Prediction of Chemical Composition in Distillers Dried Grain with Solubles and Corn Using Real-Time Near-Infrared Reflectance Spectroscopy

Sung Won Choi[1], Chang Hee Park[1], Chang Sug Lee[1], Dong Hee Kim[1], Sung Kwon Park[2], Beob Gyun Kim[3] and Sang Ho Moon[4]

[1]Korea Spectral Products, Seoul 152-779, Korea, [2]National Institute of Animal Science, RDA, Suwon 441-706, Korea, [3]Department of Animal Science and Technology, Konkuk University, Seoul 143-701, Korea, [4]School of Food Bio Science, Konkuk University, Chungju 380-701, Korea

ABSTRACT

This work was conducted to assess the use of Near-infrared reflectance spectroscopy (NIRS) as a technique to analyze nutritional constituents of Distillers dried grain with solubles (DDGS) and corn quickly and accurately, and to apply an NIRS-based indium gallium arsenide array detector, rather than a NIRS-based scanning system, to collect spectra and induce and analyze calibration equations using equipment which is better suited to field application. As a technique to induce calibration equations, Partial Least Squares (PLS) was used, and for better accuracy, various mathematical transformations were applied. A multivariate outlier detection method was applied to induce calibration equations, and, as a result, the way of structuring a calibration set significantly affected prediction accuracy. The prediction of nutritional constituents of distillers dried grains with solubles resulted in the following: moisture (R^2=0.80), crude protein (R^2=0.71), crude fat (R^2=0.80), crude fiber (R^2=0.32), and crude ash (R^2=0.72). All constituents except crude fiber showed good results. The prediction of nutritional constituents of corn resulted in the following: moisture (R^2=0.79), crude protein (R^2=0.61), crude fat (R^2=0.79), crude fiber (R^2=0.63), and crude ash (R^2=0.75). Therefore, all constituents except for crude fat and crude fiber were predicted for their chemical composition of DDGS and corn through Near-infrared reflectance spectroscopy.

(**Key words :** Chemical composition, Corn, DDGS, NIRS, PLS)

Ⅰ. INTRODUCTION

Corn is an important energy source of animal feed and is used as an important food resource in some countries. Distillers dried grain with solubles (DDGS) whose demand is increasing with the expansion of fuel ethanol industry is a good alternative to feed in response to increasing feed price in the international grain market.

Analyses of nutritional constituents of these animal feed are very important in evaluating the growth rates of animals taking relevant feed and the cost of feed. However, the laboratory chemical analysis is very complicated in terms of a procedure and a process. Furthermore as its analysis method varies depending on nutritional component, it takes expensive and time-consuming, which means it does not have price competitiveness in response to the international grain market. Therefore, it is essential to develop an accurate but more simple, efficient and cost effective technique to analyze various nutritional constituents of feed, and, at the same time, to reduce an analytical procedure, time and cost.

The analysis technique using Near Infrared spectroscopy (NIRS) is not destructive and does not requires various kinds of reagents. It enables the analysis of many constituents at the same time. Since the 1970s, the technique has widely been used not only for analysis of feed, but also in many areas, including medical, pharmaceutical, biochemical, and petrochemical areas. Abroad, the near-infrared based practical technology of evaluating for feed nutrients has already been researched and applied (Jocelyne Aufrere et al., 1996; Sandra et al., 2002; Norris et al., 1976; Shenk et al., 1977; Blosser, 1985; Randall, 1984; Isaac, 1982).

In the United States, some states set up a consortium to share with one another the chemical data and other data

---

* Corresponding author : Sung Won Choi, The industrial R&D center, Korea Spectral Products, Seoul 152-779, Korea, Tel: +82-2-2109-8871, Fax: +82-2-2109-8872 E-mail: choi0903@ksp.co.kr

analyzed through near-infrared and to induce more accurately analyzed results. In addition, they are attaching the data to various types of harvesting equipment for real-time analysis (Dinamica generale, AKEY Nutrition & Research Center, USA).

However, domestically, there is a lack of NIRS based research on feed (Park et al., 2004; Lee et al., 2009). In particular, it is rare to see research on statistical analysis of near-infrared data in chemometrics. Therefore, this work was intended to use the chemometrics based on near-infrared analysis and statistical analysis to determine nutritional constituents of distillers dried grain with solubles (DDGS) and corn, mostly used for feed, and thereby to investigate the prediction accuracy of near-infrared analysis technique and to review the possibility of the nutritional analysis of feed to reduce time and cost and analyze many different constituents at the same time.

Ⅱ. MATERIALS AND METHODS

1. Sampling

The forage samples used in this study were obtained from various countries. Each sample weighted approximately 500g. The production day was recorded and forage samples obtained were stored in a freezer. Most samples were produced in the United States and other countries including China, India, Brazil, Argentina, Indonesia and Australia. The selection of the samples is very important in completing a robust prediction model. Therefore, this study tried to reflect various characteristics of moisture, protein, fat, fiber, and ash.

2. Spectroscopic analysis

The samples used for analysis were ground by a home shredder to minimize the effect of particle size, and then were physically pre-processed by 1mm sieve to make particle size even. A sample cup was used to measure the ground feed samples. The structure of the sample cup in this study was based on the round-shaped structure of Optical window suggested by Rosenthal (1988) and USDA. For a method of making the sample cup, this work used

the optical window made from Material (Corning 7980 fused silica), mostly used in high quality optics these days, instead of Material (GE-124), previously used a lot, and evenly coated the inside of the cup with gold, a material which has high reflectance. About 10g sample was put in a sample cup, and then was pressed constantly by stand typed gold coated reflector. After that, reflectance Near-infrared (NIR) spectra (log 1/R) were obtained with model DAF2000 NIR Instrument, between 1100 nm and 2000 nm (Korea Spectral Products, Korea). The NIR instrument (DAF2000) has silicon and indium gallium arsenide (InGaAs) detector arrays with two stage thermoelectric (TE) cooling system and an intense broadband light source.

3. Reference analysis

Reference analysis referred to laboratory chemical analysis. It means the standard analysis used for predicting results with the use of reflectance NIR spectra of samples obtained from spectroscopic analysis and statistical analysis. Since reflectance NIR spectra rely on reference analysis results at the time of statistical analysis, the quality of reference analysis plays a critical role to accuracy of predicted results (Williams, 1987). No matter how good NIR Instrument, calibration methods and procedure, and high dimensional algorithm are, if the quality of reference analysis is poor, predicted results are bad. Therefore, for DDGS and Corn used for analysis, reference analysis results with high quality were drawn with the help of Korea Feed Association. In accordance with Association of Official Analytical Chemists (AOAC) method, this work measured moisture, crude protein, crude fat, crude fiber and crude ash.

4. Calibration

Chemometrics is the science of statistics and chemistry, and has been used for NIR based analysis of feed. After the application of mathematical models, statistics principles and other logic-based methods have been applied since 1971 (Gemperline, 2006). These methods have been grown as one area of analytical chemistry.

Reflectance NIR spectra of forages has multivariate characteristic and serves as a method of predicting an

amount of constituents by using NIR calibration equations to analyze chemical information quickly and accurately. It used Partial Least Squares (PLS) (Martens and Naes, 1989), one of analytical methods suggested in chemometrics. For all of the statistical analysis, R ver. 2.15.2 was used. Various mathematical transformations were applied to each chemical composition in calibration set to write NIR calibration equations.

As evaluation criteria of the written NIR calibration equations, coefficient of determination (R^2), standard errors of calibration set (SEC), and standard errors of robust calibration set (SERC) were used. For NIR calibration equations test, validation set being comprised of NIR spectra of forages which are not included in calibration set was used, and evaluated by standard errors of prediction (SEP).

1) Mathematical transformation of spectra data

When milled samples are measured by NIR diffuse reflectance spectroscopy, particle size, light scatter, and path-length variation badly affect the reflectance NIR spectra obtained. As a result, when results are predicted by statistical analysis, they hinder accuracy. As effective methods to remove or minimize the effect, Multiplicative Scatter Correction (MSC), De-trending, and Standard Normal Variate (SNV) (Barnes et al., 1989) were used for removing unnecessary noise in NIR spectra data, correcting damaged spectra, and improving the accuracy of NIR calibration equations.

2) Reduction methods of the outlier samples in Calibration set

As a method to improve the accuracy of NIR calibration equations, it is very important to structure a robust and highly perfect calibration set prior to statistical analysis. To do that, it is necessary to select and remove abnormal samples hindering analysis and to structure a calibration set with a proper size. Therefore, this work used multivariate outlier detection method (Filzmoser, 2003). Outliers are source of the main causes of errors when data are analyzed. When data are unvariate following normal distribution, data out of standard deviation from the mean is selected as outlier, and for standard deviation, 2 or 1.96 is mainly used. However, the reflectance NIR spectra obtained from spectroscopic analysis are multivariate data with a variety of wavelength, so that mahalanobis distance (Dagnelie, 1975) is used for selecting outlier. Multivariate outliers detection method must use principle component scores (PC) to measure a distance in PC-Space, and outlier is selected by cut off value, and its criteria are based on the ones suggested by Shenk and Westerhaus (1993).

Ⅲ. RESULT AND DISCUSSION

1. NIR Spectra characteristics

The NIR region of the electromagnetic spectrum lies between the visible and infrared regions, and is usually defined by the wavelength range 700~3,000 nm (Norris, 1989). However, the most analytical use of NIR is between 1,100 and 2,500 nm. Reflectance NIR spectra are essentially plots of reciprocal $\log_{10}$ reflectance (log 1/R) versus wavelength. The spectrum looks like a smooth line (Murray, 1988) but it consists of many overlapped bands. Reflectance NIR spectra of feed have brief important information about chemical constituents. Therefore, various mathematical transformations and statistical analysis are used for extracting useful chemical information of feed sample from Reflectance NIR spectra.

As shown in Fig. 1, the wavelength of the measured reflectance NIR spectra of DDGS and corn ranged between 1,100 and 2,000 nm. Two types of reflectance NIR spectra showed multiple apparent major and minor molecular absorption peaks. In the region of 1,100~1,800 nm, overtone band appeared, and near 1,800~2,000 nm, combination band appeared (Shenk et al, 1992). In the region of 1,400, 1,725, and 1,210 nm, carbon hydrogen band (C-H, lipids) appeared, and near 1,600 nm, oxygen bands suggested by Murray (1983) appeared. Near 1,510 nm, the range useful to measure protein, first overtone region appeared.

2. Robust Calibration set structuring

For a calibration set with high perfection, it is important to select proper samples and a proper sample size for a calibration set (Dardenne, 1996). When a calibration set that includes chemical and physical characteristics of the
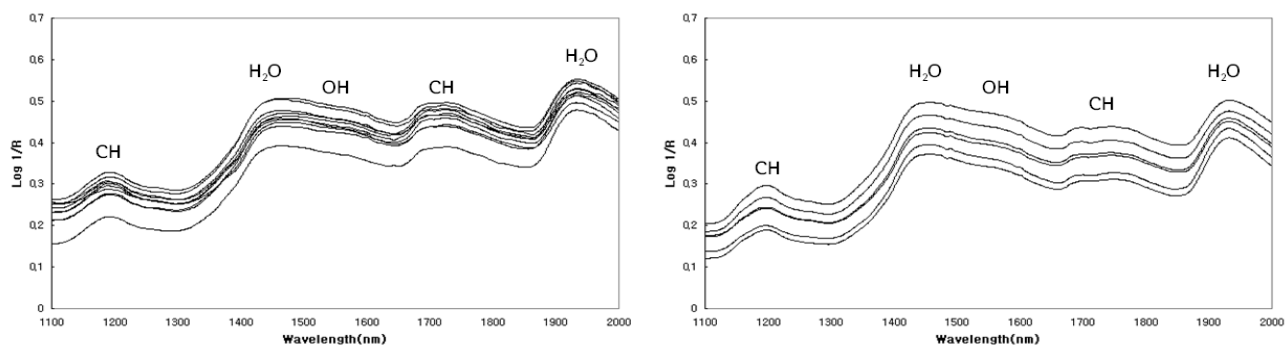
Fig. 1. NIR log 1/R spectra of DDGS and corn samples showing locations of absorptions due to major functional groups.

constituents to be measured is structured, it is possible to achieve stable and accurate prediction. To structure a calibration set, the distance based method measures mahalanobis distance (Dagnelie, 1975) between neighboring PCs in PC-space and then removes samples when outlier is selected at cut off value (>3).

Table 1 presented chemical composition distribution by constituents of DDGS. The total number of samples was 132. The samples had a wide range of chemical composition

and came from various countries, including USA, Brazil, India, China, and Australia.

Robust Calibration set (n = 89) was structured after multivariate outlier detection method was used for finding and removing outlier (Cut off value = distance H > 3) 7 samples and readjusting to a proper calibration set size.

The calibration set to write NIR calibration equations should have more physical and chemical characteristics than the validation set. Therefore, it is needed to structure samples

Table 1. Distribution of chemical compositions for DDGS used in the calibration set and robust calibration set and validation set

| Constituent | No. of Sample | Range | | Mean | SD |
		Min.	Max.		
Calibration set					
Moisture	96	6.9	12.36	10.08	1.25
Crude Protein	96	21.61	29.11	27.21	1.23
Crude fat	96	5.85	14.54	11.32	1.77
Crude fiber	96	5.84	7.49	6.75	0.43
Crude ash	96	4.1	5.6	4.53	0.34
Robust Calibration set					
Moisture	89	6.9	12.36	10.10	1.29
Crude Protein	89	21.61	29.11	27.16	1.25
Crude fat	89	5.85	14.54	11.37	1.81
Crude fiber	89	5.84	7.49	6.76	0.42
Crude ash	89	4.1	5.6	4.54	0.35
Validation set					
Moisture	36	8.38	11.94	10.10	0.78
Crude Protein	36	26.1	28.93	27.66	0.77
Crude fat	36	5.85	13.12	11.37	1.36
Crude fiber	36	6.0	7.36	6.70	0.37
Crude ash	36	3.92	5.6	4.47	0.31

USA (n=108), Brazil (n=6), India (n=6), China (n=6), Australia (n=6), n = number of samples, Min = minimum, Max = maximum, SD = standard deviation.

in diverse ways, and distribute chemical composition widely, and distribute samples evenly as shown in Fig. 2. In so doing, it is required to produce high accuracy of prediction in whole regions.

The chemical composition distribution by constituents of corn is presented in Table 2. The total number of samples was 63. Whereas DDGS mainly came from the USA, corn

mainly came from South America.

The same selection criteria with DDGS were applied so that outlier (Cut off value = distance H > 3) 2 samples were found and removed. After that, robust calibration set (n = 41) was finally structured.

As shown in Table 1 and Table 2, the number of corn samples used for calibration equations and validation fell to
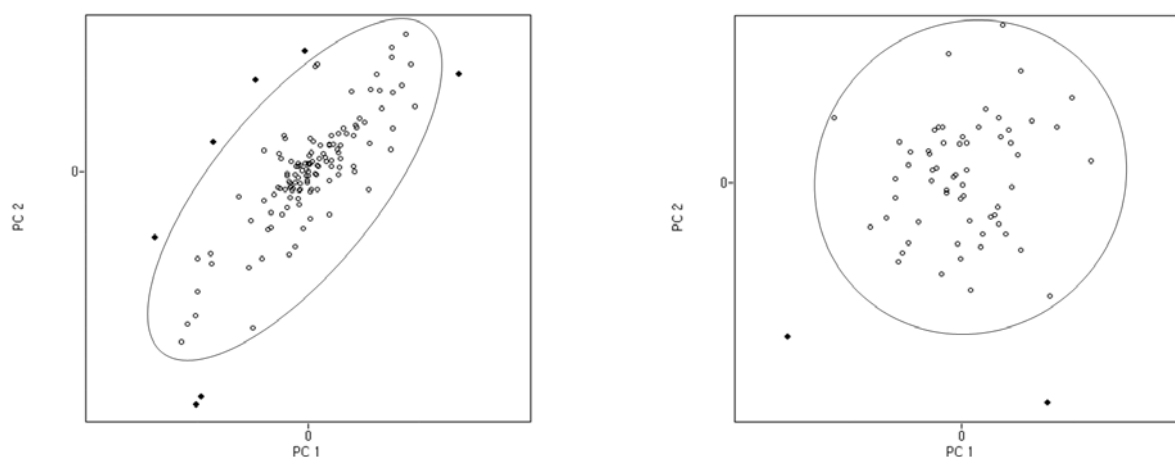

Fig. 2. Resulting outlier (dark points) according cut off value (>3) in DDGS Calibration set (left figure) and Corn Calibration set (right figure).

Table 2. Distribution of chemical compositions for corn used in the calibration set and robust calibration set and validation set

Constituent	No. of Sample	Range		Mean	SD
		Min.	Max.		
Calibration set					
Moisture	43	12.88	14.54	13.53	0.52
Crude Protein	43	6.95	8.05	7.43	0.34
Crude fat	43	3.3	4.03	3.77	0.20
Crude fiber	43	2.03	2.58	2.22	0.15
Crude ash	43	0.96	1.24	1.11	0.09
Robust Calibration set					
Moisture	41	12.88	14.54	13.54	0.52
Crude Protein	41	6.95	8.05	7.43	0.35
Crude fat	41	3.3	4.03	3.78	0.20
Crude fiber	41	2.03	2.58	2.23	0.15
Crude ash	41	0.96	1.24	1.12	0.09
Validation set					
Moisture	20	12.88	14.54	13.47	0.49
Crude Protein	20	6.95	7.91	7.30	0.30
Crude fat	20	3.3	4.03	3.71	0.19
Crude fiber	20	2.05	2.58	2.28	0.15
Crude ash	20	0.96	1.24	1.08	0.08

USA (n=9), Brazil (n=39), Argentina (n=12), Indonesia (n=3), n = number of samples, Min = minimum, Max = maximum, SD = standard deviation.

a half of that of DDGS samples. A size of population for calibration and validation plays a critical role as much as outlier selection does (Windham et al., 1989). A large number of samples do not mean that it is always good, but, a large size means that it is highly likely to reflect various physical and chemical characteristics and therefore leads to structuring a calibration set with higher perfection for validation.

3. Prediction

For the results of laboratory chemical analysis of the reflectance NIR spectra and general constituents of forages, statistical analysis was used to write NIR calibration equations for prediction. For better prediction accuracy, multivariate outlier detection method was used for finding and removing outlier and therefore for structuring a more robust calibration set. Also, various mathematical transformations were applied to find the optimal conditions of high prediction accuracy and use them as NIR calibration equations for feed analysis using reflectance NIR system. As a method of evaluating the written NIR calibration equations, the calibration set and the validation set being comprised of unknown samples were used. And it was

defined that the higher R^2 value is, and the lower SEC, SERC, and SEP values are, the better calibration equations are.

1) The prediction of constituents of DDGS

Table 3 presented the results of evaluation of NIR calibration equations of general constituents of DDGS and the results of validation evaluation of unknown samples. The optimal mathematical transformation method is different depending on each component. According to Barnes et al. (1989), although mathematical transformation method is instrumental in writing calibration equations, it is not always necessary, or any specific method does not lead to good results. As such, for crude protein, better results were produced when mathematical transformation method was not applied. As shown in Table 3, the optimal conditions of mathematical transformation were different depending on constituents.

The calibration equations of calibration set of moisture resulted in R^2=0.82 and SEC=0.70, and the examination of validation set resulted in R^2=0.72 and SEP=0.53. The calibration equations of crude protein resulted in R^2=0.72 and SEC=0.86, and the examination of validation set resulted in R^2=0.69 and SEP=0.78. The calibration equations

Table 3. Calibration and validation statistics by PLS in DDGS

Constituent	Mathematical transformation	Calibration set (96 samples)		Validation set (36 samples)	
		R^2	SEC	R^2	SEP
Moisture	B	0.82	0.70	0.72	0.53
Crude Protein	A	0.72	0.86	0.69	0.78
Crude fat	C	0.83	0.95	0.78	0.85
Crude fiber	C	0.67	0.32	0.27	0.40
Crude ash	B	0.71	0.24	0.72	0.21
		Robust Calibration set (89 samples)		Validation set (36 samples)	
		R^2	SERC	R^2	SEP
Moisture	B	0.87	0.62	0.80	0.40
Crude Protein	A	0.73	0.88	0.71	0.76
Crude fat	C	0.87	0.86	0.80	0.83
Crude fiber	C	0.68	0.30	0.32	0.38
Crude ash	B	0.74	0.23	0.72	0.21

R^2: coefficient of determination, SEC: standard errors of calibration set, SEP: standard errors of prediction.
A: log 1/R, B: MSC log 1/R, C: SNV with Detrending log 1/R.
Calibration set: reflectance nir spectra of forage samples.
Robust Calibration set: reflectance nir spectra of forage samples using Multivariate outlier detection method.

of crude fat resulted in $R^2=0.83$ and SEC=0.95, and the examination of validation set resulted in $R^2=0.78$ and SEP=0.85. The calibration equations of crude ash resulted in $R^2=0.71$ and SEC=0.24, and the examination of validation set resulted in $R^2=0.72$ and SEP=0.21. The calibration equations of crude fiber resulted in $R^2=0.67$ and SEC=0.32, and the examination of validation set resulted in $R^2=0.27$ and SEP=0.40, which means bad result. Therefore, it was found that NIR was not appropriate for predicting crude fiber. Nevertheless, the NIR calibration equations resulted in high prediction of all constituents except crude fiber. This work wrote calibration equations with the robust calibration set structured stronger than a calibration set and examined them with a validation set. As a result, all constituents except crude fiber which failed to improve prediction accuracy brought about better results than before. In particular, the examination of moisture resulted in $R^2=0.80$, SERC=0.62, and SEP=0.40, which means an improvement in prediction accuracy. The examination of crude fat resulted in $R^2=0.80$, SERC=0.86, and SEP=0.83 so that the method brought about better prediction accuracy than the method of writing calibration equations with a calibration set.

As shown in Table 3, other constituents also showed an improvement, though not large, in prediction accuracy. This work proved that removal of outlier of a calibration set contributed to improving predicted results. Therefore, it was found that the way of structuring a population of a calibration set plays a critical role in writing good NIR calibration equations.

2) The prediction of constituents of corn

Table 4 presents the results of evaluation of NIR calibration equations of general constituents of corn and the results of validation evaluation of unknown samples. The optimal conditions of mathematical transformation were different depending on each constituent. Unlike the constituents of DDGS in Table 3, all constituents of corn needed mathematical transformation. The calibration equations of moisture resulted in $R^2=0.84$ and SEC=0.27, and the validation evaluation resulted in $R^2=0.70$ and SEP=0.45. The calibration equations of crude protein resulted in $R^2=0.68$ and SEC=0.25, and the validation evaluation resulted in $R^2=0.55$ and SEP=0.29, which means somewhat low accuracy of prediction. The calibration equations of crude fat resulted in $R^2=0.82$ and SEC=0.10, and validation evaluation of $R^2=0.68$ and SEP=0.15. The calibration equations of crude fiber resulted in $R^2=0.73$ and SEC=0.10, and the validation evaluation resulted in $R^2=0.58$ and SEP=0.12. The the validation evaluation resulted in crude ash resulted in $R^2=0.87$ and SEC=0.044, and the validation

Table 4. Calibration and validation statistics by PLS in corn

Constituent	Mathematical transformation	Calibration set (43 samples)		Validation set (20 samples)	
		R^2	SEC	R^2	SEP
Moisture	C	0.84	0.27	0.70	0.45
Crude protein	B	0.68	0.25	0.55	0.29
Crude fat	B	0.82	0.10	0.68	0.15
Crude fiber	C	0.73	0.10	0.58	0.12
Crude ash	C	0.87	0.044	0.74	0.061
		Robust Calibration set (41 samples)		Validation set (20 samples)	
		R^2	SERC	R^2	SEP
Moisture	B	0.81	0.29	0.79	0.47
Crude protein	C	0.86	0.17	0.61	0.28
Crude fat	C	0.90	0.08	0.79	0.13
Crude fiber	C	0.81	0.08	0.63	0.11
Crude ash	C	0.87	0.045	0.75	0.062

R^2: coefficient of determination, SEC: standard errors of calibration set, SEP: standard errors of prediction.
A: log 1/R, B: MSC log 1/R, C: SNV with Detrending log 1/R.
Calibration set: reflectance nir spectra of forage samples.
Robust Calibration set: reflectance nir spectra of forage samples using Multivariate outlier detection method.

evaluation resulted in R^2=0.74 and SEP=0.061. All constituents but crude protein were predictable through NIR system, but prediction accuracy was somewhat low.

It was found that NIR calibration equations with a robust calibration set resulted in better prediction accuracy than NIR calibration equations with a calibration set. Each prediction accuracy of moisture (R^2=0.79, SEP=0.47) and crude fat (R^2=0.79, SEP=0.13) was improved better than the prediction accuracy of other constituents. Although outlier samples were small, it was found that they served an important role to structure a robust and highly perfect calibration set. However, compared to prediction accuracy of constituents of DDGS, their prediction accuracy was somewhat low. This is attributable to a lack of a population of a calibration set, caused by a small number of samples. Therefore, this work showed that it would be necessary to use multiple samples with various characteristics to structure a calibration set.

These results indicate that NIRS can allow prediction of most chemical compositions of DDGS and corn. Further work will be necessary to explore the significance of calibration set structuring and to improve predicted results using mathematical transformation. Further research is also warranted to test NIRS for the prediction of chemical compositions of forage and roughage.

Ⅳ. ACKNOWLEDGEMENT

Ⅴ. REFERENCES

AOAC. 1995. Official methods of analysis (16th ed.) Association of Official Analytical Chemist, Washington DC.

Cowe, I.A. and McNicol, J.W. 1985. The use of principal components in the analysis of near infrared spectra. Applied Spectroscopy. 39:257-266.

Deaville, E.R. and Flinn, P.C. 2000. Near-infrared spectroscopy: an Alternative approach for the estimation of forage quality and voluntary intake. CAB International 2000. Forage Evaluation in Ruminant Nutrition. UK. pp. 301-320.

Filzmoser, P. 2004. A multivariate outlier detection method, in Proc. of the Seventh International Conference on Computer Data Analysis and Modeling. Minsk. Belarus. pp. 18-22.

Jocelyne, A., Graviou, D., Demarquilly, C., Perez, J.M. and Andrieu, J. 1996. Near-infrared reflectance spectroscopy to predict energy value of compound feeds for swine and ruminants. Animal Feed Science and Technology. 62:77-90.

Kurt, V. and Filzmoser, P. 2009. Introduction to Mutivariate Statistical Analysis in Chemometrics (1st ed.). CRC Press. pp. 103-190.

Kays, S.E. and Franklin, E.B. 2002. Rapid prediction of gross energy and utilizable energy in cereal food products using near-infrared reflectance spectroscopy. Journal of Agricultural and Food Chemistry. 50:1284-1289.

Kays, S.E., Douglas, D.A. and Sohn, M.Y. 2005. Prediction of fat in intact cereal food products using near-infrared reflectance spectroscopy. Journal of the science of food and agriculture. 85:1596-1602.

Marten, G.C., Shenk, J.S. and Barton, F.E. 1989. Near infrared reflectance analysis of forage quality. USDA Agriculture Handbook. No 643, US Government. Washington DC, USA.

Marco, R., Atkinson, A. C. and Cerioli, A. 2009. Finding an unknown number of multivariate outliers. Journal of the Royal Statistical Society. B. 71(2):447-466.

Norris, K.H., Barnes, R.E.F., Moore, J.E. and Shenk, J.S. 1976. Predicting forages quality by infrared reflectance spectroscopy. Journal of Animal Science. 43:889-897.

Park, H.S., Lee, J.K., Lee, H.W., Kim, S.G. and Ha, J.K. 2006. Prediction of the digestibility and energy value of corn silage by Near Infrared Reflectance Spectroscopy. Journal of the Korean Society of Grassland and Forage Science. 26(1):45-52.

Paul, G. 2006. Practical Guide to Chemometrics. 2nd ed, CRC Press. London. pp. 168-211.

De Maesschalck, R., Jouan-Rimbaud, D. and Massart, D.L. 2000. The Mahalanobis distance. Chemometrics and Intelligent Laboratory Systems. 50:1-18.

Shenk, J.S. and Westerhaus, M.O. 1991. Population definition, sample selection, and calibration procedures for near infrared reflectance spectroscopy. Crop Science. 31(2):469-474.

Westad, F. and Martens, H. 2000. Variable selection in near infrared spectroscopy based on significance testing in partial least squares regression. Journal of Near Infrared Spectroscopy. 8:117-124.