

Espectroscopia no infravermelho próximo (NIRS): Um método inovador e rápido de análise de biomassa para avaliar o perfil nutricional de milho.

Maria Lúcia F. Simeone^{1*}, Thaís F. Silva², Lorena P. Maciel¹, Jamilly C. Valadares²; José Angelo N. de Menezes Júnior¹, Flávio D. Tardin¹

¹Embrapa Milho e Sorgo, Sete Lagoas, Brasil (marialucia.simeone@embrapa.br)

²Universidade Federal de São João del-Rei - Campus Sete Lagoas, Sete Lagoas, Brasil

Resumo: A espectroscopia no infravermelho próximo (NIRS) com regressão por mínimos quadrados parciais (PLS), foi usada para prever a composição bromatológica e a digestibilidade *in vitro* da biomassa de milho. Os modelos preditivos foram calibrados e validados para avaliar os teores de proteína, fibra detergente neutro, fibra detergente ácido, extrato etéreo, lignina, cinzas, umidade, carbono e digestibilidade *in vitro*, mostrando-se viáveis para fenotipagem em larga escala.

Palavras-chave: quimiometria; PLS, composição bromatológica; qualidade nutricional

INTRODUÇÃO

O milho (*Cenchrus americanus*) é uma cultura anual da família *Poaceae* que se desenvolve melhor em clima quente. É um cereal que se destaca por sua resistência à seca, alto valor nutricional e versatilidade de uso, tanto na alimentação animal quanto na produção de biomassa. Em programas de melhoramento genético e em processos industriais, o monitoramento da qualidade do milho requer métodos analíticos eficientes e precisos. Nesse contexto, a espectroscopia no infravermelho próximo é uma técnica rápida e capaz de prever a composição química de biomassas (Decker et al., 2018). A implementação dessa tecnologia, baseada em calibração multivariada, permite uma triagem de alto rendimento, superando as limitações operacionais dos métodos laboratoriais convencionais, que costumam ser lentos e destrutivos. Esse avanço metodológico facilita a avaliação sistemática de grandes conjuntos de amostras, possibilitando a identificação de genótipos com digestibilidade e perfis nutricionais superiores. Além disso, essas abordagens quimiométricas contornam as restrições ambientais associadas às análises tradicionais de química por via úmida, pois dispensam reagentes perigosos e não geram resíduos químicos (Pereira et al., 2023). O sucesso desses modelos depende da aquisição espectral na faixa do infravermelho próximo, na qual

sobretons e combinações de vibrações fundamentais de grupos funcionais — como -CH, -NH e -OH — fornecem assinaturas químicas únicas (Ferreira et al., 2010). Para estabelecer uma estrutura preditiva robusta, esses dados espectrais são correlacionados com os valores de referência obtidos por meio de análises convencionais de química por via úmida e de técnicas de regressão multivariadas (Trópia et al., 2023).

O objetivo deste estudo foi desenvolver uma metodologia analítica robusta e eficiente, utilizando dados de espectroscopia NIR e calibração multivariada, para determinar a qualidade da biomassa de milho (avaliação dos teores de proteína (PTN), fibra detergente neutro (FDN), fibra detergente ácido (FDA), lignina (LDA), cinzas (CZ), extrato etéreo (EE), umidade (UM), carbono (CBN) e digestibilidade *in vitro* da matéria seca (DIVMS) para fenotipagem em larga escala da biomassa de diferentes genótipos de milho.

MATERIAL E MÉTODOS

2.1 Preparo das amostras

Amostras de genótipos foram obtidas a partir de experimentos do programa de melhoramento de milho da Embrapa Milho e Sorgo, em Sete Lagoas, MG, nas safras 23/24 e 24/25. As amostras recém-



colhidas foram colocadas em sacos de papel e secas por 72 horas a 55 °C. Posteriormente, foram moídas em moinho de facas até atingirem granulometria de 1 mm, em preparação para as análises químicas laboratoriais.

A biomassa moída foi então submetida a análises laboratoriais de química úmida para quantificar os teores de PTN, FDN, FDA, LDA, CZ, EE, UM, CBN e DIVMS, conforme os métodos descritos por Nogueira et al., 1995.

2.2 Obtenção dos espectros NIR

Para a coleta dos espectros NIR, 15 g de amostras de milho moídas foram colocadas em placas de Petri de borossilicato, em triplicata, para a obtenção dos espectros. Os espectros foram obtidos utilizando o equipamento NIRFlex 500 na faixa de 1.000 a 2.500 nm (4.000 a 10.000 cm^{-1}) com 32 varreduras (*scans*) por espectro. Posteriormente, os dados brutos de refletância foram convertidos em absorbância, e a correção de linha de base foi aplicada para minimizar os efeitos físicos da variação do tamanho de partícula na aquisição spectral.

2.3 Desenvolvimento e validação dos modelos NIR

Um conjunto de 1054 amostras de milho foi dividido em amostras de calibração (70%) e de validação (30%) para cada componente químico avaliado, utilizando o algoritmo de Kennard-Stone, (Andrés-Camelo et al., 2025). Para as análises de LDA, EE, CBN e DIVMS, foram selecionadas 126, 195, 126, 119 e 200 amostras, respectivamente, com maior variância espectral, de acordo com o algoritmo de Kennard-Stone. Essa seleção ocorreu em decorrência do custo da análise de referência. O desempenho dos modelos foi avaliado com base no conjunto de dados de validação, cujas amostras não participaram da calibração.

Os dados espectrais foram pré-processados utilizando os métodos SNV (Standard Normal Variate) e de derivada primeira (DP) ou de derivada segunda (DS), com filtro Savitzky-Golay, variando-se o tamanho das janelas de 5, 9 e 11 pontos. Essas etapas de pré-processamento foram aplicadas para remover variações, como desvios da linha de base e espalhamento da luz, melhorando, assim, a qualidade da informação espectral. Os dados espectrais pré-processados foram centralizados em torno da média antes da construção dos modelos pelo algoritmo de

mínimos quadrados parciais (PLS – Partial Least Squares), com todas as variáveis espectrais e seleção de variáveis baseada nos coeficientes de regressão significativos. Os modelos obtidos para todos os constituintes físico-químicos e DIVMS analisados foram validados internamente (validação cruzada) e externamente com amostras que não foram utilizadas na construção do modelo de calibração.

A detecção de amostras anômalas foi realizada utilizando gráficos de resíduos versus o *leverage*, na matriz X (espectros) e em y (propriedade de interesse). As amostras que apresentaram resíduos elevados e alto *leverage* foram excluídas do conjunto de dados. Em seguida, novos modelos foram construídos e os parâmetros, como o coeficiente de correlação (R^2) e a raiz quadrada do erro médio de validação cruzada (RMSECV, do inglês, *Root Mean Squared Error of Cross Validation*), foram reavaliados. O número ideal de variáveis latentes de cada modelo PLS foi obtido a partir da análise do gráfico da variância explicada versus RMSECV, do gráfico dos *loadings* e do resíduo dos modelos. O software Unscrambler® 10.5 (Camo Software, Oslo, Norway) foi empregado no pré-processamento dos dados e na obtenção dos modelos PLS.

A capacidade preditiva dos modelos PLS foi avaliada pelos erros de previsão RMSEP (do inglês, *Root Mean Squared Error of Prediction*), para o conjunto de validação e pelos seguintes parâmetros estatísticos: raiz quadrada do erro médio de calibração, RMSEC (do inglês, *Root Mean Squared Error of Calibration*); raiz quadrada do erro médio de calibração cruzada, RMSECV (do inglês, *Root Mean Squared Error of Cross Validation*); R^2 para o conjunto de calibração e validação e relação de desempenho do desvio, RPD (do inglês, *Residual Prediction Deviation*) (Serafim et al., 2021).

O teste *t* pareado foi utilizado para verificar se havia diferença significativa entre os valores de referência e os valores previstos dos constituintes nos conjuntos de validação dos modelos otimizados. Nesse teste, calculou-se a probabilidade (p-valor) de obter uma diferença de erro absoluto maior do que a observada experimentalmente. Um valor-p inferior ao nível de significância indica uma diferença estatisticamente significativa nas capacidades preditivas do modelo.

RESULTADOS E DISCUSSÃO



A Tabela 1 apresenta os valores máximos, mínimos, médios e desvios-padrão das variáveis de composição físico-química dos conjuntos de calibração e validação da biomassa de milheto. Os teores de PTN variaram entre 4,94 e 17,88; FDN (49,27 e 77,79); FDA (20,11 e 48,42); LDA (45,35 e 49,70); CZ (0,32 e 18,7); EE (1,02 e 2,89); UM (1,54 e 10,19); CBN (45,35 e 49,70); e DIVMS (52,44 e 70,03), com valores médios e desvios-padrão semelhantes. Tais variações de composição podem ser devidas à diversidade genética, ao estágio de crescimento, à época de colheita e a fatores ambientais.

Tabela 1. Resultados das análises estatísticas dos conjuntos de calibração e validação do milheto.

Teor (%)	N	Max	Min	Med	DP
Calibração					
PTN	430	17,88	4,94	12,4	1,93
FDN	700	77,79	49,27	60,25	6,15
FDA	700	48,42	20,11	32,92	5,31
LDA	85	49,7	45,35	47,4	0,76
CZ	700	18,7	0,32	9,69	4,02
EE	131	2,89	1,02	1,92	0,48
UM	700	10,19	1,54	6,33	1,59
CBN	85	49,7	45,35	47,4	0,76
DIVMS	80	70,03	52,44	61,34	4,07
Validação					
PTN	210	15,31	7,81	12,42	1,31
FDN	353	76,67	51,02	60,56	5,75
FDA	354	46,57	21,46	32,86	4,56
LDA	41	69,36	53,66	61,34	3,88
CZ	352	17,26	0,54	9,54	3,38
EE	64	2,83	1,07	1,9	0,49
UM	354	9,13	1,47	6,47	1,31
CBN	41	48,92	46,18	47,41	0,68
DIVMS	39	69,36	53,66	61,34	3,88

Valores máximos (MAX), mínimos (MIN), médios (MED) e desvios-padrão (DP); PTN: Proteína, FDN: Fibra Detergente Neutro; FDA: Fibra Detergente Ácido; LDA: Lignina; EE: Extrato Etéreo; UM: Umidade; Carbono: CBN; DIVMS: Digestibilidade *in vitro* da Matéria Seca,

As tabelas 2, e 3 apresentam os resultados obtidos para os conjuntos de calibração e validação dos modelos de calibração NIR-PLS da biomassa de milheto.

Tabela 2. Resultados das análises estatísticas dos conjuntos de calibração dos constituintes da biomassa de milheto,

Teor (%)	Pré-Trat	VL	Calibração			
			R ² _c	RMSEC	R ² _{cv}	RMSECV
PTN	DS9	8	0,90	0,62	0,88	0,68
FDN	DP5	9	0,92	1,72	0,91	1,83
FDA	DP5	5	0,91	1,63	0,90	1,66
LDA	DS9	4	0,89	0,87	0,87	0,69
CZ	DS5	6	0,95	0,90	0,95	0,93
EE	DP9	6	0,91	0,14	0,90	0,14
UM	DS5	7	0,91	0,47	0,90	0,48
CBN	DS5	7	0,91	0,23	0,80	0,41
DIVMS	DS11	7	0,86	1,49	0,69	2,30

VL: Variável Latente; R²_c: Coeficiente de determinação do conjunto de calibração; RMSEC: Raiz Quadrada do Erro Médio de Calibração; R²_{cv}: Coeficiente de determinação do conjunto de validação cruzada; RMSECV: Raiz Quadrada do Erro Médio de Calibração Cruzada; DP: Derivada Primeira com janela de 5 pontos; DS: Derivada Segunda com janela de 5, 9 ou 11 pontos, FDN: Fibra Detergente Neutro; FDA: Fibra Detergente Ácido; LDA: Lignina; EE: Extrato Etéreo; UM: Umidade, CBN: Carbono, DIVMS: Digestibilidade *in vitro* da Matéria Seca,

Tabela 3. Resultados das análises estatísticas do conjunto de validação dos constituintes da biomassa de milheto,

Teor (%)	Validação			
	R ² _p	RMSEP	RPD	p-valor
PTN	0,81	0,58	4,17	0,99
FDN	0,91	1,68	3,4	0,36
FDA	0,90	1,69	3,3	0,11
LDA	0,85	0,84	2,4	0,70
CZ	0,94	0,81	4,2	0,95
EE	0,94	0,81	3,0	0,92
UM	0,88	0,41	2,4	0,95
CBN	0,80	0,29	2,0	0,15
DIVMS	0,66	2,24	2,7	0,86

R²_c: Coeficiente de determinação do conjunto de previsão; RMSEP: Raiz Quadrada do Erro Médio de Previsão; RPD: Relação de Desempenho do Desvio, PTN: Proteína, FDN: Fibra Detergente Neutro; FDA: Fibra Detergente Ácido; LDA: Lignina; EE: Extrato Etéreo; UM: Umidade, CBN: Carbono, DIVMS: Digestibilidade *in vitro* da Matéria Seca.

p-valor: ao nível de 5% de significância.

Obteve-se uma boa correlação entre os valores previstos pelo modelo e os valores obtidos pelo método de referência, tanto para as amostras do conjunto de calibração (RMSEC) quanto para as do conjunto de validação (RMSECV), conforme a Figura 1. O coeficiente de determinação (R^2) obtido para o conjunto de amostras de calibração foi superior a 0,86 para os constituintes analisados, indicando uma boa correlação entre os valores previstos pelo modelo multivariado desenvolvido e os valores obtidos pelos métodos de referência. Os erros (RMSEC) dos modelos de calibração foram próximos aos erros (RMSEP) obtidos nos conjuntos de validação externa.

O RPD é a razão entre o desvio padrão das amostras de previsão (SD) e o erro padrão de previsão (Williams et al., 1987), e representa a relação entre o desvio padrão dos valores da propriedade medidos pelo método convencional, baseado na química úmida, e o erro padrão das amostras do conjunto de previsão. Os valores de RPD superiores a 2 indicam que o SEP é inferior a 0,5 vezes o SD das amostras de referência. As razões RPD de todos os modelos NIR-PLS avaliados foram superiores a 2,0 (Tabela 3), o que os torna adequados para a quantificação dos constituintes da biomassa de milheto.

Para comparar os resultados obtidos entre o método de referência e o modelo desenvolvido com o algoritmo PLS, utilizou-se também o teste *t* pareado (Tabela 3). O teste *t* pareado não revelou diferenças estatisticamente significativas entre os dois métodos avaliados (referência e NIR) ao nível de 5% de significância.

Assim, em menos de 1 minuto, é possível qualificar a biomassa de milheto em 9 características distintas por meio de modelos de calibração NIR-PLS.

Além da expressiva redução do tempo de análise, a tecnologia NIR também oferece vantagens econômicas em relação aos métodos analíticos convencionais. Os métodos tradicionais de química úmida são reconhecidamente onerosos, pois envolvem a aquisição contínua de reagentes e solventes em larga escala, demandam infraestrutura laboratorial específica e geram resíduos químicos que requerem tratamento adequado, o que limita o número de amostras que podem ser processadas no curto prazo (Rushing et al., 2016). Estimativas obtidas em estudos de biomassa lignocelulósica indicam que a análise por métodos de química úmida pode custar aproximadamente 300 dólares por amostra, enquanto a análise equivalente por NIR, incluindo custos de aquisição de equipamento, suporte técnico e operação, custa apenas 5 dólares por amostra (Vogel et al., 2010). Quando projetada para um conjunto amostral de larga escala, essa diferença torna-se ainda mais relevante: por exemplo o trabalho de Vogel et al.,

2010, estimou que em um experimento com centenas de amostras, os custos totais da análise convencional podem ultrapassar 110 mil dólares, ao passo que a análise equivalente por NIR totaliza cerca de 2 mil dólares, representando uma redução superior a 98% nos custos por amostra (Vogel et al., 2010).

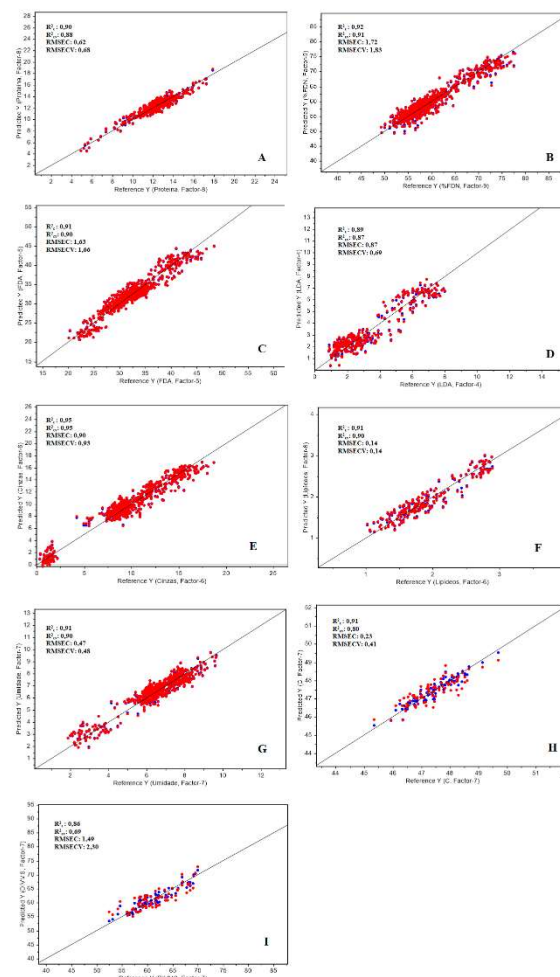


Figura 1, Gráfico dos valores de referência versus valores preditos para a composição bromatológica e digestibilidade *in vitro* do milheto: A-PTN; B-FDN; C-FDA; D-LDA; E-Cinzas; F-Lipídeos (EE) G-UM; H-CBN; I-DIVMS. RMSEC: azul, RMSECV: vermelho.

Adicionalmente, como a técnica NIR não requer o uso de reagentes químicos ou solventes, evitam-se custos recorrentes com a aquisição e o descarte de insumos perigosos (Simanjuntak et al., 2020), o que reforça a economicidade do método em operações de rotina (Rushing et al., 2016). A literatura destaca ainda que, embora o investimento inicial para a implementação de um laboratório NIR possa ser significativo, a operação torna-se relativamente barata em escala, configurando-se como uma alternativa alinhada à química verde e de baixo custo operacional para programas que demandam alta capacidade de



fenotipagem, como os de melhoramento genético, (Monrroy et al., 2017; Rushing et al., 2016).

CONCLUSÃO

A espectroscopia NIR, combinada com modelos de calibração PLS, demonstrou ser uma ferramenta robusta e eficiente para a predição da composição físico-química da biomassa de milho para os teores de proteína (PTN), fibra detergente neutro (FDN), fibra detergente ácido (FDA), lignina (LDA), cinzas (CZ), extrato etéreo (EE), umidade (UM), carbono (CBN) e digestibilidade in vitro da matéria seca (DIVMS), indicando seu potencial de aplicação em programas de melhoramento genético e de controle de qualidade da biomassa vegetal.

AGRADECIMENTOS

Embrapa Milho e Sorgo 20,23,01,010,00,02,002, Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG, APQ-01275-22) e pelas bolsas do CNPq (T.F.S., J.C.V.).

REFERÊNCIAS

Andrés-Camelo, R., Mazabel, J., Espitia-Buitrago, P., Chávez-Jáuregui, R. N., & Cardoso, J. A. (2025). Near-Infrared Spectroscopy and Wet Chemistry Dataset for Forage Nutritional Quality Assessment in *Urochloa humidicola*. *Data in Brief*, 60, 111651–111651.

Decker, S. R., Harman-Ware, A. E., Happs, R. M., Wolfrum, E. J., Tuskan, G. A., Kainer, D., Oguntimein, G., Rodríguez, M., Weighill, D., Jones, P., & Jacobson, D. (2018). High Throughput Screening Technologies in Biomass Characterization. *Frontiers in Energy Research*, 6.

Monrroy, M., Gutiérrez, D., Miranda, M., Hernández, K. R., & García, J. R. (2017). Determination Of *Brachiaria* Spp. Forage Quality By Near-Infrared Spectroscopy And Partial Least Squares Regression. *Journal of the Chilean Chemical Society*, 62(2), 3472–3477.

Nogueira, A. R. A., Matos, A. O., Carmo, C. A. F. S., Silva, D. J., Monteiro, F. L., Souza, G. B., Pita, G. V. E., Carlos, G. M., Oliveira, H., Comastri Filho, J. A., Miyazawa, M., & Oliveira Neto, W. (2005). Tecido vegetal. In A. R. A. Nogueira & G. B. Souza (Eds.), *Manual de laboratórios: solo, água, nutrição vegetal, nutrição animal e alimentos* (pp. 145–199). Embrapa Pecuária Sudeste.

Pereira, M. M. S., Santos, L. S., Silva, F. F. da, Silva, J. W. D., Peruna, A. B., Lisboa, M. de M., Santos, L. V., Lima-Júnior, D. M. de, & Silva, R. R. (2023). Prediction of the physical-chemical composition of tropical grasses through NIR spectroscopy. *Chilean Journal of Agricultural Research*, 83(5), 635–642.

Rushing, J. B., Saha, U., Lemus, R., Sonon, L., & Baldwin, B. S. (2016). Analysis of Some Important Forage Quality Attributes of Southeastern Wildrye (<i>Elymus glaberrimus</i>) Using Near-Infrared Reflectance Spectroscopy. *American Journal of Analytical Chemistry*, 7(9), 642–662.

Serafim, C. C., Guerra, G. L., Mizubuti, I. Y., Castro, F. A. B. de, Prado-Calixto, O. P., Galbeiro, S., Poveda-Parra, A. R., Bumbieris, V. H., Pértile, S. F. N., & Rêgo, F. C. de A. (2021). Use of near-infrared spectroscopy for prediction of chemical composition of Tifton 85 grass. *Semina Ciências Agrárias*, 42(3), 1287–1302.

Simanjuntak, L. E., Soesilo, T. E. B., & Herdiansyah, H. (2020). Analysis of Near Infra Red (NIR) Impactas the Control Process of a Palm Oil Milland Refinery Plant for EnvironmentalSustainability. *Polish Journal of Environmental Studies*, 29(5), 3013–3021.

Trópia, N. V., Silva, F. A. S., Andrade, D. R. de, Cidrini, F. A. A., Ebani, Y. C., Matos, E. M. de A., Borges, K. M., Roque, J. V., Zanetti, D., & Filho, S. de C. V. (2023). Construção e validação de modelos de regressão a partir de espectros NIR para predição da composição da cana-de-açúcar, farelo de soja e fubá de milho. *Semina Ciências Agrárias*, 44(2), 859–880.

Vogel, K. P., Dien, B. S., Jung, H., Casler, M. D., Masterson, S. D., & Mitchell, R. B. (2010). Quantifying Actual and Theoretical Ethanol Yields for Switchgrass Strains Using NIRS Analyses. *BioEnergy Research*, 4(2), 96–110.

Williams, P.C. Variables affecting near-infrared reflectance spectroscopic analysis, In: Williams P, Norris K (eds) *Near-infrared technology in the agriculture and food industries*, 1st Ed, Am Cereal Assoc Cereal Chem, St, Paul, MN, p 143–167, 1987.