Calibrating fecal NIRS equations for predicting botanical composition of diets

JOHN W. WALKER, SCOTT D. McCOY, KAREN L. LAUNCHBAUGH, MERRITA J. FRAKER, AND JEFF POWELL

Authors are Resident Director of research, Texas Agricultural Experiment Station, San Angelo, Tex. 76901; Research Technician, USDA U. S. Sheep Experiment Station, Dubois, Ida. 83423; Assistant Professor, University of Idaho, Moscow, Ida, 83844; Research Associate, Montana State University, Bozeman, Mont. 59717; and Professor, University of Wyoming, Laramie Wyo. 82071.

Abstract

The objectives of this study were to investigate the use of near infrared spectroscopy (NIRS) of fecal samples for predicting the percentage of mountain big sagebrush (*Artemisia tridentata* Nutt. ssp. vaseyana (Rydb) Beetle) in sheep diets and to quantify the limitations of using NIRS of fecal samples to predict diet composition. Fecal material from a sheep feeding trial with known levels of sagebrush and several background forages was used to develop fecal NIRS calibration equations validated with fecal material from 2 other sheep feeding trials with known levels of sagebrush in the diets. The 1996 calibration trial varied the level of sagebrush, alfalfa, and grass hay in the diets. The 1998 trial compared frozen to air-dried sagebrush. The Wyoming trial was a metabolism study using frozen sagebrush. Trials used different levels of sagebrush varying from 0 to 30% of the diet in increments of 4 to 10 percentage points. Internal validation of the 1996 trial with a subset of the samples not used for calibration showed that when predicted samples are from the same population as the calibration samples, this procedure can accurately predict percent sagebrush ($R^2 = 0.96$, SEP = 1.6). However, when predicted samples were from a different population than calibration samples, accuracy was much less, but precision was not affected greatly. Low accuracy was caused by a compression of the range of data in the predicted values compared to the reference values, and the predicted sagebrush levels in the diet should be considered to represent an interval scale of measurement. Modified partial least squares regression resulted in better calibration than stepwise regression, and calibration data sets with only high, low, and no sagebrush resulted in calibrations almost as good as data sets with several intermediate levels of sagebrush. High values of the H statistic were related to low precision but did not affect the accuracy of predictions. We believe the interval scale of measurement will contain sufficient information for the purpose of addressing many questions on rangelands.

Key Words: *Artemisia*, diet selection, sagebrush, sheep

Since the initial report of using near infrared spectroscopy (NIRS) of fecal material to predict nutritional parameters of grazing animals (Coleman et al. 1989), the use of this procedure has expanded rapidly. The NIRS of feces can accurately predict dietary crude protein, digestible organic matter (Lyons and Stuth

Resumen

El objetivo de este estudio fué investigar el uso de la Espectroscopía de Reflectancia en el Infrarrojo Cercano (NIRS) en muestras de heces fecales para predecir el porcentaje de Artemisa Grande de la Montaña (*Artemisia tridentata* ssp. vaseyana [Rydb.] Beetle) en dietas de borregos y para cuantificar las limitaciones del uso de Espectroscopía de Reflectancia en el Infrarrojo Cercano en muestras fecales para predecir la composición de la dieta. Se usaron heces fecales de borregos bajo dietas con niveles conocidos de artemisa y varios forrajes complementarios para calibrar las ecuaciones de Espectroscopía de Reflectancia en el Infrarrojo Cercano, las cuales se validaron con material fecal de otros dos ensayos nutricionales en borregos con niveles conocidos de artemisa en las dietas. En el ensayo de calibración de 1996 los niveles de artemisa, alfalfa y paja de pasto en las dietas variaron. En el ensayo de 1998 se comparó artemisa congelada y artemisa secada al aire. El ensayo de Wyoming fué un estudio de metabolismo en el que se uso artemisa congelada. En los ensayos se usaron niveles del 0 al 30% de artemisa en la dieta con incrementos de 4 a 10 puntos porcentuales. La validación interna del ensayo de 1996 con un subgrupo de muestras no usadas para la calibración mostro que cuando las muestras predecidas eran de la misma población que las muestras de calibración, este procedimiento puede predecir correctamente el porcentaje de artemisa ($R^2 = 0.96$, SEP = 1.6). Sin embargo cuando las muestras predecidas provenían de una población diferente a las muestras de calibración, la exactitud fué mucho menor, pero la precisión no se afectó grandemente. La baja exactitud fué causada por una compresión del rango de datos en los valores predecidos cuando se compararon con los valores de referencia, y los niveles predecidos de artemisa en la dieta deben ser considerados para representar un intervalo en la escala de medición. La regresión linear o procedimiento PLS (partial least-squares regression) modificada resulto en una mejor calibración que la regresión múltiple (stepwise regression), y la calibración de los grupos de datos con niveles altos, bajos y sin artemisa resultaron en calibraciones casi tan buenas como los grupos de datos con varios niveles intermedios de artemisa. Niveles altos de estadística-H (H statistic) estuvieron relacionados a una baja precisión, pero no afectaron la exactitud de las predicciones. Creemos que el intervalo en la escala de medición que representa la Espectroscopía de Reflectancia en el Infrarrojo Cercano en la predicción de la composición de la dieta contiene suficiente información para responder muchas preguntas en relación a praderas.

1992, Leite and Stuth 1995), and botanical composition (Walker et al. 1998) when the validation samples are a subset of the calibration population. Currently, NIRS of fecal material is used to predict dietary composition of samples whose relationship to the calibration population is not well known (Snowder et al. 2001, USDA-NRCS 1999), even though Coleman et al. (1995) warned that extrapolation beyond the conditions represented in the calibration samples is risky.

Because the material being predicted (i.e., the diet) is not available for standard laboratory analysis, the use of NIRS to predict dietary parameters of grazing animals presents unique challenges compared to other agricultural applications of this technique. In other applications of NIRS technology, if samples are spectrally different from the calibration population, or if the predicted value is outside the expected range, the true reference value of the sample can be determined by standard laboratory procedures.

One of the statistics used in NIRS analysis to determine if a sample is from the same spectral population as the samples that the calibration equation was developed from is called H (Shenk 1989). The H statistics is the distance of the sample from the population centroid (i.e., the population mean in multi-dimensional space). The H statistics is standardized to have unit variance so it represents the number of standard deviations that a sample is from the multi-dimensional population mean. Shenk (1989) suggested that if the H of a predicted sample is greater than 3 standard deviates from the average spectrum of the calibration samples, the results should be reported with caution. Whereas the use of H for typical NIRS agricultural applications was investigated by Shenk (1989) and Shenk and Westerhaus (1991a,1991b), most previous research on using fecal NIRS equations did not examine the relationship of H to prediction accuracy (Coleman et al. 1989, Lyons and Stuth 1992, Coleman et al. 1995, Leite and Stuth 1995, and Coleman et al. 1999).

The NIRS technique can be calibrated to predict botanical composition of clipped forages composited to create samples of plant mixtures (Coleman et al. 1985, Pitman et al. 1991, Petersen et al. 1987, Moore et al. 1990). The calibration samples in these studies were created by compositing mixed samples with known amounts of the different botanical species. Coleman et al. (1990) demonstrated that

calibrations based on botanically pure samples were as acceptable as calibrations developed from mixtures. Volesky and Coleman (1996) showed that the addition of botanically pure samples to esophageal extrusa calibration data sets was beneficial. However, pure diets of many of the plant species of interest to the study of dietary habits of ruminants cannot be fed because of problems with toxicosis at high levels of intake. Nonetheless, it may be possible to predict botanical composition of diets with just 2 levels of the component of interest, which would provide a great improvement in the design of calibration trials by reducing the number of samples required for calibration.

Development of calibration equations using stepwise regression techniques (with a limited number of individual wavelengths) has not been compared with multivariate procedures using information from all wavelengths. Multivariate procedures use a data reduction technique such as principal component analysis (PCA), which reduces the original data to sets of scores that have 2 properties 1) the variance of the first set of scores is maximized and successive sets of scores explain successively less of the variation in the original data; and 2) the sets of scores are uncorrelated to each other. This is accomplished by multiplying the original log 1/R data by a weighting coefficient (i.e., eigenvector) and summing these values across all wavelengths to create a score for each sample. This is analogous to the coefficients in a multiple regression equation transforming a set of independent variables to a single dependent variable, i.e., predicted Y variable. The principle component scores produced can then be used in multiple regression to predict the value of the dependent variable (i.e., botanical composition in the current study). In this study we used modified partial least squares regression, which is similar to PCA regression except that information from the dependent variable is included in the transformation so that the resultant scores have maximum correlation with the dependent variable and the residuals obtained after each factor is calculated are standardized (divided by the mean residual value) before calculating the next factor. Pitman et al. (1991) reported there were no consistent differences between modified partial least squares regression and modified stepwise regression when predicting the botanical composition of tropical grass-legume mixes.

The objective of this study was to determine if NIRS of fecal samples could predict the percentage of mountain big sagebrush (*Artemisia tridentata* Nutt. ssp. *vaseyana* (Rydb) Beetle) in sheep diets. Four specific objectives were evaluated: 1) stepwise regression was compared to modified partial least squares regression; 2) calibrations based on maximum, minimum, and no sagebrush in the diets were compared to calibrations that included intermediate levels of sagebrush; 3) the ability of calibrations based on diets composed of dried sagebrush to predict diets from frozen sagebrush were investigated; and 4) the relationship of H with the precision and accuracy of predictions was investigated.

Methods[1]

This research was conducted at the USDA, U.S. Sheep Experiment Station located in Dubois, Ida. (112°10'W, 44°21'N). Fecal materials from a feeding trial with sheep fed known levels of sagebrush with several background diets were used to develop NIRS fecal calibration equations. Equations were validated with samples from 2 independent feeding trials with known levels of sagebrush with a single background forage. Because the composition of the base diet did not vary in the validation trials, calibration equations for sagebrush were identical to equations for other components, and samples from these trials could not be used to develop unambiguous calibration equations.

1996 Feeding Trial

The 1996 feeding trial consisted of feeding varying levels of mountain big sagebrush, early bloom alfalfa hay (*Medicago sativa* L.) and a grass hay composed of smooth brome (*Bromus inermis* Leyss.) and timothy (*Phleum pratense* L.). The sagebrush was collected in late September 1996 and air-dried in the shade (maximum daily temperature 21 to 30° C). After drying, sagebrush branches were flailed against a wire panel over a large tub to remove leaves and small current year's stems.

The flailed sagebrush was chopped using a small flail-type garden shredder;

other forages were ground through a hammer mill with a 2.5 cm screen. Diets were mixed to contain 0, 4, 8, 12, 16, and 24% sagebrush with a base diet of hays in the following alfalfa/grass proportions: 0:100, 20:80, 40:60, 60:40, 80:20, and 100:0 for a total of 36 different diets. Molasses was added to the diets (5% DM) during the mixing process to cause feed particles to adhere together and reduce sorting when eaten.

Diets were fed to 36 mature, white-faced, western range ewes housed in individual pens. One animal was fed each diet. Diets were fed at 0800 and 1600 hours; refusals (orts) were removed at 0800 hours. The trial consisted of a 6-day adaptation period and 4-day collection period. Fecal samples were collected from the floor of the pens at 1600 and 0800 hours and dried in a forced air oven at 60°C.

1998 Feeding Trial

The 1998 feeding trial consisted of varying levels of mountain big sagebrush stored either air-dried or frozen before the study (Fraker 1999). Sagebrush for this trial was current season's growth of vegetative stems (no flowering stalks) collected the last week of August 1997. Half of the sagebrush was air-dried (maximum temperature 30 to 33°C) and stored in plastic bags; half was frozen promptly after collection and stored in a freezer. Diets were mixed to contain 0, 8, 16, and 24% of either dry or frozen sagebrush with a base diet of 1:1 alfalfa/grass (grass was a mixture or smooth brome and timothy hay) for a total of 7 different diets. Molasses was added to the diets (5% DM) during the mixing process to cause feed to adhere together and reduce sorting.

An intake and digestion trial was conducted in January 1998 using 6 to 7-month-old, white-faced crossbred lambs with an average weight of 47 kg. The animals were penned individually and 5 animals were assigned to each of the 7 diets for a total of 35 animals (34 wethers and 1 ewe). Animals were fitted with fecal bags to determine total fecal production. Feed was offered in excess from 0800 to 1100 hours and from 1300 to 1600 hours. Every 30 min, while feed was offered, feed bunks were checked and more feed was added if needed. This procedure limited the potential for sorting and minimized the amount of refusal. The trial consisted of a 5-day adaptation period and a 6-day collection period. Fecal bags were emptied

twice daily, and a sub-sample of feces was collected. Afternoon samples were composited with collections made the following morning and dried in a forced air oven at 60° C.

Wyoming Feeding Trial

The Wyoming trial used 16 Rambouillet wether lambs (28 to 41 kg body weight) fed diets containing mixtures of hand-harvested current year's growth of mountain big sagebrush leaves and native grass hay (Ngugi et al. 1995). Sagebrush leaves were harvested in September from the western edge of the Medicine Bow Range, Carbon County, Wyo., and stored in sealed plastic bags in a freezer until fed. Four diets in the following proportions of grass hay:sagebrush: 100:0, 90:10, 80:20, and 70:30 were fed. Four animals received each diet. The trial consisted of a 9-day adjustment period followed by a 6-day collection period when all urine and feces were collected. Fecal samples were composited by animal for the 6-day collection period and ground through a 1-mm screen of a Wiley mill. Samples were reground through a cyclone mill before being scanned to collect NIRS reflectance. Details of this trial were presented by Ngugi et al. (1995).

Equation Development

All fecal samples were ground in a cyclone mill to pass a 1-mm screen, packed into sample cells with a near-infrared transparent quartz cover glass, and scanned 32 times using a NIR Systems, Inc. (Silver Spring, Md.) model 6500, scanning reflectance monochromator. Reflected energy (log 1/R) was measured, averaged over the 32 scans and recorded at 2-nm intervals from 400 to 2,500 nm.

Calibration equations were developed using stored NIRS spectra from fecal samples as the independent variables and percent sagebrush fed in the diets as the dependent reference data. Scanned samples were daily individual animal feces in the 1996 and 1998 trials, whereas for the Wyoming trial, they were individual animal feces composited across 6-days. Before calibration, each spectrum was transformed with a (1,8,8) derivative. The first number in parenthesis is the order of the derivative, the second number is the gap (number of data points over which derivative is calculated), and the third number is the smooth (number of data

points in a moving average). Scatter correction was done with the standard normal variance and detrend procedure (Barnes et al. 1989).

Data Analyses

Before further analyses, data from the 1996 and 1998 trials were individually examined to identify and remove outlier samples. To eliminate outliers, separate calibration equations were done for the 1996 and 1998 trials. Samples that had an H larger than 3 or a residual "t" statistic greater than 3 were eliminated. This resulted in the elimination of 6 and 10 samples from the 1996 and 1998 data sets, respectively, or approximately 5% of the samples from each of the data sets.

Calibration equations were evaluated for usefulness based on validation statistics for samples that were not used for calibration. Calibration refers to the development of a regression equation using the reflected energy (log 1/R) of the near infrared spectra as independent variables to predict the botanical component (i.e., dependent variable or reference value). Validation refers to the ability of the calibration equation to predict the reference value of samples not part of the data set used in the development of the calibration equation. Validation samples are referred to as 'internal' if they are a subset of a uniform group of samples used to develop a calibration equation and independent if they were from a set of samples from a different trial or treatment. The statistics evaluated included standard error of prediction (SEP), standard error of prediction corrected for bias (SEPc), coefficient of determination (R^2), slope, and bias.

These statistics indicate the relationship between the predicted percentage sagebrush in the diets and the percentage sagebrush that was fed. The coefficient of determination is the amount of variation in sagebrush fed that is explained by fecal NIRS predictions and is an indicator of the precision of the calibration equation. Slope and bias are estimates of the accuracy of the calibration equation; in general accuracy decreases as the deviation from 1 and 0, respectively, increases. However, slopes may not always be directly comparable because of the differences in the Y intercept. The SEP and SEPc statistics are estimates of error in units that are related to the unit of measurement; in this study the percentage sagebrush in the diet. The SEP includes error that results from poor

precision (i.e., lack of fit) and systematic error as a result of slope and bias deviations from 1 and 0. The SEPc has much of the systematic error (lack of accuracy) removed and is an indicator of precision.

Objective 1: This objective compared stepwise (SW) and modified partial least squares (MPLS) calibrations. Fecal samples from days 1 and 3 of the 1996 trial (96 1&3) were used to develop calibration equations using either SW, or MPLS regressions and validated with either days 2 and 4 (96 2&4) from the same trial (i.e., internal validation), or the 1998 Dry samples (i.e., independent validation). The 1998 Dry samples were used for independent validation because they most closely duplicated the procedures used for the feeding trial from which samples for calibration were produced. The best calibration method was determined by comparing validation statistics and then was used for all further analyses.

Objective 2: This objective examined the effect of eliminating diets with intermediate levels of sagebrush on the precision and accuracy of resultant calibration equations. The purpose of this objective was to determine if the efficiency of feeding trials for developing fecal NIRS calibration equations could be improved by reducing the number of levels fed without affecting the precision or accuracy of predictions. The MPLS calibrations were developed for all 1996 samples and compared to calibrations from 3 subsets: HiLo All contained all samples with 0, 4, and 24% sagebrush; HiLo Mixed contained samples with 60:40 or 40:60 proportions of alfalfa and grass as the background and 0, 4 and 24% sagebrush; HiLo Pure contained samples with 100% alfalfa or 100% grass and 0, 4, and 24% sagebrush. The diets with no sagebrush were included in the calibration data sets because, in practice, treatments with only the base diets require relatively less labor to conduct because the target plant species does not need to be hand-collected. Thus the additional level can be added to the calibration data set at a relatively low cost. The calibrations for this objective were validated using the 1998 Dry samples.

Objectives 3: The purpose of this analysis was to evaluate the ability of calibration equations developed from feeding trials using target forages that were dried to predict diets from feeding trials where the target forage was kept frozen until fed. The latter case more closely represents a normal grazing situation, whereas the use of dried forages represents a simpler method for conducting trials. Calibrations were developed using all samples from the 1996 trial. Frozen samples from 1998 and Wyoming samples were used for validation and dried samples from 1998 served as a control.
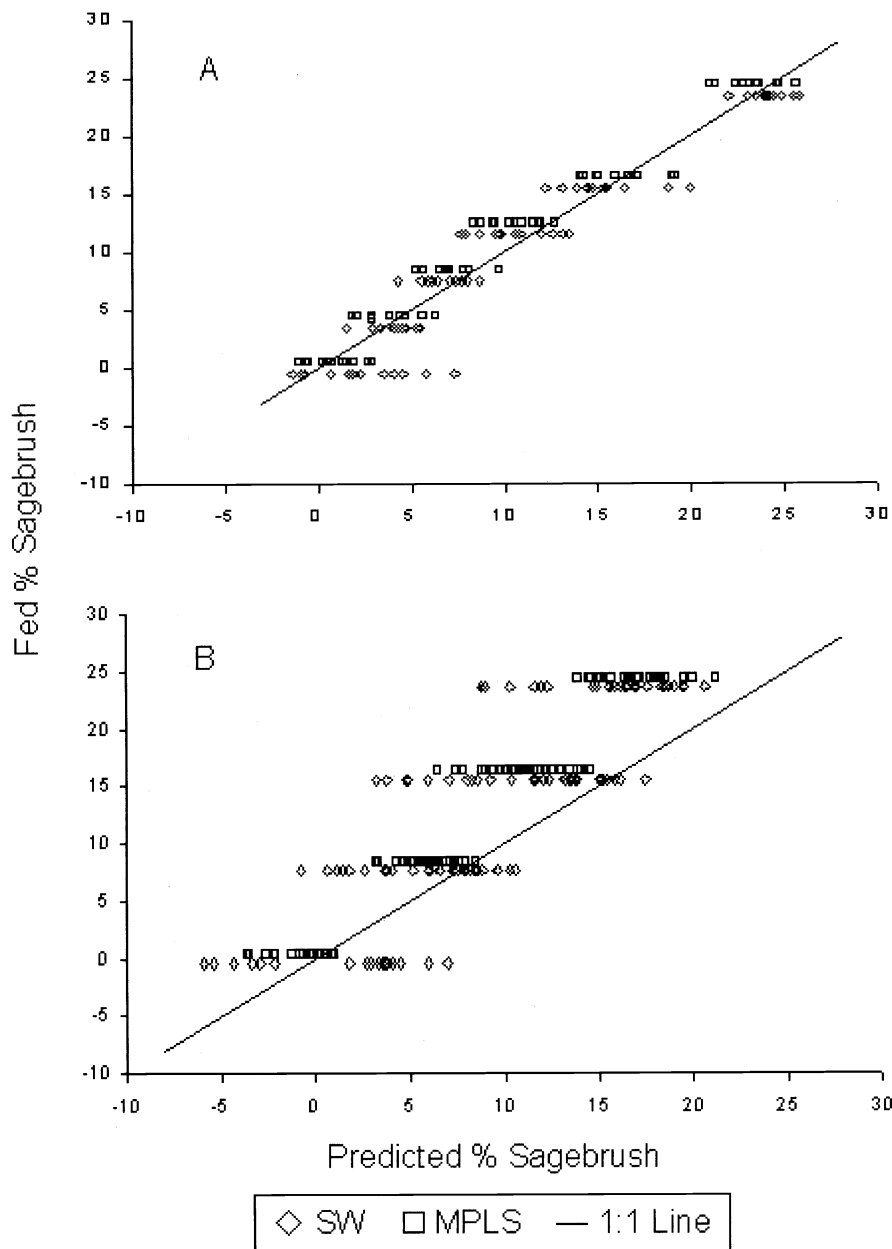
Objective 4: The effect of the value of H on the precision and accuracy of predictions was evaluated for the 1998 Frozen and Wyoming data sets. The value of H is the number of standard deviates that a particular sample is from the mean of the population of samples that were used to develop the calibration equation. Thus it represents the spectral similarity between the predicted samples and the calibration samples. The simple correlation between H and the absolute value of the residual was calculated to determine if the precision of predictions (i.e., difference between actual and predited values) was related to the spectral closeness of the sample to the calibration population. Each data set was divided in half based on the median value of H in the data set. The predictions were compared between the 2 subsets to determine the effect of H on accuracy of predictions. If the slope of the line relating the predicted and actual levels of sagebrush in the diets was the same for both of the subsets of data, it would indicate that accuracy was not affected by the spectral distance (i.e., H) between the sample and the calibration population.

Results and Discussion

Objective 1: Comparison of stepwise (SW) and modified partial least squares (MPLS) calibrations for days 1 and 3 of the 1996 trial showed that calibration using MPLS improved predictions for both the internal and independent valida-

tion data sets (Table 1; Fig. 1). Predictions for the internal validation set (96, 2&4) showed that MPLS compared to SW slightly increased the precision of predictions as indicated by the higher R2 and lower SEP(c), but not the accuracy of predictions (i.e., bias and slope). The value of MPLS compared to SW calibrations for improving precision were more apparent for the independent validation of the 1998 Dry samples. In this comparison R^2 and SEP(c) showed much improvement in MPLS compared to SW, the similar bias for the 2 predictions showed that accuracy was about equal. The improved precision of the MPLS calibration compared to the SW calibration indicates that the wavelengths selected by the stepwise procedure for the 1996 calibration data set did not include all of the relevant wavelengths for the independent 1998 Dry validation data set. The average H value was higher for the MPLS predictions of the 1998 Dry samples compared to the SW prediction of these samples. The higher H statistic is a result of the use of all wavelengths in the MPLS equation compared to only 5 wavelengths in the SW equation. This apparent increase in spectral distance between independent predictions and the mean calibration spectra, and the assumption that it reduces the robustness of the resultant equations, caused most previous fecal NIRS calibrations to use SW equations. Because of the improved precision of predictions using the MPLS equation, this calibration procedure was used for subsequent analysis.

Objective 2: The effect of calibrations based on the 1996 trial after removing the samples with intermediate levels of sagebrush (i.e., 8, 12, and 16%) on precision and accuracy for predictions of the 1998 Dry samples, is shown in Table 2 and Figure 2. Reducing the number of samples reduced the accuracy of predictions, but

Table 1. Comparison of stepwise to modified partial least squares regression equations on validation statistics for calibrations developed from the 1996 trial on days 1 and 3.

Validation data set	R^2	Slope[a]	Bias[b]	SEP[c]	SEP(c)[d]	Average H[e]
96 2&4						
Stepwise	0.92	1.00	0.1	2.2	2.2	1.4
MPLS	0.96	1.01	0.4	1.6	1.6	1.5
98 Dry						
Stepwise	0.66	1.10	4.1	6.4	5.0	2.0
MPLS	0.93	1.26	3.9	4.8	2.8	2.7

[a]Slope = slope of the line between reference and predicted values (indicator of accuracy).
[b]Bias = systematic error (i.e., mean of differences, and an indicator of accuracy).
[c]SEP = standard error of prediction (error caused by both lack of precision and lack of accuracy).
[d]SEP(c) = standard error of prediction corrected for bias (indicator of precision).
[e]Average H = standardized Mahalanobis distance (mean spectral distance of validation data set from centroid of calibration data set in multiples of standard deviates).

Fig. 1. Comparison of stepwise regression (SW) and modified partial least squares regression (MPLS) predictions of percent sagebrush fed using 1996 feeding trial days 1 and 3 for calibration and either internal validation with 1996 days 2 and 4 for validation (A) or independent validation using 1998 Dry samples for validation (B).

mixtures compared to the 100% alfalfa or 100% grass background mixtures.

As the number of samples in the calibration data set was reduced, the average H of the validation data decreased, even though the validation data set was the same for the 4 different calibration data sets. This would normally be interpreted to mean that the validation data were more spectrally similar to the calibration data. However, in this instance it indicates that as the number of samples in the calibration data set decreased the variance of the multidimensional centroid increased and because H is standardized to have unit variance the distance of the predicted population from the calibration population is smaller in terms of multiples of the unit variance.

Comparison of the MPLS validation statistics for 1998 Dry samples between Tables 1 and 2 also demonstrates the effect of size of calibration data set on the average value of H. In this instance the slope and R^2 are essentially the same for calibrations based on days 1 and 3 of the 1996 samples (Table 1) or all days (Table 2), but the average H statistic increased from 2.7 to 3.8 when all of the samples were used in the calibration data set.

This analysis indicates that increased efficiency of experimental design can be obtained by reducing the number of intermediate levels and that there is an advantage of using spectrally distinct backgrounds, i.e., pure rather than mixed background diets.

Objective 3: The use of calibrations based on dried samples to predict diets of animals eating an independent source of frozen forage is shown in Table 3 and Figure 3. Compared to the 1998 Dry samples, the validation statistics for the 1998 Frozen samples showed that the slope for the frozen samples (1.43) was greater than that for the dried samples (1.27), resulting in a larger bias and consequently a larger SEP. The SEP(c) was increased only slightly for the frozen samples, and the R^2 was similar for the dried (0.94) and frozen (0.93) predictions. These statistics, in conjunction with the increased deviation from unity of the slope, indicate that using an equation based upon dried forages from the independent 96 trial to predict percentage 98 frozen sagebrush in the diet, decreased the accuracy of predictions with little effect on the precision.

Validation using the Wyoming samples was much worse than the 1998 Frozen samples, but was similar to the 1998

this reduction was less than the proportional decrease in number of samples. For instance the HiLo ALL subset represented a 50% reduction in the number of samples but only a 3% reduction in the R^2 between actual and predicted diets, a slope with a 4% greater deviation from unity, a 27% increase in the SEP, and a 16% increase in SEP(c).

Further reducing the calibration data sets to 23 samples by using only samples from diets with mixed alfalfa and grass (HiLo Mixed) background or 100% alfalfa or 100% grass (HiLo Pure) as a background further reduced the precision and accuracy of predictions. Validation statistics from calibrations based on HiLo Mixed samples showed that this subset resulted in a much higher bias and a slope with a greater deviation from unity. The reduced performance of HiLo Mixed calibrations presumably was caused by a lack of unique spectral information between the 60:40 and 40:60 alfalfa:grass background

Table 2. Effect of reducing the number of intermediate treatment levels in the Calibration data set on validation statistics for 1998 Dry samples.

Calibration data set	n[a]	R[2]	Slope[b]	Bias[c]	SEP[d]	SEP(c)[e]	Average H[f]
96 All[g]	137	0.94	1.27	4.8	5.6	2.7	3.8 *
HiLo All[h]	69	0.91	1.33	6.9	7.6	3.2	2.2
HiLo Mixed[i]	23	0.78	1.61	16.4	17.2	4.9	2.3
HiLo Pure[j]	23	0.82	1.38	8.6	9.5	4.2	1.3

[a]Number of samples in calibration data set.
[b]Slope = slope of the line between reference and predicted values (indicator of accuracy).
[c]Bias = systematic error (i.e., mean of differences, and an indicator of accuracy).
[d]SEP = standard error of prediction (error caused by both lack of precision and lack of accuracy).
[e]SEP(c) = standard error of prediction corrected for bias (indicator of precision).
[f]Average H = standardized Mahalanobis distance (mean spectral distance of validation data set from centroid of calibration data set in multiples of standard deviates).
[g]96 All = all samples.
[h]HiLo All = all samples with 0, 4, or 24% sagebrush.
[i]HiLo Mixed = samples with background matrix of 40:60 or 60:40 alfalfa and grass hay with 0, 4, or 24% sagebrush.
[j]HiLo Pure = samples with background matrix of 100% alfalfa or 100% grass hay with 0, 4, or 24% sagebrush.
*Exceeded value recommended by Shenk et al. (1989).

Frozen in that most of the error was a result of reduced accuracy. The low accuracy was caused primarily by the high slope (2.63), which indicates that the range of the predicted values was about one-third the range of the actual values.

The compressed range of data in these predicted data sets indicates that when the predicted samples are from a different population than the calibration samples, the ratio scale of measurement for the reference value of the samples may be reduced to an interval scale for the predicted values. Measurement scale refers to the different properties (i.e., relationships and operations) of the numbers that can be applied to the measurements. The 4 levels of measurements in order of increasing amount of information that can be determined from them are 1) nominal, 2) ordinal, 3) interval, and 4) ratio. Nominal measurements name the attribute; ordinal measurements can be rank ordered; for interval measurements the distance between attributes has meaning; and ratio measurements have an absolute zero and a meaningful ratio can be constructed. The effect of different scales of measurements on the information that can be extracted from measured attributes can be demonstrated using the reference values of sagebrush fed, which are ratio scale measurements, and the fecal NIRS predicted values, which are interval scale measurements. The mean predicted values for the 8, 16 and 24% diets in the 1998 Frozen data were 0.6, 6.3 and 10.9%, respectively. For the reference ratio data it is appropriate to state that the difference between 8 and 24% is twice as large as the difference between 8 and 16%. It is also appropriate to state that 16% is 2 times larger and 24% is 3 times larger than 8%. At the

interval level of measurement for the fecal NIRS predicted values, an increase from 0.6 to 10.9% is about twice as much as an increase from 6.3 to 10.9%, but it is not appropriate to state that 6.3% is about 10 times larger than 0.6%. In practice this means that fecal NIRS predictions could be used to estimate the difference in botanical composition of the diets of dif-

ferent treatment groups or individuals within a group of animals. But it could not be used to determine how much of a plant species of interest should be fed to increase their level of consumption to a desired level.

Objective 4: For typical agricultural applications, the size of H (i.e., the standardized spectral distance of the predicted sample from the calibration data spectral mean) normally is used to determine if a sample is from the same spectral population as the samples used to develop calibration equations. This analysis investigates the relationship between H and the precision and accuracy of predictions when calibration equations from the 1996 trial were used to predict the 1998 Frozen or Wyoming diets. The value of H was positively correlated to the absolute value of the residual (Fig. 4), which indicates that prediction error increased as H increased. However, when the data sets were split into two groups based on the median value of H, the slope (P > 0.30) and intercept (P > 0.30) of the predictions for the 2 groups were not affected by the
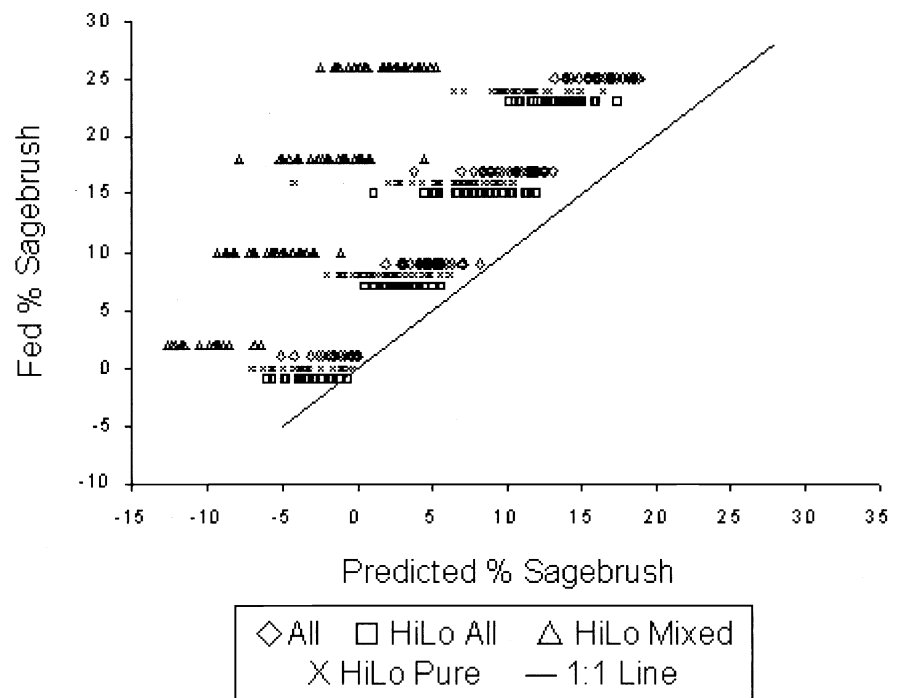
Fig. 2. A comparison of modified partial least squares calibrations using reduced subsets of calibration samples from the 1996 trial was used to predict samples from animals fed dry sagebrush in the 1998 trial. (All = all samples; HiLo = 24, 4, and 0% sagebrush with all backgrounds; Mixed = 24, 4, and 0% sagebrush with 60:40 and 40:60 alfalfa:grass backgrounds; and Pure = 24, 4, and 0% sagebrush with 100% alfalfa or 100% grass backgrounds).

Table 3. Validation statistics for predicting fresh samples using calibration from 1996 samples with 1998 Dry samples included for comparison.

Validation data set	R^2	Slope[a]	Bias[b]	SEP[c]	SEP(c)[d]	Average H[e]
1998 Dry	0.94	1.27	4.8	5.6	2.7	3.8 *
1998 Frozen	0.93	1.43	6.8	7.5	3.1	4.1 *
Wyoming	0.9	2.63	14.2	16	7.7	12.8 *

[a]Slope = slope of the line between reference and predicted values (indicator of accuracy).
[b]Bias = systematic error (i.e., mean of differences, and an indicator of accuracy).
[c]SEP = standard error of prediction (error caused by both lack of precision and lack of accuracy).
[d]SEP(c) = standard error of prediction corrected for bias (indicator of precision).
[e]Average H = standardized Mahalanobis distance (mean spectral distance of validation data set from centroid of calibration data set in multiples of standard deviates).
*Exceeded value recommended by Shenk et al. (1989).

size of H (Fig. 5). Walker et al. (1998) reported that when NIRS fecal predictions for percent leafy spurge (*Euphorbia esula* L.) in the diet were divided into groups based on the size of H, validation statistics were similar between the groups.

The positive relationship between H and magnitude of the residual and the lack of effect of H on the relationship between actual and predicted percent sagebrush in the diets show that for these data, the value of H is positively related to precision but not to accuracy. This interpretation of the relationship of H to the precision and accuracy of predictions is also supported by comparison of the average H for the 1998 Frozen and Wyoming samples to their validation statistics. The Wyoming data set had a much higher average H (12.8) compared to the 1998

Frozen samples (4.1) and a much lower accuracy (i.e., slope = 2.63 vs. 1.43), but good precision as shown by the high R^2 (0.90 vs. 0.93; Table 3). Finally, in both trials the treatments with the highest levels of sagebrush had the highest H values, which may have implications on the source of this spectral variation.

The results of this study show that when predicted samples are from the same population as the calibration samples, such as the internal validation of the 1996 trial (Table 1), NIRS of feces can accurately predict botanical composition of diets. However, when predicted samples are from a different population, accuracy is low and the predicted values should be considered to represent an interval scale of measurement. Of interest is that the precision of the predictions, as indicated by

coefficients of determination between actual and predicted values, consistently accounted for 90% or more of the variation. The consistently high precision but low accuracy may indicate that the wavelengths that best predicted the calibration set were not the optimal set for the validation data sets. However, the wavelengths from the calibration data sets apparently covaried closely with the appropriate wavelengths for the independent validation set.

The suspected shift in optimal wavelengths between the calibration and validation data sets could be caused by the fact that the 1996 calibration diets used one source of sagebrush for all diets, which caused low spectral variation in the feces and may be responsible for the poor accuracy of prediction of independent samples. Use of a single source of sagebrush means that all diets regardless of the proportion of sagebrush in them had the same spectral information relative to sagebrush. Had more sources, e.g., locations, season and years of sagebrush been used the resultant equation would have been more global.

However, animal variation between trials also may have been an important contribution to the lack of accuracy. Mature ewes were used for the 1996 calibration trial, whereas approximately 6 month old whether lambs were used for all of the independent validations.

Age, sex, or physiological status of the animal could affect fecal composition via the effect these factors have on intake and digestibility. Furthermore, the amount of sagebrush in the diet affects both intake and digestibility (Ngugi et al. 1995), and these effects are greater for fresh compared to dried sagebrush (Fraker 1999). Typically 60% or more of ruminant feces is composed of metabolic products including bacteria and endogenous waste (Van Soest 1982). The remainder of the feces composed of residual forage matter would consist of undigested plant cell walls (i.e., neutral detergent fiber).

Comparison of the slopes of the independent validation sets shows that as the effect of treatments on intake and digestibility increased, the accuracy (but not precision) of the predictions decreased. In the 1998 trial, frozen diets caused a small decrease of forage intake (Fraker 1999), which was associated with an increase in average H and greater deviation from unity of the slope (i.e, 1.27 and 1.43 for the dry and frozen diets, respectively).
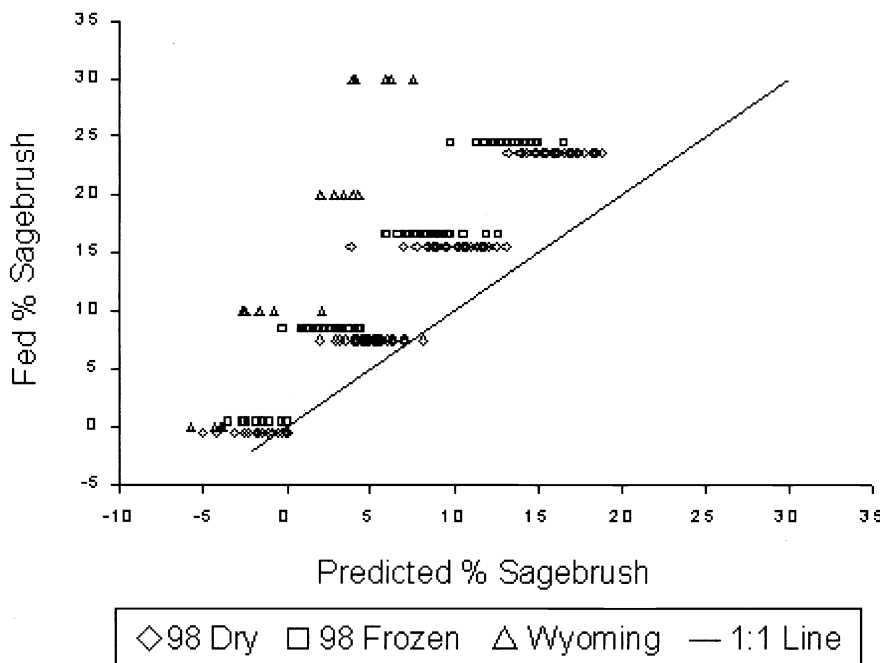
Fig. 3. Prediction of diets from frozen sagebrush using calibration equations based on diets from dried sagebrush; predictions of dry samples are included for comparison.
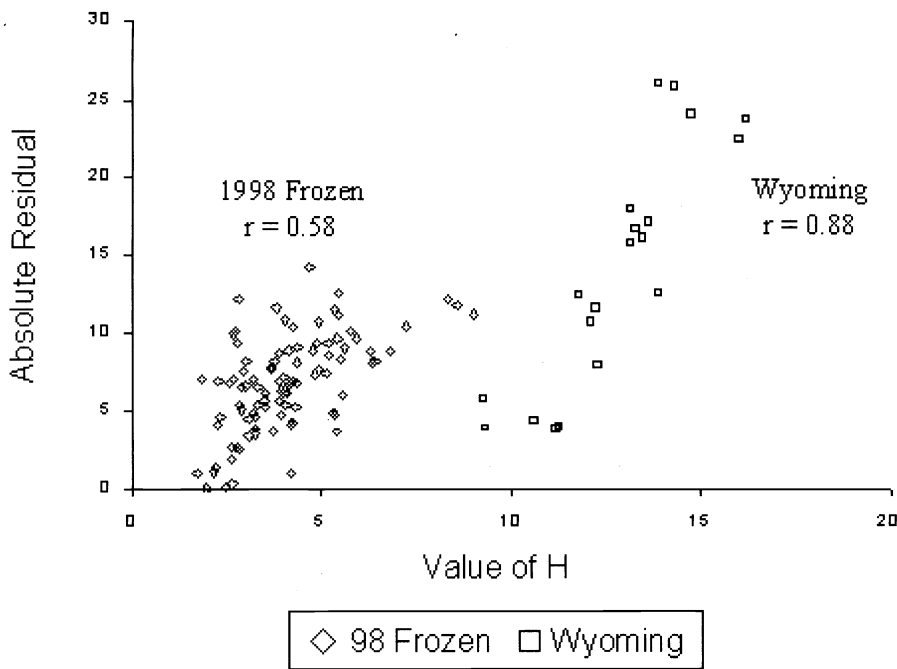
Fig. 4. Relationship between the H value of samples and the absolute value of the residual for independent validation samples and modified partial least squares regression using all 1996 samples.

Ngugi et al. (1995) reported a much greater effect of percent sagebrush in the diet on intake and digestibility for the Wyoming trial than was reported for the 1998 trial. Intake declined from 38 g/kg BW on the 0% sagebrush diet to 9 g/kg BW on the 30% sagebrush diet with a concomitant decline in apparent dry matter digestibility of from 59 to -3% in this study. The greater effect of sagebrush on
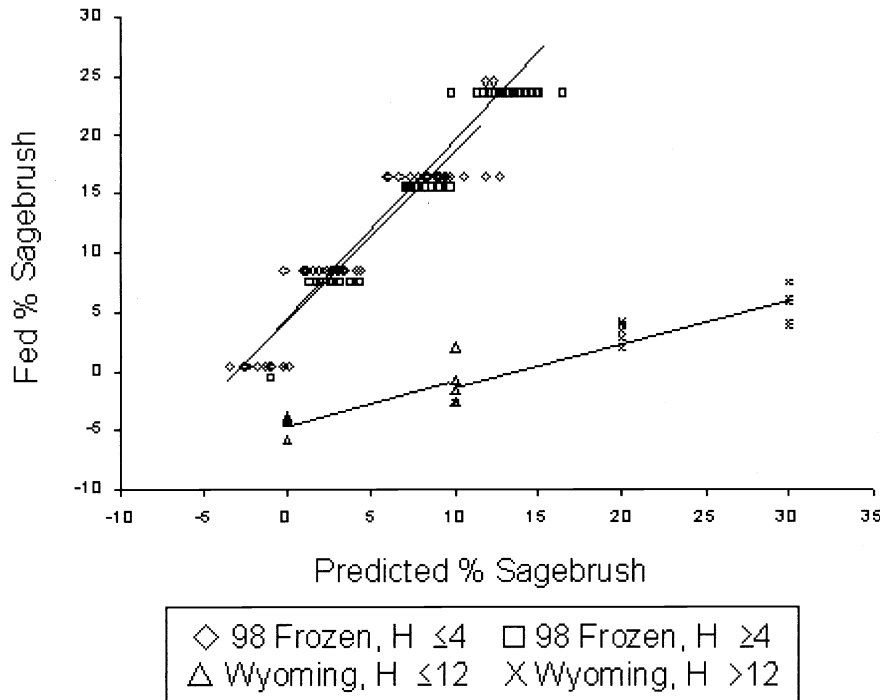
intake and digestibility was associated with a much larger average H and lower slope than the 1998 predictions. This may indicate that when fecal NIRS is used to predict the percent of a chemically defended plant in the diet, the spectra on which the calibrations are based may result from the effect of secondary plant metabolites on the digestive process rather than the plant residues in the feces. If this is the case, the loss of accuracy for independent validations noted in this study may be somewhat lessened when fecal NIRS is used to predict the diets of grazing herbivores because animals normally will limit their intake of chemically defended plants below a level that restricts the rate of nutrient capture (Provenza 1995.)

Conclusions

Relative to the design of future trials for developing fecal NIRS calibration equations, several recommendations can be made from these results. Variation in the source of a target plant species, the background forages, and the animals used for the trial should be maximized to the extent possible. The composition of diets should consist of the 'high' level of the target plant that is slightly above the amount expected to be selected by free grazing herbivores, the 'low' level of about 5% of the target plant, and a diet that consists of only the background forage. Furthermore, we recommend the use of MPLS calibrations to increase the spectral information considered in the calibration equation compared to the limited number of wavelengths that are used in SW calibrations.

When feeding trials are conducted to develop fecal NIRS calibration equations, the samples used for calibration rarely can be considered to be from the same population as the predicted samples. Even if the target plant species is well represented in the diets, other botanical components normally will not be, and the potential associative effects on digestion indicate that assumptions about the populations being the same would be tenuous.

Furthermore, the H statistic may not always indicate if a sample is from the same population as the calibration set. Therefore, until evidence to the contrary is available, predictions of dietary botanical composition using fecal NIRS should be considered to represent an interval scale of measurement. This scale of measurement

Fig. 5. Effect of size of H on the validation accuracy.

is adequate for most problems of interest on rangelands such as comparing treatment effects on diet preference or measuring the phenotypic variation in preference for a target plant in a population of animals, but would not be sufficient if it is necessary to know the actual composition of the diet.

The reduced information in the interval scale of measurement compared to a ratio scale of measurement will be more than offset by the ability to greatly increase sample size by using fecal NIRS compared to other methods for determining botanical composition of diets. For instance, Harniss et al. (1975) reported that for sheep grazing sagebrush-bunchgrass rangelands, 150 fistula diet determinations would be necessary to estimate the percent of plant species that comprise 20% of the diet to within ±10% of the mean with a 95% probability. Such sampling intensity exceeds the practical capacity of any other available method for determining botanical composition of diets except fecal NIRS.

Literature Cited

Barnes, R.J., M.S. Dhanoa, and S. J. Lister. 1989. Standard normal variate transformation and detrending of near-infrared diffuse reflectance spectra. Appl. Spectroscopy. 43:772–777.

Coleman, S.W., R.E. Barton, II, and R.D. Meyer. 1985. The use of near-infrared reflectance spectroscopy to predict species composition of forage mixtures. Crop Sci. 25:834–837.

Coleman, S.W., S. Christiansen, and J.S. Shenk. 1990. Prediction of botanical composition using NIRS calibrations developed from botanically pure samples. Crop Sci. 30:202–207.

Coleman, S.W., H. Lippke, and M. Gill. 1999. Estimating the nutritive potential of forages. p. 647–695. In: H.G. Jung and G.C. Fahey, Jr. (eds.) Proc. V[th] Int. Symp. on the Nutrition of Herbivores. San Antonio, Tex.

Coleman, S.W., J.W. Stuth, and J.W. Holloway. 1989. Vol. 2 p. 881-882. In: Monitoring the nutrition of grazing cattle with near-infrared analysis of feces. Proc. XVI Int.. Grassld. Congr. Nice, France.

Coleman, S.W., J.W. Stuth, and J.W. Holloway. 1995. Prediction of intake by near-infrared spectroscopic analysis of fecal samples. p. 145–155. In: F.N. Owens, D. Gill, K. Lusby, and T. McCollum (Ed.) Symp: Intake in Feedlot Cattle. Oklahoma Exp. Sta. P-942.

Fraker, M.J. 1999. Examination of digestive and behavioral mechanisms to explain individual variation in sagebrush consumption by

sheep. M.S. Thesis, Univ. Idaho, Moscow, Ida.

Harniss, R.O., D.A. Price, and D.C. Tomlin. 1975. Number of fistula samples needed for determination of sheep diet on sagebrush-grass range. J. Range. Mange. 28:417–419.

Leite, E.R. and J.W. Stuth. 1995. Fecal NIRS equations to assess diet quality of free-ranging goats. Small Ruminant Res. 15:223–230.

Lyons, R.K. and J.W. Stuth. 1992. Fecal NIRS equations for predicting diet quality of free-ranging cattle. J. Range Manage. 45:238–244.

Moore, K.J., C.A. Roberts, and J.O. Fritz. 1990. Indirect estimation of botanical composition of alfalfa-smooth bromegrass mixtures. Agron. J. 82:287–290.

Ngugi, R.K., F.C. Hinds, and J. Powell. 1995. Mountain big sagebrush browse decreases dry matter intake, digestibility, and nutritive quality of sheep diets. J. Range Manage. 48:487–492.

Petersen, J.C., F.E. Barton, II, W.R. Windham, and C.S. Hoveland. 1987. Botanical composition definition of tall fescue-white clover mixtures by near infrared reflectance spectroscopy. Crop Sci. 27:1077–1080.

Pitman, W.D., C.K. Piacitelli, G.E. Aiken, and F.E. Barton II. 1991. Botanical composition of tropical grass-legume pastures estimated with near-infrared reflectance spectroscopy. Agron. J. 83:103–107.

Provenza. F.D. 1995. Postingestive feedback as an elementary determinant of food preference and intake in ruminants. J. Range Manage. 48:2–17.

Shenk, J.S. 1989. Monitoring analysis results. p. 27–28. In: G.C. Maren, J.S. Shenk, and F.E. Barton II (eds.). Near Infrared Reflectance Spectroscopy (NIRS): Analysis of Forage Quality. USDA Agr. Handb. No. 643.

Shenk, J.S. and M.O. Westerhaus. 1991a. Population definition, sample selection, and calibration procedures for near infrared reflectance spectroscopy. Crop Sci. 31:469–474.

Shenk, J.S. and M.O. Westerhaus. 1991b. Population structuring of near infrared spectra and modified partial least squares regression. Crop Sci. 31:1548–1555.

Shenk, J.S., M.O. Westerhaus, and S.M. Abrams. 1989. Supplement 2. Protocol for NIRS calibration: Monitoring analysis results and recalibration. p. 104–110. In: G.C. Maren, J.S. Shenk, and F.E. Barton II (eds.). Near Infrared Reflectance Spectroscopy (NIRS): Analysis of Forage Quality. USDA Agr. Handb. No. 643.

Snowder, G.D., J.W. Walker, K.L. Launchbaugh, and L.D. Van Vleck. 2001. Genetic and phenotypic parameters for dietary selection of mountain big sagebrush (Artemisia tridentata Nutt. ssp. vaseyana [Rydb] Beetle) in Rambouillet sheep. J. Anim. Sci. 79:486–492.

USDA-NRCS, Grazing Lands Technol. Inst. 1999. Technical Support for the NIRS/NUTBAL Nutritional Management System.

Van Soest, P.J. 1982. The nutritional ecology of the ruminant. O & B Books, Corvallis, Ore. 374 pp.

Volesky, J.D. and S.W. Coleman. 1996. Estimation of botanical composition of esophageal extrusa samples using near infrared reflectance spectroscopy. J. Range Manage. 49:163–166.

Walker, J.W., D.H. Clark, and S.D. McCoy. 1998. Fecal NIRS for predicting percent leafy spurge in diets. J. Range Manage. 51:450–455.