Evaluation of *Leymus chinensis* quality using near-infrared reflectance spectroscopy with three different statistical analyses

Jishan Chen, Ruifen Zhu, Ruixuan Xu, Wenjun Zhang, Yue Shen, Yingjun Zhang

Due to a boom in the dairy industry in Northeast, China, the hay industry has been developing rapidly. Thus, it is very important to evaluate the hay quality with a rapid and accurate method. In this research, a novel technique that combines near infrared spectroscopy (NIRs) with three different statistical analyses (MLR, PCR and PLS) was used to predict the chemical quality of sheepgrass (Leymus chinensis) in Heilongjiang Province, China including the concentrations of crude protein (CP), acid detergent fiber (ADF), and neutral detergent fiber (NDF). Firstly, the linear partial least squares regression (PLS) was performed on the spectra and the predictions were compared to those with laboratory-based recorded spectra. Then, the MLR evaluation method for CP has a potential to be used for industry requirements, as it needs less sophisticated and cheaper instrumentation using only a few wavelengths. Results show that in terms of CP, ADF and NDF, (i) the prediction accuracy in terms of CP, ADF and NDF using PLS was obviously improved compared to PCR algorithm, and comparable or even better than results generated using MLR algorithm; (ii) the predictions were worse compared to laboratory-based spectra with MLR algorithmin, and poor predictions were obtained (R2 , 0.62, RPD,0.9) using MLR in terms of NDF; (iii) a satisfactory accuracy with R2 and RPD by PLS method of 0.91, 3.2 for CP, 0.89, 3.1 for ADF and 0.88, 3.0 for NDF, respectively, was obtained. Our results highlight the use of the combined NIRs-PLS method could be applied as a valuable technique to rapidly and accurately evaluate the quality of sheepgrass hay.

Evaluation of *Leymus chinensis* quality using near-infrared reflectance spectroscopy with three different statistical analyses

Jishan Chen[1,2], Ruifen Zhu[2], Ruixuan Xu[1], Zhulin Xue[1], Wenjun Zhang[1], Yue Shen[1], Yingjun Zhang[1*]

1 Department of Grassland Science, China Agricultural University, Beijing 100193, China; 2 Institute of Pratacultural Science, Heilongjiang Academy of Agricultural Science, Harbin, Heilongjiang 150086, China;

[*]Corresponding author: Y. Zhang, zhangyj@cau.edu.cn; Tel & fax: 86-10-62733380.

ABSTRACT

Due to a boom in the dairy industry in Northeast, China, the hay industry has been developing rapidly. Thus, it is very important to evaluate the hay quality with a rapid and accurate method. In this research, a novel technique that combines near infrared spectroscopy (NIRs) with three different statistical analyses (MLR, PCR and PLS) was used to predict the chemical quality of sheepgrass (*Leymus chinensis*) in Heilongjiang Province, China including the concentrations of crude protein (CP), acid detergent fiber (ADF), and neutral detergent fiber (NDF).

Firstly, the linear partial least squares regression (PLS) was performed on the spectra and the predictions were compared to those with laboratory-based recorded spectra. Then, the MLR evaluation method for CP has a potential to be used for industry requirements, as it needs less sophisticated and cheaper instrumentation using only a few wavelengths. Results show that in terms of CP, ADF and NDF, (i) the prediction accuracy in terms of CP, ADF and NDF using PLS was obviously improved compared to PCR algorithm, and comparable or even better than results generated using MLR algorithm; (ii) the predictions were worse compared to laboratory-based spectra with MLR algorithmin, and poor predictions were obtained (R^2, 0.62, RPD, 0.9) using MLR in terms of NDF; (iii) a satisfactory accuracy with R^2 and RPD by PLS method of 0.91, 3.2 for CP, 0.89, 3.1 for ADF and 0.88, 3.0 for NDF, respectively, was obtained. Our results highlight the use of the combined NIRs-PLS method could be applied as a valuable technique to rapidly and accurately evaluate the quality of sheepgrass hay.

53
54
55
56
57
58
59

INTRODUCTION

As an important perennial forage grass across the Eurasian Steppe, sheepgrass [*Leymus chinensis* (Trin.) Tzvel.] is known for its adaptability to various environmental conditions (*Chen et al., 2013*), high palatability, rich in nutrients and all kinds of livestock feeding (*Liu et al., 2004; Yang et al., 1995*). *Leymus chinensis* has tender stems and leaves, and high forage yield without irrigation about 3,000 to 4,500 kg/hm^2, and the yield with irrigation reaches 6,000 kg/ hm^2 in the Northeastern Plain and east of the Inner Mongolian Plateau (*Chen, 2001*), and it may contribute to good balanced diet for cows fed with sheepgrass hay on milk production and composition (*Yan et al., 2011*).

In the last 20 years, the population of Northeast China increased considerably, resulting in steeply increase in numbers of livestock. Hence, sheepgrass in this region is regarded as an important productive grass for the hay industry developing rapidly due to a prosperous status in the dairy industry. A great deal of animal farms with approximate 20,000 sheeps or 10,000 cattles have been or are being established in Northeast China, resulting in an urgent need for forage including commercial forage and natural herbage. Annually, more than 30 companies produce hay over 117.19 million tons of natural herbage from sheepgrass in Heilongjiang Province, China (http://www.caaa.cn/association/grass/). Consequently, a large number of sheepgrass hay produced by personal goes into the market and become a commodity. The difference between commercial forage and natural herbage dominanted by sheepgrass is that the former is attached a detailed trademark to the hay productors entered into the market, while the latter is absence of the quality indices of natural herbage dominanted by sheepgrass during selling. Nevertheless, sheepgrass is one of important hay in everyday life of animal, and the final purchase decision by buyer is often according to the feeding value, which is well related with terms of chemical and biological components. Because of the impact of factors such maturity period at harvest, botanical components, and cutting techniques on the production process in different climate, the principal difference between natural herbage and commercial forage

is the complexity of raw materials of the former (dominanted by sheepgrass), and measurements of internal quality indices of sheepgrass hay are yet costing time and destructive. Therefore, the establishment of a rapid and accurate method with nondestructive for evaluate chemical qualities of sheepgrass hay is extremely important to the hay industry before selling.

Assessing the quality of natural herbage dominanted by sheepgrass is very important for high quality forage with various parameters, namely the concentrations of crude protein (CP) , acid detergent fiber (ADF) and neutral detergent fiber (NDF), which are commonly used to assess the forage quality. However, all analytical procedures are time-consuming or expensive when a great quantity of samples are involved (*Yu et al., 2010; Yu et al., 2011; Albrecht et al., 2009*). Near infrared reflectance spectroscopy (NIRs) with wavelength range of 750-2500 nm has much superiority over chemical analyses in a laboratory, for example its ease of sample preparation, rapid spectrum acquisition, non-destructive nature of the analysis, and the portability of the technology (*Albrecht et al., 2009*). Many works have proved that NIRs is widely used to assess the quality of forage dominanted by alfalfa, oats, silage corn, ryegrass and so on (*Nie et al., 2007; Ding et al., 2009; En et al., 2009; Liu et al., 2014; Hu et al., 2014; Bai et al., 2004; Chen et al., 2007*), and it is also being implemented in sheepgrass (*Shi et al., 2011*). However, it remains unclear whether NIRs is applied to search the valuable information about NIRs prediction models and rapidly and accurately assess the quality of the hay dominanted by sheepgrass for CP, ADF and NDF.

In addition, the spectrum of sheepgrass hay obviously shows some peaks and valleys in the wavelengths from 950 to 1650 nm, which includes hidden information of different components, and this does not mean that some useful information cannot be extracted in other wavelengths (*Liu et al., 2008*). To predict and determine the quality parameters of sheepgrass hay, multivariate statistics analysis techniques, such as multiple-linear regression (MLR), principal component regression (PCR) and partial least squares regression (PLS), are applied to establish the prediction models by analyzing correlations between measured chemical values and the spectrum measurements of sheepgrass hay in this study.

The objective of this study was to search the prediction models of NIRs to determine the essential quality indices of sheepgrass hay. A total of 203 samples of sheepgrass hay were collected from 37 sampling sites distributed throughout Heilongjiang Province, China. Our purpose of this study were to evaluate the performance of NIRs in measuring CP, ADF, and NDF of sheepgrass hay, and compare the prediction potentiality of different methods (MLR, PCR and PLS) for rapidly and accurately evaluate the quality of sheepgrass hay.

MATERIALS AND METHODS

Sample collection and pretreatment

The sampling sites were from the west to east across the grassland in Heilongjiang Province of Northeast China. A total of 203 samples of sheepgrass hay were randomly collected from sheepgrass fields of the hay factories in 2013, with latitudes ranging from 44.475° N to 51.728° N, longitudes from 123.209° E to 132.944° E. The sites were chosen to be representative of sheepgrass production fields and contain a range of soils and climate are described in **Appendix S1**. These locations produce approximately 117.19 million tons of sheepgrass hay per year. All samples of sheepgrass hay were sampled at blooming stages, identified and collected before being clipping and packaging. The collected samples (not involve endangered or protected species), a total sample size of 203 from 21 sampling sites distributing over 8 regions (not privately owned or protected in any way) in Heilongjiang Province (see Fig. 1 and Appendix S1), were then forwarded to the lab and stored at 4°C for further analyses.

To be representative, each sample consisted of one quarter square meter clipped at 4 cm, transported to the lab, oven dried (65°C, 48 h), ground (1 mm sieve), and mixed. Meanwhile, the mixed samples were divided by the quartile method into half for duplicate chemical analyses and another half for the near infrared reflectance spectra.

In fact, all samples (203) were used to evaluate and develop NIRs models in this study. To ensure the adaptability of the calibration models, some samples (51) were used for prediction set and the rest of samples (152) for calibration set. All tasks including spectral measure and chemical analyses were finished on the same day or the next day.

Chemical properties analyses

The first experiment was designed to develop a database to evaluate relationships between the quality indices of sheepgrass hay and NIRs measurements. In order to accomplish this goal, three characteristics were measured at blooming stages of growth. These included CP, ADF, and NDF, which were usually regarded as the principal forage quality paremeters, serving as the primary nutrition source in the diet of dairy cattle (*Yan et al., 2011; Mertens, 1992*). All chemical analyses were conducted in duplicate using analytical grade chemicals are described in **Appendix S2**. The CP was quantified using the Kjeldahl procedure for nitrogen (N) determination (% CP = % N × 6.25) by the methods of Association of Official Analytical Chemists-AOAC (*AOAC, 1990*). The ADF and NDF were conducted by using the method described by Van Soest *et al*. (*Van Soest et al., 1991*).

137

138 **Spectroscopic measurement**

139 By using an near infrared spectrometer (DA7200, Perten Corporation, Sweden) to collect for the spectra of samples,
140 the NIR spectra were recorded at 5 nm intervals from 950 to 1650 nm in **Appendix S2**. Approximately 50 g of dried
141 sample was scanned in a 7.5 cm diameter sample cell with a quartz window in the room temperature maintained at
142 25°C. With the working parameters of DA7200 software, samples were scanned from 950 to1650 nm in reflectance
143 mode (R mode). Data form was converted to the absorbance (the logarithms of the reflectance reciprocal). In order
144 to eliminate an error caused by loading the sample, each sample were repeated to scan three times and each time
145 with scanning background light three times. The average spectral scanning were as for the final spectra of samples
146 deposited in the computer for calculating using the Grams32 software (Perten Corporation, Sweden).

147

148 **Statistical Analysis**

149 To obtain linear correlations of the NIRs with the chemical values, spectra were expressed as the absorbance A
150 (A=log1/R). The NIRs were conducted by Grams32 software and Microsoft Office Excel 2003. In performing
151 measurements with the NIR System DA7200 spectrometer, a number of data were generated for each sample and it
152 is obvious that some data reduction method is needed to facilitate data interpretation. Thus, there is a need for data
153 processing methods that transform the measured spectral data into the sample properties of interest. For producing
154 such models that determine the equations describing the relationship between spectral data and chemical values,
155 three different methods (MLR, PCR and PLS) were used. MLR uses a multiterm linear polynomial to describe this
156 relationship using only several spectral data measured as “characteristic” wavelengths (by the way of correlation
157 spectrum). With PCR, all spectral data are used, reducing the dimensionality of the data set by looking for
158 orthogonal directions in spectral data space along which the variance of the data set is maximised. These directions
159 are called principal components (PC). Thus, the first principal component (PC 1) is determined as the direction in
160 spectral data space that corresponds to the largest variation in the data set. The second principal component (PC 2) is
161 calculated as the direction perpendicular to PC 1 along which the remaining variation in the data set is the largest.
162 This procedure is repeated until no variation is left in the data set. PLS uses not only the spectral information in the
163 data set but also incorporates information about sample properties, e.g. concentrations to determine the most useful
164 orthogonal directions in spectral space.

An ideal equation would have a coefficient of determination (R^2) of 1.0 and root mean squared error (RMSE). Approximately 75% of the data were used in creating the equation, while the remaining 25% were randomly removed for validation. Validation procedures for the created equations were the same as described above, using the remaining 25% of the data. In this study, the calibration set (152 samples) and the validation set (51 samples) were randomly divided by the method applied by several publications (Miriam et al., 2011; *Peltre et al., 2011; Hayes, 2012*). To restrain invariable background signals and to improve the visual resolution, the Savitsky-Golay 2nd order derivations were also applied in this study. The selection of the models developed was largely dependent upon determination coefficient for the calibration set (R^2) and validation set (r^2), root mean squares error of calibration (RMSEC), the RPD (ratio of standard error of performance to standard deviation) and root mean squares error of prediction (RMSEP) used to evaluate the practicability of NIRs for determining to the quality indicators of sheepgrass hay.

The formula for RMSEC and RMSEP are described briefly below:

$$\text{RMSEC}=\sqrt{\frac{1}{n_c}\sum_{i=1}^{n_c}\left(\overline{y_i}-y_i\right)^2}$$

$$\text{RMSEP}=\sqrt{\frac{1}{n_p}\sum_{i=1}^{n_p}\left(\overline{y_i}-y_i\right)^2}$$

Where $\overline{y_i}$ is the predicted value of the i-th observation; y_i is the measured value of the i-th observation; n_c is the number of observations in calibration set; n_p is the number of observations in validation set.

RESULTS

Chemical value

The obtained data (chemical value) were calculated and analyzed by the min, max, average, and standard deviation (SD), while the results of the calibration sets and validation sets were summarized in Table 1. These data sets indicated a wide variability in the chemical indices. For all the total sheepgrass sampled, the CP of 203 samples ranged from 6.2% to 14.33% (SD = 1.24), while the ADF ranged from 35.13% to 42.34% (SD = 2.33) and NDF from 50.71% to 71.08% (SD = 2.67). The datum of CP and NDF were nearly normally distributed around the

190 average (10.54% and 60.89%) but the distribution of ADF measurements was skewed to the right, with the right tail
191 stretching further than the left tail (mean=38.74%).
192 This wide heteromorphosis in the quality indices of the sheepgrass hay was beneficial to successfully establish a
193 relationship between the NIRs and the quality indices (*Yu et al., 2010*). Outliers in the collected sample were
194 considered and thus excluded during the process of calibration. The elimination of outliers was based on the
195 criterion that if mahalanobis distance for a sample was 3 SD or more (*Yang et al., 2006*). For the value of chemical
196 indices in this study, 2 outliers were removed for CP, 8 for ADF, and 1 for NDF (Table 1).
197
198 **NIR spectra**
199 Some peaks and valleys were obviously shown in the spectra (Fig. 2), which represent the characteristics of
200 sheepgrass hay including hidden information of different components and its quantities. All the near infrared
201 reflectance spectra of the collected sample were separated the two groups of spectra with different slopes, one group
202 displayed an obvious absorbance peak at wavelengths of nearly 1, 100 nm, another group showed significant
203 spectral peak was at approximately1, 450 nm (Fig. 2). From the NIRs of the total sheepgrass samples (N=203), the
204 absorbance peaks were very overlapped, primary reason that the spectrum includes combinations and overtones of
205 vibration such as stretching and bending of hydrogen-bearing functional groups such as –CH, –OH, and –NH (*Saeys*
206 *et al., 2005*). Meanwhile, other interference information affects the near infrared reflectance spectra (*Meissl et al.,*
207 *2008*). Therefore, the quantitative evaluation are of difficulty through NIRs alone. Multivariate methods might be
208 necessary to analyse the response of quality of sheepgrass from spectral characteristics with the support of
209 chemometric methods, e.g., PLS, PCR and MLR analysis (*Meissl et al., 2008; Saeys et al., 2005*).
210
211 **Multivariate calibration analysis**
212 For quantitative evaluation methods multiple linear regression (MLR), principal component regression (PCR) and
213 partial least squares regression (PLS) were used to for data modelling and for predicting the investigated chemical
214 constituents. For model validation, full cross-validation was used; each case was predicted by the model derived
215 from all other remaining spectra. The results of full cross-validation is the average of the standard error of prediction
216 values produced during each cross-validation step. The optimum number of Orthogonal factors (PC) for the models
217 was obtained by using the leave-one-out cross validation technique for the calibration set. The recorded log (1/R)

spectra were smoothed and transformed to second derivative before the analysis. The optimum models were achieved in the Table 2 for the spectral processing method of CP, ADF, and NDF. PLS and PCR methods are based on the regression of the full spectra while MLR is based on discrete parts of the spectra (2 wavelengths) in this study. The table 2 showed that the comparison of the accuracy of the calibration results achieved by near infrared spectroscopy using PLS, PCR, and MLR methods, respectively, for evaluation the chemical indices of samples selected.

The results of the calibration procedures are presented in the Table 3. It can be seen in the table that the smallest value for prediction error (RMSEC) with the highest coefficient of determination for the calibration set (R^2) was provided by the PLS method. The results were less accurate for CP, NDF and ADF provided by the PCR method. The MLR had the worst prediction. Therefore, predictions were moderately successful for CP, NDF, and ADF was provided by the PLS method with second derivative of log(1/R) spectra.

The results of the cross validation are presented in the Table 4. It can be seen in the table that the smallest value for prediction error (RMSEP value for CP, NDF and ADF is 1.24, 1.41 and 1.37, respectively) with the highest the coefficient of determination for the validation set (r^2 value for CP, NDF and ADF is 0.91, 0.90 and 0.90, respectively) was provided by the PLS method. The results were less accurate for CP (RMSEP=2.63, r^2=0.86), NDF (RMSEP=4.50, r^2=0.82) and ADF (RMSEP=5.86, r^2=0.82) provided by the PCR method. The MLR had the worst prediction for CP (RMSEP=6.74, r^2=0.74), NDF (RMSEP=8.59, r^2=0.62) and ADF (RMSEP=9.91, r^2=0.60). Therefore, predictions were moderately successful for CP, NDF, and ADF was provided by the PLS method. The application of PCR using the whole wavelength region requiring scanning type spectrometers as with PLS, resulted in an almost twice as high prediction error. The same order of magnitude could be obtained for RMSEP value by using MLR. Compared to PCR and PLS the MLR model uses only two characteristic wavelengths for the calculation, creating the conditions for constructing cheaper single-purpose filter instruments.

Similarly as RPD determined by the above three methods in the Table 4, the accuracy was considered good for RPD > 2, acceptable for 1.4 < RPD < 2, and unreliable for RPD < 1.4 according to Albrecht (*Albrecht et al., 2009*) and Chang *et al*. (*Chang et al., 2001*). So the highest RPD was provided by the PLS method for CP (RPD =3.2), NDF(RPD =3.0), and ADF(RPD =3.1) while the worst RPD provided by the PCR and MLR method are also shown in Table 4.

DISCUSSION

Samples selected for calibration (Table 1) comprised a wide range from 6.20 to 14.33% for crude protein (CP), from 50.71 to 71.08% for neutral detergent fiber (NDF) and from 35.13 to 42.34 for acid detergent fiber (ADF). This is confirmed by a certain instance in our study as an obvious difference among samples from the different distributing sites is the wide variability of sheepgrass. Hence, obtaining good relationship was more difficult for sheepgrass employed in this study.

Because the near-infrared spectrum contains all strength information of the chemical bond, chemical composition, electronegativity, *etc*. Meanwhile, other interference information, such as scattering, diffusion, special reflection, surface gloss, refractive index, and reflected light polarization, affects the near-infrared spectrum. Thus, to eliminate this effects of the later, there are three types of spectral normalization method. min/max normalization, vector normalization and zero correction. One common method such as the vector normalization was applied. Meanwhile, to restrain invariable background signals and to improve the visual resolution, the Savitsky-Golay 2nd order derivations were also applied in this study.

PLS regression analysis of the spectral data with CP content resulted in the highest coefficient of determination (R^2) compared with PCR and MLR method (Table 3). Also, differences in standard error of full cross-validation (RMSEC) for CP prediction show the RMSEC achieved by PLS (0.74%) was more than 20% lower compared with PCR (1.63%) or MLR methods (5.74%). The RMSEC for CP of 0.74 gained by PLS is comparable with the range of 0.4-1.41 reported by Shi *et al*. (*Shi et al., 2011*). There were not about PCR and MLR methods, respectively. The two wavelengths chosen for calibration by the MLR method were 1366 and 1418 nm for CP; 1142 and 1380 nm for NDF, and 1148 and 1566 for ADF. They are not within the range of 2100-2164 quoted by Redshaw *et al*. (*Redshaw et al., 1986*) and Murray (*Murray, 1986*). The MLR evaluation method for CP indicates a strong relationship (R^2 of 0.85) with spectroscopic data, which suggests that this trait may be also accurately estimated by inexpensive filter instruments. Several authors compared results obtained by MLR and PLS in different products and concluded PLS and MLR give nearly the same prediction errors (*Windham et al., 1992; Rosental et al., 1996*). In our study, PLS was more accurate than MLR method. One of the reasons might be that 9 components were used for PLS compared with 5 PC used for PCR and 2 wavelengths used for MLR.

In our paper the RMSEP values were all higher than RMSEC results, similar to the ones reported by Stimson *et al.* (*Stimson et al., 1991*). The lowest RMSEP achieved for CP content among the methods applied may be due to the

fact that CP content was recalculated to the dry weight while CP was expressed on DM basis. Differences between RMSEC and RMSEP values may be due to the limited number of samples. This clearly indicates the problem of obtaining representative samples in practice. Also, sheepgrass samples, which are very heterogeneous, were scanned with pretreatment while for wet chemistry analysis the same samples were dried and ground. Although errors are often slightly higher for samples scanned in their natural state than for dried and milled samples, this is balanced by the ability to scan much larger samples, the avoidance of compositional losses and changes due to oven drying and a major reduction in analysis time and cost due to no sample preparation being necessary (*Wilman et al., 2000*).
The findings indicate spectroscopic data evaluated by PLS method were strongly related to reference values and had lower RMSEC values and highest RPD compared with PCR and MLR methods. By studying the figure 3, 4 and 5, the same conclusion can be drawn as in case of the prediction of CP, ADF and NDFcontent by PLS with the highest correlations of determination (r^2) for the validation set (n=51). Our results in this study indicated for the first time that the quality of sheepgrass hay could be successfully evaluated by the NIRs with PLS regression method.

CONCLUSION

The NIRs prediction of forage chemical value is a relatively inexpensive, rapid and reliable method compared with reference methods, requiring a relatively small quantity of sample and predicts several concentrations of components simultaneously. An important advantage of NIRs is its ability to analyze samples without chemical treatment, hence costs and chemical wastes can be reduced by using accurate NIRs model. The success of NIRs analysis depends almost entirely on the reliability of the primary reference data used at calibration. Using NIRs can predict CP, NDF, and ADF contents in sheepgrass hay samples without costly or lengthy pretreatment as shown in this paper. Nevertheless, a satisfactory accuracy with an average standard error of prediction by PLS method of 0.74, 1.31 and 1.25 for CP, NDF and ADF, respectively, was obtained. The comparison of validation statistics (r^2 and RMSEC) among PLS, PCR and MLR equations showed PLS to be the most accurate (However, it does use 9 factors compared with 5 for PCA and 2 for MLR.). The MLR evaluation method for CP has the potential to be used for industry requirements, as it needs less sophisticated and cheaper instrumentation using only a few wavelengths.

Acknowledgments

The authors thank Prof. Wu jingzhu, Li lujun and Yang Gaowen for their assistance in modification, and Liu Qiqi and Wu shuheng for the help with the laboratory-based NIR measurements

Supporting Information

Appendix S1 Summary of sample sites and its properties in Heilongjiang Province, China. (XLS)

Appendix S2 The measured spectra of NIR of a total 203 sheepgrass samples. To every fifth wavelength was retained to reduce the size of the file. (XLS)

REFERENCES

Chen SY, H X,Yan XQ , Liang Y, Wang YZ, Li XF. 2013. Transcriptome analysis in sheepgrass (*Leymus chinensis*): A dominant perennial grass of the Eurasian Steppe. *PLoS One* **8**: e67974 DOI 10.1371/journal.pone.0067974.

Liu GS, Qi DM. 2004. Research progress on the biology of *Leymus chinensis*. *Acta Prataculturae Sinica* **13**: 6–11.

Yang YF, Liu GC, Zhang BT. 1995. An analysis of age structure and the strategy for asexual propagation of *Aneurolepidium chinense* (Trin.) Kitag population. *Acta Botanical Sinica* **37**: 147–153.

Chen BS. 2001. Forage crop cultivation. Pages 449–454 in cultivation of forage grasses and forage crops. 1st ed. China Agriculture Science and Technology Press, Beijing, China.

Yan R, Chen S, Zhang X, Han J, Zhang Y, Undersander D. 2011. Effects of replacing part of corn silage and alfalfa hay with *Leymus chinensis* hay on milk production and composition. *American Dairy Science Association* **94(7)**: 3605–3608 DOI 10.3168/jds.2010-3536.

Yu GH, Luo YH, Wu MJ, Tang Z, Liu DY. 2010. PARAFAC modeling of fluorescence excitation-emission spectra for rapid assessment of compost maturity. *Bioresour Technol* **101**: 8244–8251 DOI 10.1016/j.biortech.2010.06.007.

Yu GH, Wu MJ, Luo YH, Yang XM, Ran W. 2011. Fluorescence excitation-emission spectroscopy with regional integration analysis for assessment of compost maturity. *Waste Manage* **31**: 1729–1736 DOI 10.1016/j.wasman.2010.10.031.

Albrecht R, Joffre R, Le Petit J, Terrom G, Pe´rissol C. 2009. Calibration of chemical and biological changes in cocomposting of biowastes using nearinfrared spectroscopy. *Environ Sci Technol* **43**: 804–811.

Nie ZD, Han JG, Yu Z. 2007. Applications of near infrared reflectance spectroscopy technique (NIRS) to grassland ecology research, *Spectroscopy and Spectral Analysis* **27**:691–696.

Ding WR, Gan YM, Guo XS, Yang FY. 2009. A review on studies and applications of near infrared spectroscopy technique (NIRS) in detecting quality of hay, *Spectroscopy and Spectral Analysis* **29**:358–361.

En XZ, Guo HR, Jia YS, Ge GT, Wang K. 2009. Application and prospect of near infrared reflectance spectroscopy in forage analysis, *Spectroscopy and Spectral Analysis* **29**:635–640.

Liu W, Zhang YJ, Xu RX. 2014. Comparative study on quality of alfalfa with near infrared reflectance spectroscopy and conventional chemical assay, *Chinese Agricultural Science Bulletin* **30**:269–272.

Hu C, Bai SQ, Zhang Y, Yan JU. 2014. Near infrared reflectance spectroscopy analytic model established for the IVDMD of *Cichorium intybus* L., *Spectroscopy and Spectral Analysis* **34**:2085–2088.

Bai QL , Chen SJ , Dong XL. 2004. Prediction of NDF and ADF concentrations with near infrared reflectance spectroscopy (NIRS), *Spectroscopy and Spectral Analysis* **24** : 1345–1347.

Chen PF, Rong YP, Han JG. 2007. Evaluation of fresh sample of alfalfa silage through near infrared reflectance spectroscopy (NIRS), *Spectroscopy and Spectral Analysis* **27**:1304–1307.

Shi D, Zhang YJ. 2011. Determination of *Leymus Chinensis* Quality by Near Infrared Reflectance Spectroscopy. *Spectroscopy and Spectral Analysiss* **31**: 2730–2733.

Liu YD, Chen XM, Ouyang A. 2008. Nondestructive determination of pear internal quality indices by visible and near-infrared spectrometry, *Food Science and Technology* **41**: 1720–1725.

Mertens DR. 1992. Nonstructural and structural carbohydrates. Pages 219–235. In H.H. Van Horn and C.J. Wilcox (ed.). Large dairy herd management. Management Services, American Dairy Science Association, Champaign, IL.

AOAC.1990. Official methods of analysis. 15th ed. AOAC, Arlington, USA.

Van Soest PJ, Robertson JB, Lewis BA. 1991. Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. *Journal of Dairy Science* **74**: 3583–3597 DOI 10.3168/jds.S0022-0302(91)78551-2

Miriam F, Malcolm RB, Louise RW, Natalie AM. 2011. Predicting glycogen concentration in the foot muscle of abalone using near infrared reflectance spectroscopy (NIRS). *Food Chemistry* **126**: 1817–1820.

Peltre C, Thurie`s L, Barthe`s B, Brunet D, Morvan T. 2011. Near infrared reflectance spectroscopy: A tool to characterize the composition of different types of exogenous organic matter and their behaviour in soil. *Soil Biology and Biochemistry* **43**: 197–205 DOI 10.1016/j.soilbio.2010.09.036.

Hayes DJ. 2012. Development of near infrared spectroscopy models for the quantitative prediction of the lignocellulosic components of wet Miscanthus samples. *Bioresour Technology* **119**: 393–405 DOI 10.1016/j.biortech.2012.05.137.

Yang ZL, Han LJ, Fan X. 2006. Rapidly estimating nutrient contents of fattening pig manure from floor scrapings by near infrared reflectance spectroscopy. *Journal of near infrared spectroscopy* **14**: 261–268.

Meissl K, Smidt E, Schwanninger M, Tintner J. 2008. Determination of humic acids content in composts by means of near- and mid-infrared spectroscopy and partial least squares regression models. *Applied Spectroscopy* **62**: 873–880. DOI 10.1366/000370208785284277

Saeys W, Mouazen AM, Ramon H. 2005. Potential for onsite and online analysis of pig manure using visible and near infrared reflectance spectroscopy. *Biosystems Engineering* **91**: 393–402 DOI 10.1016/j.biosystemseng.2005.05.001

Chang CW, Laird DA, Mausbach MJ, Hurburgh CR. 2001. Near-infrared spectroscopy-principal components regression analyses of soil properties. *Soil Science of American Journal* **65**: 480–490 DOI 10.1007/s00374-007-0192-z.

Murray I. 1986. Near infrared reflectance analysis of forages. In: Haresign W. and Cole D.J.A. (ed). Recent Advances in Animal Nutrition, London, Butterworths.

Redshaw ES, Mathison GW, Milligan LP, Weisenburger RD. 1986. Near infrared reflectance spectroscopy for predicting forage composition and voluntary consumption and digestibility in cattle and sheep. *Canadian Journal of Animal Science* **66**:103–115.

Windham WR, Flinn PC. 1992. Bridging the gap between data analysis and NIR appliations. In: Near infrared spectroscopy, Ed. by K.I. Hildrum, T. Isaksson, T.Naes and A. Tandberg. Ellis Horwood, Chichester, UK, p. 459.

Rosental B, Wlliams P. 1996. The future waves. In: Near infrared spectroscopy. Ed. by A.M.C. Davies and P. Williams. NIR Publications, Chichester, UK, p.1.

Stimson C, Kellawy RC, Tassell RJ, Ison RL. 1991. Prediction of the nutrient content of botanical fractions from annual legumes by near infrared reflectance spectroscopy. *Grass and forage science* **46**: 99–105.

Wilman D, Field M, Lister SJ, Givens DI. 2000. The use of near infrared spectroscopy to investigate the composition of silages and the rate and extent of cell-wall degradation. *Animal feed science and technology* **88**: 139–151.

Figures

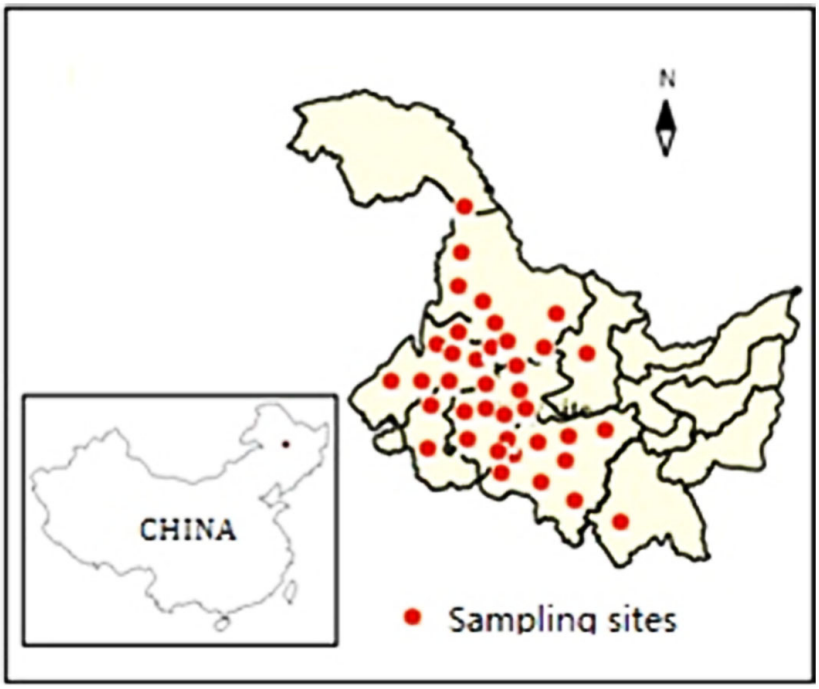


Figure 1. Summary of sampling sites distribution in Heilongjiang Province, China.

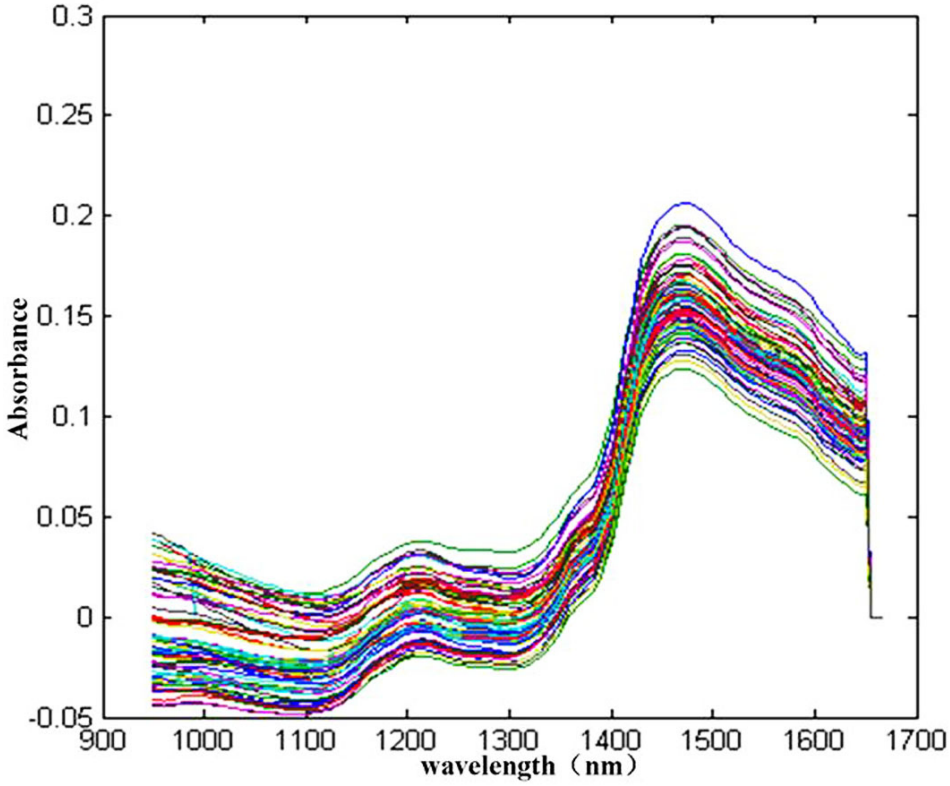


Figure 2. Spectra of NIR of a total 203 sheepgrass samples.

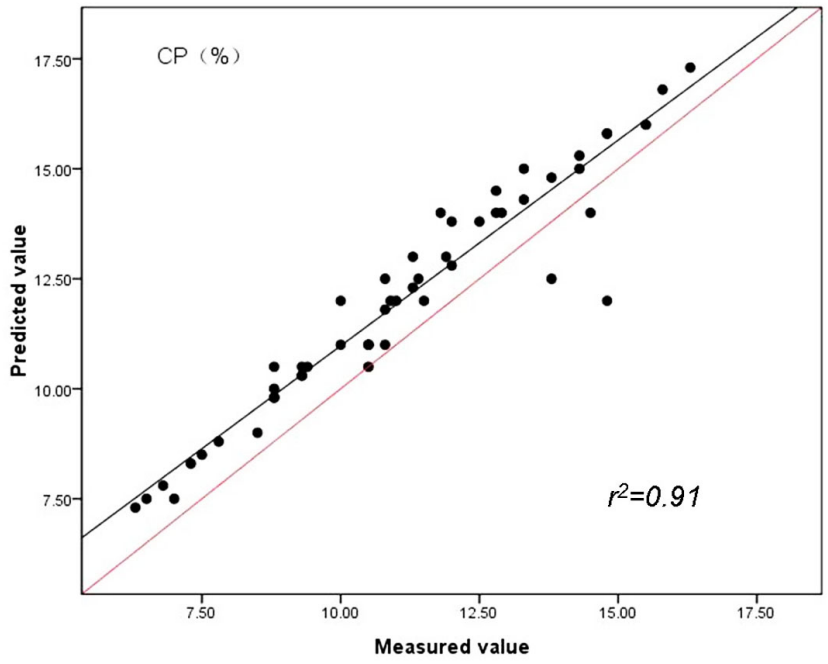


Figure 3. Relationships between the measured and predicted values of the crude protein content (CP) of sheepgrass hay for the validation data set. The red line represents the best fit.

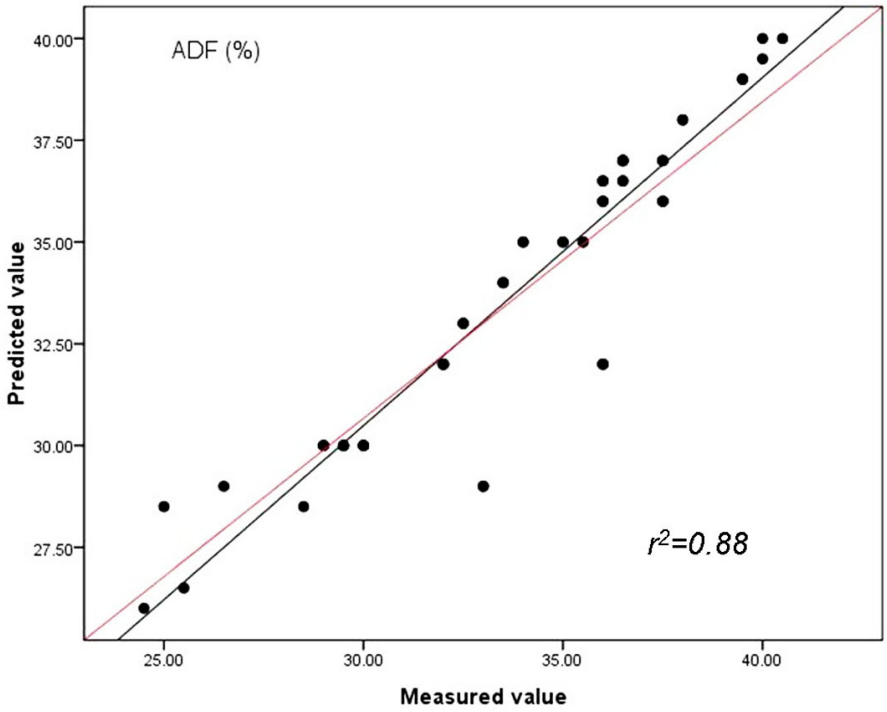
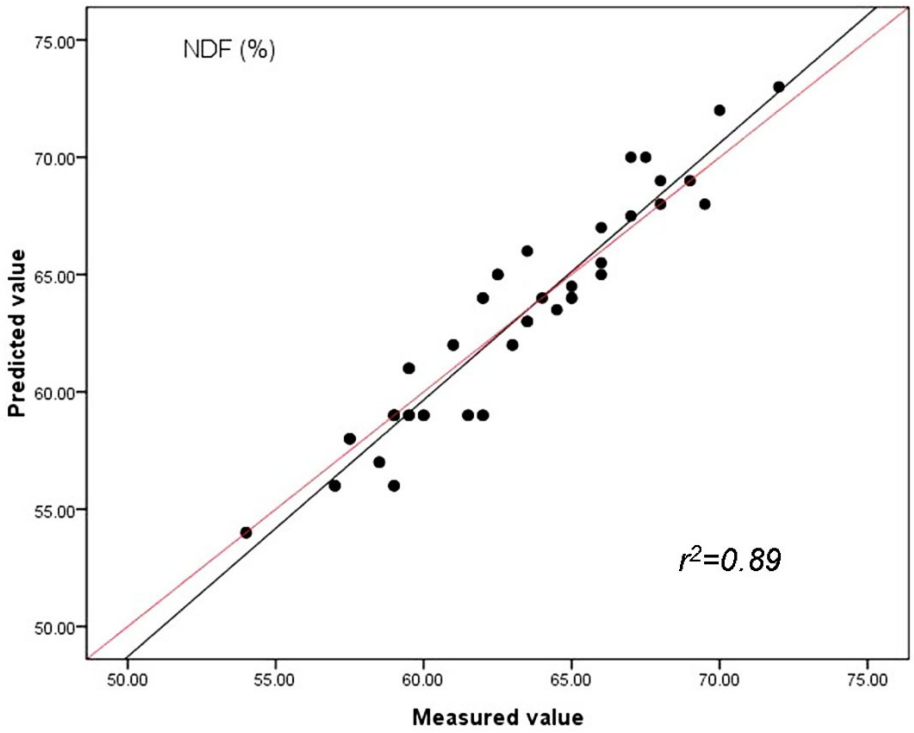


Figure 4. Relationships between the measured and predicted values of the acid detergent fiber content (ADF) of sheepgrass hay for the validation data set. The red line represents the best fit.

473
474
475
476
477
478
479
480
481
482
483
484
485
486



487

488 **Figure 5.** Relationships between the measured and predicted values of the neutral detergent fibre content (NDF) of
489 sheepgrass hay for the validation data set. The red line represents the best fit.
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505 **Tables**
506 **Table 1.** Summary statistics calibration and prediction sets for CP, ADF and NDF of sheepgrass hay by laboratory
507 reference methods (DM %).

Parameters	Data set	N	Min	Max	Mean	SD
CP (%)	Total samples	201	6.20	14.33	10.54	1.24
	Calibration set	150 (2)	6.25	14.33	10.50	1.22
	Prediction set	51	6.15	14.32	10.57	1.25
ADF (%)	Total samples	195	35.13	42.34	38.74	2.33
	Calibration set	144 (8)	35.62	42.30	38.96	2.35
	Prediction set	51	34.63	42.38	38.51	2.30
NDF (%)	Total samples	202	50.71	71.08	60.89	2.67

Calibration set	151(1)	50.20	70.66	60.43	2.69
Prediction set	51	51.21	71.50	61.34	2.68

Abbreviations: CP, crude proteinelectrical; NDF, neutral detergent fiber; ADF, acid detergent fiber; SD, standard deviation, DM, dry matter. The bracketed numbers are outliers during the calibration process.

Table 2. Main parameters of three best calibrations for three different methods

Parameters	PC	Data format	Filtering method	filtering parameter
CP (%)	8	Second derivative	Savitzky-Golay filter	3,2
ADF (%)	7	Second derivative	Norris derivative filter	5,2
NDF (%)	10	Second derivative	Norris derivative filter	5,2

Note: For Sacitzky-Golay filter, the paramerers include data points and polynomial order; For Norris derivative filter, it means segment length and gap between segment; Abbreviations: PC, number of principal component.

Table 3. Comparison of the accuracy of the calibration results (n=152) achieved by NIRS using three different methods for evaluation

	PLS		PCR		MLR	
	R^2	RMSEC	R^2	RMSEC	R^2	RMSEC
CP (%)	0.95	0.74	0.86	1.63	0.85	5.74
ADF (%)	0.93	1.25	0.84	4.96	0.70	8.91
NDF (%)	0.94	1.31	0.85	3.60	0.72	4.59

Abbreviations: CP, crude protein; NDF, neutral detergent fiber; ADF, acid detergent fiber; R^2, the coefficient of determination for the calibration set; RMSEC, the root mean square error of calibration.

Table 4. Comparison of the accuracy of the validation results (n=51) achieved by NIRS using three different methods for evaluation

	PLS			PCR			MLR		
	r^2	RMSEP	RPD	r^2	RMSEP	RPD	r^2	RMSEP	RPD

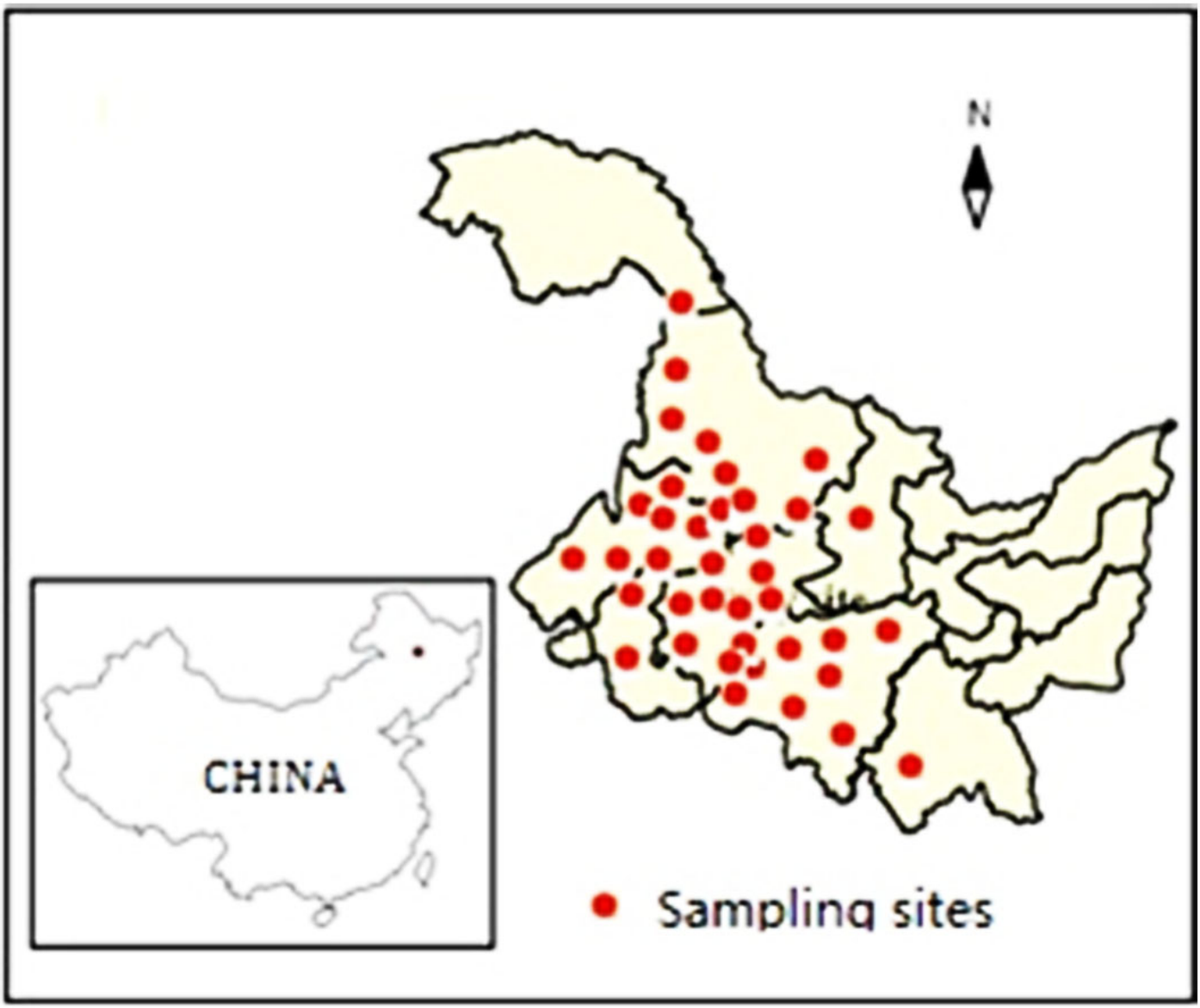
CP (%)	0.91	1.24	3.2	0.86	2.63	1.3	0.74	6.74	0.4
ADF (%)	0.89	1.37	3.1	0.82	5.86	1.2	0.60	9.91	1.1
NDF (%)	0.88	1.41	3.0	0.82	4.50	1.0	0.62	8.59	0.9

Abbreviations: CP, crude protein; NDF, neutral detergent fiber; ADF, acid detergent fiber; r^2, the coefficient of determination for the validation set; RMSEP, room mean squared error of prediction; RPD, the ratio of the standard deviation in the validation set over the room mean squared error of prediction.

1

Figure 1

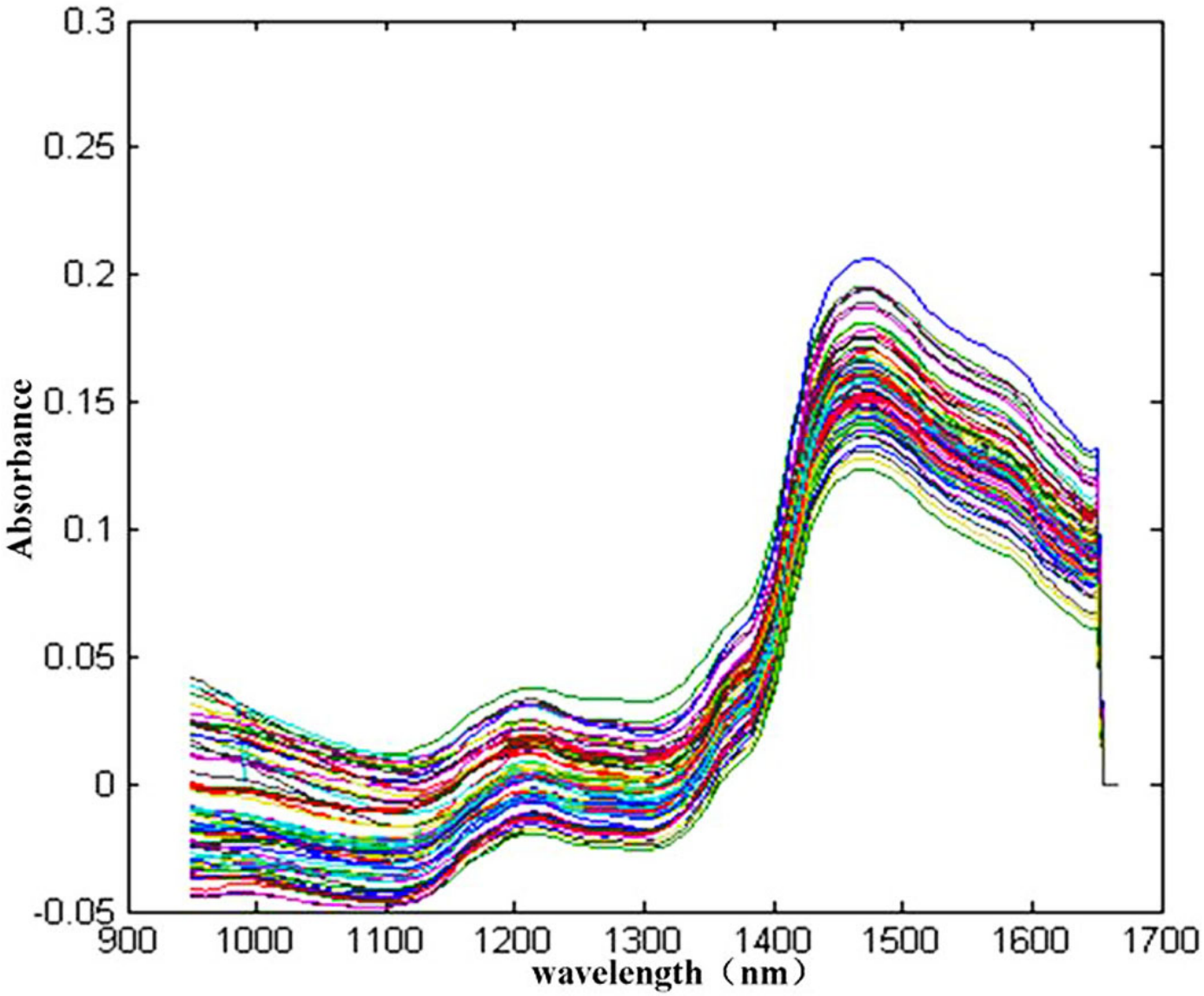
Summary of sampling sites distribut ion in Heilongjiang Province , China .



2

Figure 2

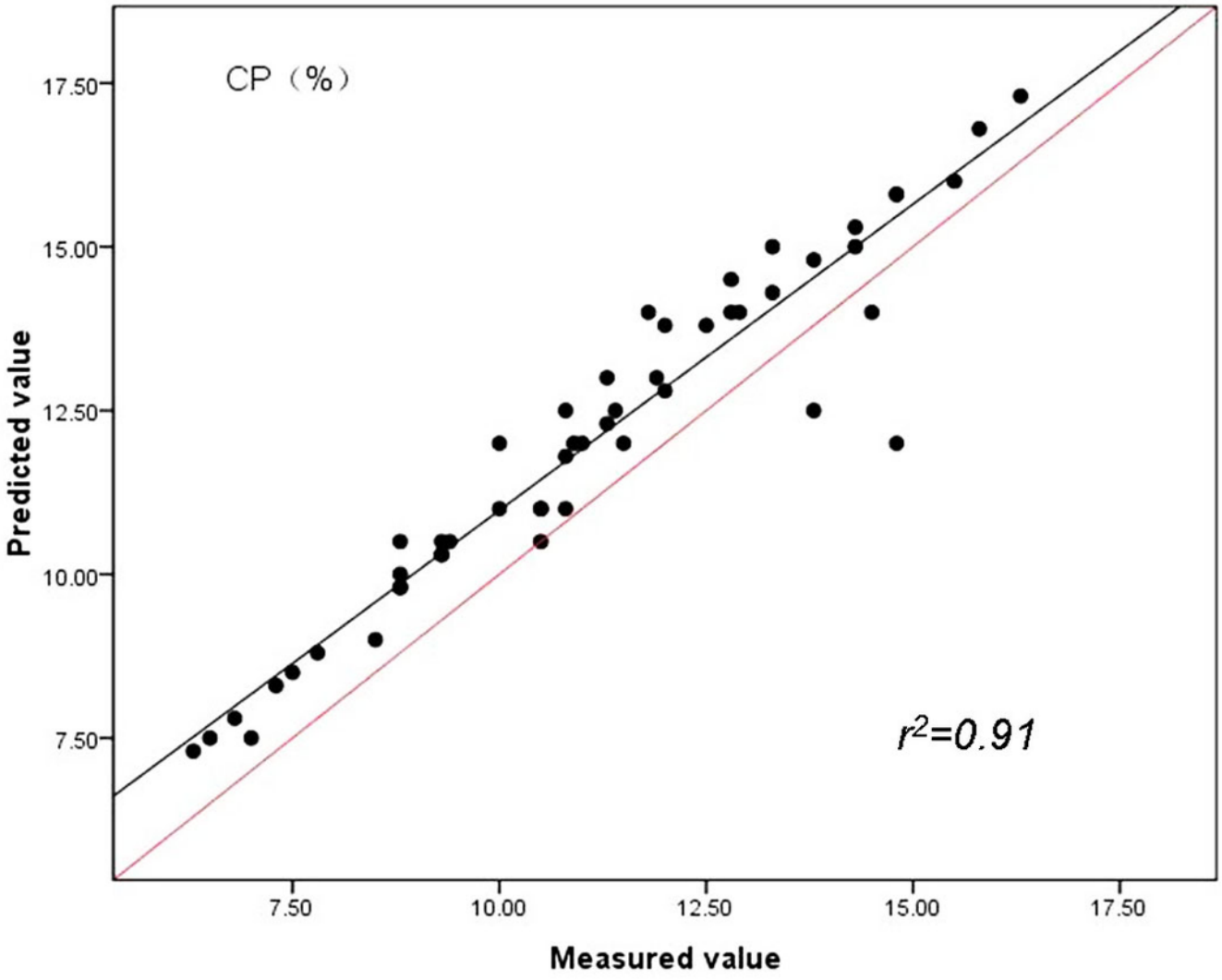
Spectra of NIR of a total 203 sheepgrass samples.



3

F igure 3

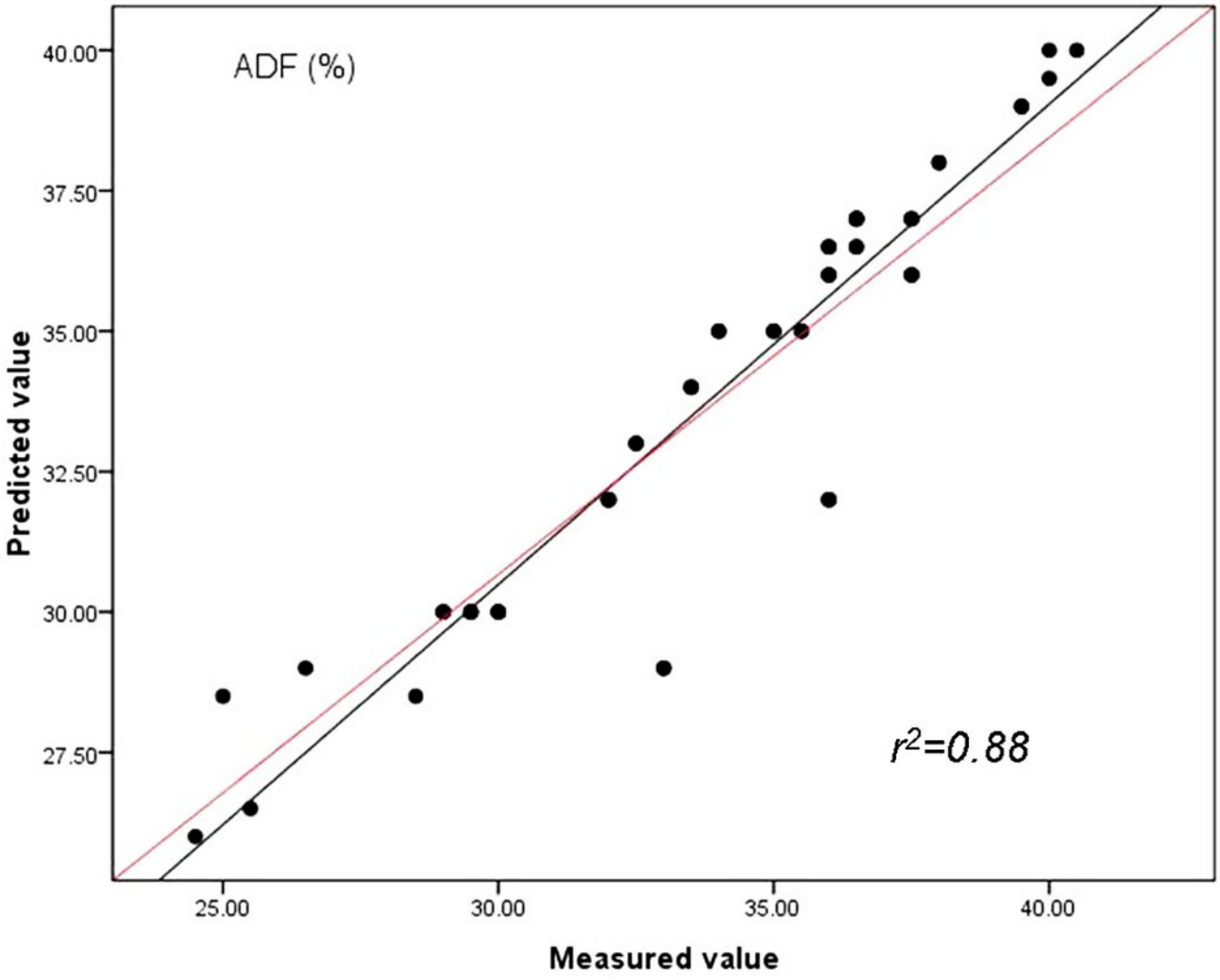
Relationships between the measured and predicted values of the c rude protein content (CP) of sheepgrass hay for the validation data set. The red line represents the best fit.



4

F igure 4

Relationships between the measured and predicted values of the acid detergent fiber content (ADF) of sheepgrass hay for the validation data set. The red line represents the best fit.



5

F igure 5

Relationships between the measured and predicted values of the n eutral detergent fibre content (NDF) of sheepgrass hay for the validation data set. The red line represents the best fit.

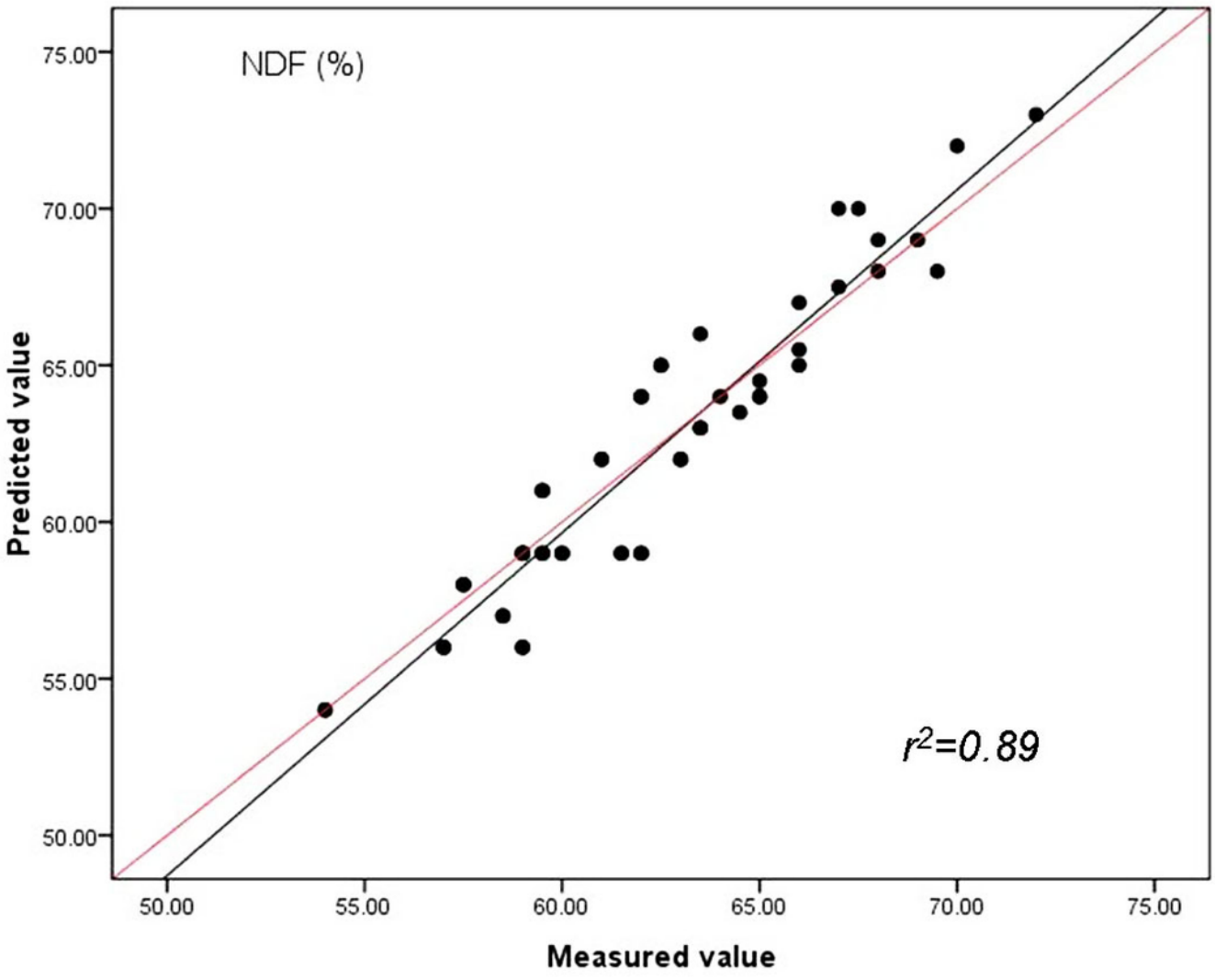


Table 1(on next page)

Table 1

Summary statistics calibration and prediction sets for CP, ADF and NDF of sheepgrass hay by laboratory reference methods (DM %).

Table 1. Summary statistics calibration and prediction sets for CP, ADF and NDF of sheepgrass hay by laboratory reference methods (DM %).

Parameters	Data set	N	Min	Max	Mean	SD
CP (%)	Total samples	201	6.20	14.33	10.54	1.24
	Calibration set	150 (2)	6.25	14.33	10.50	1.22
	Prediction set	51	6.15	14.32	10.57	1.25
ADF (%)	Total samples	195	35.13	42.34	38.74	2.33
	Calibration set	144 (8)	35.62	42.30	38.96	2.35
	Prediction set	51	34.63	42.38	38.51	2.30
NDF (%)	Total samples	202	50.71	71.08	60.89	2.67
	Calibration set	151(1)	50.20	70.66	60.43	2.69
	Prediction set	51	51.21	71.50	61.34	2.68

Abbreviations: CP, crude proteinelectrical; NDF, neutral detergent fiber; ADF, acid detergent fiber; SD, standard deviation, DM, dry matter. The bracketed numbers are outliers during the calibration process.

Table 2(on next page)

Table 2

Main parameters of three best calibrations for three different methods

Table 2. Main parameters of three best calibrations for three different methods

Parameters	PC	Data format	Filtering method	filtering parameter
CP (%)	8	Second derivative	Savitzky-Golay filter	3, 2
ADF (%)	7	Second derivative	Norris derivative filter	5, 2
NDF (%)	10	Second derivative	Norris derivative filter	5, 2

Note: For Sacitzky-Golay filter, the paramerers include data points and polynomial order; For Norris derivative filter, it means segment length and gap between segment; Abbreviations: PC, number of principal component.

Table 3(on next page)

Table 3

Comparison of the accuracy of the calibration results (n=152) achieved by NIRS using three different methods for evaluation

Table 3. Comparison of the accuracy of the calibration results (n=152) achieved by NIRS using three different methods for evaluation

	PLS		PCR		MLR	
	R^2	RMSEC	R^2	RMSEC	R^2	RMSEC
CP (%)	0.95	0.74	0.86	1.63	0.85	5.74
ADF (%)	0.93	1.25	0.84	4.96	0.70	8.91
NDF (%)	0.94	1.31	0.85	3.60	0.72	4.59

Abbreviations: CP, crude protein; NDF, neutral detergent fiber; ADF, acid detergent fiber; R^2, the coefficient of determination for the calibration set; RMSEC, the root mean square error of calibration.

Table 4(on next page)

Table 4

Comparison of the accuracy of the validation results (n=51) achieved by NIRS using three different methods for evaluation

Table 4. Comparison of the accuracy of the validation results (n=51) achieved by NIRS using three different methods for evaluation

	PLS			PCR			MLR		
	r^2	RMSEP	RPD	r^2	RMSEP	RPD	r^2	RMSEP	RPD
CP (%)	0.91	1.24	3.2	0.86	2.63	1.3	0.74	6.74	0.4
ADF (%)	0.89	1.37	3.1	0.82	5.86	1.2	0.60	9.91	1.1
NDF (%)	0.88	1.41	3.0	0.82	4.50	1.0	0.62	8.59	0.9

Abbreviations: CP, crude protein; NDF, neutral detergent fiber; ADF, acid detergent fiber; r^2, the coefficient of determination for the validation set; RMSEP, room mean squared error of prediction; RPD, the ratio of the standard deviation in the validation set over the room mean squared error of prediction.