



Revista Agrária Acadêmica

[Agrarian Academic Journal](#)



doi: 10.32406/v5n5/2022/80-88/agrariacad

Effect of near infrared spectroscopy instrumentation on forecasting protein and digestibility of cottonseed¹. Efeito da instrumentação de espectroscopia de infravermelho próximo na previsão de proteína e digestibilidade de caroço de algodão¹.

[Sueli Freitas dos Santos](#)², [Marco Aurélio Delmondes Bomfim](#)³, [Everaldo Paulo de Medeiros](#)⁴

¹ Part of the first author's postdoctoral research; Project financed by the Brazilian Agricultural Research Company – Embrapa; National Council for Scientific and Technological Development – CNPq; Ceará Foundation for the Support of the Scientific and Technological Development – Funcap

² Postdoctoral candidate Embrapa Goats and Sheep, Sobral – Ceará State (CE), Brazil: sfsantoszootecnia@gmail.com

³ Researcher, Embrapa Goats and Sheep, Sobral – Ceará State CE, Brazil: marco.bomfim@embrapa.br

⁴ Researcher, Embrapa Cotton, Campina Grande – Paraíba State (PB), Brazil: everaldo.medeiros@embrapa.br

Abstract

The objective was to evaluate the effect of near infrared spectroscopy instrumentation with NIRS FOSS and Perten on predictions of crude protein (CP) and *in vitro* dry matter digestibility (IVDDM) and *in vitro* organic matter digestibility (IVDOM) of cottonseed. The models for the CP parameter obtained better results in the NIR Perten and FOSS instruments than the models for IVDDM and IVDOM of cottonseed. In addition, the model with the best performance observed for CP was attributed to NIR Perten, considering its best RPD index.

Keywords: Alternative food. Food analysis. Spectroscopy. NIRS.

Resumo

O objetivo foi avaliar o efeito da instrumentação de espectroscopia no infravermelho próximo com NIRS FOSS e Perten nas previsões de proteína bruta (PB) e digestibilidade *in vitro* da matéria seca (DIVMS) e digestibilidade *in vitro* da matéria orgânica (DIVMO) do caroço de algodão. Os modelos para o parâmetro PB obtiveram melhores resultados nos instrumentos NIR Perten e FOSS do que os modelos para DIVMS e DIVMO de caroço de algodão. Além disso, o modelo com melhor desempenho observado para PB foi atribuído ao NIR Perten, considerando seu melhor índice RPD.

Palavras-chave: Alimento alternativo. Análise de alimentos. Espectroscopia. NIRS.

Introduction

The adequate nutrition of herds of livestock is of fundamental importance to an efficient system of animal production. In addition to providing livestock with protein and energy, pastures are also a primary source of fiber. Fiber promotes chewing and rumination and consequently helps maintain rumen health (VAN SOEST, 1967). In addition to the supply of nutrients required for the production and considering that food is the costliest element in most production units, a properly balanced diet is beneficial to animal health, as well as to the economic and environmental efficiency of production systems.

In Brazil, most of the diet of ruminant livestock is based on pasture, whether it comprises native or cultivated species. However, forage in adequate amounts and quality is not available throughout the year. The seasonality in forage production usually results in a nutritional deficit among livestock and negatively affects the production of ruminants in Brazil. An alternative that seeks to minimize the detrimental effects resulting from the low productivity of pastures and food shortages is nutritional supplementation, which seeks to provide the most crucial nutrients, such as protein and energy (NRC, 2007).

A diet that meets the nutritional needs of livestock is an essential pillar in production systems. The agribusiness industry has been generating large amounts of by-products resulting from the processing of cereal grains, oilseeds, and other commercially significant plants. Cottonseed (NIDA et al., 1996; MOHAMED et al., 1988) stands out from these as a co-product with high contents of protein, fat, and fiber. To this end, it is essential to know the precise nutritional value of the food employed in order to develop feeding strategies and optimize an efficient and economic system for animal production. Therefore, the search for faster and cheaper technologies and methodologies for analyzing potential food sources is essential to prompt decision-making regarding the administration of livestock feeding. For instance, these decisions must consider the strategic use of food supplements and enable the access of livestock producers to food science services.

Hence, as an alternative to traditional methodologies, near-infrared spectroscopy (NIRS) has been applied to optimize the time required to assess the nutritional quality of food without destroying the samples and or generating pollutants. This spectroscopy technology may produce different outcomes due to the differences in wavelength ranges of instruments such as the NIR Perten[®] DA 7250 (PerkinElmer, Inc., USA), with a spectral range from 950 to 1650 nm and a spectral interval of 5 nm, and the NIR FOSS 5000 Nirsystem II using the **ISIScan[®] software**, with a spectral range from 1100 to 2500 nm and spectral interval of 2 nm. Therefore, our research sought to evaluate the effect of NIR instrumentation in the transfer of protein calibration models and cottonseed digestibility.

Materials and methods

Cottonseed samples were harvested monthly for 12 months in the Brazilian states of Ceará (150 samples) and Mato Grosso (150 samples). To this end, collaborative relations were established with teachers, researchers, and agricultural sciences students, as well as cereal farmers, professionals of the agricultural business, and staff of animal feed factories, who made these collections possible. This network of collaborations resulted in the collection of 300 samples, which were sent by mail to the

Embrapa Goats and Sheep in the town of Sobral, Ceará, in order to undergo analytical procedures at the Animal Nutrition Laboratory.

The samples were subjected to an initial stage of spectra collection using a Perten® DA 7250 (PerkinElmer, Inc., USA) NIR instrument. This initial step was conducted to select a sample bank for calibration. After harvested spectra from the samples, the multiplicative scatter correction (MSC), using the software The Unscrambler® v.10.5.1 (Camo Inc). In addition, a principal component analysis (PCA) (HOTELLING, 1933), was carried out in order to observe classes or categories for the distribution of the set of samples. In PCA, the spectra were centered on the mean for the exploratory analysis of the samples.

Following the pre-processing stage mentioned above, a set of 109 samples was selected through the X matrix of the spectra using the “Evenly Distributed Samples” selection tool of The Unscrambler® v.10.5.1 (Camo Inc), observing the greater variability between the samples to be destined to the chemical analyses and to compose the calibration banks (75% of the samples) for the construction of the models and validation (25% of the samples), for independent validation. The highest variability between these samples was used to build the calibration bank and perform the chemical analyses for constructing the models.

The samples selected to constitute the calibration set were dried in a forced ventilation oven at 55°C until their masses stabilized. Subsequently, the dried samples were ground in a Wiley-type mill equipped with a 1.0 mm mesh sieve and were then stored in containers that were adequately identified and the harvested of spectra in the instruments using the NIR Perten® DA 7250 (PerkinElmer, Inc., USA) and NIR FOSS 5000 Nirsystem II instruments using the **ISIScan® software**.

The following chemical analyses were performed: dry matter (DM) content; mineral matter (MM) content; organic matter (OM) content, calculated as the difference between DM and MM; and concentration of total nitrogen (N), determined using a combustion system (Leco FP-628, Leco Corp., St. Joseph, MI, USA). In order to convert the (N) values into crude protein (CP) content, the conversion factor of 6.25 was applied. The *in vitro* digestibility of dry matter (IVDDM) and *in vitro* digestibility of organic matter (IVDOM) were determined using an MA443 automatic incubator (MA443, Marconi Equipment's for Laboratories Ltda., Piracicaba, SP, Brazil), in accordance with previously established technical procedures (TILLEY; TERRY, 1963). All spectral measurements were performed under controlled conditions of sample humidity (room temperature of 25°C, and relative air humidity of 55%) to avoid possible interferences to the spectra harvested (LYONS; STUTH, 1992).

With the chemical analyzes after the harvested of the spectra of the samples in NIR instruments, through the software The Unscrambler® v. 10.5.1 (Camo Inc), the models were built. The regression method used was partial least squares, one variable at a time (PLS - 1) (KOURTI; MACGREGOR, 1995), considering the reference values obtained by chemical laboratory analyzes as a dependent variable and the latent variables created from the spectra as independent variables of the multiple regression models.

Models were elaborated for parameters of crude protein (CP) and *in vitro* digestible matter of IVDMD and *in vitro* digested organic matter (IVOMD) for NIRS Perten and FOSS instruments. For each modeled constituent, the models were generated by submitting the original spectra to different mathematical pre-treatments, such as multiplicative signal correction (MSC), normal variance transformation (SNV), first and second derivatives (Savitzky-Golay) (KOURTI; MACGREGOR, 1995), with windows varying from 1 to 4 points (SAVITZKY; GOLAY, 1964; BROWN et al., 2000). And the number of PLS factors of the models was determined by the cross-validation procedure “leave-one-out”

(GELADI; KOWALSKI, 1986). In the independent validation, the separate database was used initially (25% of the samples).

After preparing the models, for all pre-treatments used, the best models for each parameter were selected, according to the criteria: model determination coefficient in calibration, cross-validation (R^2), square root of the standard error of the calibration mean and cross validation (RMSE) (LEITE; STUTH, 1995; LANDAU et al., 2006), in addition to the number of factors used in calibration as suggested by (PASQUINI, 2003). Another parameter used in the evaluation of the performance of the models was the R_{cal}/R_{val} (R^2 of the calibration and the R^2 of the validation), which represents the division between the coefficients and the RPD (Ratio of Performance to Deviation), which represents the division between the standard deviation of the analyzes and reference and the average forecast error (WILLIAMS; SOBERING, 1993; CHANG et al., 2001).

Results and discussion

Table 1 shows descriptive statistics covering the number of samples (N); the average, minimum, and maximum values; standard deviation (SD); and coefficient of variation (CV) of the parameters used as a reference for developing the calibration models.

Table 1 - Descriptive statistics of crude protein (CP), *in vitro* digestibility of dry matter (IVDDM), and *in vitro* digestibility of organic matter (IVDOM) of cottonseed.

Parameter	N	Mean (%)	Min, Max	SD	CV (%)
CP	75	26.0	20.5–32.8	2.8	10.6
IVDDM	70	61.2	51.0–75.4	4.8	7.8
IVDOM	65	62.8	52.0–76.3	4.8	7.6

N (number of samples); Min/Max (minimum and maximum values); SD (standard deviation); CV (coefficient of variation).

The average value observed for cottonseed CP was 26.0%, which is higher than values reported by other studies (NIDA et al., 1996; MOHAMED et al., 1988). For IVDDM, the average value was 61.2%, while for IVDOM, the values remained at 62.8%; these results come close to those reported by other studies (NIDA et al., 1996). However, the fluctuation in the minimum and maximum reference values obtained indicates that the strategy used to build the database (sample collection throughout 12 months) was effective in obtaining a wide range of variations, which contributed to the robustness of the models. It is important to emphasize that the chemical makeup of the food varies depending on some factors inherent to food composition and crop management. These inherent factors were considered while conducting the sample collection, and this may be observed through the variation in results, as evidenced by the chemical analysis of the samples.

Table 2 shows the models with the best performance in terms of cottonseed CP, IVDDM, and IVDOM.

Table 2 - Calibration and validation models using partial least squares (PLS) regression for crude protein (CP), *in vitro* digestibility of dry matter (IVDDM), and *in vitro* digestibility of organic matter (IVDOM) of cottonseed.

Dried and ground cottonseeds (Perten)									
PARAMETER	N	TREATMENT	FPLS	Calibration		Validation			
				R ²	RMSEC	R ²	RMSEV	Rcal/Rval	RPD
CP	75	MSC	5	0.82	1.15	0.78	1.29	1.05	2.15
IVDDM	70	SNV	7	0.46	3.15	0.36	3.79	1.28	1.27
IVDOM	65	MSC/SG2.2	7	0.52	3.16	0.30	4.77	1.73	1.00
Dried and ground cottonseeds (FOSS)									
PARAMETER	N	TREATMENT	FPLS	Calibration		Validation			
				R ²	RMSEC	R ²	RMSEV	Rcal/Rval	RPD
CP	72	MSC/SG2.1	6	0.95	0.51	0.61	1.51	1.56	1.71
IVDDM	76	SMOOTHING	3	0.36	4.25	0.30	4.54	1.20	1.20
IVDOM	76	SMOOTHING	3	0.37	3.89	0.30	4.14	1.23	1.20

FPLS (number of PLS factors); N (number of samples); SG1 and SG2: 1 to 4 (first and second Savitzky-Golay derivatives, points 1 to 4); MSC (multiplicative scatter correction); SNV (standard normal variate); SMOOTHING (Smoothing); R² (coefficient of determination); RMSEC (root-mean-square error of calibration); RMSEV (root-mean-square error of validation); Rcal/Rval (R²calibration/ R²validation); RPD (ratio of performance to deviation).

The outcomes for the models developed for CP were an R² of 0.82 and 0.95, RMSEC of 1.15 and 0.51, and RMSEV of 1.29 and 1.51 for NIR Perten and NIR FOSS instruments, respectively. Regarding the values for RMSEC and RMSEV, they were relatively low, which indicates the model to be moderately accurate, and demonstrates conformity between the estimated and the reference value (LANDAU et al., 2006). It is important to note that, depending on the number of factors determining the models' complexity, overfitting (excessive number of factors) or underfitting (insufficient number of factors) (PASQUINI, 2003) can characterize some models. Hence, the observed approximation between the values of RMSEC and RMSEV concerning the CP parameter indicates that an adequate number of factors (varying between 5 and 6) was used for developing the models.

The performance of the validation models evaluated by the RPD followed a standardized classification (SAVITZKY; GOLAY, 1964). The models for CP had an RPD of 2.15 for NIR Perten and 1.71 for NIR FOSS and were classified as excellent and fitted, respectively. On the contrary, for IVDDM and IVDOM parameters, the models were classified as not fitted for both instruments. Thus, models for CP displayed a performance superior to the models for IVDDM and IVDOM, which were characterized by overfitting in NIR Perten and underfitting in NIR FOSS. Because it is a more laborious analysis, digestibility aggregates greater errors due to the difficulty of its estimation, which may lead to the insertion of systematic and random errors into the models. Furthermore, the observed results may be due to the structure of the analyzed parameters, since proteins contain N-H, C-H, and C=O (SHENK et al., 2008) bonds, all of which can be absorbed in the near-infrared region, besides having the simplest reference method when compared to the IVDDM and IVDC models.

Table 3 describes the most important wavelengths for measuring the CP, IVDDM, and IVDOM parameters obtained in the present study.

Table 3 - Wavelengths related to crude protein content (CP), *in vitro* digestibility of dry matter (IVDM), and *in vitro* digestibility of organic matter (IVDOM).

	Wavelength (nm)																
	900	1000	1100	1200	1300	1400	1500	1600	1700	1800	1900	2000	2100	2200	2300	2400	2500
Model performance (Dried and ground Perten 950–1650nm)																	
CP		•	•	•	•	•	•	•									
IVDDM	•	•	•	•	•	•											
IVDMO	•	•	•	•	•												
Model performance (Dried and ground FOSS 1100–2500 nm)																	
CP			•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
IVDDM			•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
IVDMO			•	•	•	•	•	•	•	•	•	•	•	•	•	•	•

• (apparent organic bonds at wavelengths); nm (nanometers); 3rdOT (third overtone: 700 to 1100); 2ndOT (second overtone: 1200 to 1500); 1stOT (first overtone: 1600 to 2000); combination bands: 2100 to 2400).

The spectral bands related to the protein parameter comprised the ranges of 1000–1100 nm (third overtone), 1200–1500 nm (second overtone), and 1600 nm (first overtone) for the NIR Perten instrument, and 1100 nm (third overtone), 1200–1500 nm (second overtone), 1600–2000 nm (first overtone), and 2100–2400 nm (combination bands) for the NIR FOSS instrument. In most of these spectral regions, the protein organic bonds also occurred along with bond vibrations (N-H, C-H, C=O, C-N-C) (SHENK et al., 2008), indicating the efficiency of these regions in predicting the CP content of cottonseed.

For IVDDM and IVDOM parameters, the spectral bands comprised the ranges of 900–1100 nm (third overtone) and 1200–1400 nm (second overtone); for IVDOM, the range was limited to 1300 nm in the NIR Perten instrument. However, for NIR FOSS, the most expressive bands for IVDDM and IVDOM were 1100 nm (third overtone), 1200–1500 nm (second overtone), 1600–2000 nm (first overtone), and 2100–2400 nm (combination bands).

Digestibility is correlated to food composition, mainly with regards to cellulose and lignin. For cellulose, the composition has been associated with information collected at the wavelengths of 1490, 1780, 1820, 2335, 2347, 2352, and 2488 nm, whereas for lignin, informative wavelengths are 1100, 1170, 1410, 1417, 1420, and 1440 nm (SHENK et al., 2008).

It should be noted that the wavelengths observed in the present study remained within those reported in the literature. Furthermore, although the near-infrared region covers a range from 780 to 2500 nm (SHENK et al., 2008), these end values are not commonly used since it is possible to ascertain the scope of wavelengths from the NIR Perten and NIR FOSS instruments. It is likely that the models will not predict the digestibility of the cottonseed in these extreme regions; the values observed may instead be related to the high RMSEC and RMSEV values, which classify the models for IVDDM and IVDOM parameters as overfitting and underfitting models, respectively.

Conclusion

The models for the CP parameter had better results in the NIR Perten and FOSS instruments than models for IVDDM and IVDOM of cottonseed. Moreover, the model with the best performance observed for CP was attributed to NIR Perten, considering its best RPD index.

Authors' contribution

Sueli Freitas dos Santos - conducting the experimental research, data collection and interpretation, writing; Marco Aurélio Delmondes Bomfim - Postdoctoral supervision of the first author, original idea, correction and revision; Everaldo Paulo de Medeiros - correction and revision.

Declaration of Conflicting Interests

The authors declare that there is no conflict of interest.

Acknowledgments

The authors are grateful to Embrapa Goats and Sheep for support and assistance with the development of our research in its facilities.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: this work was supported by a Regional Scientific Development grant (DCR) from the National Council for Scientific and Technological Development (CNPq) [DCR-0024-01687.01.00/16 SPU N° 0543037/2016] and the Ceará Foundation for the Support of Scientific and Technological Development (FUNCAP) [DCR-0024-01687.01.00/16 SPU N° 0543037/2016].

References

- BROWN, C. D.; VEGA-MONTOTO, L.; WENTZELL, P. D. Derivative preprocessing and optimal corrections for baseline drift in multivariate calibration. **Applied Spectroscopy**, v. 54, n. 7, p. 1055-1068, 2000. <https://doi.org/10.1366/0003702001950571>
- CHANG, C-W.; LAIRD, D. A.; MAUSBACH, M. J.; HURBURGH, C. R. Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties. **Soil Science Society of American Journal**, v. 65, n. 2, 480-490, 2001. <https://doi.org/10.2136/sssaj2001.652480x>
- GELADI, P.; KOWALSKI, B. R. Partial least-squares regression: a tutorial. **Analytica Chimica Acta**, v. 185, p. 1-17, 1986. [https://doi.org/10.1016/0003-2670\(86\)80028-9](https://doi.org/10.1016/0003-2670(86)80028-9)
- HOTELLING, H. Analysis of a complex of statistical variables into principal components. **Journal of Educational Psychology**, v. 24, n. 6, p. 417-441, 1933. <https://doi.org/10.1037/h0071325>
- KOURTI, T.; MACGREGOR, J. F. Process analysis, monitoring and diagnosis, using multivariate projection methods. **Chemometrics and Intelligent Laboratory Systems**, v. 28, n. 1, p. 3-21, 1995. [https://doi.org/10.1016/0169-7439\(95\)80036-9](https://doi.org/10.1016/0169-7439(95)80036-9)
- LANDAU, S.; GLASSER, T.; DVASH, L. Monitoring nutrition in small ruminants with the aid of near infrared reflectance spectroscopy (NIRS) technology: a review. **Small Ruminant Research**, v. 61, n. 1, p. 1-11, 2006. <https://doi.org/10.1016/j.smallrumres.2004.12.012>
- LEITE, E. R.; STUTH, J. W. Fecal NIRS equations to assess diet quality of free-ranging goats. **Small Ruminant Research**, v. 15, n. 3, p. 223-230, 1995. [https://doi.org/10.1016/0921-4488\(94\)00026-4](https://doi.org/10.1016/0921-4488(94)00026-4)
- LYONS, R. K.; STUTH, J. W. Fecal NIRS equations for predicting diet quality of free-ranging cattle. **Journal of Range Management**, v. 45, n. 3, p. 238-244, 1992. <https://doi.org/10.2307/4002970>
- MOHAMED, O. E.; SATTER, L. D.; GRUMMER, R. R.; EHLE, F. R. Influence of dietary cottonseed and soybean on milk production and composition. **Journal of Dairy Science**, v. 71, n. 10, p. 2677-2688, 1998. [https://doi.org/10.3168/jds.S0022-0302\(88\)79861-6](https://doi.org/10.3168/jds.S0022-0302(88)79861-6)
- NRC. National Research Council. **Nutrient Requirements of Small Ruminants: Sheep, Goats, Cervids, and New World Camelids**. 1st ed. Washington, D. C.: National Academy Press, 2007, 362p.
- NIDA, D. L.; PATZER, S.; HARVEY, P.; STIPANOVIC, R.; WOOD, R.; FUCKS, R. L. Glyphosate-tolerant cotton: the composition of the cottonseed is equivalent to that of conventional cottonseed. **Journal of Agricultural and Food Chemistry**, v. 44, n. 7, 1967-1974, 1996. <https://doi.org/10.1021/jf950565s>
- PASQUINI, C. Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. **Journal of the Brazilian Chemical Society**, v. 14, n. 2, p. 198-219, 2003. <https://doi.org/10.1590/S0103-50532003000200006>

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. **Analytical Chemistry**, v. 36, n. 8, p. 1627-1639, 1964. <https://doi.org/10.1021/ac60214a047>

SHENK, J. S.; WORKMAN JUNIOR, J. J.; WESTERHAUS, M. O. Application of NIR spectroscopy to agricultural products. *In*: BURNS, D. A.; CIURCZAK, E. W. (Eds). **Handbook of Near-Infrared Analysis**. 3rd ed. Boca Raton: CRC Press, Chapter 17, 2008, 40p. <https://doi.org/10.1201/9781420007374>

TILLEY, J. M. A.; TERRY, R. A. A two-stage technique for the *in vitro* digestion of forage crops. **Grass and Forage Science**, v. 18, n. 2, p. 104-111, 1963. <https://doi.org/10.1111/j.1365-2494.1963.tb00335.x>

VAN SOEST, P. J. Development of a comprehensive system of feed analysis and its application to forages. **Journal of Animal Science**, v. 26, n. 1, p. 119-128, 1967. <https://doi.org/10.2527/jas1967.261119x>

WILLIAMS, P. C.; SOBERING, D. C. Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds. **Journal of Near Infrared Spectroscopy**, v. 1, n. 1, p. 25-32, 1993. <https://doi.org/10.1255/jnirs.3>

Recebido em 10 de outubro de 2022
Retornado para ajustes em 9 de janeiro de 2023
Recebido com ajustes em 9 de janeiro de 2023
Aceito em 10 de janeiro de 2023