https://doi.org/10.22319/rmcp.v16i2.6637

Technical note

Estimation of the chemical composition of grains and protein meals by spectroscopy (NIRS-FTIR)

Ricardo Basurto Gutiérrez [a]*

Ericka Ramírez Rodríguez [a]

Gerardo Mariscal Landín [a]

[a] Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias. Centro Nacional de Investigación Disciplinaria en Fisiología y Mejoramiento Animal, km 1 Carretera a Colón, Ajuchitlán, 76280, Colón, Querétaro, México.

*Corresponding author: basurto.ricardo@inifap.gob.mx

Abstract:

This study aimed to use near-infrared reflectance spectroscopy (NIRS) to estimate the chemical composition of ingredients for pigs. Samples (n=118) of grains (corn and sorghum) and meals (soybeans and rapeseed) were collected. Samples were analyzed for dry matter (DM), organic matter (OM), crude protein (CP), ethereal extract (EE), neutral detergent fiber (NDF) or acid detergent fiber (ADF), total starch (starch), and gross energy (GE), and were scanned with an FTIR spectrophotometer. This information was used to develop calibration equations (CES) by employing a program package (TQ Analyst; Thermo Fisher Scientific). The selection of the CE was based on the minimization of standard errors of calibration (SEC), cross-validation (SECV) and prediction (SEP), and the maximization of determination coefficients of calibration (R^2cal) and validation (R^2val). The predictive power of the CE was assessed using the RPD (ratio of prediction to deviation) statistic. Except for DM, the statistics for OM (0.99, 0.21, 0.39, 0.999, 0.34, and 7.3), CP (0.99, 0.62, 0.95, 0.998, 0.99, and 28.2), EE (0.98, 0.15, 0.31, 0.963, 0.22, and 3.7), NDF (0.98, 0.60, 1.82, 0.962, 1.38, and 3.6), ADF (0.99, 0.70, 1.43, 0.979, 1.04, and 4.6), starch (0.99, 2.31, 3.34, 0.994, 3.03, and 9.7), and GE (0.98, 22.8, 34.9, 0.976, 36.0, and 4.2), R^2cal, SEC, SECV, R^2val,

SEP, and RPD, respectively, support the conclusion that CES have an acceptable predictive capacity of the chemical composition of pig feed ingredients.

Keywords: Spectroscopy, Feed, Pigs.

Received: 03/02/2024

Accepted: 22/04/2025

The tables of the National Research Council and the Institut Nationtal de la Recherche Agronomique report the average chemical composition of the ingredients commonly used in the formulation of rations; however, due to different agronomic, environmental or climate factors under which the plants are grown, the composition of the local ingredients may differ substantially from what the nutritional tables show. Although using tabular values allows for quick estimates of chemical composition, their use would be limited to common ingredients and ingredients of known composition[1]. This makes it imperative to use the laboratory to know the actual chemical composition of local ingredients in order to know the quality of the inputs that are acquired, and to formulate rations that precisely meet the nutritional requirements of the animals. This will undoubtedly have an impact on improving feed efficiency and reducing feed costs.

Nonetheless, the costs of chemical analysis and the need to frequently use the laboratory with the entry of each new batch of ingredients can substantially increase production costs. A viable option is the analysis of ingredients using NIRS technology, which has several advantages since, if there is a large collection of samples, it is possible to generate a calibration equation that accurately predicts the chemical composition of the ingredients expeditiously and at a lower cost[2]. Thus, NIRS technology has predicted the moisture, crude protein, and EE contents of ground or whole soybeans[3] or the chemical composition (ash, crude protein, EE, fiber fractions, lignin, and starch) of sorghum grains[4] with excellent accuracy. In addition, with NIRS technology, it is possible to preselect intact sorghum grains for the starch, amylose, or amylopectin contents of grains[5]. NIRS technology has usually been used to determine the composition of specific ingredients, e.g., sorghum[4,5,6] or soybeans[3]; nevertheless, even though the set of ingredients is broad with respect to the matrix, i.e., several ingredients are included, robust calibrations that predict the chemical contents accurately have been obtained[2,7]. These capabilities of NIRS technology have been widely exploited in the feed processing industry as spectroscopy allows various constituents from a large number of samples to be rapidly evaluated. Finally, there are few studies aimed at the analysis of ingredients for use in the pig industry.

The objective was to develop calibrations by using near-infrared reflectance spectroscopy (NIRS) based on FT-IR technology to estimate the chemical composition of ingredients intended for pig feeding. From various states of Mexico, 118 samples of ingredients used for pig feed were collected, of which 28, 30, 30, and 30 samples corresponded to corn grain, sorghum grain, rapeseed meal, and soybean meal, respectively. The samples were taken to the Animal Nutrition Laboratory of the National Center for Research in Animal Physiology and Improvement for processing.

Each ground sample was placed in a cup (4 inches in diameter) with a quartz window. Thus, the samples were exposed to a Nicolet 6700 FT-IR spectrophotometer (Thermo FisherScientific, 2000-2007[8]) equipped with an integrating sphere to generate the absorbance spectra in the electromagnetic region between 4,000 and 10,000 waves cm^{-1} at a resolution of 4 nm and with 72 scans. With the OMNIC™ 8.0 Software (Thermo Fisher Scientific, 2000-2007[8]), the absorbance values were transformed to log(1/Reflectance) and then the spectra were stored in the computer.

The individual calibrations for each analytical component were obtained through the use of the TQ Analyst program (v8.0; Thermo Fisher Scientific, 2000-2007[8]) and a multivariate partial least squares (PLS) model. The mathematical pretreatment of the spectra was based on the degree of the derivative (first or second) and on the Savitzky-Golay filter, which aims to reduce (smooth) the spectral noise. This filter is based on the number of points in the segment (NP) and the order of the polynomial (OP) to be smoothed. Likewise, the samples were divided into two groups: one set was used for the development of the model, calibration group (CG), and the second set to validate the model, validation group (VG). To ensure uniformity of the dispersion of the chemical composition between the sets, the samples were ordered in ascending order within ingredients. Two samples were selected for the CG, and the next sample was integrated into the VG, repeating this procedure until the end of the list[9,10].

The selection of the calibration equation or model for each component was based on maximizing the determination coefficient of the calibration (R^2cal) and minimizing the standard error of the calibration (SEC), which evaluates the accuracy of the model[11]:

$$SEC = \sqrt{\frac{\sum_{=ij}(\hat{y}ij - yij)^2}{Ncal}}$$

Where: $\hat{\mathbf{y}}_i$ is the predicted value of the *i-th* sample of the CG, $\mathbf{y}_i$ is the actual or laboratory value of the *i-th* sample of the CG, and $Ncal$ is the number of observations of the CG.

In a similar approach to evaluate the prediction power of the model[12], the maximization of the coefficient of determination of the validation (R^2val), and the minimization of the standard error of prediction (SEP) were considered, which was calculated as follows:

$$SEP = \sqrt{\frac{\sum_{i=1}^{Iv}(yi - \tilde{y}i)^2}{N\ val}}$$

Where: $\tilde{\mathbf{y}}_\mathbf{i}$ is the predicted value of the *i-th* sample of the CG, $\mathbf{y}_\mathbf{i}$ is the actual or laboratory value of the *i-th* sample of the VG, and $N\ val$ is the number of observations of the VG.

In addition, a second approximation of the internal validation of the model[11,13,14] was performed, which consists of determining the standard error of cross-validation (SECV) and removes some of the bias when the program calculates the number of recommended factors (Thermo Fisher Scientifc Inc, 2000-2007[8]). To calculate SECV, the calibration procedure is performed without one of the samples in the set and the statistics are estimated; the process is repeated as many times as the number of samples in the calibration set. At the end, the average of the squared deviations is calculated and the square root is applied. The equation of the SECV[12] is the following:

$$SECV = \sqrt{\frac{1}{Ic - 1}\sum_{i=1}^{Ic}(\hat{y}i - yi)^2}$$

Where: $\hat{\mathbf{y}}_\mathbf{i}$ is the value predicted with the *i-th* observation removed, $\mathbf{y}_\mathbf{i}$ is the actual or laboratory value of the *i-th* observation, and ***Ic*** is the number of observations in the calibration set.

The TQ Analyst v8.0 program (Thermo Fisher Scientific Inc.) limits the maximum number (10) of factors or components in the equation. In addition, the program selects the model with the lowest SECV value.

To assess the robustness and accuracy of calibration equations, RPD [ratio of (standard error) prediction to (standard) deviation], specifically the relationship between the standard error of prediction (SEP) and the standard deviation of VG (SD val)[9,15,16,17]:

$$RPD = \frac{SD\ val}{SEP}$$

Where, SD val= standard deviation of laboratory values for VG, and SEP= standard error of prediction (SEP).

Thus, if RPD < 2.3, the model is unreliable and not recommended for use; if RPD > 2.4 and < 3.0, it indicates that the model has poor predictive power and is only useful for a very rough selection; if RPD >3.0 and >4.9, it means that the calibration has a good predictive capacity and can be used for an acceptable selection; if RPD is between 5.0 and 6.4, it suggests that calibration has good predictive power and can be used for quality control processes; if RPD is >6.5 and > 8.0, it indicates that the calibration equation has a very good predictive capacity and can be used for analytical purposes or control processes[9].

With the procedures and functions of the SAS statistical package[18], the slope between the actual and the predicted values of the VG was evaluated, under the null hypothesis that the slope is equal to one, using the REG procedure and the PROBT function. In addition, as a measure of the bias of the NIRS prediction, it was assessed whether the expected difference between the actual and predicted values of the VG differs from zero with the use of the TTEST procedure and the PAIRED function. Finally, the normality (Shapiro-Wilk) of the distribution of prediction errors was evaluated with the UNIVARIATE procedure.

In the laboratory, the samples were ground with a Wiley mill (Thomas Scientific, Swedesboro, NJ) through a 1 mm Ø screen for storage, and then analyzed for dry matter at 100 °C (DM; method 967.03, AOAC[19]), ash (ASH, method 942.05, AOAC[19]; 600°C x 4 h); organic matter (OM; calculated as OM=100-ASH), crude protein (CP; method 976.05, AOAC[19]), ethereal extract (EE; AOAC method 920.85, AOCS[20]), fiber fractions (Van Soest *et al*[21]): neutral (NDF) and acid detergent fiber (ADF), reported on an organic matter basis. The total starch content was determined with the KTSTA Megazyme-L kit (AOAC[19], method 996.11). The determination of the gross energy content (GE, kcal/kg) was performed using an adiabatic bomb calorimeter (model 1281; Parr, Moline, IL, USA).

Table 1 shows the descriptive statistics of the study variables of the total samples collected (T). Several databases were consulted: National Animal Nutrition Program[22], NRC[23], Tejada[24], Ramírez *et al*[6], and Feedipedia[25], which report the number of samples analyzed, the mean, and the standard deviation for DM, OM, CP, EE, fiber fractions, starch, and GE of the ingredients analyzed in the present study. Most of the means of components of the ingredients of the present study were different (P<0.05) from the means of Feedipedia[25], except for the means for OM, EE, and starch of corn grain and OM of rapeseed meal. Ramírez *et al*[6] analyzed 216 samples of sorghum grain for CP (mean= 8.2 % and a standard deviation of 1.09 %); the mean of this group of samples was 0.88 % percentage units lower (P<0.05) than the mean CP contents of the sorghum grains in the present study. These comparisons could suggest that tabular values should be used with caution since several factors can explain the differences, such as different analysis techniques between laboratories, differences in the

varieties available over time, and progress in the genetic improvement of the varieties and agronomic management, among others.

Table 1: Descriptive statistics for the chemical composition of grains (corn and sorghum) and protein meals (soybeans and rapeseed) used for the development of the NIRS equations

Variable	Chemical analysis			
	Number	**Mean**	**Range**	**SD**
DM	117	90.11	8.24	1.75
OM	115	95.90	8.59	2.86
CP	114	24.21	41.78	17.20
EE	111	2.04	4.44	1.15
NDF	117	12.00	20.16	6.60
ADF	117	8.34	22.86	6.61
Starch	117	31.40	75.66	31.52
GE	115	4115	601	148

SD= standard deviation; DM= dry matter (%); OM= organic matter (%); CP= crude protein (%); EE= ethereal extract (%); NDF= neutral detergent fiber (% OM); ADF= acid detergent fiber (% OM); total starch (% DM); GE= gross energy (kcal/kg).

The DM, OM, CP, EE, NDF, ADF, starch, and GE components of corn and sorghum grains and rapeseed and soybean meals for CG and VG groups are shown in Table 2. There were no differences ($P>0.05$) in the mean value and the variance between the sets; this suggests that the selection method allowed to get two sets, CG and VG, statistically similar.

Table 2: Descriptive statistics for the actual chemical composition of grains (corn and sorghum) and protein meals (soybeans and rapeseed) by calibration group (CG) or validation group (VG) used in the development of the NIRS equations

Variable	Chemical analysis				
	Group	Number	Mean	Range	SD
DM	CG	79	90.21	8.24	1.76
	VG	38	89.92	7.02	1.72
OM	CG	77	95.81	8.53	2.90
	VG	38	96.08	7.42	2.83
CP	CG	78	24.04	41.78	17.16
	VG	36	24.58	41.45	17.53
EE	CG	75	2.06	4.12	1.15
	VG	36	1.99	4.35	1.16
NDF	CG	79	12.04	20.16	6.68
	VG	38	11.91	18.63	6.52
ADF	CG	79	8.34	22.30	6.63
	VG	38	8.35	19.71	6.64
Starch	CG	78	31.17	75.66	31.34
	VG	38	30.75	70.95	31.97
GE	CG	79	4,114	601	148
	VG	36	4,116	456	148

SD= standard deviation; DM= dry matter (%); OM= organic matter (%); CP=crude protein (%); EE= ethereal extract (%); NDF= neutral detergent fiber (% OM); ADF= acid detergent fiber (% OM); total starch (% DM); GE= gross energy (kcal/kg).

The statistics of the calibration equations developed are shown in Table 3. Except for the model for the dry matter variable, it is noteworthy that the calibration models explained more than 96 % of the variation. The calibration model for dry matter explained only 79 % of the variation in the samples.

Table 3: Statistics of NIRS calibration and validation equations for estimating the chemical composition of ingredients used in pig nutrition

	Calibration equation			Internal calibration	Validation			
Component	**N**	**R^2cal**	**SEC**	**SECV**	**N**	**R^2val**	**SEP**	**RPD**
DM	79	0.98	0.26	1.15	38	0.791	0.99	1.74
OM	77	0.99	0.21	0.39	38	0.999	0.34	7.26
CP	78	0.99	0.62	0.95	36	0.998	0.99	28.24
EE	75	0.98	0.15	0.31	36	0.963	0.22	3.74
NDF	79	0.99	0.60	1.82	38	0.962	1.38	3.58
ADF	79	0.99	0.70	1.43	38	0.979	1.04	4.64
Starch	78	0.99	2.31	3.34	39	0.994	3.03	9.67
GE	79	0.98	22.80	34.9	36	0.976	36	4.24

[1]N= number of samples; [2]R^2cal= coefficient of determination of the calibration equation; SEC= standard error of calibration; SECV= standard error of cross-validation; R^2val=coefficient of determination of the validation equation; SEP= standard error of prediction; RPD= ratio of prediction error to deviation; DM= dry matter (%); OM= organic matter (%); CP= crude protein (%); EE= ethereal extract (%); NDF= neutral detergent fiber (% OM); ADF= acid detergent fiber (% OM); total starch (% DM); GE= gross energy (kcal/kg).

The statistic of the cross-validation (SECV) is shown in Table 3. According to Williams[9], SECV is useful for detecting atypical samples or outliers during the calibration process. If SEP is large but smaller than SEC, this indicates the existence of atypical samples that should be identified and removed before recalibration. In the present study, the SECVs[9] were higher than the SEC of each of the calibrations. Likewise, the SECV[9] indicates that it is useful to optimize the mathematical pretreatment of the spectra by determining the best model of the calibration models evaluated, the best model being the one with the highest value of R^2cal, SEP, and RPD.

Table 4 shows the factors and algorithms that were used in the mathematical pretreatment of the spectra of the selected models, and which were considered to be the best in the present study. The analysis of the variation contained in the spectra determines the orthogonal factors or independent sources of variation[8] that explain the largest proportion of the variation. In addition, it was established that the number of factors in the model should not be greater than 10 to avoid overparameterization of the model[26]. The SEP expresses the difference between the predicted values and the actual values of the validation samples, so low values are related to models of greater precision[27]. Before generating the calibration models, several mathematical procedures were used to process the spectra to correct the dispersion, baseline intensity, and peak overlap to improve the accuracy of the calibration models[28].

In the present study, the Saviztky-Golay filter and the multiplicative scatter correction and standard normal vary procedures were used. The selection of methods for spectra correction was carried out empirically with trial and error, evaluating the accuracy of the calibration and

the number of factors in the calibration[17,28]. Table 4 shows the values of the filters and methods used for the calibration of each component and the values of the predicted residual errors sum of squares (PRESS).

Table 4: Factors and mathematical pretreatment of spectra in the development of NIRS calibration equations and PRESS statistics

Component	Derivative[1]	Filter[2]	NP[3]	OP[4]	Factors[5]	PRESS[6]
DM[7]	1	SG	7	5	7	101.00
OM[8]	1	SG	5	1	6	13.72
CP[9]	1	SG	35	5	8	71.89
EE[10]	1	SG	9	2	9	6.98
NDF[11]	1	SG	7	3	7	268.58
ADF[12]	1	SG	11	3	7	154.62
Starch[13]	1	SG	21	3	7	818.06
GE[14]	1	SG	13	3	10	84.60

[1]Derivative=1 first derivative; 2= Savitzky-Golay (mathematical filter);[3]NP= number of points in the segment; [4]OP= order of the polynomial; [5]Factors= independent components of the equation; [6]PRESS= predicted residual errors sum of squares; [7]DM= dry matter (%); [8]OM= organic matter (%); [9]CP= crude protein (%); [10]EE= ethereal extract (%); [11]NDF= neutral detergent fiber (% OM); [12]ADF= acid detergent fiber (% OM); [13]total starch (% DM); [14]GE= gross energy (kcal/kg).

Ideally, the predicted value should vary by one unit when the actual value in the samples changes by one unit, i.e., the slope of the model would be equal to one. Therefore, the slope of linear regression between the actual or laboratory and the predicted values for each component was evaluated through a t-test with a confidence level of 95 % (Table 5). If the slope is statistically different from one, it suggests that the calibration model has a bias in the estimation[29,30]. In the actual test, the regression slopes between the actual and predicted values for DM and NDF were detected to be different from one, indicating that there is a bias, specifically in these cases, the underestimation of DM and NDF.

Table 5: 95 % confidence limits of the slope between actual (XX_{LAB}[1]) and predicted values (XX_{NIRS}[2]) and the t-test to determine whether the slope (ρ) differs statistically from one

Component	ρ[3]	SEM[4]	*t* cal[5]	LCL[6]	UCL[7]	Valor *p*[8]	Opn[9]
DM_{LAB} - DM^{10}_{NIRS}	0.66	0.077	4.36	0.51	0.82	0.001	H_A: ρ≠1
OM_{LAB} - OM^{11}_{NIRS}	0.96	0.019	1.99	0.92	1.00	0.055	H_0[1]
CP_{LAB} - CP^{12}_{NIRS}	1.01	0.010	0.79	0.99	1.03	0.434	H_0
EE_{LAB} - EE^{13}_{NIRS}	0.94	0.031	1.81	0.88	1.01	0.079	H_0
NDF_{LAB} - NDF^{14}_{NIRS}	0.91	0.032	2.88	0.84	0.97	0.007	H_A: ρ≠1
ADF_{LAB} – ADF^{15}_{NIRS}	0.95	0.025	1.90	0.90	1.00	0.065	H_0
$Starch_{LAB}$ - $Starch^{16}_{NIRS}$	0.98	0.015	1.30	0.95	1.01	0.202	H_0
GE_{LAB} - GE^{17}_{NIRS}	0.92	0.040	1.97	0.84	1.00	0.056	H_0

[1]XX_{LAB}= actual value of the sample; [2]XX_{NIRS}= value estimated by NIRS; [3]ρ= slope; [4]SEM= standard error of the mean, [5]*t* cal= value of the t-statistic; [6]LCL= lower confidence limit; [7]UCL= upper confidence limit; *P*8= value of α at which H_0: ρ=1 is rejected; [9]Opn= opinion: H_0= insufficient evidence to reject the null hypothesis (ρ=1); H_A: ρ≠1. [10]DM= dry matter (%); [11]OM= organic matter (%), [12]CP= crude protein (%), [13]EE= ethereal extract (%); [14]NDF= neutral detergent fiber (% OM); [15]ADF= acidic detergent fiber (% OM); [16]total starch (% DM); [17]GE= gross energy (kcal/kg).

Table 6 presents the comparison of the means of the actual and predicted values using a paired mean test. In none of the comparisons was it detected (*P*>0.05) that there were differences between the means of the actual and predicted values. According to John *et al*[30], this suggests that the models are applicable for prediction. On the other hand, the RPD statistic aims to relate the standard error of prediction (SEP) to the standard deviation of the CG[9,17], with which the predictive capacity of the calibration model can be evaluated. Thus, models with RPD values < 2.3 are considered to have a low predictive power, so calibration to estimate DM (RPD=1.7) cannot be recommended for use. In contrast, calibration models for OM, CP, and starch have very good predictive power because all their statistics (R^2 and RPD and SEP and SECV) have high and low values, respectively. The models for EE, GE, NDF, and ADF have good predictive power.

Table 6: Paired mean test to evaluate the difference between methods (XX_{LAB}[1] *vs.* XX_{NIRS}[2]) to determine the chemical contents of ingredients used in pig nutrition

Component	Paired means (percentage units)					
	Difference	SEM[3]	LCL[4]	UCL[5]	Df[6]	α-value[7]
DM_{LAB} - DM_{NIRS}[8]	-0.031	0.160	-0.36	0.29	37	0.85
OM_{LAB} - OM_{NIRS}[9]	-0.045	0.056	-0.16	0.07	36	0.43
CP_{LAB} - CP_{NIRS}[10]	-0.211	0.164	-0.54	0.12	35	0.21
EE_{LAB} - EE_{NIRS}[11]	0.019	0.037	-0.06	0.09	35	0.62
NDF_{LAB} - NDF_{NIRS}[12]	0.040	0.228	-0.42	0.50	37	0.86
ADF_{LAB} - ADF_{NIRS}[13]	-0.049	0.189	-0.39	0.29	38	0.78
$Starch_{LAB}$ - $Starch_{NIRS}$[14]	0.248	0.623	-1.01	1.51	38	0.69
	Paired means (kilocalories / kg)					
GE_{LAB} - GE_{NIRS}[15]	-0.954	6.086	-13.31	11.40	35	0.88

[1]XX_{LAB}= actual value of the sample; [2]XX_{NIRS}= value estimated by NIRS; [3]SEM= standard error of the mean; [4]LCL= lower confidence limit; [5]UCL= upper confidence limit; [6]Df= degrees of freedom of the test; [7] α-value= at which H_0: difference between means is not different from 0, is rejected; [8]DM= dry matter (%); [9]OM= organic matter (%), [10]CP= crude protein (%), [11]EE= ethereal extract (%); [12]NDF= neutral detergent fiber (% OM); [13]ADF= acid detergent fiber (% OM); [14]total starch (% DM); [15]GE= gross energy (kcal/kg).

Table 3 shows the statistical evaluation (null hypothesis= 1) of the regression slopes between the actual (laboratory) and the predicted values of each chemical component in the validation samples. It was detected that the slopes of the equations for DM and NDF were statistically (P<0.01) different from 1 (0.662 and 0.907, respectively), indicating that the equations are underestimating the actual value of the validation samples.

Based on the modified partial least squares (MPLS) algorithm, NIRS prediction equations were developed to estimate the protein and starch contents, among other components, of accessions of cowpea (*Vigna unguiculata*) crops[10]. The statistics of the calibration equations were R^2= 0.800 and 0.997 and their standard calibration errors were 1.23 and 0.063; in contrast, the external validation statistics were R^2val= 0.903 and 0.997, SEP= 0.598 and 0.528, and RPD= 2.80 and 5.32 for CP and starch, respectively; the authors[10] concluded that their calibrations have good predictive power. Nonetheless, considering the RPD ranges proposed by Williams[9], the calibration model for crude protein would have poor predictive power[10]. Noel *et al*[7] collected (n= 858) feed samples that were grouped into three different types: cereals, cereal byproducts, and rations for pigs, to build near-infrared spectroscopy-based calibration equations to estimate CP, EE, NDF, and ADF contents; the equations developed were either for each type of ingredient set or a single general equation that included all ingredients. Calibration equations built with only cereal samples (n= 212), which included wheat, barley, rye, sorghum, rice, and triticale grains, had acceptable statistics, R^2cal (0.95 and 0.97) and SEC (0.04 and 0.08 %) for CP and EE, respectively. In contrast, the calibration equations[7] for starch, NDF, ADF, and gross energy explain relatively low

percentages of the variation (R^2= 0.75, 0.89, 0.86, and 0.45, respectively) and their SECs (3.0 %, 1.24 %, 0.73 %, and 59.7 kcal/kg, respectively) were similar to those of the present study. Nevertheless, only the validation R^2 (0.91, 0.89, 0.46, 0.76, 0.49, and 0.46, respectively) and SEP (0.55 %, 0.40 %, 3.39 %, 1.5 %, 0.99 %, and 62.11 kcal/kg, respectively) for crude protein, EE, starch, NDF, ADF, and GE, respectively, with only the statistics for the variables of CP and EE contents being acceptable.

In the present study, the slopes between the actual and predicted values (Table 5), the slopes for DM and NDF were less than one (underestimate), indicating that the increase of one percentage unit of dry matter or NDF in the sample than the increase in the predicted value; however, for dry matter, the underestimation is much higher. In contrast, other authors[7] reported that the slopes for crude protein and gross energy were similar to one (1.01 and 0.99, respectively); the rest of the slopes were probably different from one (1.07, 0.87, 1.10, and 0.80 for EE, starch, NDF, and ADF, respectively) since they do not show the slope confidence intervals. Considering the classification of RPD to measure the predictive power of the calibration equations[9], only the equation for CP[7] would have good predictive capacity (3.26); however, this RPD value is much lower than reported in the present study (28.24). For EE and NDF[7] calibrations, RPDs of 2.35 and 2.49, respectively, were reported, indicating that calibrations are useful for low- or high-content selection. The calibrations for EE and NDF in the present study have good predictive power according to RPD values of 3.74 and 3.58, respectively. When Noel *et al*[7] built a general calibration equation for all their groups, the increase in the predictive capacity (RPD= 3.54 to 5.11) of their equations improved substantially, except for the estimation of GE (RPD= 2.27).

The number of samples, the diversity of the matrices, or a wide variation in the nutritional contents of the samples determine the construction of calibration equations with good predictive power. For instance, the standard error of the validation decreased curvilinearly when the calibrations included 15, 21, 36, and 43 % of the available samples[31]. These authors[31] suggested that the increase in variation was due to the overparameterization of the model when the number of samples in the calibration set was too small; they[31] recommended that calibrations include more than 100 samples so that the error and bias are close to the values obtained with the total samples (n= 467). Although other authors[32] point out that determining the number of samples to build calibration models is not easy, they mention that calibrations based on sets with fewer than 50 samples may not be reliable. Likewise, the number of samples required for calibration also depends on the type of population on which inferences are to be made; for example, in a closed population that includes only samples of the same type (e.g., sorghum grains), 50 to 100 samples may be needed. In an open population, as in the present study, it suggests that it is necessary to include at least 150 samples to build reliable calibrations [31]. Although the calibration sets included 75 to 79 samples, the calibration equations generated have good predictive power, which could be explained by the difference in the technology of the equipment used: monochromators[7,31]

vs. the FTIR of the present research. In a study with 328 samples[33], they developed a calibration equation to estimate the CP contents of five cereal grains (corn, sorghum, and three species of millets), and the sample set had a range of CP contents between 6.0 and 21.5 %. Using the partial least squares model, these authors[33] report that the statistics of their calibration equation were R^2cal= 0.90 and SEP= 0.9, and for the validation test: R^2val= 0.86, with a slope of 0.82, SEP= 1.09, and an RPD= 3.08, concluding that their calibration has an excellent quantitative predictive power. Although the present study used a smaller number of samples with a wider range of values (41 percentage units) and different matrices, most of the statistics in this study were better (R^2cal= 0.99 and SEP= 0.62) for the calibration and the validation equations (R^2val= 0.99 and SEP= 0.62 and RPD= 28.24); only the prediction error of the present research was greater (SEP= 0.91 *vs*. 0.99) than that of the study mentioned previously study[33]. In another study with 400 samples[2], calibration equations were developed and evaluated in the region of 4,000 to 25,000 waves cm^{-1} to determine the chemical composition of a wide group of ingredients and rations, protein and energy supplements, grains and protein meals; Cozzolino[2] reported the following statistics, R^2, SECV and *SD val*/SECV (0.91, 0.82 %, and 2.56; 0.98, 1.3 %, and 6.6; 0.93, 1.4, and 3.45; 0.80, 5.3 %, and 1.81; 0.92, 1.3 %, and 2.94) for dry matter, crude protein, ADF, NDF, and ash contents, respectively, and concluded that NIRS is a technique that can determine crude protein, ADF, and ash adequately; however the calibration for dry matter could be useful only to select samples with high or low dry matter contents, and finally, the technology could not estimate the NDF of the samples. Something similar happened in the present study. The calibration equation to estimate the dry matter has a poor predictive power (R^2cal= 0.98 and SEC= 0.26, R^2val= 0.79, SEP= 0.99, and RPD= 1.74), so its use is not recommended.

Likewise, in another study that was carried out to use NIRS technology to predict the composition of ileal digesta and feces from pigs[34], the calibration equation had no predictive power (R^2cal= 0.87; SEC= 0.83, SEP= 1.15, and RPD= 1.78). This inability of NIRS technology in the estimation of dry matter could be explained by the fact that the presence of water produces two broad and dominant absorption bands in the regions 5,180 and 6,945 waves/cm^{-1} in the NIRS spectrum[35], which explains why the predictive capacity for other chemical components in fresh samples decreases[36]. Even in the spectra of dry samples at 55°C, the two water-related bands are visible, so it is not easy to explain why the predictive power of NIRS technology for dry matter is low, being specific to moisture. Other components have also not been accurately predicted. Ramírez *et al*[6] evaluated the use of NIRS to predict the dry matter, crude protein, and neutral detergent fiber contents of 216 sorghum grain samples collected over three years. The prediction equations explained 96 %, 93 %, and 74 % of the variation in the population studied and with SEP (0.372, 0.186, and 0.838) for dry matter, crude protein, and neutral detergent fiber, respectively. Although the authors[6] conclude that NIRS has good predictive capacity, the percentage of the variation explained for NDF could question their conclusion.

The energy of poultry and pig diets comes primarily from cereal grains, with starch being the main constituent of cereals, hence the importance of its determination. However, starch measurement techniques are laborious and expensive, so analytical methods that produce reliable predictions are important to formulate rations for monogastric species. Although the sample population of the present study had two extremes with significant differences in starch contents, the calibration equation for starch explains 99 % of the variation, SEC (2.31 %) and SEP (3.03 %) are low, the slope does not differ statistically from 1, and the RPD (9.67) indicates that it has a strong predictive power (Table 3). In a study with 400 samples of cereals and byproducts (wheat, corn, barley, wheat bran, oats, rye, corn germ, and sorghum), Nieto-Ortega *et al*[37] developed a global calibration equation to estimate the total starch contents with 275 samples and reported that the statistics for the calibration equation were R^2= 0.99 and SEC= 1.53; in contrast, the statistics for the validation equation were R^2val= 0.99, SEP= 1.70, and RPD= 9.53. These statistics reported by these authors[37] are better than those reported in the present study; nevertheless, the RPDs of the calibration equations of both studies indicate that they can be used for quantitative estimation.

Except for dry matter, it is concluded that NIRS-FTIR spectroscopy accurately and confidently predicts the organic matter, crude protein, ethereal extract, gross energy, and neutral detergent fiber and acid detergent fiber fractions of the main energy and protein ingredients used in pig farms.

NIRS is a tool that can be routinely used to make accurate estimates of chemical composition at a low cost and in a short time. The information released can be used to formulate precise rations, which benefit consultants, producers of swine, and, possibly, production units of other monogastric species.

Funding source

The results presented here are part of the project (SIGI: 13202534765) “Mitigation of the climate impact of pig production through feeding and management strategies”, which has fiscal resources from INIFAP.

Literature cited:

1. Paternostre L, Millet S, De Boever J. Comparison of feed tables, empirical models and near-infrared spectroscopy to predict chemical composition and net energy of pelleted pig feeds. Anim Feed Sci Technol 2023;297:115578. https://doi.org/10.1016/j.anifeedsci.2023.115578.

2. Cozzolino D. Uso de la espectroscopia de reflectancia en el infrarrojo cercano (NIRS) en el análisis de alimentos para animales. Agrociencia 2002;6(2):25-32.

3. Takahash M, Hajika M, Igita K, Sato T. Rapid estimation of protein, oil and moisture contents in whole-grain soybean seeds by near-infrared reflectance spectroscopy. In (eds Davies AMC, Williams PC) Near-infrared Spectroscopy: Future Waves. 7th Proc Int Conf Near-infrared Spectrosc 1996:494–497.

4. Ejaz I, He S, Li W, Hu N, Tang C, Li S, Li M, Diallo B, Xie G, Yu K. Sorghum grains grading for food, feed, and fuel using NIR spectroscopy. Front Plant Sci 2021;12:720022. doi:10.3389/fpls.2021.720022.

5. Peiris KHS, Wu X, Bean SR, Perez-Fajardo M, Hayes C, Yerka MK, Jagadish SVK, Ostmeyer T, Aramouni FM, Tesso T, *et al*. Near infrared spectroscopic evaluation of starch properties of diverse sorghum populations. Processes 2021;9;1942. https://doi.org/10.3390/pr9111942.

6. Ramírez RE, Anaya EAM, Mariscal LG. Predicción de la composición química del grano de sorgo mediante espectroscopía de reflectancia en el infrarrojo cercano (NIRS). Téc Pecu Méx 2005;43(1):1-11.

7. Noel SJ, Jørgensen HJH, Knudsen KEB. The use of near-infrared spectroscopy (NIRS) to determine the energy value of individual feedstuffs and mixed diets for pigs. Anim Feed Sci Technol 2022;283:115156. https://doi.org/10.1016/j.anifeedsci.2021.115156.

8. Thermo Fischer Scientific. TQ Analyst. User's Guide. 2000-2007. Thermo Fischer Scientific Inc. Madison, WI, USA.

9. Williams P. Near Infrared Technology: Getting The best out of Light. SUN PRESS. 2019.

10. Padhi SR, John R, Bartwal A, Tripathi K, Gupta K, Wankhede DP, Mishra GP, Kumar S, Rana JC, Riar A, Bhardwaj R. Development and optimization of NIRS prediction models for simultaneous multi-trait assessment in diverse cowpea germplasm. Front Nutr 2022;9:1001551. doi:10.3389/fnut.2022.1001551.

11. Tahmasbian I, Morgan NK, Hosseini BS, Dunlop MW, Moss AF. Comparison of hyperspectral imaging and near-infrared spectroscopy to determine nitrogen and carbon concentrations in wheat. Remote Sens 2021:13:1128. https://doi.org/10.3390/rs13061128.

12. Forina M, Lanteri S, Casale M. Multivariate calibration. J Chromat A 2007;1158(1–2): 61-93. https://doi.org/10.1016/j.chroma.2007.03.082.

13. Stone M. Cross-validatory choice and assessment of statistical predictions. J Royal Stat Soc: Series B (Methodological). 1974;36(2):111-133. https://doi.org/10.1111/j.2517-6161.1974.tb00994.x.

14. Zeaiter M, Rutledge D. Preprocessing methods. In: Brown SD, Tauler, Walczak B, editors. Comprehensive chemometrics. Elsevier. 2009;3.04:121-231. https://doi.org/10.1016/B978-044452701-1.00074-0.

15. Nie Z, Tremblay GF, Bélanger G, Berthiaume R, Castonguay Y, Bertrand A, *et al*. Near-infrared reflectance spectroscopy prediction of neutral detergent-soluble carbohydrates in timothy and alfalfa. J Dairy Sci 2009;92(4):1702-11. doi:10.3168/jds.2008-1599. PMID:19307652.

16. Andueza D, Muñoz F, Murray I. The prediction of chemical composition and *in vitro* digestibility of samples of *Atriplex halimus* by NIR spectroscopy. In: Ben SH, *et al* editors. Nutrition and feeding strategies of sheep and goats under harsh climates. Zaragoza: CIHEAM, 2004:165-168 (Options Méditerranéennes: Série A. Séminaires Méditerranéens; n. 59).

17. Manley, M. Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. Chem Soc Rev 2014;43(24):8200-8214. doi =10.1039/C4CS00062E.

18. SAS Institute. 2002-2010. SAS User's Guide: Statistics. Version 9.4 TS Level 1M7 Edition. SAS Institute Inc., Cary, NC.

19. AOAC. Official methods of analysis. 17th ed. Association of Official Analytical Chemists, Arlington, VA. 2000.

20. AOAC. Official methods of analysis. 18th ed. Association of Official Analytical Chemists, Gaithersburg, Maryland, USA. 2005.

21. Van Soest PJ, Robertson JB, Lewis BA. Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. J Dairy Sci 1991;74(10):3583–3597. doi:10.3168/jds.S0022-0302(91)78551-2.

22. National animal nutrition program. Nat Res Supp Project (NRSP-9), USDA https://animalnutrition.org.

23. National Research Council. United States-Canadian tables of feed composition: Nutritional data for United States and Canadian feeds. Third Rev. Washington, DC: The National Academic Press. 1982. https://doi.org/10.17226/1713.

24. Tejada HI. Control de calidad y análisis de alimentos para animales. México, DF: Editorial Sistema de Educación Continua en Producción Animal; 1992.

25. Feedipedia. Animal feed resources information system. https://feedipedia.org.

26. Belanche A, Weisbjerg MR, Allison GG, Newbold CJ, Moorby JM. Measurement of rumen dry matter and neutral detergent fiber degradability of feeds by Fourier-transform infrared spectroscopy. J Dairy Sci 2014;97(4):2361-75. doi:10.3168/jds.2013-7491. Epub 2014 Feb 6. PMID: 24508438.

27. Parrini S, Acciaioli A, Crovetti A, Bozzi R. Use of FT-NIRS for determination of chemical components and nutritional value of natural pasture. Italian J Anim Sci 2018; 17(1):87-91, doi:10.1080/1828051X.2017.1345659.

28. Díaz-Cruz JM, Esteban ME, Ariño C. Exploratory data analysis. In: Chemometrics in electroanalysis. Monographs in electrochemistry. Springer, Cham: 33-67. https://doi.org/10.1007/978-3-030-21384-8_3.

29. Givens D, De Boever J, Deaville E. The principles, practices and some future applications of near infrared spectroscopy for predicting the nutritive value of foods for animals and humans. Nutr Res Review 1997;10(1):83-114. https://doi:10.1079/NRR19970006.

30. John R, Bhardwaj R, Jeyaseelan C, Bollinedi H, Singh N, Harish GD, *et al*. Germplasm variability-assisted near infrared reflectance spectroscopy chemometrics to develop multi-trait robust prediction models in rice. Front Nutr 2022:946255.doi: 10.3389/fnut.2022.946255.

31. Abrams SM, Shenk JS, Westerhaus MO, Barton FE. Determination of forage quality by near infrared reflectance spectroscopy: Efficacy of broad-based calibration equations. J Dairy Sci 1987;70:806-813.

32. Deaville ER, Flinn PC. Near-infrared (NIR) Spectroscopy: an alternative approach for the estimation of| forage quality and voluntary intake. Givens DI, Owen E *et al*, editors. Forage evaluation in ruminant nutrition. CABI, Pub, New York. 2000:301-320.

33. Chadalavada K, Anbazhagan K, Ndour A, Choudhary S, Palmer W, Flynn JR, *et al*. NIR instruments and prediction methods for rapid access to grain protein content in multiple cereals. Sensors 2022;22:3710. https://doi.org/10.3390/s22103710.

34. Ramos CR, Basurto GR, Ramírez RE, Reis de Souza TC, Mariscal LG. Predicción de la composición química de las heces y digesta ileal de cerdos mediante espectroscopia de reflectancia en el infrarrojo cercano (NIRS). Rev Mex Cienc Pecu 2023;14(3):488-503. doi:https://doi.org/10.22319/rmcp.v14i3.6175.

35. Büning-Pfaue H. Analysis of water in food by near infrared spectroscopy. Food Chemistry 2003;82(1):107-115. https://doi.org/10.1016/S0308-8146(02)00583-6.

36. Parrini S, Acciaioli A, Franci O, Pugliese C, Bozzi R. Near infrared spectroscopy technology for prediction of chemical composition of natural fresh pastures, J Appl Anim Res 2019;47(1):514-520. doi:10.1080/09712119.2019.1675669.

37. Nieto-Ortega B, Arroyo JJ, Walk C, Castañares N, Canet E, Smith A. Near infrared reflectance spectroscopy as a tool to predict non-starch polysaccharide composition and starch digestibility profiles in common monogastric cereal feed ingredients. Anim Feed Sci Technol 2022;285:115214. https://doi.org/10.1016/j.anifeedsci.2022.115214.