



PREDIÇÃO DO TEOR DE ÓLEO EM SEMENTES DE SOJA PELA TÉCNICA DE ESPECTROSCOPIA DE REFLECTÂNCIA NO INFRAVERMELHO PRÓXIMO

FABIANA C. GIORA¹; DANIELA MATSUMOTO²; REGINA H. G. PRIOLLI³;
CÁSSIA R. L. CARVALHO⁴

Nº 12157

RESUMO

A soja é um produto agrícola de grande interesse mundial graças à versatilidade de aplicação de seus produtos em diferentes segmentos agroindustriais. É uma rica fonte de proteínas e apresenta um nível relativamente elevado de óleo. Esses constituintes são as principais qualidades estimuladoras do complexo agroindustrial da soja. A qualidade da semente é um dos principais fatores na determinação do sucesso de uma cultura, pois é responsável pela transferência das inovações tecnológicas e dos ganhos genéticos até o campo. Para isso, métodos apropriados de avaliação que possam facilitar a seleção dessas características são necessários. Dessa forma, foi desenvolvido um modelo de calibração multivariada, método analítico não destrutivo, utilizando a espectroscopia na região do infravermelho próximo para dosagem do teor de óleo em soja, empregando grãos inteiros, com o propósito de utilizar as próprias sementes analisadas para uso no campo.

ABSTRACT

Soybean has attracted worldwide interest due to the versatility of soy products in different agri-business segments. It is a rich source of protein and has a relatively high oil level. These constituents are the main good points of the agro industrial complex of soybeans. The seed quality is a major factor in determining the success of a crop, once it is responsible for transferring new technologies and genetic gains to the field. For this purpose, appropriate methods of assessment that may facilitate the selection of these features are needed. Therefore, we developed a multivariate calibration model, a non-destructive analytical method, using Near-Infrared Reflectance Spectroscopy to measure the oil content in soybeans using the whole grains in order to sow their own seeds into the field.

¹Bolsista FUNDAG: Graduanda em Ciências Biológicas pela Pontifícia Universidade Católica de Campinas, SP, fabianacgiora@gmail.com.

²Colaboradora: Química, Centro de Recursos Genéticos Vegetais/IAC, Campinas, SP.

³Colaboradora: Pós-doutoranda, Departamento de Genética, ESALQ/USP, Piracicaba, SP.

⁴Orientadora: Pesquisadora, Centro de Recursos Genéticos Vegetais/IAC, Campinas-SP, climonta@iac.sp.gov.br.



INTRODUÇÃO

A soja é um produto agrícola de grande interesse mundial graças à versatilidade de aplicação de seus produtos na alimentação humana e animal e, mais recentemente, com o resurgimento do uso de energias renováveis, tornou-se uma das principais fontes para a extração de óleo vegetal, visando a produção de biocombustível.

Estudos têm demonstrado que o consumo dos grãos, ou derivados de soja, está associado à redução do risco de várias doenças. Com uma composição química quase completa, a soja é uma rica fonte de proteínas (33 a 45%, base seca) e apresenta um nível relativamente elevado de óleo (de 16 a 25%, base seca), contendo nesta fração ácidos graxos poli-insaturados, os quais tem ação na redução de doenças crônico-degenerativas. Essas características e as aplicações variadas em diferentes setores proporcionam à cultura um elevado valor econômico nos mercados nacional e internacional. O Brasil é o maior produtor e exportador mundial, sendo a oleaginosa cultivada em várias regiões do País, principalmente nas regiões Sul e Centro-Oeste (SILVA et al., 2006).

Segundo Vieira (1980), a qualidade da semente é um dos principais fatores na determinação do sucesso de uma cultura, pois é responsável pela transferência das inovações tecnológicas e dos ganhos genéticos até o campo, resultantes dos trabalhos de melhoramento. Embora a composição da soja possa sofrer influências ambientais durante o seu cultivo, ela é dependente principalmente dos fatores genéticos e, como os teores de proteína e de óleo nas sementes determinam o seu valor comercial, o desenvolvimento de genótipos com maior concentração desses nutrientes torna-se um alvo importante a ser alcançado pelos melhoristas. Para isso, métodos apropriados de avaliação que possam facilitar a seleção dessas características são necessários.

Dessa forma, o objetivo do presente trabalho foi desenvolver um modelo de calibração multivariada, método analítico não destrutivo, utilizando a espectroscopia na região do infravermelho próximo para dosagem do teor de óleo em soja, empregando grãos inteiros, com o propósito de utilizar as próprias sementes analisadas para uso no campo.

MATERIAL E MÉTODOS

Amostras usadas na construção de curva de calibração:

Empregou-se neste trabalho, para a construção de uma curva de calibração para dosagem do teor de óleo, um conjunto de sementes composto por 87 genótipos de soja, oriundas de ensaios agrônômicos realizados pelo Departamento de Genética da ESALQ, Piracicaba, SP, nos anos agrícolas de 2009/2010 e 2010/2011.



Coletas de espectros no infravermelho próximo:

Medidas de reflectância difusa dos grãos de soja foram coletadas em espectrofotômetro FOSS NIRSystems, modelo 6500, equipado com detector PbS e lâmpada de filamento de tungstênio-halogênio. Os espectros foram registrados pelo software ISScanTM, versão 3.10 (Infrasoft Internacional, 2007), na faixa espectral de 1100 a 2498 nm, com os sinais expressos em $\log(1/R)$, sendo coletados de grãos inteiros, em cela *opaque liquid*, à temperatura de 20-22°C, com 128 varreduras e resolução espectral de 2 nm. Foram coletados dois espectros por genótipo e cada espectro correspondeu à média de duas leituras espectrofotométricas, totalizando, desse modo, 348 espectros registrados. Leituras de referência da fonte de luz foram realizadas nas mesmas condições antes de cada leitura das amostras.

Métodos de referência:

Após a coleta dos espectros, parte das sementes foi moída e os teores de umidade e de lipídeos foram quantificados nos grãos, realizando-se duas análises por amostra. A umidade foi determinada com secagem do material em estufa ventilada, a 100°C, até obtenção de massa constante. Os lipídeos foram extraídos e quantificados por refluxo, utilizando-se aparelho de *Butt* e hexano como solvente de extração, seguindo os procedimentos descritos em Carvalho et al. (1990).

Construção da curva de calibração e validação interna do modelo matemático:

Utilizou-se o método de regressão por Quadrados Mínimos Parciais (PLS, *Partial Least Squares*) para a regressão multivariada dos espectros, empregando o *software Pirouette 4.0* (Infometrix, 1990 -2007) e tomando como base os trabalhos de BRAGA e POPPI (2004) e LIMA et al. (2008). Para a validação interna e escolha do número ótimo de variáveis latentes (k) utilizou-se o método da validação cruzada “deixe uma fora” (*leave one out*). A seleção do modelo de calibração multivariada foi definida avaliando-se os erros padrões de validação interna (SEV) e de calibração (SEC), e os valores do coeficiente de correlação linear (r), correspondentes aos teores de óleo estimados pelo modelo em construção e aos valores obtidos pelo método de referência das amostras, que pertencem ao conjunto de calibração (LIMA et al., 2008).

Transformações e pré-processamento matemáticos:

Visto que os espectros de reflectância difusa coletados apresentaram variações indesejáveis, antes da construção do modelo propriamente dito, tratamentos matemáticos foram empregados para correção dos espectros, a saber:



1. Correção multiplicativa de sinal (MSC) – Inicialmente calcula-se o espectro médio (x_M) do conjunto de amostras a ser usado na construção do modelo; para cada espectro individual (x_i) dois parâmetros a_i e b_i são estimados por regressão contra a média dos espectros de calibração;
2. Primeira derivada, com janela de 5 pontos – a transformação foi determinada usando o algoritmo de Savitzky-Golay;
3. Dados centrados na média – após as transformações citadas acima, a matriz de dados foi centrada na média, que consistiu em determinar a média de cada variável (coluna); em seguida, a média foi subtraída do valor de cada elemento da coluna para gerar a matriz centrada na média.

Validação externa do modelo matemático:

Dentre as sementes analisadas em relação ao teor de óleo e com os seus respectivos espectros registrados, um conjunto, composto por 15 amostras, foi selecionado a fim de ser usado para validação externa do modelo matemático construído, levando em consideração a amplitude do teor de óleo encontrada nos grãos.

RESULTADO E DISCUSSÃO

Os grãos de soja usados neste estudo apresentaram de 4,64 a 10,72% de umidade, com valor médio de 6,87% e os valores de lipídeos, determinados pelo método de referência, variaram de 12,20 a 23,74%, com conteúdo médio de 19,00%, em base úmida. As 15 amostras de sementes usadas para a validação externa do modelo a ser proposto foram selecionadas neste intervalo de abrangência, com teores médios de óleo variando de 14,57 a 23,60%.

Os espectros originais registrados no espectrofotômetro do infravermelho próximo (NIRS) são mostrados na Figura 1, os quais não apresentaram ruídos. Porém, as medidas apresentaram deslocamento e inclinação da linha de base, devido respectivamente aos efeitos de espalhamentos aditivos e multiplicativos (BRAGA e POPPI, 2004).

Desse modo, para garantir uma boa relação entre as informações espectrais e os valores de referência (teor de óleo), transformações e pré-processamento matemáticos foram realizados nos espectros. Antes, contudo, para excluir amostras anômalas, a matriz de dados espectrais de calibração, composta por 318 linhas (amostras) por 700 variáveis (leituras de absorbância aparente), foi submetida à técnica de componentes principais (PCA), com auxiliar inspeção da análise de

agrupamentos hierárquicos (HCA). Em seguida, para correção dos espectros, foram aplicadas as transformações: correção multiplicativa de sinal (MSC) para corrigir o efeito de espalhamento de luz, causada pela falta de homogeneidade das amostras, e primeira derivada (janela de 5 pontos) para correção de deslocamento de linha de base. A Figura 2 mostra os espectros corrigidos, após os tratamentos matemáticos.

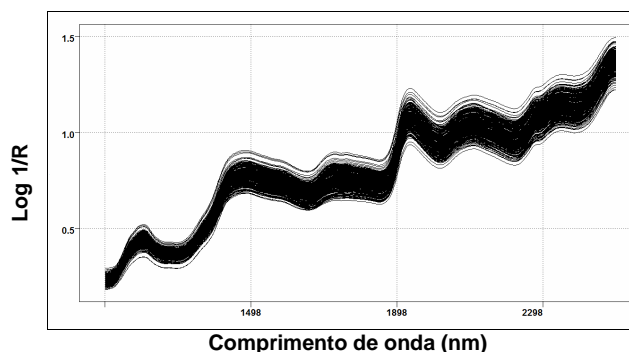


Figura 1. Espectros de reflectância difusa originais de 174 amostras de grãos de soja obtidos na região do NIR, na faixa de 1100 a 2498 nm, resolução de 2 nm e com sinais de resposta da reflectância em log (1/R).

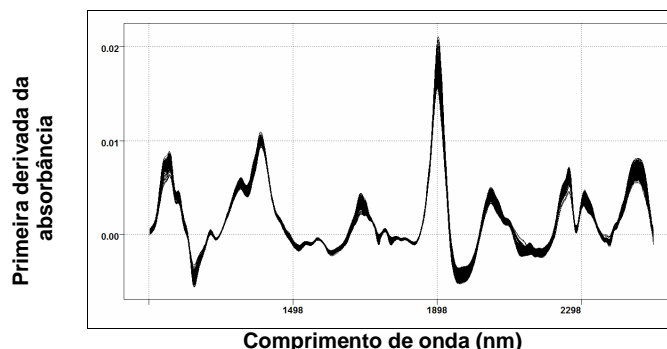


Figura 2. Conjunto de espectros após as transformações aplicadas: correção multiplicativa de sinal (MSC) e primeira derivada.

Empregando as mesmas transformações, a Figura 3 traz os espectros de amostras de farinha de soja desengordurada e seca, de farinha apenas desengordurada e de farinha integral. Observa-se que os lipídeos deram respostas mais intensas em 3 faixas espectrais (linha em azul): de 1120 a 1288 nm; de 1662 a 1838 nm e 2230 a 2394 nm, sendo que qualquer uma das variáveis dessas regiões poderia ser selecionada para a construção do modelo, objeto do presente trabalho. Porém, como o objetivo foi o de usar sementes integrais, sendo estas de diferenciados tamanhos e características externas, optou-se por usar a faixa integral dos espectros, visando alcançar resultados robustos.

A matriz de dados, após as correções necessárias, foi centrada na média para melhorar a visualização das relações entre as amostras. O modelo matemático

desenvolvido para a dosagem de lipídeos, com os menores erros padrões de validação interna e de calibração (respectivamente, SEV e SEC) e respectivos coeficientes de correlação linear (r), está demonstrado na Figura 4.

O modelo PLS construído utilizou sete variáveis latentes, que corresponderam aproximadamente a 95% da variância dos dados originais (espectrais e concentrações de referência). Durante a construção do modelo matemático, ao redor de 1/3 dos 318 espectros originais foram descartados, permanecendo 214 medidas espectrais na curva de calibração (Figura 4); as medidas anômalas ou com elevados resíduos foram excluídas na etapa de calibração, sendo identificadas pelos valores de Student, por estarem acima do limite de confiança de 95%, e pelo valor limite de *leverage*, considerado três vezes o número de variáveis latentes, dividindo-o pelo número de amostras (MORGANO, 2005). O modelo apresentou baixos erros padrões e coeficientes de correlação (r) da validação interna e da calibração próximos de 1 (Figura 4).

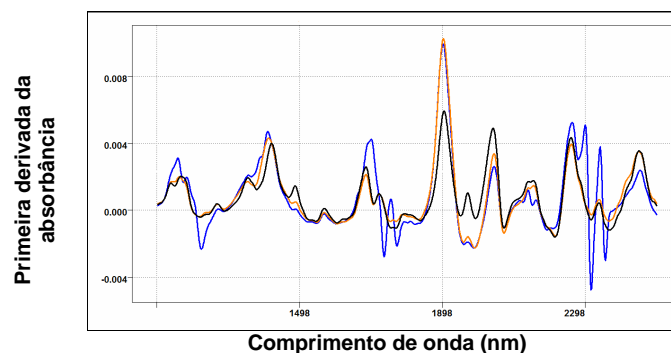
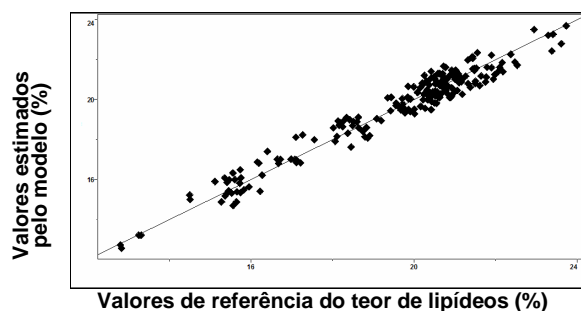


Figura 3. Espectros transformados (MSC e primeira derivada) de amostras de soja moídas: em preto – amostra desengordurada e seca; em laranja – amostra desengordurada; em azul – amostra integral; absorção máxima da água a 1898nm.



| Parâmetros de construção e avaliação da modelagem | |
|---|-------|
| n | 214 |
| k | 7 |
| SEV (%) | 0,499 |
| r_{SEV} | 0,975 |
| SEC (%) | 0,460 |
| r_{SEC} | 0,979 |

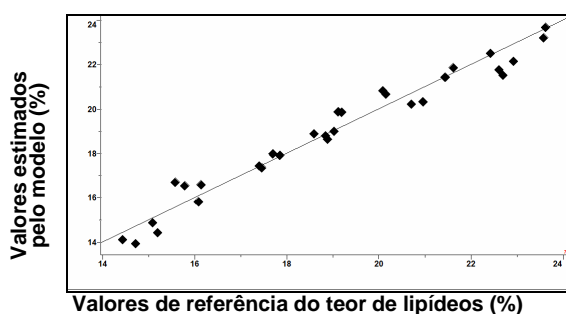
Figura 4. Curva de calibração para dosagem do teor de óleo em grãos de soja: valores de referência versus valores previstos pelo modelo empregando o infravermelho próximo, com os respectivos números de amostras (n) e de variáveis latentes (k) usados no modelo, erros de validação interna e calibração e coeficientes de regressão obtidos na construção do modelo.

Para verificar o desempenho de previsão do modelo, os teores de óleo de 15 amostras foram determinados usando a curva de calibração proposta. Na Tabela 1 encontram-se os valores médios de óleo e os desvios padrões determinados por ambos os métodos, de referência e modelo NIRS/PLS, para o conjunto de amostras de validação, com as respectivas diferenças residuais e erros relativos percentuais entre os métodos aplicados. Avaliando os resultados da Tabela 1, observa-se que o modelo construído apresentou boa capacidade de prever os teores de óleo das sementes, uma vez que os erros relativos entre os métodos aplicados foram inferiores a $\pm 5,00\%$, exceto para uma amostra (161), que apresentou valor acima (6,04%). Igualmente, a Figura 5 mostra um bom ajuste do modelo, boas correlação e linearidade entre os valores experimentais versus os valores preditos pelo modelo NIRS/PLS para o conjunto de validação externa, confirmados pelo baixo erro padrão de previsão (SEP de 0,549%), pelo elevado coeficiente de correlação ($r=0,980$) e pela distribuição aleatória dos pontos na regressão linear.

TABELA 1. Valores médios¹ de predição dos teores de óleo para as amostras usadas no processo de validação externa, com respectivos resíduos e erros percentuais.

| Amostras | Teores e desvios padrões ¹ - método de referência (%) | CV ² (%) | Teores e desvios padrões ¹ - método NIRS/PLS (%) | CV ² (%) | Resíduo | Erro (%) |
|----------|--|---------------------|---|---------------------|---------|----------|
| 8 | 23,60 \pm 0,03 | 0,12 | 23,46 \pm 0,34 | 1,45 | 0,14 | 0,60 |
| 9 | 22,67 \pm 0,35 | 1,56 | 22,34 \pm 0,25 | 1,10 | 0,33 | 1,44 |
| 21 | 20,12 \pm 0,05 | 0,25 | 20,75 \pm 0,10 | 0,49 | -0,64 | -3,16 |
| 29 | 22,65 \pm 0,06 | 0,25 | 21,65 \pm 0,18 | 0,83 | 1,00 | 4,42 |
| 56 | 15,14 \pm 0,08 | 0,51 | 14,66 \pm 0,31 | 2,14 | 0,48 | 3,14 |
| 101 | 17,43 \pm 0,04 | 0,20 | 17,40 \pm 0,07 | 0,43 | 0,02 | 0,13 |
| 102 | 17,77 \pm 0,11 | 0,64 | 17,96 \pm 0,04 | 0,23 | -0,19 | -1,05 |
| 135 | 14,57 \pm 0,20 | 1,36 | 14,02 \pm 0,13 | 0,96 | 0,55 | 3,77 |
| 161 | 15,68 \pm 0,15 | 0,95 | 16,62 \pm 0,11 | 0,68 | -0,95 | -6,04 |
| 184 | 18,86 \pm 0,03 | 0,15 | 18,72 \pm 0,12 | 0,64 | 0,14 | 0,72 |
| 194 | 16,10 \pm 0,04 | 0,22 | 16,21 \pm 0,55 | 3,41 | -0,11 | -0,67 |
| 245 | 19,15 \pm 0,06 | 0,30 | 19,87 \pm 0,02 | 0,09 | -0,72 | -3,74 |
| 267 | 20,83 \pm 0,18 | 0,85 | 20,28 \pm 0,08 | 0,40 | 0,55 | 2,65 |
| 301 | 18,81 \pm 0,30 | 1,62 | 18,95 \pm 0,08 | 0,41 | -0,15 | -0,78 |
| 306 | 21,53 \pm 0,12 | 0,56 | 21,66 \pm 0,30 | 1,37 | -0,13 | -0,61 |

¹ Valores médios de duas análises. ² CV– coeficiente de variação dos valores, expressos em porcentagem.



| Parâmetros de erro padrão da previsão | |
|---------------------------------------|-------|
| n | 30 |
| k | 7 |
| PRESS | 9,046 |
| SEP (%) | 0,549 |
| r | 0,980 |

Figura 5. Correlação entre valores de referência versus os valores estimados pelo modelo NIR, erro padrão de previsão e coeficientes de correlação obtidos na etapa de validação externa.

CONCLUSÃO

Embora nesse estudo tenha utilizado sementes inteiras, foi possível construir uma curva de calibração para dosagem de óleo por NIRS/PLS com bom ajuste, boa exatidão e precisão em relação ao método de referência. Por meio de teste-F, com 5% de probabilidade, não houve diferenças significativas entre os dois métodos aplicados. Os grãos após serem analisados podem ser selecionados e encaminhados ao campo com transferência, para as próximas gerações, da característica de interesse avaliada.

AGRADECIMENTOS

A FUNDAG pela concessão de bolsa de iniciação científica e a FAPESP por financiamento de projeto (R.H.G. PRIOLLI, Proc. BIOEN: 2008/56249-9).

REFERÊNCIAS BIBLIOGRÁFICAS

- BRAGA, J.W.B.; POPPI, R.J. Validação de modelos de calibração multivariada: uma aplicação na determinação de pureza polimórfica de carbamazepina por espectroscopia no infravermelho próximo. **Química Nova**, Vol. 27, No. 6, p. 1004-1011, 2004.
- CARVALHO, C. R. L; MANTOVANI, D.M.B.; CARVALHO, P. R. N.; MORAES, R.M. de. **Análises Químicas de Alimentos**. Instituto de Tecnologia de Alimentos, Campinas, 1990, 121p. (Manual Técnico).
- Worwitz, W. Evaluation of analytical methods used for regulation of foods and drugs. **Analytical Chemistry**, Vol. 54, p.67–76, 1982.
- LIMA, K.M.G., TREVISAM, M.G., POPPI, R.J., ANDRADE, J.C. de. Determinação não destrutiva do nitrogênio total em plantas por espectroscopia de reflectância difusa no infravermelho próximo. **Química Nova**, Vol. 31, No. 3, p. 700-703, 2008.
- MORGANO, M.A. **Aplicação de métodos quimiométricos em análise de alimentos**. Campinas, 2005. 110 f. Tese (Doutorado em Ciências) – Instituto de Química, Universidade Estadual de Campinas, Campinas.
- SILVA, M.S., NAVES, M.M.V., OLIVEIRA, R.B de, LEITE, O de S.M. Composição Química e Valor Protéico do Resíduo de Soja em Relação ao Grão de Soja. **Ciênc. Tecnol. Aliment.**, Campinas, Vol. 26, No. 3, p. 571-576, 2006.
- VIEIRA, R. D. **Avaliação da qualidade fisiológica de sementes de quatorze cultivares de soja (*Glycine max* (L) Merrill)**. 1980. 76f. Dissertação (Mestrado em Fitotecnia) - Universidade Federal de Viçosa, Viçosa, 1980.