한작지(Korean J. Crop Sci.), 55(4): 357~364(2010)

Application of Near-Infrared Reflectance Spectroscopy (NIR) Method to Rapid Determination of Seed Protein in Coarse Cereal Germplasm

Young-Yi Lee*†, Jung-Bong Kim**, Ho-Sun Lee*, Sok-Young Lee*, Jae-Gyun Gwag*, Ho-Cheol Ko*, Yun-Chan Huh*, Do-Yoon Hyun*, and Chung-Kon Kim*

*Agrobiodiversity Center, NAAS, RDA. Suwon 441-717, Korea
**Dept. of Agro-Food Resources, NAAS, RDA. Suwon 441-857, Korea

ABSTRACT Kjeldahl method used in many materials from various plant parts to determine protein contents, is laborious and time-consuming and utilizes hazardous chemicals. Near-infrared (NIR) reflectance spectroscopy, a rapid and environmentally benign technique, was investigated as a potential method for the prediction of protein content. Near-infrared reflectance spectra(1100-2400 nm) of coarse cereal grains(n=100 for each germplasm) were obtained using a dispersive spectrometer as both of grain itself and flour ground, and total protein contents determined according to Kjeldahl method. Using multivariate analysis, a modified partial least-squares model was developed for prediction of protein contents. The model had a multiple coefficient of determination of 0.99, 0.99, 0.99, 0.96 and 0.99 for foxtail millet, sorghum, millet, adzuki bean and mung bean germplasm, respectively. The model was tested with independent validation samples (n=10 for each germplasm). All samples were predicted with the coefficient of determination of 0.99, 0.99, 0.99, 0.91 and 0.99 for foxtail millet, sorghum, millet, adzuki bean and mung bean germplasm, respectively. The results indicate that NIR reflectance spectroscopy is an accurate and efficient tool for determining protein content of diverse coarse cereal germplasm for nutrition labeling of nutritional value. On the other hands appropriate condition of cereal material to predict protein using NIR was flour condition of grains.

Keywords : NIR, cereal, germplasm, protein

Coarse cereal grains role as a supplementary food for human since it has not only functional traits distinguishable to major grain food but also essential constituents including protein. Protein is essential to a healthy diet and has important properties in food processing. The World Health Organization estimate that more than 2 billion people have deficiency in key micronutrients such as Zn and Fe and more than 160 million children under the age of 5 lack adequate protein (Uauy *et al.*, 2006). While it mainly supplied by meat and beans, minor cereal crop grain contributes to protein supply also. But many studies have focused on main components in major crop while minor crops such as millet, sorghum, adzuki bean and mung bean have not been spotlighted.

RDA genebank, Agrobiodiversity center of Korea, conserves total 156 thousands of plant seed germplasm and over 75% (118,045 accessions) of those is occupied by grain food crops (RDA 2009). But coarse cereal crop is conserved with accessions of 23,070 in Agrobiodiversity center which is included adzuki bean and mung bean germplasm. In case of rice and soybean, germplasm has been focused as main resources for breeding of food crop for a lot of time in Korea. So characterization and evaluation of those germplasm have been performed actively. However other grain crop germplasm remained has been relatively little highlighted than in case of rice and soybean. For user or distribution of germplasm, a rapid characterization of breeding-targeted characteristics is very important for conservation, management and utilization of plant germplasm (Lee *et al.*, 2009, Vines *et al.*, 2005).

Near-infrared reflectance spectroscopy (NIR) has been developed for measurement of many quality-related traits that routinely tested in cereal breeding program. This technique has several well-known advantaged as compared to conventional laboratory method. For some major components of rice grain such as starch, protein, amino acid and fat, NIR calibration models have been well-developed. For other components such as phenolic and tocopherol compounds or for other crop grain, such as adzuki bean, mung bean and

†Corresponding author: (Phone) +82-31-299-1814
(E-mail) youngyi@korea.kr <Received November 4, 2010>

foxtail millet, almost no NIRs models have been previously developed (Kim *et al.*, 2004). Some literatures just show that NIRs methodology is feasible for determining the contents of phenolic compounds in other system or in other food. For example, phenolic compound in red wine fermentation could be accurately predicted by NIRs (Cozzolino *et al.*, 2004). In this study, we tried to examine and compare the protein contents in 5 species of coarse cereal grains, and find out method to predict grain protein rapidly in massive seed gerplasm of coarse cereal crops using NIR also. As the coarse cereal crops, the collections of foxtail millet (*Setaria italica* L. P. Beauv), sorghum (*Sorghum tricolor*), common millet (*Panicum miliaceum*), adzuki bean (*Vigna angularis*), and mung bean (*Vigna radiata*) germplasms were subject plant resources since characterization and evaluation of those were relatively little and no NIR model was developed for this germplasm. Successful prediction will contribute to more effective application of NIRs in coarse cereal germplasm utilization and crop breeding program.

MATERIALS & METHODS

Five coarse cereal species were germplasm samples for this study. Those were 433 accessions of foxtail millet (*Setaria italica*), 942 accessions of sorghum (*Sorghum tricolor*), 516 accessions of millet (*Panicum miliaceum*), 641 accessions of adzuki bean (*Vigna angularis*), and 1125 accessions of mung bean (*Vigna radiata*). For scanning samples as flour condition and analyzing the protein using Kjeldahl method, all 100 samples of each germplasm were ground as the size of 200 sieve mesh using ball mill instrument.

To obtain NIR reflectance spectra, samples were scanned immediately after grinding in triplicate spectral analysis cells using a dispersive NIR Systems 6500 spectrometer (FOSS North America Inc.). The instrument was fitted with a spinning cup sampling device. Each cell was scanned 16 times, and the spectra were averaged and transformed to log (1/*R*). After visual inspection, the spectra of the triplicate subsamples were averaged. Protein contents of each 100 accessions of germplasm were found by the standard Kjeldahl Procedure (N × 6.25) using 0.2 g of ground whole seed.

The NIR reflectance model was developed using a commercial spectral analysis program (ISI40 NIRS 2 version 4.01 and WINISI software, FOSS North America Inc.). Preprocessing of the spectral data (1104-2494 nm) for 100 samples consisted of a normal multiplicative scatter. The data were centered on the mean spectrum and mean reference value using modified partial least-squares (PLS) regression. The PLS regression used to develop the model is modified in that the reference values and reflectance data are scaled at each wavelength to have a standard deviation of 1.0 before each PLS regression term. The preprocessing methods chosen for the model were optimum to obtain minimum error following cross-validation (4 cross-validation groups, in each collection of germplasm). The optimum number of PLS regression terms for the calibration was also determined by cross-validation and was the number of factors that gave the minimum error between predicted and reference values (standard error of cross validation, SECV) .The modified PLS regression model was tested using independent validation samples (n = 10). Statistics used to assess the model were standard error of performance (SEP, not bias corrected), coefficient of determination (r^2), slope, and intercept of the linear regression of NIR predicted versus analyzed values.

Mathematic treatments were conducted to modify the spectrum. For better coefficient between analyzed value and NIR absorbance value, the waves of spectra were corrected in various ways. Mathematical treatment is expressed as 1,4,4,1 for the default condition of the WINSIS program. It means change of derivative, gap, smooth 1 and smooth 2 conditions of spectrum opposite to the numeral, respectively.

RESULTS AND DISCUSSION

Spectra collection through NIRs scanning

In the present study, the NIR spectra of 433 accessions of foxtail millet, 942 accessions of sorghum, 516 accessions of millet, 641 accessions of adzuki bean, and 1125 accessions of mung bean germplasm were obtained as grain status which is not ground. And then about 100 accessions of each of those were selected by spectral data, origin and color of seed coat. Figure 1 shows the spectra of selected samples are almost similar to those of whole spectra in

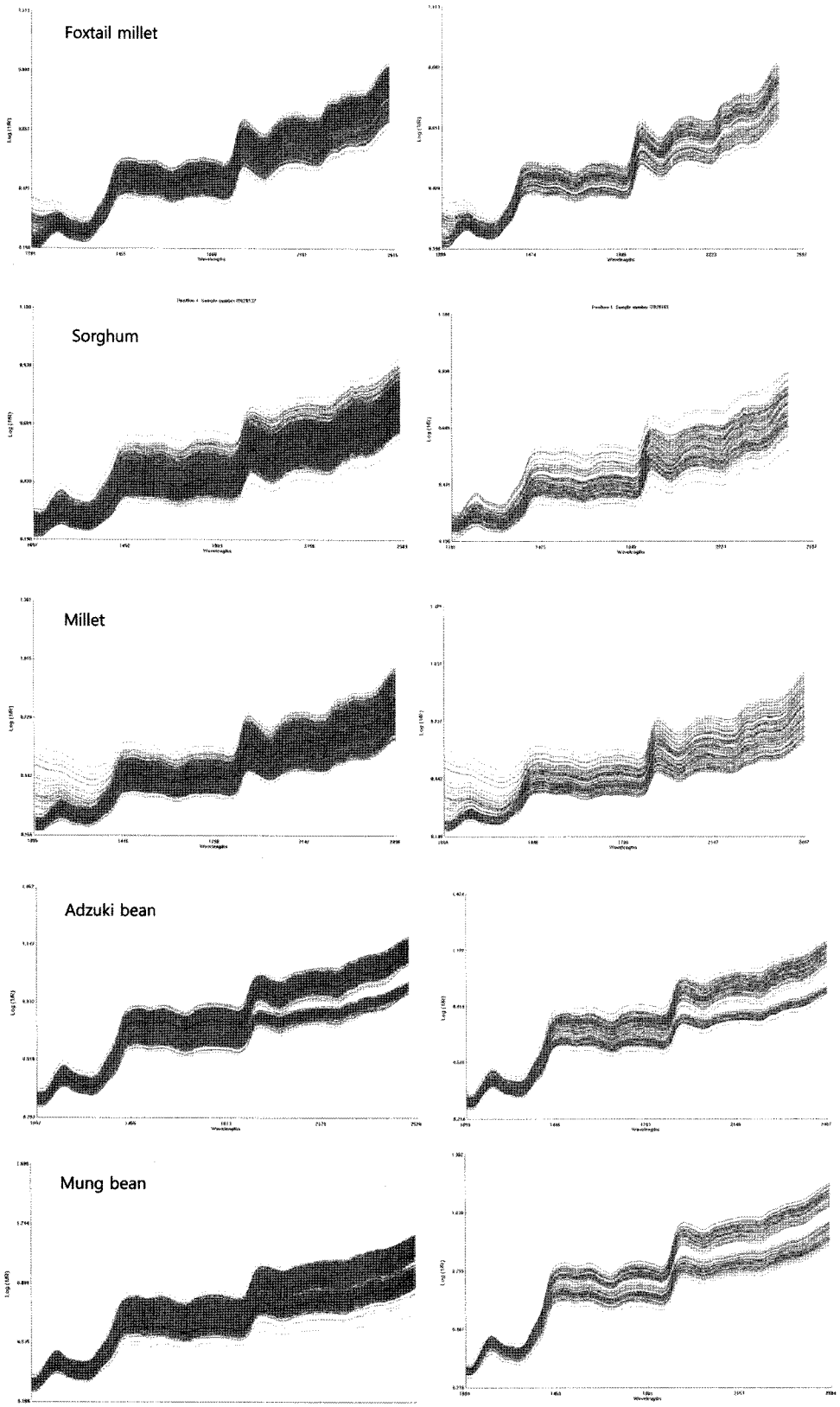

Fig. 1. NIR spectra for grain condition of foxtail millet, sorghum, millet, adzuki bean, and mung bean germplasm. Left, spectra for whole samples of 433, 942, 516, 644, and 1125 accessions of foxtail millet, sorghum, millet, adzuki bean and mung bean germplasm, respectively; Right, spectra for 100 accessions of each germplasm selected for NIR calibration.

diversity that we can see as the width of spectra collection. It means the selected samples could be representative to whole collections of germplasm. Also figure 1 shows the comparison of spectra with status of grain in each germplasm. By and large patterns of spectra seemed to be similar each other but about at 1974 nm, the height of peaks have some differences by germplasms. So the absorbance at the wavelength of 1974 nm is thought to be decisive value that concern chemical composition including protein composition. Because little distinct peaks were observed in raw spectra without math treatment, these spectra were modified to find out more correlative wavelength with determination of protein by mathematic treatment in variable peak and shape. After mathematic treatment, some absorption bands displayed clearer peak (data is not shown). Subsequently selected 100 samples were ground and NIR scanned again before Kjeldahl analysis of protein was started. In the spectra of flour ground samples (Figure 2), their peaks are sharper and clearer than in those of grain samples not ground. So these two patterns of spectrum could be compared and investigated which one was more appropriated for NIR model regression. Zhang *et al.* (2008) reported that the spectra of the dehulled rice grain were very similar to those of milled rice grain, but the grain

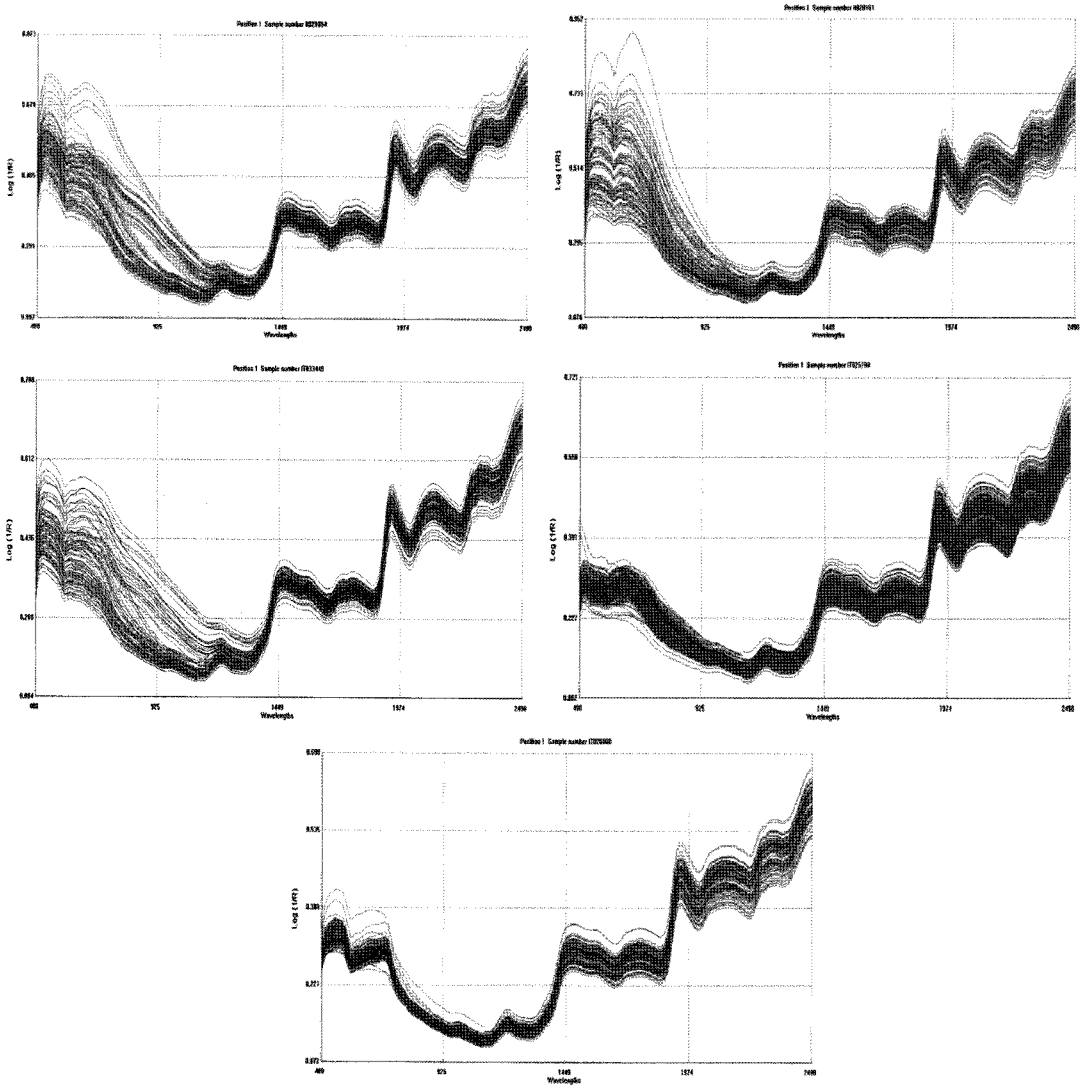

Fig. 2. NIR spectra for flour condition of foxtail millet, sorghum, millet, adzuki bean and mung bean germplasm. Spectra are for 100 accessions of each germplasm selected for NIR calibration.

sample clearly had stronger energy absorption than flour samples.

Determination of reference values

Protein contents in examined grain were examined shown in Table 1 and Table 2. The average of seed protein contents was the highest in mung bean germplasm with 25.22% and the lowest in foxtail millet with 1.58%. Sorghum, millet and adzuki bean contain the averages of 10.79, 11.96 and 22.11% of seed protein, respectively. Among the germplasm collections, IT201177, a mung bean accession, had especially high content of protein as 28.62% and IT209140, a foxtail millet accession, had especially low content of protein as 1.22%.

NIRs calibration for modeling

The reference values of all samples were used to make mathematical treatments in order to make the prediction equation with a modified PLS. The equations were used to calculate predicted values. And then external validation was accomplished with independent sample set (n=10) to make sure whether the equations developed in this study were applicable to predict seed protein contents. The validation graphics were shown in Figure 3 and the predictive values were given in Figure 4.

Table 1 shows the equation statistics for NIR prediction of protein as grain condition of germplasm samples. The coefficient of determine between the predicted values as grain and the reference values were 0.97 and 0.94 for foxtail millet and millet respectively. For the rest species of germplasm, the coefficient values were comparatively not significant (r^2=0.03, 0.55, 0.06 for sorghum, adzuki bean, mung bean). These results show the seed size is important for non-destructive prediction by NIR. Especially the values of r^2 for sorghum and mung bean were too low, that was thought to be for the grain size of those was comparatively big. The big grain could make big size of apertures also which were interruptions to NIR scanning.

Table 1. Equation statistics using regression model (MPLS) and scatter correction for NIRS prediction of protein contents in the calibration set of coarse cereal germplasm as grain samples.

Cereal grain	N	Mean	SD	Calibration		Cross-validation		Math treatment
				SEC	R^2	SECV	1-VR	
Foxtail millet	92	10.58	1.57	0.25	0.97	0.40	0.93	2,4,4,1
Sorghum	98	10.79	1.57	1.54	0.03	1.59	-0.02	1,4,4,1
Millet	93	11.96	1.74	0.41	0.94	0.50	0.92	2,6,4,1
Adzuki bean	113	22.11	1.31	0.88	0.55	1.06	0.38	3,5,5,1
Mung bean	99	25.22	1.11	1.08	0.06	1.12	0.00	2,6,4,1

N, number of samples used to develop the model; SD, standard deviation; SEC, standard error of calibration; R^2, coefficient of determination of calibration; SECV, standard error of cross-validation; 1-VR, coefficient of determination for cross-validation.

Table 2. Equation statistics using regression model (MPLS) and scatter correction for NIRS prediction of protein contents in the calibration set of coarse cereal germplasm as flour samples.

Cereal flour	N	Mean	SD	Calibration		Cross-validation		Math treatment
				SEC	R^2	SECV	1-VR	
Foxtail millet	92	10.58	1.57	0.19	0.99	0.23	0.98	2,6,4,1
Sorghum	96	10.75	1.55	0.16	0.99	0.26	0.97	2,4,4,1
Millet	96	11.86	1.76	0.14	0.99	0.22	0.99	2,6,4,1
Adzuki bean	109	22.16	1.39	0.26	0.96	0.30	0.96	3,5,5,1
Mung bean	97	25.31	1.17	0.12	0.99	0.26	0.95	2,4,4,1

N, number of samples used to develop the model; SD, standard deviation; SEC, standard error of calibration; R^2, coefficient of determination of calibration; SECV, standard error of cross-validation; 1-VR, coefficient of determination for cross-validation.

Table 2 shows the equation statistics for NIR prediction of protein as flour condition of germplasm samples. The coefficient of determine between the predicted values as flour and the reference values had a high significance with 0.99 for every species of germplasm except adzuki bean that had a coefficient of determine with 0.96 and it is enough significant also. Additionally the SD/SECV ratio was considered as the factor to evaluate the reliability of calibration model (Moschner & Biskuper-Korell 2006). The SD/SECV values for protein were all above 3 (9.54, 6.01, 8.04, 4.70 and 4.54 for foxtail millet sorghum, millet, adzuki bean, and mung bean, respectively). These higher value of index (SD/SECV > 3.0) indicate a significant coefficient between reference values and NIRs estimated values similar to those of calibration. Subsequently in validation set (n=10), the coefficient of determine seed protein level were still high as 0.99, 0.99, 0.99, 0.91 and 0.99 for foxtail millet sorghum, millet, adzuki bean, and mung bean, respectively in flour condition (Figure 3). Therefore the developed equations could be applied for

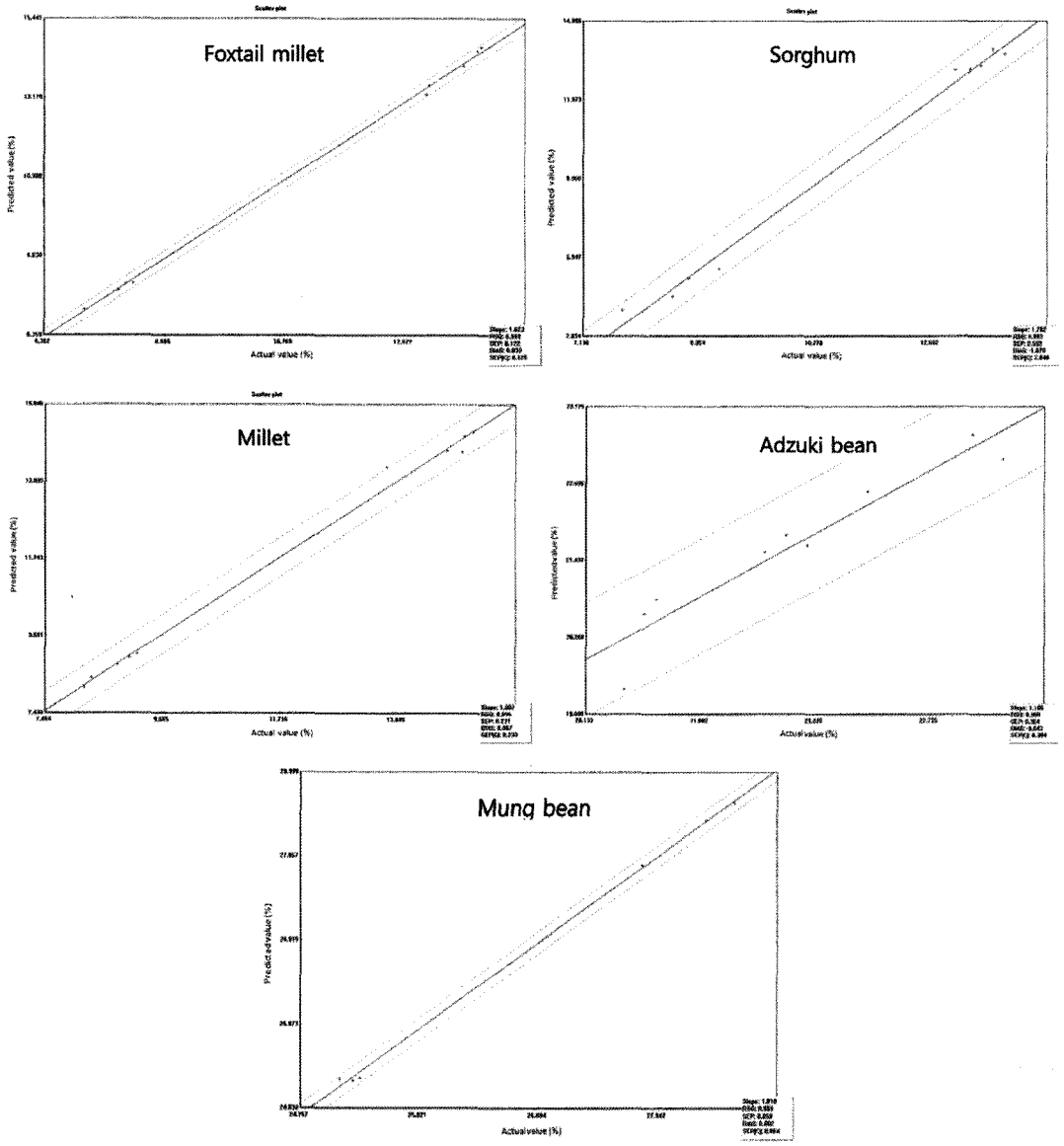

Fig. 3. Scatter plots of NIR predicted by developed equation versus actual-analyzed reference values for protein (R^2=0.993, 0.999, 0.995, 0.908 and 0.999 for foxtail millet, sorghum, millet, adzuki bean, and mung bean) in the external validation set (n=10) of flour samples of each germplasm.

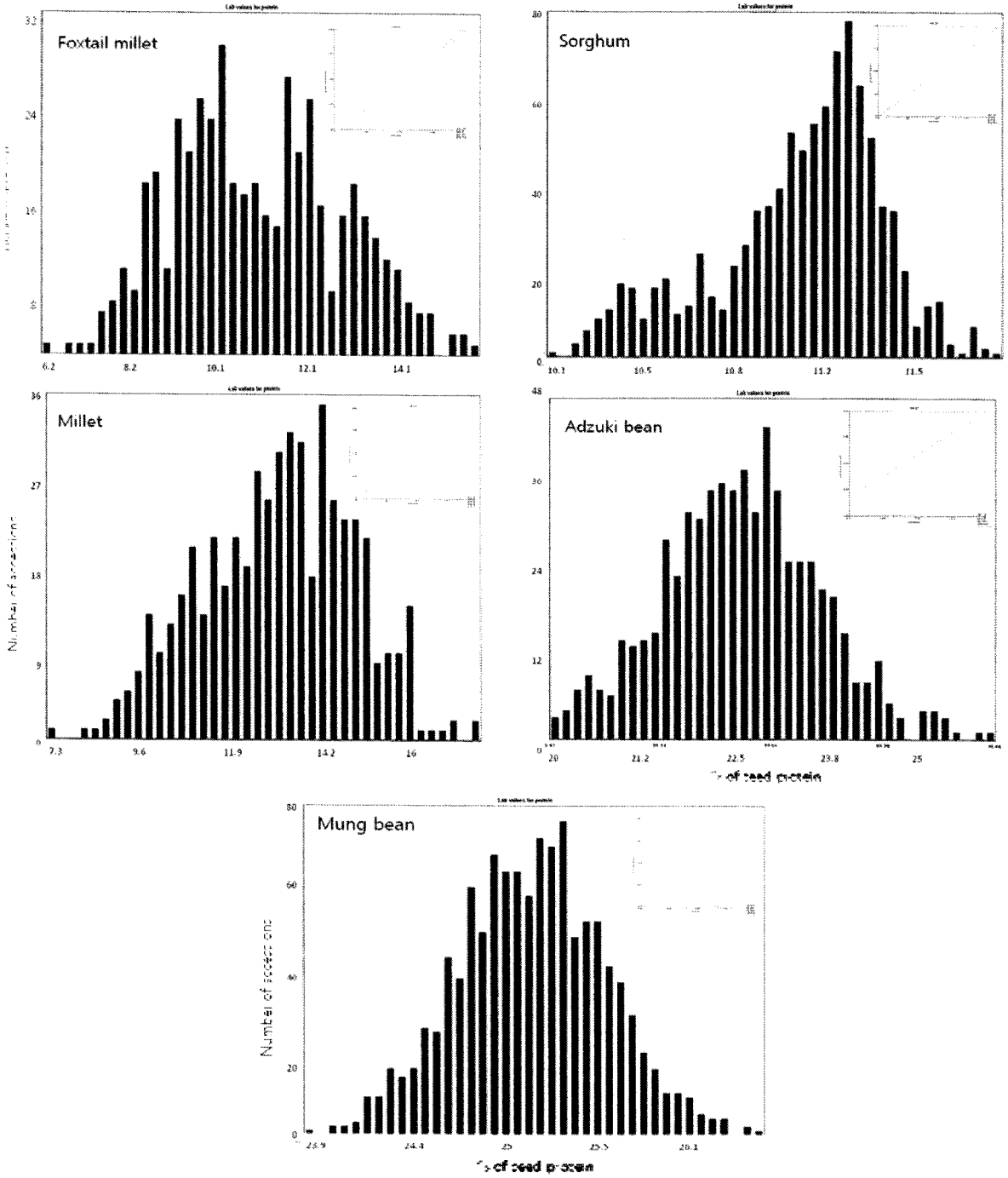

Fig. 4. Histograms for the distribution of whole used accessions according to NIR predicted value of protein contents by the equation developed in this study.

prediction of seed protein in all species of germplasm and as that results the histograms of germplasm distribution were obtained like Figure 4.

On the other hand the effective status of sample scanned by NIRs was investigated as grain versus flour status. In all cases flour samples were better than whole grain samples for NIR prediction while in some cases for non-destructive, rapid and convenient prediction of seed protein contents, raw grain condition can be recommended.

NIRs absorbance is affected by OH-group of chemical a

lot. Germplasm conserved in National agrobiodiversity center was considered as in a same condition of moisture. Because the germplasm materials have been conserved in same mid-term conservation of 4℃ in National agrobiodiversity center. In furthermore study investigation of seed moisture on each accession would be necessary.

In conclusion, the contents of seed protein could be predicted with reliable accuracy using NIRs analysis of intact seeds of foxtail millet sorghum, millet, adzuki bean, and mung bean. And considering SD/SECV values also, the contents of seed protein could be predicted with reliable accuracy using NIRs analysis of intact seeds of foxtail millet sorghum, millet, adzuki bean, and mung bean. Generally, despite the presence of larger population of germplasm covering a wide range of chemical values was required to obtain more accurate prediction of chemical composition (Gotor *et al.*, 2007), the results on this study may have enough significant coefficient of determine as above 0.99 in almost species of germplasm.

REFERENCES

Cozzolino D., M. J. Kwiatkowskia, M. Parkera, W. U. Cynkara, R. G. Dambergsa, M. Gishena, and M. J. Herderich. 2004. Prediction of phenolic compounds in red wine fermentations by visible and near infrared spectroscopy. Anal. Chim. Acta 513(1): 73-80.

Gotor A. A., M. Berger, E. Farkas, F. Labalette, S. Centis, and A. Calmon. 2007. Quantification of sunflower minor components by near-infrared spectrometry (NIR). Helia. 30(47): 183-190.

Kim Y. H., C. S. Kang, and Y. S. Lee. 2004. Quantification of Tocopherol and Tocotrienol Content in Rice Bran by Near Infrated Reflectance Spectroscopy. Korean J. Crop Sci. 49(3): 155-275.

Lee S. Y., Y. Y. Lee, S. K. Lee, Y. H. Cho, K. H. Ma, H. K. Kang, J. G. Gwag, Y. H. Park, and H. S. Lee. 2009. Selecting high amylose lines of rice varieties using NIR spectroscopy at the RDA gene bank conserved. Korean journal of breeding science, 41(1): 108.

Moschner, C. and B. Biskupek-Korel. 2006. Estimating the content of free fatty acids in high oleic sunflower by near-infrared spectroscopy. Eur. J. Lipid Sci. Technol. 108(7): 606-613.

RDA. 2009. The status of plant genetic resource conservation.

Uauy C., A. Distelfeld, T. Fahima, A. Blechl, and J. Dubcovsky. 2006. A NAC Gene regulation senescence improve grain protein, zinc, and iron content in wheat. Science 314: 1298-1301.

Vines L., S. E. Kays, and P. E. Koehler. 2005, Near-infrared reflectance model for the rapid prediction of total fat in cereal foods, J. Agric. Food Chem. 53: 1550-1555.

Zhang C., Y. Shen, J. Chen, P. Xiao, and J. Bao. 2008. Nondestructive prediction of total phenolics, flavonoid contents, and antioxidant capacity of rice grain using near-infrared spectroscopy. J. Agric. Food Chem. 56: 8268-8272.