seven wavelengths can be selected in this case, because 36 samples were used for developing the calibration equations. Further, in the prediction process, different samples from the calibration set were used to check the overfitting. Both calibration equations for the estimation of the total isoflavone content obeyed this rule. Further, as for the powder analysis, the calibrations for some of the respective components of isoflavone in the powder were also adequate. On the other hand, for intact plural soybean seeds, the contribution ratio (r^2) was low, even when many wavelengths were selected for developing calibration equations for the respective components.

Figure 1 shows the prediction results of the total isoflavone analyses for the powder (Fig.1-a)), and for the intact plural seeds (Fig.1-b)). The correlation coefficient (r) between the chemical method and NIR method, and the standard error of prediction (SEP) were also described. The SEP value was one third to one half of SD described in Table 1. Comparing from the SEPs in Fig.1 with the standard deviation of the samples (about 110 mg(100 gDW)$^{-1}$ as described in Table 1), the author consider that the total isoflavone content could be estimated.

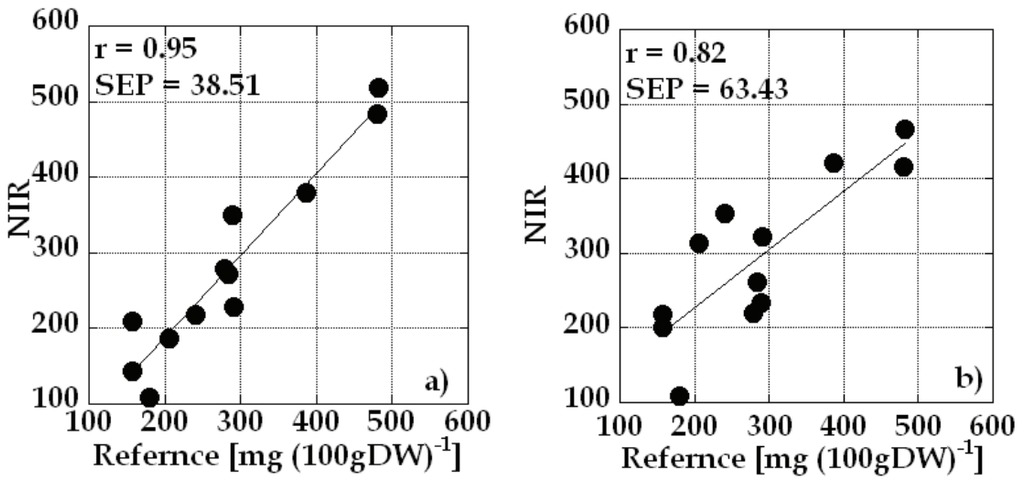


Fig. 1. The results of NIR analysis developed for total isoflavone: a) powdered soybean; b) intact plural soybean seeds with MLR analysis on IA 500 data.

3.1.2 PLSR/PCR analysis with the Unscrambler on IA 500 data.

Table 4 describes the results of PLSR/PCR analysis obtained using the Unscrambler, the calibration process (left side) and the prediction results (right side) developed for powdered soybean. The treatment on the original spectra, the number of factors, the correlation coefficient (r), the standard error of calibration (SEC), root mean squared error of prediction (RMSEP), the standard error of prediction (SEP), and bias were described. The better cases were described among PLSR and PCR. For the total isoflavone content, SEP was adequate for the estimation. Further, especially as for powdered soybean, the respective component, glycoside and malonyl glycoside, also could be successfully estimated separately, as described in Table 4, where the contribution ratios of the respective components, glycosides and malonyl glycosides, exceeded 0.5. The respective analysis of the total of the malonyl

glycosides was drastically improved from the results of MLR analysis (Table 2). On the other hand, acetyl glycoside and aglycon contents were poorly estimated as in MLR analysis.

		Calibration				Validation			
		Treatment	Factors	r	SEC	r	RMSEP	SEP	Bias
glycoside	**daidzin**	**d1**	**pls-6**	**0.78**	**7.45**	**0.80**	**6.88**	**6.79**	**2.27**
	glycitin	**d1**	**pls-9**	**0.89**	**1.16**	**0.77**	**2.40**	**2.32**	**0.92**
	genistin	**raw**	**pls-8**	**0.85**	**8.64**	**0.90**	**6.63**	**6.49**	**2.31**
	total	**raw**	**pls-8**	**0.85**	**15.22**	**0.90**	**12.21**	**11.99**	**4.14**
malonyl glycoside	**malonyl daidzin**	**d1**	**pls-7**	**0.91**	**17.25**	**0.91**	**17.68**	**17.95**	**4.16**
	malonyl glycitin	**d2**	**pls-12**	**0.93**	**1.68**	**0.74**	**4.25**	**4.37**	**0.76**
	malonyl genistin	**d2**	**pls-7**	**0.90**	**22.31**	**0.90**	**21.78**	**22.71**	**-1.25**
	total	**d1**	**pls-7**	**0.92**	**36.59**	**0.93**	**34.31**	**35.68**	**3.13**
acetyl glycoside	acetyl daidzin	raw	pls-6	0.71	0.25	0.51	0.40	0.41	-0.08
	acetyl genistin	d1	pls-1	0.42	0.37	0.69	0.39	0.40	0.02
	total	d1	pls-1	0.42	0.67	0.61	0.77	0.80	0.01
aglycone	daidzein	d2	pls-3	0.60	0.48	0.37	0.54	0.47	0.30
	genistein	d1	pls-8	0.74	0.44	0.45	0.61	0.55	0.31
	total	d1	pls-9	0.77	0.77	0.50	1.23	1.02	0.74
total isoflavone	**total**	**raw**	**pls-8**	**0.91**	**48.83**	**0.95**	**40.01**	**38.88**	**14.66**
Vitamin B	thiamine	d2	pls-1	0.22	0.09	0.58	0.07	0.08	0.01
	riboflavin	raw	pls-2	0.50	0.01	0.61	0.01	0.01	0.00
tocopherol (toc)	α-toc	raw	pls-1	0.19	2.47	0.45	1.49	1.55	-0.39
	β-toc	**d2**	**pls-3**	**0.87**	**0.50**	**0.70**	**1.53**	**1.14**	**1.11**
	γ-toc	d2	pls-2	0.57	2.38	0.25	3.45	3.71	-0.28
	δ-toc	d2	pcr-3	0.71	2.18	0.51	2.22	2.20	0.90
	total	raw	pls-1	0.13	3.78	-0.42	5.95	5.44	3.17
	α-toc equivalence	raw	pla6-1	0.18	2.64	0.48	1.51	1.62	-0.15

see footnotes in Table 2.

Table 4. The calibration and the prediction results (powdered soybean) with the Unscrambler on IA500 data.

Table 5 describes the calibration process and the prediction results developed for intact plural seeds. As for the total isoflavone content, SEP was fair enough for the estimation. As for intact plural soybean seeds, the results were improved: some of the respective component also could be estimated. The respective analysis of the total of the malonyl glycosides was also drastically improved. However, acetyl glycoside and aglycon contents were still poorly estimated.

The total isoflavone content could be estimated not only with powdered soybean but also with intact plural soybean seeds. It is the similar level as shown in Fig. 1. However, the content of the respective isoflavone component could be estimated for powdered soybean as described in Table 2-5. The present findings suggest that the total isoflavone content of the soybean seeds could be estimated for simple, rapid, and nondestructive breeding selection by the NIRS method. The respective elements in the powder could be estimated. PLSR and PCR analyses were also tried, and the results were similar to those obtained by MLR analysis. Further, for total malonyl glycoside, the bias was drastically improved by PLSR analysis.

The estimations for some of the contents of Vitamin B, and tocopherol were fair for rough estimation despite of their small range fluctuations. As for Vitamin B, considering from the

SEPs in Table 2-5 with comparing the standard deviation in Table 1: 0.08 for thiamin, and 0.01 for riboflavin. These contents might be fairly estimated.

		Calibration				Validation			
		Treatment	Factors	r	SEC	r	RMSEP	SEP	Bias
glycoside	**daidzin**	**d1**	**pls-11**	**0.96**	**3.34**	**0.81**	**6.98**	**7.28**	**0.33**
	glycitin	d2	pls-20	0.99	0.26	0.65	2.56	2.67	-0.13
	genistin	**d2**	**pls-12**	**0.97**	**4.10**	**0.84**	**8.33**	**7.84**	**-3.63**
	total	**d2**	**pls-12**	**0.97**	**6.96**	**0.84**	**16.19**	**15.43**	**-6.62**
malonyl glycoside	**malonyl daidzin**	**d2**	**pls-13**	**0.98**	**7.44**	**0.94**	**14.83**	**14.87**	**-4.15**
	malonyl glycitin	**raw**	**pls-12**	**0.99**	**0.75**	**0.73**	**4.16**	**4.33**	**-0.40**
	malonyl genistin	**d2**	**pls-11**	**0.96**	**13.43**	**0.93**	**15.91**	**16.36**	**-2.81**
	total	**d2**	**pls-12**	**0.98**	**19.86**	**0.95**	**26.64**	**26.52**	**-8.07**
acetyl glycoside	acetyl daidzin	raw	pls-6	0.69	0.26	0.63	0.36	0.38	-0.05
	acetyl genistin	raw	pcr-6	0.68	0.30	0.69	0.36	0.37	-0.02
	total	raw	pcr-6	0.67	0.55	0.69	0.68	0.71	-0.07
aglycone	daidzein	d2	pls-7	0.81	0.35	0.36	0.55	0.49	0.29
	genistein	raw	pls-8	0.83	0.36	0.40	0.63	0.52	0.39
	total	d2	plst-6	0.75	0.80	0.36	1.14	0.98	0.66
total isoflavone	**total**	**d2**	**pls-12**	**0.97**	**26.57**	**0.96**	**32.80**	**30.28**	**-15.34**
Vitamin B	thiamine	d2	pcr-1	0.14	0.09	0.24	0.08	0.08	0.01
	riboflavin	d1	pls-3	0.63	0.01	0.54	0.01	0.01	0.00
tocopherol (toc)	α-toc	d2	pcr-1	0.02	2.52	0.03	1.62	1.73	-0.24
	β-toc	**d2**	**pls-1**	**0.46**	**0.91**	**0.74**	**0.46**	**0.28**	**0.38**
	γ-toc	raw	pls-1	0.13	2.87	0.35	3.17	3.27	0.95
	δ-toc	raw	pls-7	0.95	0.95	0.40	2.52	2.47	1.04
	total	raw	pcr-1	0.25	3.70	0.16	5.35	4.90	2.82
	α-toc equivalence	d1	pls-1	0.28	2.18	0.01	1.85	1.97	0.31

see footnotes in Table 2.

Table 5. The calibration and the prediction results (intact plural soybean seeds) with the Unscrambler on IA500 data.

3.1.3 Statistical analysis for total isoflavone content on a single seed analysis.

As for a single seed analysis, considering from SEP, NIRS may be available for nondestructively estimating the total isoflavone content in both MLR- and PLSR-analysis cases (Table 6, upper columns). The scattering graphs of these results were shown in Fig.2.

		Calibration		**Prediction**			
		r	**SEC**	**r**	**SEP**	**bias**	**MC-SEP**
single seed analysis	**MLR Analysis by IDAS**	**0.77**	**79.62**	**0.80**	**69.92**	**23.25**	**68.87**
	PLS Analysis by Unscrambler	**0.89**	**53.70**	**0.79**	**69.01**	**19.33**	**69.20**
plural seeds analysis adapted on a single seed spectrum	**MLR Analysis by IDAS**	**------**	**------**	**0.83**	**62.99**	**18.21**	**62.98**
	PLS Analysis by Unscrambler	**------**	**------**	**0.96**	**158.84**	**155.33**	**34.70**

r : Correlation coefficient between chemical method and NIR method.
SEC: Standard error of calibration; SEP: Standard error of prediction; MC-SEP: Mean-corrected SEP
The unit is mg (100gDW)$^{-1}$

Table 6. Results of statistical analysis for total isoflavone content by NIR on a single seed analysis.

The same level of SEP was obtained in this single seed analysis case as plural soybean seeds analysis case (3-1-2 section). Kudou et al. (1991) reported that isoflavone distributed mainly in hypocotyls of soybean seeds, i.e., on the surface of a seed, and this might be why the SEP was not so bad for an intact seed. (Sato et al., 2009b)

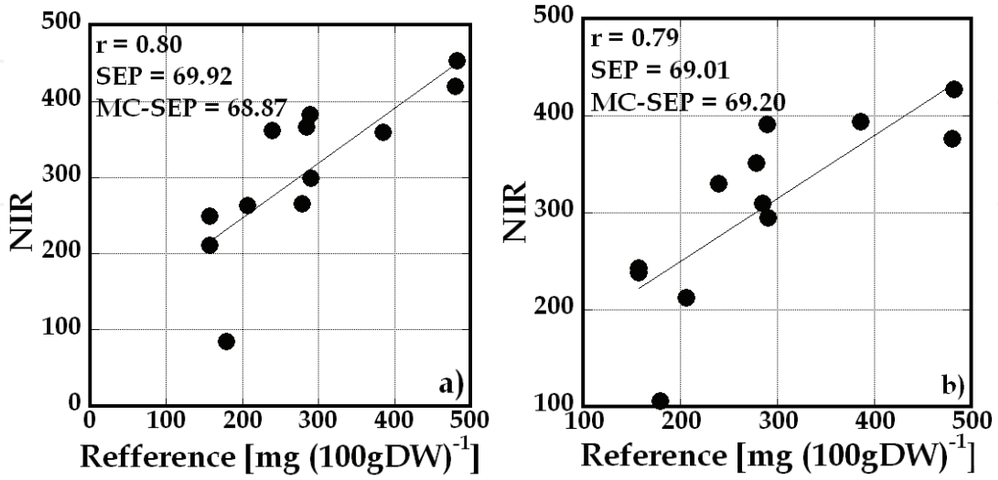


Fig. 2. The scattering graphs of the prediction results on a single seed analysis estimating the total isoflavone content by a) MLR- and b) PLSR-analysis.

3.1.4 Adaptation of calibrations obtained by plural seed analysis to single seed spectra

On the other hand, when the calibrations developed for plural seeds were adapted on a single seed spectrum case, the results of both MLR- and PLSR-analysis cases were described in the lower columns of Table 6. The scattering graphs of these results were shown in Fig.3. In this case, the bias and skew emerged as shown in Fig.3 (right). The reason is that its reflectance spectrum is a similar one as plural one, but its level is lower in a single seed case

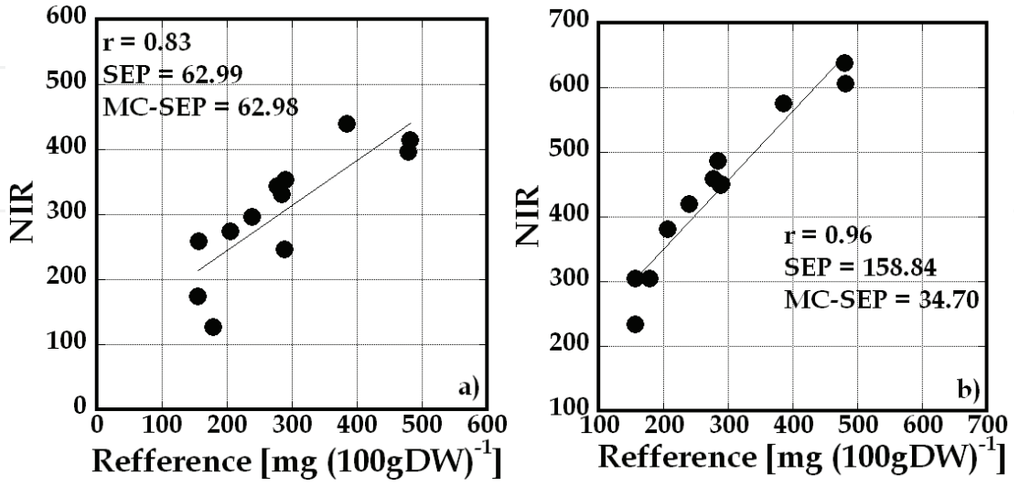


Fig. 3. The scattering graphs of the results adaptation of plural seeds analysis on a single seed for estimating the total isoflavone content by a) MLR- and b) PLSR-analysis.

than in plural seeds case. However, considering from MC-SEP, the level of the total isoflavone content could be estimated. The same level of MC-SEP was obtained in this single seed analysis case as plural soybean seeds analysis case. NIRS may be available for the nondestructive estimation of the total isoflavone content by both MLR- and PLSR-analysis cases on a single seed spectrum (Sato et al., 2009a).

3.2 Analysis on SpectraStar 2400 data(Sato et al., 2008)

3.2.1 MLR analysis

Table 7 describes the calibration and the prediction results developed for powdered soybean by MLR analysis with the SensoLogic: spectral treatment, selected wavelengths for calibration equations, the correlation coefficient (r), the standard error of estimate (SEE), root mean square error of prediction (RMSEP), bias, and the standard error of prediction (SEP). These calibrations provided the best prediction among the prediction results. The figures in bold letters mean the contribution ratio, r^2, exceeds 0.5. As for the total isoflavone content described in the last column of the isoflavone section, SEP was good enough for the estimation. The selected wavelengths were mainly due to C-H bonds (Osborne et al., 1993). The counts of the selected wavelengths in the calibration were a few and proper. Further, especially as for powdered soybean, the respective components, such as glycosides and malonyl glycosides, also could be estimated separately as described in Table 7. The estimations for acetyl glycoside,, and aglycon contents were not good enough because of their small range fluctuations. The estimation for riboflavin content was fair.

Table 8 describes the calibration and the prediction results developed for intact plural soybean seeds analysis. As for the total isoflavone content, SEP was also good enough for the estimation. The selected wavelengths were mainly due to C-H bonds, and the counts of wavelengths were a few and proper. As for intact plural soybean seeds, some of the respective components, such as malonyl daidzin and malonyl genistin, still could be estimated. On the other hand, the estimations for acetyl glycosides, aglycons contents were also not good enough because of their small range fluctuations.

Figure 4 shows the scattering graphs, which are the prediction results of the total isoflavone content between chemical method and NIR method: a) as for powdered soybean, and b) as for intact plural soybean seeds. Considering from the SEPs in Fig.4 with comparing its standard deviation (107.79 [mg $(100g\ DW)^{-1}$] described in Table 1, the isoflavone content might be fairly estimated in both cases.

3.2.2 PLSR/PCR analysis on SpectraStar 2400 data

Table 9 describes the calibration and the prediction results developed for powdered soybean by PLSR/PCR analysis. The number of factors was described instead of wavelengths. The better case among PLSR and PCR was described. As for the total isoflavone content at the last column of the isoflavone section, SEP was good enough for the estimation. Further, especially as for powdered soybean, the respective components, such as glycosides, malonyl gylucosides, and acetyl genistin, also could be estimated separately. On the other hand, the estimations for other acetyl glycosides and aglycons contents were not good enough because of their small range fluctuations. The estimation for thiamine or total tocophrol content was fair.

Table 10 describes the calibration and the prediction results developed for intact plural soybean seeds analysis. As for the total isoflavone content, SEP was fair enough for the

		Calibration				Prediction			
		treatment	calibration equations or wavelengths	r	SEE	r	RMSEP	bias	SEP
glycoside	**daidzin**	**d2**	**1684, 1766, 2223**	**0.79**	**7.62**	**0.85**	**5.97**	**-0.10**	**6.24**
	glycitin	**raw**	**1710, 2284, 2320**	**0.76**	**1.72**	**0.77**	**2.33**	**0.30**	**2.41**
	genistin	**d22**	**1820, 1887, 2002, 2264**	**0.86**	**8.80**	**0.87**	**7.97**	**3.58**	**7.44**
	total	**d22**	**1488, 1798, 2260**	**0.82**	**17.53**	**0.90**	**12.90**	**6.11**	**11.86**
malonyl glycoside	**malonyl daidzin**	**d2**	**1355, 1488, 1756, 1780, 2176, 2262**	**0.93**	**17.10**	**0.93**	**15.56**	**1.78**	**16.15**
	malonyl glycitin	**d1**	**1458, 1807, 2361**	**0.81**	**2.78**	**0.74**	**4.41**	**0.59**	**4.56**
	malonyl genistin	**d2**	**1314, 1329, 1737, 2063**	**0.88**	**25.43**	**0.92**	**17.66**	**-1.56**	**18.37**
	total	**d2**	**1315, 1336, 1738, 1757, 2065**	**0.90**	**42.88**	**0.94**	**28.70**	**-1.78**	**29.91**
acetyl glycoside	acetyl daidzin	raw	1705, 2328, 2336	0.63	0.29	0.62	0.36	0.03	0.38
	acetyl genistin	d1	1431, 1655, 1664, 1828	0.71	0.30	0.56	0.40	-0.04	0.42
	total	**d22**	**1542, 1809, 1886, 2000**	**0.80**	**0.47**	**0.76**	**0.61**	**-0.12**	**0.63**
aglycon	daidzein	raw	1344, 1368, 1376, 1396, 1790, 1813, 1824, 1835, 1873	0.86	0.36	0.64	0.50	0.14	0.50
	genistein	d1	1330, 1389, 1437, 1617, 1824, 1836, 2005, 2308	0.90	0.33	0.52	0.55	0.22	0.52
	total	d2	1299, 1612, 1829, 2037, 2112	0.87	0.65	0.49	1.08	0.40	1.05
total isoflavone	**total**	**d2**	**1685, 1767, 2022**	**0.87**	**61.11**	**0.94**	**43.63**	**-8.38**	**44.72**
Vitamin B	thiamine	d22	1230, 2142	0.59	0.08	0.26	0.09	0.02	0.09
	riboflavin	d2	1410, 1496, 1636, 1818, 2030, 2320	0.79	0.01	0.63	0.01	0.00	0.01
tocopherol (toc)	**α-toc**	**d2**	**1888, 2153, 2343**	**0.92**	**1.12**	**0.87**	**1.23**	**0.15**	**1.31**
	β-toc	d1	1272, 1822, 2196, 2314, 2359	0.95	0.41	0.54	0.45	-0.09	0.48
	γ-toc	raw	2287, 2330, 2364, 2372, 2380, 2400	0.98	0.65	0.60	2.81	-0.98	2.84
	δ-toc	d1	1655, 2339	0.91	1.42	0.44	2.24	0.81	2.25
	total	d2	1415, 1499, 1820, 1889, 2069, 2282	1.00	0.38	0.64	4.31	1.90	4.18
	α-toc equivalence	raw	2369, 2377, 2395	0.90	1.30	0.69	1.66	0.29	1.77

r : Multiple correlation coefficient between chemical method and NIR method.
SEE : Standard error of estimate
RMSEP : Root mean square error of prediction
SEP : Standard error of prediction.

Table 7. The calibration process and the prediction results for powdered soybean with MLR analysis on SpectraStar 2400 data.

		Calibration				Prediction			
		treatment	wavelength	r	SEE	r	RMSEP	bias	SEP
glycoside	**daidzin**	**d2**	**1623, 1741, 2293**	**0.80**	**7.51**	**0.73**	**7.44**	**1.62**	**7.59**
	glycitin	d22	1568, 1745, 2204	0.82	1.52	0.44	3.02	0.08	3.16
	genistin	d2	1621, 1743, 2293	0.80	10.32	0.65	10.45	1.73	10.77
	total	**d22**	**1548, 1713, 2187**	**0.87**	**15.24**	**0.79**	**16.51**	**4.24**	**16.66**
malonyl glycoside	**malonyl daidzin**	**d2**	**1301, 1554, 2170**	**0.84**	**23.39**	**0.88**	**18.62**	**-0.86**	**19.42**
	malonyl glycitin	d2	1256, 1365, 1473, 1527, 1629, 1841, 1974, 2115, 2158, 2216, 2350	0.95	1.68	0.63	5.03	1.76	4.92
	malonyl genistin	**d22**	**1350, 1369, 1671, 1706, 1893, 1982**	**0.94**	**18.34**	**0.86**	**22.06**	**-3.52**	**22.75**
	total	**d2**	**1260, 1316, 1662, 1745, 2291**	**0.89**	**44.12**	**0.88**	**40.23**	**0.80**	**42.01**
acetyl glycoside	acetyl daidzin	d22	1566, 1600, 1751, 1830, 1846, 2185	0.91	0.16	0.68	0.35	0.02	0.36
	acetyl genistin	d1	1937, 2102, 2347	0.72	0.30	0.55	0.41	0.02	0.42
	total	d2	2051, 2255	0.63	0.59	0.56	0.76	0.02	0.79
aglycon	daidzein	d22	1464, 1682, 2256, 2375	0.80	0.37	0.41	0.51	0.24	0.47
	genistein	d22	1542, 1682,2376	0.78	0.43	0.21	0.58	0.19	0.58
	total	d2	1746, 2166	0.48	1.09	0.31	1.10	0.45	1.05
total isoflavone	**total**	**d22**	**122 1 , 1550, 1708, 2187**	**0.91**	**51.88**	**0.88**	**57.26**	**2.66**	**59.75**
Vitamin B	thiamine	d1	1636, 1824	0.52	0.08	0.63	0.07	0.00	0.07
	riboflavin	d22	1352, 1643, 2060, 2371	0.62	0.01	0.29	0.01	0.00	0.01
tocopherol (toc)	**α-toc**	**d22**	**1226, 1733, 2355**	**0.86**	**1.45**	**0.72**	**1.15**	**-0.26**	**1.21**
	β-toc	raw	1716, 2350, 2363	0.84	0.63	0.35	0.54	0.17	0.56
	γ-toc	raw	1259, 1272	0.75	2.06	0.46	3.03	0.31	3.25
	δ-toc	raw	2268, 2295, 2323	0.91	1.46	0.04	3.65	1.61	3.55
	total	d22	1537, 2186, 2187, 2400	0.89	2.01	0.53	4.11	1.30	4.21
	α-toc equivalence	**d22**	**1226, 1733, 2355**	**0.86**	**1.53**	**0.80**	**1.09**	**0.00**	**1.18**

see footnotes in Table 7.

Table 8. The calibration and the prediction results for intact plural soybean seeds with MLR analysis on SpectraStar 2400 data.

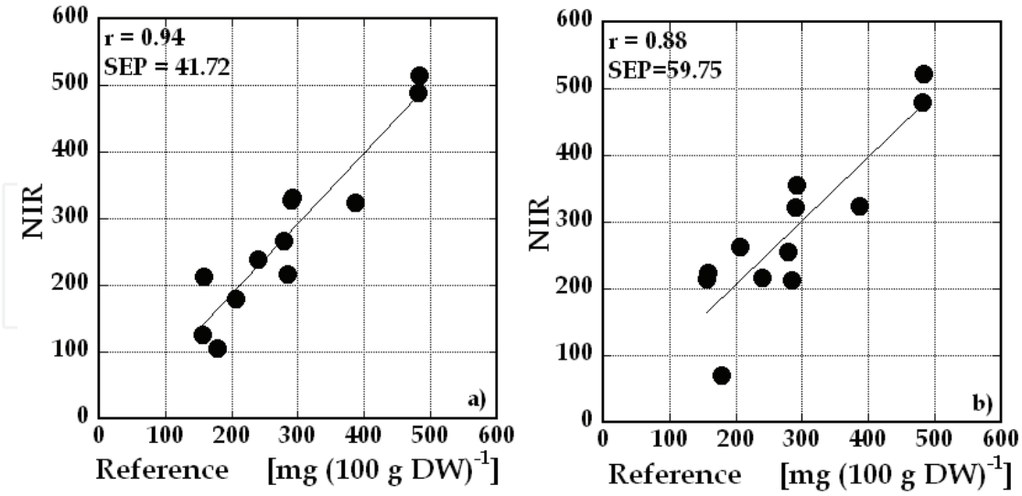


Fig. 4. The results of NIR analysis developed for total isoflavone: a) powdered soybean; b) intact plural soybean seeds with MLR analysis on SpectraStar 2400 data.

		Calibration				Prediction			
		treatment	factors	r	SEE	r	RMSEP	bias	SEP
glycoside	**daidzin**	**d22**	**pls-10**	**0.94**	**4.82**	**0.90**	**4.82**	**-1.18**	**4.88**
	glycitin	**raw**	**pls-4**	**0.72**	**1.87**	**0.82**	**2.41**	**0.41**	**2.48**
	genistin	**d22**	**pcr-18**	**0.83**	**13.72**	**0.93**	**5.53**	**0.25**	**5.78**
	total	**d1**	**pcr-11**	**0.84**	**18.93**	**0.82**	**16.12**	**4.16**	**16.26**
malonyl glycoside	**malonyl daidzin**	**d1**	**pls-6**	**0.90**	**19.32**	**0.94**	**14.28**	**5.90**	**13.58**
	malonyl glycitin	**raw**	**pls-4**	**0.72**	**3.36**	**0.71**	**4.77**	**1.12**	**4.85**
	malonyl genistin	**d2**	**pls-7**	**0.90**	**25.36**	**0.95**	**15.18**	**0.60**	**15.84**
	total	**d2**	**pls-9**	**0.94**	**35.96**	**0.95**	**30.17**	**-0.25**	**31.52**
acetyl glycoside	acetyl daidzin	raw	pls-2	0.49	0.32	0.60	0.38	-0.02	0.40
	acetyl genistin	**d1**	**pcr-1**	**0.38**	**0.38**	**0.75**	**0.37**	**0.01**	**0.39**
	total	d1	pls-1	0.47	0.66	0.64	0.75	-0.01	0.78
aglycon	daidzein	raw	pcr-1	0.49	0.53	0.27	0.55	0.28	0.49
	genistein	d1	pcr-16	0.85	0.47	0.47	0.65	0.26	0.62
	total	raw	pcr-1	0.45	1.09	0.13	1.17	0.55	1.08
total isoflavone	**total**	**d22**	**pls-10**	**0.97**	**36.36**	**0.92**	**42.50**	**-6.06**	**43.94**
Vitamin B	**thiamine**	**raw**	**pls-1**	**0.31**	**0.09**	**0.83**	**0.06**	**0.01**	**0.06**
	riboflavin	d2	pls-1	0.52	0.01	0.34	0.01	0.00	0.01
tocopherol (toc)	α-toc	d2	pcr-4	0.65	2.23	0.69	1.42	0.73	1.31
	β-toc	**d2**	**pcr-1**	**0.07**	**1.06**	**0.73**	**0.40**	**0.28**	**0.30**
	γ-toc	d22	pcr-1	0.14	2.96	0.48	3.32	1.22	3.34
	δ-toc	d22	pcr-7	0.99	0.64	0.71	2.18	1.24	1.93
	total	**d22**	**pls-5**	**0.93**	**1.70**	**0.79**	**4.35**	**3.25**	**3.13**
	α-toc equivalence	**d2**	**pcr-4**	**0.66**	**2.38**	**0.73**	**1.59**	**1.07**	**1.28**

see footnotes in Table 7

Table 9. The calibration and the prediction results for powdered soybean with PLSR/PCR analysis on SS2400 data. [mg (100g DW)$^{-1}$]

estimation. As for intact plural soybean seeds, some of the respective component such as malonyl daidzin, and malonyl genistin, still could be estimated. On the other hand, the estimation for glycosides, acetyl glycosides, and aglycons were not good enough because of their small range fluctuations.

As for Vitamin B, considering from the SEPs in Table 7-10 with comparing the standard deviation: 0.08 [mg (100g DW)$^{-1}$] for thiamin, and 0.01 for riboflavin. These contents might be fairly estimated.

The feasibility of NIRS for the estimation of the contents of isoflavone in soybean seeds was examined. As for SpectraStar 2400 case, considering from SEP, NIRS may also be available for estimating the total isoflavone content. Even if the whole seeds analysis case, that of SEP was fair enough for their estimations. Further, especially as for powdered soybean, the respective components of the isoflavones, such as glucosides and malonyl glucosides, could be estimated separately. However, the estimations of acetyl glucosides and aglycons contents were not good because of their small range fluctuations. PLSR/PCR analysis were also tried, and the similar results were obtained and some were improved.

		Calibration				Prediction			
		treatment	factors	r	SEE	r	RMSEP	bias	SEP
glycoside	daidzin	d22	pls-11	0.95	4.33	0.63	9.06	0.80	9.43
	glycitin	d22	pls-15	0.99	0.57	0.53	3.20	1.43	2.99
	genistin	d2	pcr-10	0.78	12.34	0.60	12.06	2.96	12.21
	total	d22	pls-11	0.96	10.27	0.60	22.57	2.85	23.38
malonyl glycoside	**malonyl daidzin**	**d22**	**pls-11**	**0.98**	**10.64**	**0.92**	**15.68**	**1.39**	**16.31**
	malonyl glycitin	d22	pcr-7	0.56	4.24	0.41	5.78	1.50	5.83
	malonyl genistin	**d22**	**pls-12**	**0.97**	**14.46**	**0.87**	**21.21**	**-0.99**	**22.13**
	total	**d22**	**pls-11**	**0.97**	**24.95**	**0.91**	**33.75**	**2.48**	**35.15**
acetyl glycoside	acetyl daidzin	d1	pls-3	0.65	0.28	0.52	0.40	0.01	0.42
	acetyl genistin	d1	pls-3	0.57	0.35	0.47	0.43	0.04	0.45
	total	d1	pcr-4	0.58	0.64	0.52	0.80	0.07	0.83
aglycon	daidzein	d1	pcr-1	0.04	0.61	-0.40	0.54	0.30	0.47
	genistein	raw	pcr-1	0.10	0.66	-0.44	0.60	0.27	0.56
	total	raw	pcr-1	0.15	1.21	-0.36	1.15	0.59	1.03
total isoflavone	**total**	**d22**	**pls-11**	**0.97**	**34.70**	**0.88**	**51.34**	**5.73**	**53.28**
Vitamin B	thiamine	d22	pls-1	0.28	0.09	0.29	0.08	0.01	0.08
	riboflavin	d1	pls-1	0.29	0.02	0.07	0.01	0.00	0.01
tocopherol (toc)	α-toc	d1	pcr-1	0.02	2.60	0.11	1.62	-0.24	1.73
	β-toc	d22	pls-1	0.62	0.84	0.68	0.50	0.39	0.34
	γ-toc	d1	pls-1	0.37	2.77	0.61	2.63	0.70	2.74
	δ-toc	raw	pcr-1	0.66	2.42	0.23	3.28	1.75	2.99
	total	raw	pcr-1	0.41	3.59	0.31	5.19	2.80	4.73
	α-toc equivalence	d1	pcr-1	0.08	2.77	0.10	1.70	0.02	1.84

see footnotes in Table 7

Table 10. The calibration and the prediction results for intact plural soybean seeds with PLSR/PCR analysis on SS2400 data.

3.2.3 Data transfer (Sato et al., 2009a)

Table 11 described the calibration and prediction results, that are analyzed using SensoLogic with these NIRS data converted from InfraAlyzer 500 data into those of SpectraStar2400 data. The similar SEP results were obtained as in Table 7. Then, the obtained equations were adopted on those of prediction set of original data of SpectraStar2400. The SEP of the total isoflavone content was 70.82 [mg (100g DW)$^{-1}$] in the MLR analysis of the powder, and it shows that total isoflavone level was able to be estimated in stead of bias and skew. Further, using the original calibration set of SpectraStar2400, the bias and skew correction were carried out and the SEP was drastically improved to 45.72 [mg (100g DW)$^{-1}$]. The similar results were obtained in PLSR analyses, and also in the analysis of plural seeds.

			Calibration		Prediction			
			r	SEC	r	RMSEP	bias	SEP
powder	MLR Analysis	analysis on converted data	0.90	56.91	0.94	43.37	-1.47	45.27
		adopted on SpectraStar2400 data	before bias and skew correction		0.91	68.87	12.05	70.82
			after bias and skew correction		0.91	44.18	5.96	45.72
powder	PLSR Analysis	analysis on converted data	0.86	66.54	0.90	51.75	0.54	54.05
		adopted on SpectraStar2400 data	before bias and skew correction		0.91	83.94	-57.13	64.23
			after bias and skew correction		0.91	46.17	9.67	47.16
plural seeds	MLR Analysis	analysis on converted data	0.90	56.31	0.78	71.37	-2.03	74.52
		adopted on SpectraStar2400 data	before bias and skew correction		0.72	109.95	-68.76	89.61
			after bias and skew correction		0.72	74.70	2.81	77.97
plural seeds	PLSR Analysis	analysis on converted data	0.92	56.17	0.82	73.84	32.11	69.45
		adopted on SpectraStar2400 data	before bias and skew correction		0.88	91.62	66.80	65.50
			after bias and skew correction		0.88	56.23	18.22	55.57

see footnotes in Table 7

Table 11. The results of data transfer trial. The unit is mg (100g DW)$^{-1}$.

3.3 Analysis on Bruker Data

Table 12 described the calibration and the prediction results of the estimation of total isoflavone content with automatic analysis. Figure 5 shows the scattering plots of the prediction results of total isoflavone analysis. As for soybean powder analysis, judging from RMSEP, NIRS may be available for estimating the total isoflavone content. Even if the plural seeds analysis case, RMSEP was also good enough for the nondestructive estimation. Further, the respective components of the isoflavones, such as glycosides and malonyl glycosides, could be estimated separately as for powdered soybean and a part of plural soybean seeds cases (data abbreviated). As for a single seed analysis case, the precision was fair. Table 13 described the calibration and the prediction results of the estimation of Vitamin B and total tocopherol content with automatic analysis. The estimations for the contents of Vitamin B, and tocopherol were fair for rough estimation despite of their small range fluctuations.

	Calibration (n=36)			Prediction (n=12)			
sample type analyzed	rank	r	RMSEE	r	RMSEP	bias	RPD
soybean powder	9	0.974	31.9	0.971	24.9	6.71	3.99
plural soybean seeds	8	0.996	11.8	0.935	36.3	-5.48	2.81
a single seed soybean	8	0.814	77.1	0.886	53.1	1.91	2.16

rank: the number of PLSR/PCR vectors
r: Correlation coefficient between chemical method and NIR method.
RMSEE: Root mean square error of estimation. [mg/100gDW]
RMSEP: Root mean square error of prediction. [mg/100gDW]
RPD: ratio of standard error of prediction to standard deviation.

Table 12. The calibration and the prediction results of the estimation of total isoflavone content with automatic analysis on Bruker data.

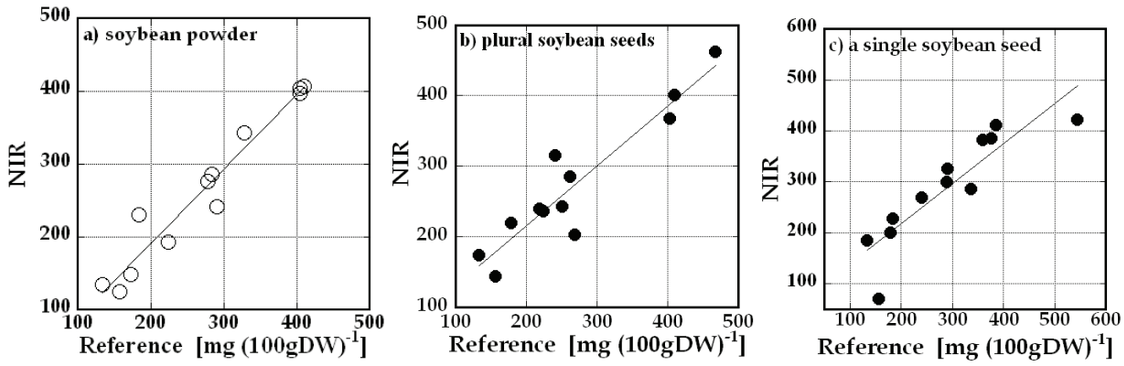


Fig. 5. Estimation of total isoflavone content. a) soybean powder; b) intact plural soybean seeds; c) a single soybean seed on Bruker data.

3.4 Analysis on Buchi Data

Table 14 described the calibration and the prediction results of the estimation of total isoflavone, Vitamin B and total tocopherol content with automatic analysis. As for soybean powder analysis, judging from SEP, NIRS may be available for estimating the total isoflavone content. Even if the plural seeds analysis case, SEP was also good enough for the nondestructive estimation. Further, the respective components of the isoflavones, such as glucosides and malonyl glucosides, could be estimated separately as for powdered soybean and a part of plural soybean seeds cases (data abbreviated). As for a single seed analysis case, the precision was fair. The estimations for the contents of Vitamin B, and tocopherol were fair for rough estimation despite their small range fluctuations.

4. Conclusions

The feasibility of some types of the near infrared spectroscopy (NIRS) for the estimations of the contents of isoflavones, thiamine (Vitamin B_1), riboflavin (Vitamin B_2), and tocopherol (Vitamin E) in soybean seeds was examined with MLR and PLSR/PCR analysis. Considering from the standard error of prediction, NIRS may be available for estimating the total isoflavone content not only as for powdered soybean but also as for intact plural soybean seeds. Vitamin B and tocopherol contents were fair enough for rough estimation

despite of their small range fluctuations. Trials of a single seed analysis, and the data transfer were also carried out.

The authors already reported the NIRS analysis of major constituents and the deterioration indices in the soybean (Sato et al. 1994), and the fatty acid composition in soybean (Sato et al. 2002). There were some trials to estimate amino acids composition in soy by NIRS method (Kovalenko et al., 2006; Pazdernik et al., 1997). In this study, some of physiologically functional elements can be estimated by NIRS method. The present findings showed that the isoflavone content could be estimated by the NIRS method, and the NIR method increase the value in soybean analysis. The NIRS method will gain greater position in the soybean analysis by these results.

sample type analyzed	constituents	Calibration (n=36 or 16)			Prediction (n=12 or 7)			
		rank	r	RMSEC	r	RMSEP	bias	RPD
soybean powder	thiamine(VB_1)	5	0.457	0.1	0.865	0.03	-0.01	1.94
	riboflavin(VB_2)	9	0.874	0.01	0.762	0.01	0.01	1.43
	total tocopherol	6	0.987	0.95	0.954	1.1	-0.02	3.07
plural soybean seeds	thiamine(VB_1)	9	0.853	0.05	0.826	0.07	-0.01	1.69
	riboflavin(VB_2)	7	0.812	0.01	0.86	0.01	0	1.85
	total tocopherol	7	0.972	1.15	0.8	3.12	-0.43	1.64
a single seed soybean	thiamine(VB_1)	4	0.371	0.09	0.741	0.06	-0.04	1.49
	riboflavin(VB_2)	1	0.228	0.01	0.537	0.02	0.01	1.11
	total tocopherol	5	0.727	3.83	0.978	1.03	-0.61	4.45

see footnotes in Table 12.

n: counts of samples (n = 36 for isoflavone, and Vitamin B, or 16 for tocopherol) and so on .

Table 13. The calibration and the prediction results of the estimation of Vitamin B and tocopherol with automatic analysis on Bruker data.

sample type analyzed	constituents	Calibration (n = 36 or 16)		Prediction (n = 12 or 7)		
		r	SEC	r	SEP	bias
soybean powder	total isoflavone	0.913	48.35	0.912	48.35	4.54
	thiamine(VB_1)	0.751	0.06	0.747	0.08	0.02
	riboflavin(VB_2)	0.738	0.01	0.758	0.02	0.00
	total tocopherol	0.803	2.67	0.804	2.82	1.02
plural soybean seeds	total isoflavone	0.941	40.13	0.934	40.61	20.83
	thiamine(VB_1)	0.860	0.04	0.873	0.06	0.02
	riboflavin(VB_2)	0.777	0.01	0.813	0.02	0.01
	total tocopherol	0.610	3.87	0.759	3.31	2.40

Table 14. The calibration and the prediction results of the estimation of total isoflavone, Vitamin B and tocopherol content with automatic analysis on Buchi data.

5. References

Choung M. G.; Baek I. Y.; Kang S. T.; Han W. Y.; Shin D. C.; Moon H. P. & Kang K. H. (2001). Non-destructive method for selection of soybean lines contained high protein and oil by near infrared reflectance spectroscopy, *Korean Journal of Crop Science*, 46, 5, May, 2001, 401-406, 0252-9777

Dyer, D. J. (2004). Analysis of Oilseeds and Coarse Grains. In: *Near-infrared Spectroscopy in Agriculture*, Roberts, C. A.; Workman Jr, J. & Reevees III, J.B. (Ed.), 321-344, American Society of Agronomy; Inc., 978-0891181552, Madison, Wisconshin.

Hollung, K. ; Overland, M.; Hrustic, M.; Sekulic, P.; Miladinovic, J.; Martens, H.; Narum, B.; Sahlstrom, S.; Sorensen, M.; Storebakken, T. & Skrede, A. (2005). Evaluation of nonstarch polysaccharides and oligosaccharide content of different soybean varieties (*Glycine max*) by near-infrared spectroscopy and proteomics. *Journal of Agricultural & Food Chemistry.* 53, 23, NOV 16 2005, 9112-9121, 0021-8561.

Hruschka, W. (2001). Data Analysis; Wavelength Selection Methods, In: *Near-Infrared Technology: In the Agricultural and Food Industries 2nd edition,* Williams, P. & Norris, K. (Ed.), 39-58, Amer Assn of Cereal Chemists, 978-1891127243, St. Pail, Minnesota.

Hymowitz, T.; Duldley, J.W.; Collins, F.I. & Brown, C.M. (1974). Estimation of Protein and Oil Concentration in Corn, Soybean, and Oat Seed by Near-Infrared Light Reflectance, *Crop Science*, 14, 5, Sep-Oct 1974, 713-715, 0011-183X

Kovalenko, I. V.; Rippke, G. R. & Hurburgh, C. R. (2006). Determination of amino acid composition of soybeans (*Glycine max*) by near-infrared spectroscopy. *Journal of Agricultural & Food Chemistry,* 54, 10, MAY 17 2006. 3485-3491, 0021-8561.

Kovalenko, V.; Rippke, G. R. & Hurburgh, C. R. (2006). Measurement of soybean fatty acids by near-infrared spectroscopy: linear and nonlinear calibration methods. *Journal of the American Oil Chemists' Society*, 83, 5, May 2006, 421-427, 0003-021X.

Kudou, S.; Fleury, Y.; Welti, D.; Magnolato, D.; Uchida, T.; Kitamura, K. & Okubo,K. (1991). Malonyl isoflavone glycosides in soybean seeds, *Agric. Biol.Chem.*, 55, 9, Sep 1991, 2227-2233, 0002-1369

Li, H.; Goulden, M.; Cocciardi, R. & Hughes. J. (2009). Fourier transform near infrared spectroscopy as a quality control tool for the analysis of lecithin and by-products during soybean oil processing. *Journal of the American Oil Chemists' Society.* 86, 9, Sep 2009, 835-841, 0003-021X

Nishiba, Y.; Suda I.; Oki T. & Sugawara T. (2007). Variations in Isoflavone, Thiamine, Riboflavin and Tocopherol Content in Soybean Seeds Grown in Japan, *Nippon Shokuhin Kagaku Kogaku Kaishi,* 54:295-303. 1341-027X. in Japanese

Osborne, B.G. & Fearn T. (1986). FOSFA Official Method, In : *Near Infrared Spectroscopy in food analysis,* Osborne, B.G. & Fearn T. (Ed.), 190-193, Longman Science and Technical, John Wiley and Sons, Inc., 0582494893, NY, USA.

Osborne, B.G.; Fearn T. & Hindle P.H. (1993). Theory of near infrared spectroscopy, In : *Practical NIR Spectroscopy with applications in food and beverage analysis, 2nd. Edn.*,

Osborne B.G. ; Fearn T. & Hindle P.H. (Ed.), 13-35. Longman Science and Technical, John Wiley and Sons, Inc., 0470221283, NY, USA.

Pazdernik, D.L.; Killam, A.S. & Orf, J.H. (1997). Analysis of Amino and Fatty Acid Composition in Soybean Seed, Using Near Infrared Reflectance Spectroscopy, *Agronomy Journal*, 89, 4, Jul-Aug, 1997, 679-685, 0002-1962.

Roberts, C. A.; Ren, C.; Beuselinck, P. R.; Benedict, H. R. & Bilyeu, K. (2006). Fatty Acid profiling of soybean cotyledons by near-infrared spectroscopy. *Applied spectroscopy,* 60, 11, Nov 2006, 1328-1333, 0003-7028.

Sato, T.; Abe, H.; Kawano, S.; Ueno, G.; Suzuki, K. & Iwamoto, M. (1994). Near Infrared Spectroscopic Analysis of Deterioration Indices of Soybeans for Process Control in Oil Milling Plant, *Journal of the American Oil Chemists' Society,* 71, 10, Oct 1994, 1049-1055, 0003-021X.

Sato, T.; Takahashi M, Matsunaga R (2002). Use of NIR spectroscopy for estimation of FA composition of soy flour, *Journal of the American Oil Chemists' Society,* 79, 6, Jun 2002, -537, 0003-021X.

Sato, T.; Eguchi, K.; Arimura, K. & Nishiba, Y. (2008). Nondestructive Estimation of Contents of Functional Constituents in Soybean Seeds by Near-infrared Spectroscopy, *Abstracts of 5th International Crop Science Congress & Exhibition (ICSC 2008, Jeju, Korea),* Apr 2008, 70.

Sato, T.; Eguchi, K.; Hatano, T. & Nishiba, Y. (2008). Use of Near-Infrared Reflectance Spectroscopy for the Estimation of the Isoflavone Contents in Soybean Seeds, *Plant Production Science,* 11, 4, Oct 2008, 481-486, 342-943X.

Sato, T ; Eguchi, K. ; Hatano, T. & Nishiba, Y. (2009a). Examination of Transformation of Near-Infrared Spectroscopic Data, *Abstracts of the 25th NIR Forum,* May 2009, 123.

Sato, T.; Eguchi, K.; Eguchi, K. & Nishiba, Y. (2009b). Use of NIR for the Estimation of Total Isoflavone Content in Single Seed Soybean, *Abstracts of the 14th International Conference on Near-Infrared Spectroscopy* (NIR-2009, Bangkok, Thailand), Nov 2009, 16.

Shenk, J. S.; Workman Jr. J. J. & Westerhaus M.O. (2007). Application of NIR Spectroscopy to Agricultural Products, In : *Handbook of Near-Infrared Analysis 3rd edition*, Burns D.A. & Ciurczak E.W. (Ed.), 347-386, CRC Press, 978-0849373930, Boca Raton, FL.

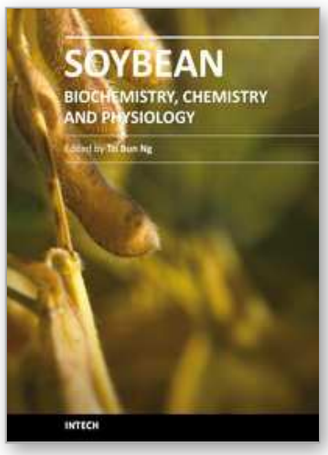
Tajuddin, T.; Watanabe, S.; Masuda, R.; Harada, K. & Kawano, S. (2002). Application of near infrared transmittance spectroscopy to the estimation of protein and lipid contents in single seeds of soybean recombinant inbred lines for quantitative trait loci analysis, *Journal of Near Infrared Spectroscopy,* 10, 4, dec 2002, 315-325, 0967-0335.

Tsujimura, T. (2004). Kinds of Vitamin and their Properties, In : *Standard Tables of Food Composition in Japan, fifth edition*, Kagawa, Y. (Ed.), 408-409,Kagawa Nutrition University Publishing Division, 478950109, Tokyo. in Japanese.

USDA FIGS (1996). Near-Infrared Transmittance (NIRT) Handbook, Washington DC,

Williams, P.C. (2006). Grains and Seeds (Application to Agricultural and Marine Products), In: *Near-infrared Spectroscopy in Food Science and Technology,* Ozaki, Y.; McClure W. F. & Christy A. A. (Ed.), 165-217,John and Wiley & Sons., Inc., 978-0471672012, Hoboken, NJ.

Yamori, Y. (2001). Osteoporosis, In : *Soybean Isoflavones,* Yamori Y., Ohta S. & Watanabe S. (Ed.), 79-87, Saiwai Shobou Co., 978-4782101810, Tokyo. in Japanese.



Soybean - Biochemistry, Chemistry and Physiology
Edited by Prof. Tzi-Bun Ng

ISBN 978-953-307-219-7
Hard cover, 642 pages
Publisher InTech
Published online 26, April, 2011
Published in print edition April, 2011

Soybean is an agricultural crop of tremendous economic importance. Soybean and food items derived from it form dietary components of numerous people, especially those living in the Orient. The health benefits of soybean have attracted the attention of nutritionists as well as common people.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Tetsuo Sato (2011). Nondestructive Estimation of the Contents of the Functional Elements in Soybean by Near Infrared Reflectance Spectroscopy, Soybean - Biochemistry, Chemistry and Physiology, Prof. Tzi-Bun Ng (Ed.), ISBN: 978-953-307-219-7, InTech, Available from: http://www.intechopen.com/books/soybean-biochemistry-chemistry-and-physiology/nondestructive-estimation-of-the-contents-of-the-functional-elements-in-soybean-by-near-infrared-ref

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen