

Capítulo 3

Uso da Espectroscopia no Infravermelho Próximo e Calibração Multivariada para Avaliar a Composição Química do Milho

Maria Lúcia Ferreira Simeone
Marco Aurélio Guerra Pimentel
Miguel Marques Gontijo Neto
Maria Cristina Dias Paes
Dagma Dionísia da Silva

Introdução

O milho é um dos cereais mais utilizados na nutrição humana e animal; entretanto, sua importância econômica não se encontra vinculada apenas à sua aplicação alimentícia, mas também nos usos de seus derivados, que se estendem às indústrias química, farmacêutica, de papéis, têxtil, combustíveis, entre outras aplicações.

Na safra 2015/2016, o Brasil produziu 96,03 milhões de toneladas de milho (Acompanhamento..., 2017), o que permitiu atender à demanda crescente dos sistemas de produção de suínos e aves, bem como permitir ao País se tornar o segundo maior exportador de milho do mundo. O milho é uma cultura que permite duas safras por ano no Brasil, uma denominada de 1ª safra (safra de verão) e outra denominada de 2ª safra, realizada normalmente em sucessão à cultura da soja. Atualmente, a 2ª safra apresenta o dobro da produção da 1ª safra.

A composição química do milho é formada predominantemente por carboidratos (amido), proteína e lipídios (óleo), tornando-o um alimento energético de grande utilização para a indústria de ração animal e de produtos para o consumo humano. Conhecer a composição química do milho é um dos fatores importantes para a sua valoração e/ou o aproveitamento ideal do grão em programas de melhoramento genético, em sistemas produtivos ou na cadeia fornecedora de insumos para diferentes áreas da indústria.

Entre as diferentes técnicas analíticas para a caracterização química, uma que possibilita a realização de análises químicas com precisão, rapidez, baixo custo e pouca manipulação de amostras é a espectroscopia no infravermelho próximo – NIRS (*Near Infrared Spectroscopy*) associada aos métodos quimiométricos (Shenk; Westerhaus, 1994).

A reflectância no infravermelho próximo é um método espectroscópico que surgiu como alternativa aos métodos analíticos tradicionais para a determinação de diferentes propriedades ou constituintes de matrizes de interesse agrícola (Williams et al., 1996). Porém, metodologias que empregam a espectroscopia NIR têm se mostrado altamente eficientes na substituição de métodos de análise laboratoriais em diversos setores, como em análises de alimentos, produtos farmacêuticos e agrícolas, entre outros, viabilizando a avaliação de um grande número de parâmetros, em um largo espectro de amostras (Pasquini, 2003). Mas, para que a técnica seja empregada com boa acurácia e precisão, a cada característica de interesse é necessário realizar a construção de um modelo de calibração multivariada associando os dados espectroscópicos NIR aos resultados analíticos realizados pelos métodos de referência. E, para a realização dos ajustes no modelo de calibração, são necessários procedimentos matemáticos e quimiométricos apropriados ao modelo multivariado escolhido para o conjunto de amostras analisadas, conjunto este que contemple a maior amplitude/variabilidade nos teores para cada um dos constituintes do grão a ser avaliado.

Alguns estudos têm avaliado o teor de proteína, óleo, amido e umidade em milho grão utilizando NIR (Baye et al., 2006; Berardo et al., 2009; Tallada et al., 2009; Fassio et al., 2015), porém os resultados apresentam diminuição da precisão em função da forma e tamanho do grão de milho (Manley et al., 2009) e da dificuldade de realizar a análise química de um único grão pelos métodos de referência, sendo necessária a utilização de amostras compostas (Agelet; Hurburgh Jr., 2014).

Já com amostras de milho grão moído, os trabalhos de Orman e Schumann Jr. (1991), Berardo et al. (2009), Ferrarini (2004) e Egesel e Kahrman (2012) apresentaram com sucesso o desenvolvimento de modelos de calibração multivariada para a realização da análise de proteína, óleo, cinzas e carboidratos. Orman e Schumann Jr. (1991) utilizaram os modos de transmissão e reflectância para a obtenção dos espectros e verificaram que os menores erros de predição dos constituintes do grão moído ocorreram quando se empregou o modo de reflectância. No trabalho de Egesel e Kahrman (2012), foram utilizadas 138 amostras de grão de milho moído para o desenvolvimento dos modelos de calibração multivariada dos constituintes: proteína, óleo, cinzas e carboidratos, utilizando os métodos multivariados quadrado mínimo parciais (PLS) e regressão linear multivariada (MLR). O método PLS resultou em

menores erros de predição no conjunto de validação externa, apresentando os menores valores de RMSEP entre os trabalhos citados acima, sendo 0,41% para proteína; 0,47% para óleo; 1,75% para carboidrato; e 0,24% para cinzas.

Neste capítulo serão apresentados os procedimentos e resultados do desenvolvimento de modelos de calibração multivariada utilizando a técnica NIR para a determinação do teor de proteína, extrato etéreo, fibra bruta, cinzas, amido e matéria seca em amostras de grãos de milho moído.

Metodologia

As amostras de grãos de milho foram selecionadas a partir de diferentes genótipos comerciais de milho e coletadas entre os anos de 2013 e 2016 em ensaios experimentais localizados em diferentes regiões de Minas Gerais, Goiás, Mato Grosso, Paraná e São Paulo.

Foram selecionadas 676 amostras de milho, buscando-se maior representatividade das amostras para a realização das análises químicas, porém nem todas as propriedades foram analisadas para todas as amostras em função dos custos de realização das análises químicas.

Os grãos de milho foram secos a 65 °C até peso constante. Em seguida, foram triturados em moinho de facas tipo Willey.

A porcentagem de proteína bruta foi obtida pelo método de combustão de Dumas em analisador Leco, modelo FP254 (Leco Instruments, Inc. St. Joseph, MI), multiplicando o teor de nitrogênio por 6,25 (Association of Official Analytical Chemists, 2005). O extrato etéreo foi determinado em extrator pressurizado Ankom (Ankom Technology Inc. Macedon, NY), (American Oil Chemists' Society, 2005); fibra bruta, matéria seca e cinzas foram determinadas de acordo com a metodologia descrita por Nogueira e Souza (2005) e amido, conforme o método descrito pela Association of Official Analytical Chemists (2005). Os resultados foram corrigidos pela porcentagem de matéria seca e as análises foram realizadas em duplicatas.

Os espectros das amostras foram obtidos em equipamento Buchi, modelo NIRFlex 500 (Buchi Labortechnik, Flawil, Switzerland) equipado com detector de InGaAs (índio, gálio e arsênio), utilizando como porta-amostra uma placa de Petri de vidro borossilicato. O equipamento foi calibrado utilizando o padrão Spectralon®. Os espectros das amostras de grãos de milho foram obtidos em triplicata, na região de 4.000 cm⁻¹ a 10.000 cm⁻¹, com resolução de 4 cm⁻¹ e 32 varreduras por espectro.

Para a construção dos modelos foi utilizado o algoritmo Kennard-Stone (Kennard; Stone, 1969) para a seleção do conjunto de amostras de calibração e validação. Dessa forma, foram obtidos dois conjuntos, um de calibração com dois terços das amostras e outro de validação contendo um terço das amostras aproximadamente, cobrindo assim toda a faixa de concentração de cada propriedade analisada.

É importante observar a necessidade de incluir um pré-tratamento aos dados para evidenciar as informações que se pretende extrair. No caso, para corrigir os efeitos de espalhamento da luz e deslocamentos de linha de base, os espectros foram pré-processados utilizando a variação normal padrão - SNV (*Standard Normal Variate*) e primeira derivada Savitzky-Golay com nove pontos à direita e nove pontos à esquerda. Os dados também foram centrados na média para que as diferenças entre os espectros fossem evidenciadas, uma vez que o resultado desse pré-processamento é apenas uma translação de eixos para o valor médio de cada um deles e, conseqüentemente, a estrutura dos dados é totalmente preservada (Ferreira, 2015).

Os teores de proteína, extrato etéreo, fibra bruta, cinzas, amido e matéria seca obtidos pelos métodos de referência foram associados à média dos espectros NIR de cada amostra, e os modelos de calibração multivariada foram desenvolvidos utilizando o algoritmo PLS (*Partial Least Square*) e a validação cruzada, utilizando o método randômico com segmentos de 33 amostras. Para o desenvolvimento dos modelos, utilizou-se o software Unscrambler® (versão 10.3, CAMO Software Inc., Norway). Foi utilizado o teste de incerteza de Martens (Martens; Martens, 2000), disponível no Unscrambler, para selecionar as variáveis mais significantes e estáveis e assim obter modelos mais simples e robustos. Amostras anômalas foram avaliadas pelo gráfico de *leverage* e *scores* e, quando necessário, foram excluídas do conjunto de amostras analisadas.

A escolha do número de variáveis latentes foi realizada por meio da resposta obtida pela soma do quadrado da diferença entre os erros residuais (PRESS) na validação cruzada, conforme Haaland e Thomas (1988). A situação ideal é aquela em que ocorre a minimização dos erros residuais.

Os indicadores estatísticos (ASTM International, 2012) utilizados para avaliar o desempenho e validação dos modelos foram: raiz quadrada do erro médio de calibração - RMSEC (*Root Mean Squared Error of Calibration*), raiz quadrada do erro médio de validação cruzada - RMSECV (*Root Mean Squared Error of Cross Validation*), raiz quadrada do erro médio de predição - RMSEP (*Root Mean Squared Error of Prediction*), R^2 (coeficiente de determinação para o conjunto de calibração e validação) e *bias* (viés). Também foram avaliadas a relação de desempenho do desvio

- RPD (*Residual Prediction Deviation*) e razão de intervalo de erro - RER (*Range Error Ratio*) (Williams, 2001).

Resultados e discussão

Todos os espectros NIR de reflectância difusa passaram por uma transformação logarítmica para a obtenção dos valores de absorvância, conduzindo assim a uma relação linear com a concentração: $A_{\lambda} = -\log R_{\lambda}$ (Figura 1).

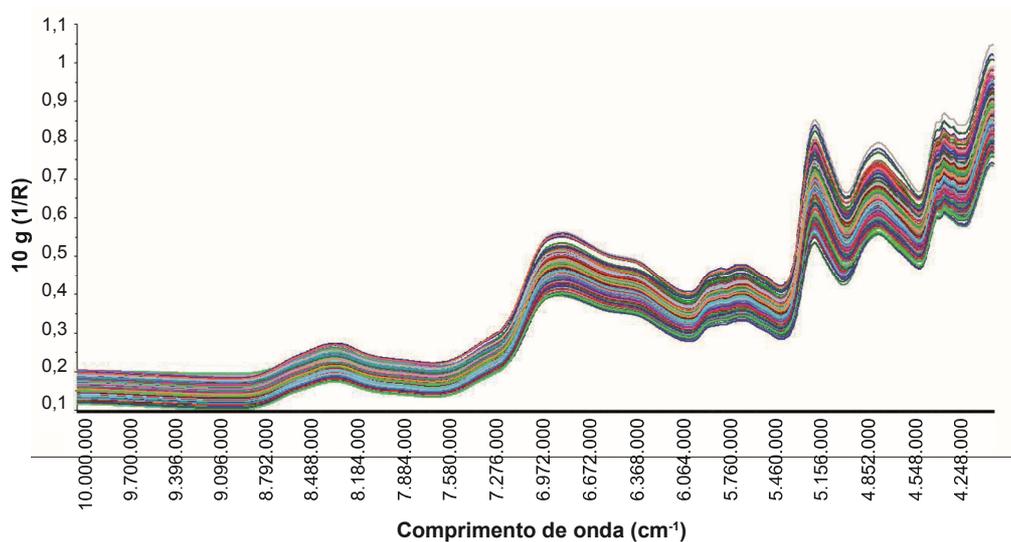


Figura 1. Espectros NIRS originais das amostras de grãos de milho moído.

Nos espectros, podemos observar um pico forte de absorção de água em 6.820 cm^{-1} , referente ao primeiro sobreton OH, e outro em 5.170 cm^{-1} , referente a uma banda de combinação de OH, para todas as amostras analisadas.

Uma região rica em informação a partir de 4.600 cm^{-1} a 4.000 cm^{-1} pode ser atribuída a combinações de estiramento de OH/ligação de hidrogênio, C-C estiramento 4.393 cm^{-1} e combinações de C-H/C-C (4.382 cm^{-1} – 4.063 cm^{-1}), características das moléculas de açúcares que compõem o amido e a fibra bruta e ácidos graxos presentes no extrato etéreo (Osborne; Douglas, 1981). O intervalo de 4.624 cm^{-1} a 4.508 cm^{-1} pode ser associado à região de bandas de combinação de estiramento C=O, deformação

de N-H no plano e estiramento C-N; a região de 4.878 cm^{-1} a 4.852 cm^{-1} está associada a bandas de combinação de estiramento N-H; em 4.988 cm^{-1} aparece a banda de combinação de estiramento C-N e deformação no plano de N-H, regiões características da absorção de ligações químicas presentes na proteína.

Para o resultado dos indicadores estatísticos obtidos na calibração e validação foi utilizado o modelo PLS e as amostras de grãos de milho apresentaram uma ampla faixa de resultados para todas as propriedades estudadas (Tabela 1).

Os resultados obtidos para o teor de proteína variaram entre 6,78% e 15,66%, com valores médios de 10,08%, cujos valores representaram a faixa de amostragem para o desenvolvimento do modelo de calibração multivariada. Para as outras propriedades os resultados obtidos variaram entre 2,43% e 5,75% (extrato etéreo); 2,78% e 0,4% (fibra bruta); 2,26% e 0,97% (cinzas); 75,35% e 65,45% (amido) e 94,96% e 88,34% (matéria seca). A média dos valores encontrados para cada propriedade analisada pelos métodos de referência está de acordo com o descrito por outros autores (Orman; Schumann Jr., 1991; Berardo et al., 2009; Ferrarini, 2004; Egesel; Kahriman, 2012).

Obteve-se uma boa correlação entre os valores previstos pelo modelo e os valores obtidos pelo método de referência para todas as propriedades avaliadas, tanto para as amostras do conjunto de calibração (RMSEC) como para as do conjunto de validação (RMSEP).

O coeficiente de determinação (R^2) obtido para o conjunto de amostras da calibração foi acima de 0,9 para todos os constituintes analisados, portanto existe uma boa correlação entre os valores previstos pelos modelos multivariados desenvolvidos e os valores obtidos pelos métodos de referência. A eficiência do modelo foi avaliada pela previsão de resultados para o conjunto de validação de amostras de grãos de milho moído (Figura 2). Nesta etapa, foram utilizadas amostras externas à etapa de calibração e o erro associado a cada modelo pôde ser avaliado pelo RMSEP. Assim, o RMSEP encontrado para proteína foi de 0,28%, extrato etéreo (0,06%), fibra bruta (0,08%), cinzas (0,05%), amido (0,35%) e matéria seca (0,43%). Os resultados obtidos no conjunto de previsão apresentaram erro relativo médio para proteína de 1,84%, extrato etéreo 1,02%, fibra bruta 4,65%, cinzas 1,70%, amido 2,5% e matéria seca 0,32%, indicando boa capacidade de previsão dos modelos.

Tabela 1. Resultados das análises estatísticas dos conjuntos de calibração e validação para o teor de proteína, extrato etéreo, fibra bruta, cinzas, amido e matéria seca (%) em grãos de milho moído. Embrapa Milho e Sorgo, Sete Lagoas, MG, 2018.

Propriedade	Proteína		Extrato etéreo		Fibra bruta		Cinzas		Amido		Matéria seca	
	Cal ^a	Val ^b	Cal	Val	Cal	Val	Cal	Val	Cal	Val	Cal	Val
Número de amostras	451	225	421	209	421	209	421	209	109	54	381	189
Valor máximo (%)	16,33	15,66	5,83	5,73	2,71	2,70	2,45	2,44	76,64	76,02	94,96	94,8
Valor mínimo (%)	7,46	8,14	2,77	2,77	0,02	0,33	0,99	1,04	66,62	67,57	88,34	88,52
Média (%)	10,26	10,26	4,74	4,74	1,67	1,64	1,56	1,56	73,02	73,01	91,56	91,57
Variáveis latentes	5	5	7	7	9	9	9	9	7	7	3	3
R ² c	0,93	0,93	0,98	0,98	0,97	0,97	0,97	0,94	0,97	0,96	0,96	0,95
RMSE ^d (%)	0,28	0,28	0,06	0,06	0,08	0,08	0,05	0,05	0,32	0,35	0,39	0,43
Bias ^e	0,001	0,001	0,000	0,001	0,000	-0,007	0,000	-0,002	0,000	-0,05	0,000	-0,048
RPD ^f	3,7	3,7	8,57	8,35	6,27	5,91	6,73	6,54	6,01	4,53	4,85	4,37
RER ^g	31,68	27,85	50,98	46,41	33,62	29,86	29,34	28,13	31,31	20,60	16,98	14,61

^a Cal: conjunto de amostras de calibração^b Val: conjunto de amostras de validação^c R²: coeficiente de determinação^d RMSE: raiz quadrada do erro médio – Cal (RMSEC); Val (RMSEP).^e Bias: viés^f RPD: relação de desempenho do desvio^g RER: razão de intervalo de erro.

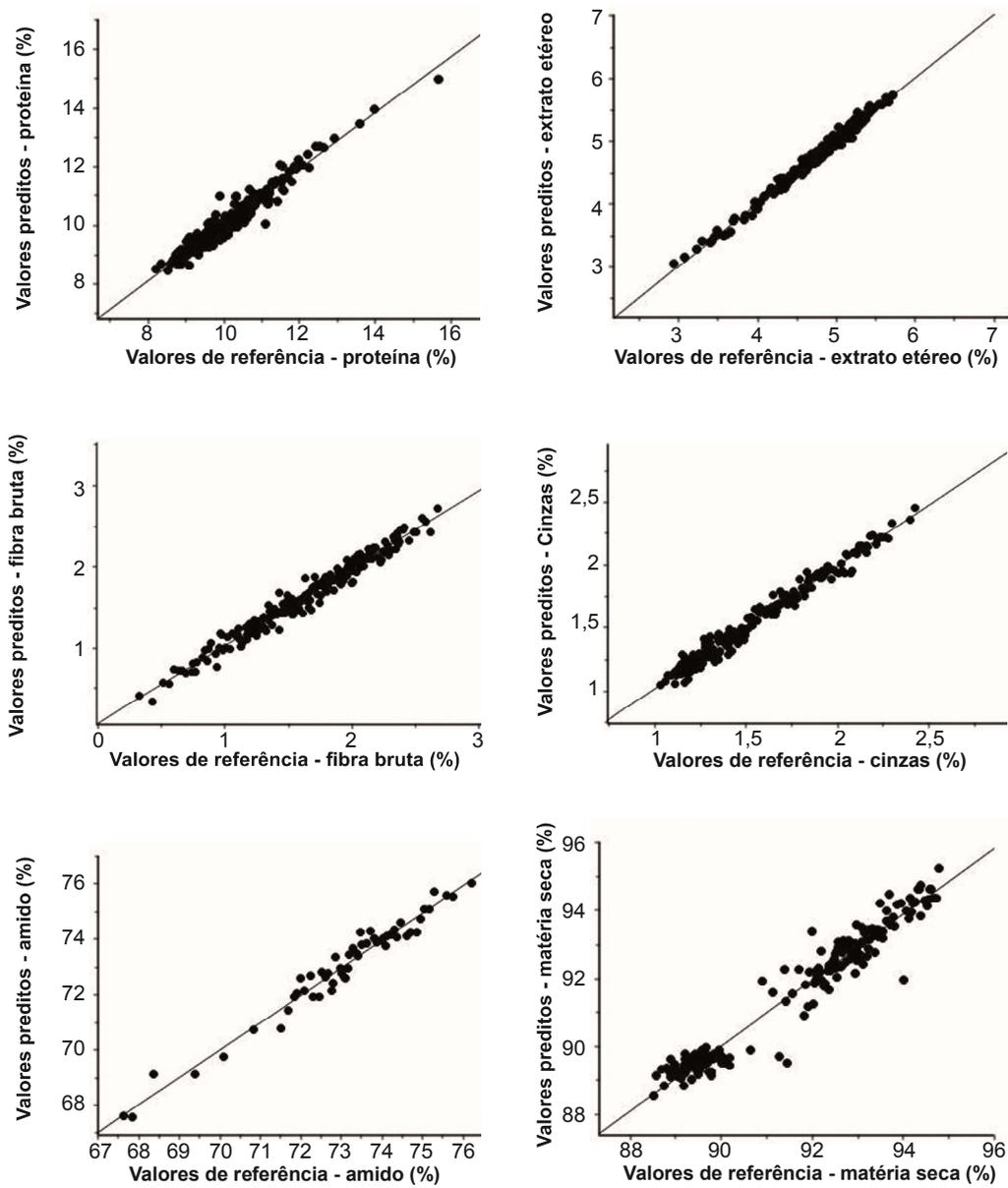


Figura 2. Valores de referência versus valores preditos em porcentagem para as propriedades (proteína, extrato etéreo, fibra bruta, cinzas, amido e matéria seca) das amostras de grãos de milho moído para o conjunto de amostras de validação externa.

As razões RPD e RER também foram utilizadas para estimar a capacidade preditiva dos modelos e relacionam o SEP à variância e à faixa utilizada nas amostras de referência (Williams; Sobering, 1993). O valor de RPD representa a relação entre o desvio padrão dos valores da propriedade medida pelo método convencional e o erro padrão das amostras contidas no conjunto de calibração ou previsão. Williams e Sobering (1993) indicam que se deve utilizar RPD com valores acima de 3 e RER acima de 10 (Starr et al., 1981) para determinações quantitativas na maioria das aplicações utilizando a tecnologia NIR em produtos agrícolas. Os valores de RPD e RER para todas as propriedades analisadas estão acima de 3 e 10 (Tabela 1), respectivamente, e foram considerados adequados para a realização das análises quantitativas de composição química do milho. Contudo, é importante ressaltar que a precisão de um modelo depende de sua aplicação e do erro de predição (RMSEP).

O acompanhamento dos modelos deve ser realizado ao longo do tempo pela verificação dos resultados obtidos para amostras de referência. Uma boa prática é a adoção de uma carta-controle de resultados, e quando necessário o modelo deve ser atualizado com a inclusão de novas amostras.

Com os modelos de calibração multivariada PLS-NIR desenvolvidos foi possível realizar a análise da composição química do milho em menos de 1 minuto. A rapidez na obtenção dos resultados possibilita que um grande número de amostras seja analisada diariamente, colaborando para que o programa de melhoramento genético e sistemas de produção possa desenvolver e validar novos materiais genéticos, novas tecnologias de cultivo e utilização para esse cereal de grande importância para o agronegócio brasileiro.

Considerações finais

O uso da espectroscopia no infravermelho próximo associado a métodos de calibração multivariada (PLS) possibilitou o desenvolvimento de um método rápido para análise das propriedades (proteína, extrato etéreo, fibra bruta, cinzas, amido e matéria seca) de grãos de milho moído. Os resultados indicam que os modelos desenvolvidos foram capazes de fornecer resultados confiáveis e que possibilitam a sua implantação em rotina laboratorial para atender às demandas de pesquisa e valoração da cadeia produtiva do milho.

Agradecimentos

Os autores agradecem à estudante Edislane de Araújo Souza pela coleta dos espectros NIR e aos técnicos Carlos Henrique de Paula Pires e José Eduardo Filho pela realização das análises químicas.

Referências

- ACOMPANHAMENTO DA SAFRA BRASILEIRA [DE] GRÃOS: safra 2016/17: décimo levantamento. Brasília, DF: Conab, v. 4, n. 10, jul. 2017. 171 p.
- AGELET, L. E.; HURBURGH JR., C. R. Limitations and current applications of near infrared spectroscopy for single seed analysis. **Talanta**, v. 121, p. 288-299, 2014.
- ASSOCIATION OF OFFICIAL ANALYTICAL CHEMISTS. **Official methods of analysis of the AOAC International**. 18th ed. Gaithersburg, 2005.
- AMERICAN OIL CHEMISTS' SOCIETY. **Rapid determination of oil/fat utilizing high temperature solvent extraction**. Urbana, 2005. Official method Am 5-04.
- ASTM INTERNATIONAL. **Standard practices for infrared multivariate quantitative analysis**. West Conshohocken, 2012. (ASTM. E1655).
- BAYE, T. M.; PEARSON, T. C.; MARK SETTLES, A. Development of a calibration to predict maize seed composition using single kernel near infrared spectroscopy. **Journal of Cereal Science**, v. 43, p. 236-243, 2006.
- BERARDO, N.; MAZZINELLI, G.; VALOTI, P.; PAOLO, L.; REDAELLI, R. Characterization of maize germplasm for the chemical composition of the grain. **Journal of Agricultural and Food Chemistry**, v. 57 n. 6, p. 2378-2384, 2009.
- EGESEL, C. O.; KAHRIMAN, F. Determination of quality parameters in maize grain by NIR reflectance spectroscopy. **Journal of Agricultural Sciences**, v. 18, p. 31-42, 2012.
- FASSIO, A. S.; RESTAINO, E. A.; COZZOLINO, D. Determination of oil content in whole corn (*Zea mays* L.) seeds by means of near infrared reflectance spectroscopy. **Computers and Electronics in Agriculture**, New York, v. 110, p. 171-175, 2015.
- FERRARINI, H. **Determinação de teores nutricionais do milho por espectroscopia no infravermelho e calibração multivariada**. 2004. 125 f. Dissertação (Mestrado) - Universidade Federal do Paraná, Curitiba, 2004.

- FERREIRA, M. M. C. **Quimiometria**: conceitos, métodos e aplicações. Campinas: Unicamp, 2015. 493 p.
- HAALAND, D. M.; THOMAS, E. V. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. **Analytical Chemistry**, v. 60, n. 11, p. 1193-1202, 1988.
- KENNARD, R. W.; STONE, L. A. Computer aided design of experiments. **Technometrics**, v. 11, n. 1, p. 137-148, 1969.
- MANLEY, M.; WILLIAMS, P.; NILSSON, P. D.; GELADI, P. Near infrared hyperspectral imaging for the evaluation of endosperm texture in whole yellow maize (*Zea mays* L.) kernels. **Journal of Agricultural and Food Chemistry**, Easton, v. 57, n. 19, p. 8761-8769, 2009.
- MARTENS, H.; MARTENS, M. Modified jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). **Food Quality and Preference**, v. 11, n. 1/2, p. 5-16, 2000.
- NOGUEIRA, A. R. A.; SOUZA, G. B. (Ed.). **Manual de laboratórios**: solo, água, nutrição vegetal, nutrição animal e alimentos. São Carlos: Embrapa Pecuária Sudeste, 2005. 313 p.
- ORMAN, B. A.; SCHUMANN JR., R. A. Comparison of near-infrared spectroscopy calibration methods for the prediction of protein, oil, and starch in maize grain. **Journal Agricultural and Food Chemistry**, v. 39, n. 5, p. 883-886, 1991.
- OSBORNE, B. G.; DOUGLAS, S. Measurement of the degree of starch damage in flour by near infrared reflectance analysis. **Journal Science Food Agriculture**, v. 32, p. 328-332, 1981.
- 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.
- SHENK, J. S.; WESTERHAUS, M. O. The application of Near Infrared Reflectance Spectroscopy (NIRS) to forage analysis. In: FAHEY JR., G. C. **Forage quality evaluation and utilization**. Madison: American Society of Agronomy, 1994. p. 406-449.
- STARR, C.; MORGAN, A. G.; SMITH, D. B. An evaluation of near infra-red reflectance analysis in some plant breeding programmes. **Journal Agricultural Science**, v. 97, n. 1, p. 107-118, 1981.
- TALLADA, J. G.; PALACIOS-ROJAS, N.; ARMSTRONG, P. R. Prediction of maize seed attributes using a rapid single kernel near infrared instrument. **Journal of Cereal Science**, London, v. 50, p. 381-387, 2009.

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**, Sussex, v. 1, p. 25-32, 1993.

WILLIAMS, P. C.; SOBERING, D.; DAVES, A. M. C. How do we do it: a brief summary of the methods we use in developing near infrared calibrations. In: WILLIAMS, P. C. (Ed.). **Near infrared spectroscopy: the future waves**. Chichester: NIR Publications, 1996. p. 185-188.

WILLIAMS, P. C. Implementation of near-Infrared technology. In: WILLIAMS, P. C.; NORRIS, K. H. (Ed.). **Near-infrared technology in agricultural and food industries**. Saint Paul: American Association of Cereal Chemist, 2001. p. 145-169.