

# Espectroscopia no Infravermelho Próximo para Avaliar Indicadores de Qualidade Tecnológica e Contaminantes em Grãos

Casiane Salete Tibola  
Everaldo Paulo de Medeiros  
Maria Lúcia Ferreira Simeone  
Marcelo Alvares de Oliveira  
Editores Técnicos



**Espectroscopia no Infravermelho Próximo  
para Avaliar Indicadores de Qualidade  
Tecnológica e Contaminantes em Grãos**



*Empresa Brasileira de Pesquisa Agropecuária  
Embrapa Trigo  
Ministério da Agricultura, Pecuária e Abastecimento*

# **Espectroscopia no Infravermelho Próximo para Avaliar Indicadores de Qualidade Tecnológica e Contaminantes em Grãos**

Casiane Salete Tibola  
Everaldo Paulo de Medeiros  
Maria Lúcia Ferreira Simeone  
Marcelo Alvares de Oliveira  
*Editores Técnicos*

**Embrapa**  
Brasília, DF  
2018

Exemplares desta publicação podem ser adquiridos na:

**Embrapa Trigo**

Rodovia BR 285, Km 294  
Caixa Postal 3081  
Telefone: (54) 3316-5800  
Fax: (54) 3316-5802  
99050-970 Passo Fundo, RS  
www.embrapa.br/fale-conosco/sac

Comitê Local de Publicações da Embrapa Trigo  
Presidente  
*Leila Maria Costamilan*

Membros

*Alberto Luiz Marsaro Júnior*  
*Alfredo do Nascimento Junior*  
*Anderson Santi*  
*Genei Antonio Dalmago*  
*Sandra Maria Mansur Scagliusi*  
*Tammy Aparecida Manabe Kiihl*  
*Vladirene Macedo Vieira*

Normalização bibliográfica

*Maria Regina Martins*

Projeto gráfico, editoração eletrônica e tratamento das ilustrações

*Fátima Maria De Marchi*

Revisão ortográfica

*Antônio Cláudio da Silva Barros*

Fotos da capa

*Renata Silva (espigas e grãos de milho), Vladimir Moreira (vagem de soja), Danilo Estevão (grãos de soja), Pedro Scheeren (espigas de trigo) e Paulo Odilon Kurtz (grãos de trigo)*

**1ª edição**

1ª impressão (2018): 500 exemplares

**Todos os direitos reservados.**

A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei nº 9.610).

**Dados Internacionais de Catalogação na Publicação (CIP)**

Embrapa Trigo

---

Espectroscopia no Infravermelho próximo para avaliar indicadores de qualidade tecnológica e contaminantes em grãos / Casiane Salete Tibola... [et al.], editores técnicos. – Brasília, DF : Embrapa, 2018.  
200 p.

ISBN: 978-85-7035-839-4

1. Grãos - Qualidade. 2. Grãos - Contaminação. 3. Trigo - Controle de qualidade. 4. Soja - Controle de qualidade. 5. Milho - Controle de qualidade. 6. Espectroscopia. 7. Infravermelho próximo. 8. NIRS. I. Tibola, Casiane Salete. II. Medeiros, Everaldo Paulo de. III. Simeone, Maria Lúcia Ferreira. IV. Oliveira, Marcelo Alvares de.

CDD: 631.56

## **Autores**

### **Ana Rita de Araujo Nogueira**

Química, Dra. em Química Analítica, pesquisadora da Embrapa Pecuária Sudeste, São Carlos, SP.

### **Casiane Salete Tibola**

Engenheira-agrônoma, Dra. em Agronomia, pesquisadora da Embrapa Trigo, Passo Fundo, RS.

### **Celio Pasquini**

Químico, Dr. em Química, professor titular convidado do Instituto de Química da Universidade Estadual de Campinas (Unicamp), Campinas, SP e professor voluntário do Departamento de Solos, Universidade Federal de Viçosa (UFV), Viçosa, MG.

### **Cristina Rocha Vieira Abucáter**

Bacharel em Ciências Biológicas, gerente de Controle de Qualidade da Vilma Alimentos, Contagem, MG.

### **Dagma Dionísia da Silva**

Engenheira-agrônoma, Dra. em Agronomia, pesquisadora da Embrapa Milho e Sorgo, Sete Lagoas, MG.

### **Eliana Maria Guarienti**

Engenheira-agrônoma, Dra. em Tecnologia de Alimentos, pesquisadora da Embrapa Trigo, Passo Fundo, RS.

### **Everaldo Paulo de Medeiros**

Químico-industrial, Dr. em Química, pesquisador da Embrapa Algodão, Campina Grande, PB.

**Francisco Fernandes Gambarra Neto**

Farmacêutico, Dr. em Química, bolsista CAPES/Embrapa. Universidade Federal da Paraíba, João Pessoa, PB.

**Gilberto Batista de Souza**

Químico, Dr. em Química Analítica, analista da Embrapa Pecuária Sudeste, São Carlos, SP.

**Igor Renato Bertoni Olivares**

Químico, Dr. em Química Analítica, professor do Instituto de Química de São Carlos/ Universidade de São Paulo (USP), São Carlos, SP.

**Jayme Garcia Arnal Barbedo**

Engenheiro-eletricista, Dr. em Engenharia Elétrica, pesquisador da Embrapa Informática Agropecuária, Campinas, SP.

**Jez Willian Batista Braga**

Químico, Dr. em Química Analítica, professor do Instituto de Química, Universidade de Brasília, Brasília, DF.

**Jobson Borges de Araújo**

Químico-industrial, especialista em Gestão, Auditoria e Perícia Ambiental, analista da Embrapa Algodão, Campina Grande, PB.

**José Elias Carvalho**

Tecnólogo em alimentos, técnico da Cooperalfa, Chapecó, SC.

**José Marcos Gontijo Mandarinó**

Farmacêutico bioquímico, M.Sc. em Ciência e Tecnologia de Alimentos, pesquisador da Embrapa Soja, Londrina, PR.

**Laiz de Oliveira Magalhães**

Química, M.Sc. em Química, professora da Secretaria de Educação do Governo do Distrito Federal, Brasília, DF.

**Maicon Andre de Paula**

Bacharel em Geografia, analista de Controle de Qualidade da Cocamar Cooperativa Agroindustrial, Maringá, PR.

**Marcelo Alvares de Oliveira**

Engenheiro-agrônomo, Dr. em Agronomia/Horticultura, pesquisador da Embrapa Soja, Londrina, PR.

**Marcelo Martins de Sena**

Químico, Dr. em Química, professor associado do Departamento de Química do Instituto de Ciências Exatas (ICEx), Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG.

**Marco Aurélio Guerra Pimentel**

Engenheiro-agrônomo, Dr. em Entomologia, pesquisador da Embrapa Milho e Sorgo, Sete Lagoas, MG.

**Maria Cristina Dias Paes**

Nutricionista, Ph.D. em Ciência de Alimentos e Nutrição Humana, analista da Embrapa Milho e Sorgo, Sete Lagoas, MG.

**Maria Imaculada Pontes Moreira Lima**

Engenheira-agrônoma, Dra. em Agronomia/Fitopatologia, pesquisadora da Embrapa Trigo, Passo Fundo, RS.

**Maria Lúcia Ferreira Simeone**

Química, Dra. em Química, pesquisadora da Embrapa Milho e Sorgo, Sete Lagoas, MG.

**Mariana Ramos de Almeida**

Química, Dra. em Química, professora adjunta do Departamento de Química do Instituto de Ciências Exatas (ICEx), Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG.

**Martha Zavariz de Miranda**

Farmacêutica bioquímica e industrial, Dra. em Tecnologia de Alimentos, pesquisadora da Embrapa Trigo, Passo Fundo, RS.

**Miguel Marques Gontijo Neto**

Engenheiro-agrônomo, Dr. em Zootecnia, pesquisador da Embrapa Milho e Sorgo, Sete Lagoas, MG.

**Oneida Vasconcelos Vieira**

Química, diretora comercial da CBO Análises Laboratoriais, Valinhos, SP.

**Rodrigo Santos Leite**

Químico, M.Sc. em Tecnologia de Alimentos, analista da Embrapa Soja, Londrina, PR.

**Sidnei Menossi**

Bioquímico, assessor técnico da CBO Análises Laboratoriais, Valinhos, SP.





## Apresentação

O setor de produção de grãos no Brasil caracteriza-se por expressivo incremento tecnológico ocorrido nos últimos anos. Isso permitiu a elevação da produtividade nos campos paralelamente à racionalização no uso de insumos. A demanda atual dos mercados é para a gestão da qualidade (aptidão tecnológica e contaminantes), a preservação da identidade do produto (segregação e rastreabilidade) e a certificação do sistema de produção. As perdas quantitativas e qualitativas na fase de pós-colheita, devido aos contaminantes e à ausência de direcionamento da matéria-prima de acordo com especificações de produto final, comprometem a qualidade e a eficiência do processo produtivo nas lavouras.

Na pós-colheita de grãos, um dos fatores limitantes na escolha do método de análise é o tempo de execução do teste, dificultando a segregação por qualidade, quando esse é demorado. Assim, uma das grandes demandas em pós-colheita de grãos é a seleção de métodos rápidos e eficazes para segregação e rastreabilidade de lotes comerciais, passíveis de ser empregados na recepção de unidades armazenadoras. Atualmente, a segregação de alguns grãos, a exemplo do trigo, é feita em função da classificação comercial da cultivar semeada. Entretanto, este indicativo de qualidade, muitas vezes, não se confirma, em função de fatores que influenciam a qualidade do produto colhido, tais como: adubação, controle fitossanitário, manejo do solo, condições climáticas, manejos pré e pós-colheita, entre outros. Portanto, somente após a adequada avaliação da qualidade na pós-colheita, pode-se definir pela segregação mais adequada.

Neste sentido, o presente livro tem por objetivo divulgar as aplicações de espectroscopia no infravermelho próximo (*Near Infraed Reflectance Spectroscopy* - NIRS) e NIRS associada a imagens hiperspectrais, na avaliação de indicadores de qualidade e de contaminantes em grãos.

Nesta publicação foram reunidas as contribuições de equipe multidisciplinar e multi-institucional, abrangendo diferentes áreas de aplicação de espectroscopia para grãos e produtos derivados. O livro conta com dez capítulos que tratam de temas como: princípios da espectroscopia no infravermelho próximo, características dos principais equipamentos disponíveis, quimiometria, calibração e validação multivariada, aplicações da espectroscopia NIR e NIR hiperespectral na avaliação de indicadores de qualidade em milho, soja e trigo, além de experiências com o uso de NIRS na cadeia produtiva de grãos.

Trata-se de uma obra muito importante para as cadeias produtivas que processam grãos no Brasil, considerando que compila diferentes aplicações da espectroscopia NIR em grãos e derivados. A equipe espera que as experiências compartilhadas no livro colaborem para viabilizar análises de indicadores de qualidade e de contaminantes em grãos, possibilitando a definição de manejo adequado dos lotes em todas as etapas nas diversas cadeias produtivas que usam grãos como matéria-prima, contribuindo para a comercialização de produtos mais seguros e saudáveis para os consumidores.

*Oswaldo Vasconcellos Vieira*  
Chefe-geral da Embrapa Trigo

# Sumário

## *Capítulo 1*

### **Princípios da Espectroscopia no Infravermelho Próximo**

Celio Pasquini..... 13

## *Capítulo 2*

### **Quimiometria Aplicada aos Dados Espectrais no Infravermelho Próximo**

Marcelo Martins de Sena, Mariana Ramos de Almeida .....31

## *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.....51

## *Capítulo 4*

### **Avaliação de Indicadores de Qualidade Tecnológica da Soja por Espectroscopia no Infravermelho Próximo**

Marcelo Alvares de Oliveira, Rodrigo Santos Leite, José Marcos Gontijo Mandarino.....63

**Capítulo 5**

**Avaliação de Parâmetros de Qualidade de Trigo por Espectroscopia no Infravermelho Próximo**

Casiane Salete Tibola, Martha Zavariz de Miranda, Eliana Maria Guarienti .....76

**Capítulo 6**

**Controle de Qualidade para Espectroscopia no Infravermelho Próximo**

Gilberto Batista de Souza, Ana Rita de Araujo Nogueira, Igor Renato Bertoni Olivares .....95

**Capítulo 7**

**Princípios da Espectroscopia no Infravermelho Próximo Associada a Imagens Hiperespectrais**

Francisco Fernandes Gambarra Neto, Joabson Borges de Araújo, Everaldo Paulo de Medeiros .....126

**Capítulo 8**

**Aplicação de Imagens Hiperespectrais na Detecção de Fungos e na Predição de Micotoxinas em Grãos de Trigo**

Jayme Garcia Arnal Barbedo, Casiane Salete Tibola, Maria Imaculada Pontes Moreira Lima .....141

**Capítulo 9**

**Equipamentos Portáteis de Espectrometria no Infravermelho Próximo**

Jez Willian Batista Braga, Laiz de Oliveira Magalhães .....159

**Capítulo 10**

**Aplicações da Espectroscopia no Infravermelho Próximo na Cadeia Produtiva de Grãos**

Maicon Andre de Paula, Oneida Vasconcelos Vieira, Sidnei Menossi, José Elias Carvalho, Cristina Rocha Vieira Abucáter .....177

# Princípios da Espectroscopia no Infravermelho Próximo

Celio Pasquini

## Introdução - Breve histórico da espectroscopia NIR e sua evolução

A radiação eletromagnética referente à faixa do infravermelho próximo (NIR) foi descoberta em 1800 por Sir William Herschel, constituindo a primeira parte do espectro eletromagnético não visível a ser identificada (Herschel, 1800). Após sua descoberta, passou-se quase um século e meio até que as primeiras aplicações analíticas da radiação NIR fossem desenvolvidas. Um dos fatores apontados para este atraso relaciona-se à falta de seletividade dos espectros de absorção na região NIR, sempre referida pelos espectroscopistas de então como uma região espectral de difícil interpretação, na qual os espectros eram constituídos por inúmeras bandas de absorção fortemente sobrepostas. O desinteresse por esta região espectral adiou também a construção de equipamentos dedicados a explorá-la analiticamente. Até a década de 1960, não havia espectrofotômetros comerciais capazes de obter um espectro com qualidade na região NIR. Além destes fatores, os métodos quantitativos clássicos de tratamento de dados exigiam alta seletividade, restringindo o aproveitamento analítico das informações geradas pela espectroscopia NIR.

Embora a literatura registre alguns trabalhos anteriores que fizeram uso da espectroscopia NIR na determinação de água em gelatina e misturas de hidrocarbonetos (Ellis; Bath, 1938; Spakowski et al., 1950), foram os trabalhos desenvolvidos por Karl Norris, no Cereal Centre, em Beltsville, nos Estados Unidos, que de fato lançaram os princípios da técnica analítica e iniciaram a sua disseminação no meio acadêmico (Hart; Norris, 1965; Bengera; Norris, 1968).

Trabalhos relevantes relacionados aos aspectos fundamentais da espectroscopia NIR e a sua instrumentação foram desenvolvidos no ambiente acadêmico por Thomas Hirschfeld, um espectroscopista reconhecido, nas décadas de 1970 e 1980 (Shapiro, 1986), permitindo aumentar o conhecimento sobre esta região espectral até então negligenciada.

A inserção da espectroscopia NIR no ambiente industrial deve ser atribuída a Phil Williams, que substituiu o método de Kjeldahl, intensivamente utilizado na Canadian Grain Commission, pelo então recentemente introduzido método de determinação de proteína baseado em espectroscopia NIR (Williams, 2015). Além de demonstrarem o uso prático da nova tecnologia analítica por meio da substituição de um método via úmida que demandava horas para realizar uma única análise, empregando reagentes tóxicos e de difícil descarte, por uma técnica rápida e limpa, os trabalhos de Phil Williams tiveram o mérito de despertar o interesse das companhias fabricantes de instrumentos, que passaram a produzi-los com as especificações necessárias para operar na faixa espectral do NIR.

Os trabalhos pioneiros desenvolvidos por Karl Norris mostraram que a espectroscopia NIR necessitava de ferramentas matemáticas para contornar o problema da falta de seletividade da informação analítica contida nos espectros NIR. De fato, a determinação de umidade em farinha de trigo, descrita no primeiro trabalho, somente pôde ser viabilizada no trabalho pioneiro de Karl Norris após ter sido estabelecida a correlação do conteúdo de água com a diferença de absorbância entre dois comprimentos de onda, com a finalidade de corrigir efeitos de espalhamento da radiação, que acompanha as medidas de sólidos pulverizados (Hart; Norris, 1965; Bengera; Norris, 1968). Este aspecto da tecnologia NIR motivou, durante a década de 1980, a introdução das técnicas quimiométricas de tratamento multivariado de dados, que estavam sendo desenvolvidas a partir de 1975. A técnica de regressão por quadrados mínimos parciais, uma das mais empregadas pelos métodos analíticos baseados em espectroscopia NIR, foi introduzida na espectroscopia NIR em 1984, por Harold Martens, durante a 2ª Conferência Internacional em Reflectância Difusa, realizada no Wilson College, Chambersburg, PA, Estados Unidos (Williams, 2015).

Portanto, desde meados da década de 1970, a tecnologia NIR passou a ser rapidamente desenvolvida tendo como suporte os fundamentos da espectroscopia vibracional, a instrumentação espectrofotométrica NIR e a quimiometria. Estes três aspectos da tecnologia NIR são igualmente importantes e sua simbiose é notável. As fibras ópticas, o uso em técnicas analíticas de processo (PAT), a miniaturização dos espectrofotômetros e as imagens hiperespectrais NIR constituem as inovações mais recentes incorporadas à tecnologia. A relevância atual dos métodos analíticos qualitativos e quantitativos baseados em espectroscopia NIR tem sido reconhecida com a sua incorporação rotineira em diversas áreas do setor produtivo e também por meio da publicação de normas oficiais elaboradas para guiar desenvolvimento, validação e manutenção destes métodos (American Society for Testing and Materials, 2012, 2016).

A abrangência da tecnologia NIR atingiu um grau tão elevado que as revisões atuais são, quase que em sua maioria, dirigidas às suas áreas específicas de aplicação. No

entanto, livros textos (Williams; Norris, 2001), revisões de caráter geral (McClure, 2003; Pasquini, 2003; Lima et al., 2009; Shao et al., 2012; Zhou et al., 2012; Chu; Lu, 2014; Wilson et al., 2015; Pasquini, 2018) e tutoriais (Agelet; Hurburgh, 2010; Ozaki, 2012) podem ser encontrados na literatura e representam fontes de informação valiosas para aqueles que desejam aprofundar seus conhecimentos sobre a tecnologia NIR.

Embora os três suportes (fundamentos, instrumentação e quimiometria) tenham a mesma importância para a tecnologia NIR, aquele que tem sido menos considerado refere-se aos seus fundamentos. Isso é inconsistente com os preceitos da metodologia analítica, os quais devem ser guiados pelo conhecimento seguro das fontes da informação empregadas pela técnica analítica. Este capítulo visa a destacar os aspectos relacionados aos fundamentos da espectroscopia NIR, assegurando que os métodos analíticos que os empregam como fonte de informação possam ser desenvolvidos, validados e aceitos de forma mais científica e menos empírica.

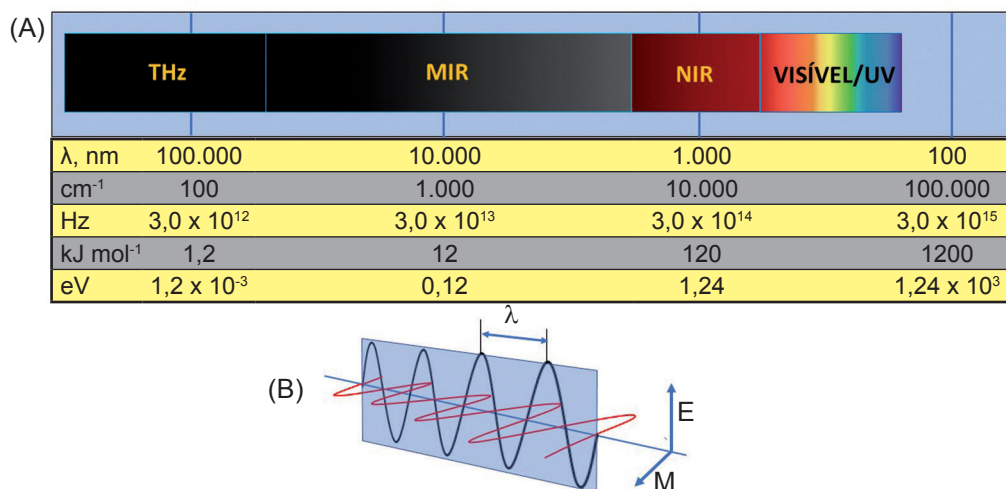
## **Introdução aos fundamentos da espectroscopia no infravermelho próximo**

A espectroscopia é o ramo da ciência voltado para a compreensão da geração da radiação eletromagnética e da sua interação com a matéria. O escopo desta disciplina é muito amplo, uma vez que o espectro eletromagnético se estende por uma larga faixa de energia, possibilitando inúmeras formas de interação da radiação com a matéria, as quais, por sua vez, induzem um número muito grande de fenômenos, muitos deles relacionados à transferência desta energia para seus constituintes. Desta forma, a espectroscopia divide-se em muitas subdisciplinas que se dedicam a estudar faixas relativamente estreitas do espectro eletromagnético, de acordo com suas energias e, conseqüentemente, com os fenômenos que elas podem produzir ao interagir com a matéria. Além disso, os instrumentos empregados para medir a intensidade destas interações apresentam limitações de ordem tecnológica que impedem sua operação em faixas muito amplas do espectro. A Figura 1A permite observar a localização da região espectroscópica NIR entre outras regiões espectrais de maior e menor energia. Dentre as diversas formas de interação da radiação com a matéria, a absorção da radiação pelos constituintes da amostra é de maior interesse para a espectroscopia analítica, pois este fenômeno gera os espectros de absorção que contêm as informações analíticas qualitativas e quantitativas a respeito de uma amostra.



A energia transportada (inclusive através do vácuo) por uma onda eletromagnética, que pode ser representada por um campo elétrico e outro magnético senoidal propagando-se de forma ortogonal (Figura 1B), é dada pela equação:  $E = h\nu$ , onde  $h$  é uma constante (constante de Plank) e  $\nu$  é a frequência Hz da onda eletromagnética.

A espectroscopia no infravermelho próximo (NIR) refere-se à faixa do espectro eletromagnético cujas ondas (ou fótons) apresentam um comprimento de onda que se situa entre 750 nm e 2.500 nm, equivalente à faixa de número de onda entre 4.000  $\text{cm}^{-1}$  e 13.333  $\text{cm}^{-1}$ , frequência entre  $4,0 \times 10^{14}$  Hz e  $1,2 \times 10^{14}$  Hz e energia entre 1,65 eV e 0,50 eV ou 160  $\text{kJ mol}^{-1}$  e 48  $\text{kJ mol}^{-1}$ , respectivamente. A energia transportada pelas ondas eletromagnéticas NIR limita e determina as suas formas de interação com a matéria que causam sua transferência em processos de absorção da radiação. Por exemplo, a energia transportada pela radiação NIR é relativamente baixa e raramente permite que a sua interação com a matéria promova o aumento da energia eletrônica das moléculas ou átomos, como acontece com as suas regiões vizinhas do visível e ultravioleta, de maior energia. Somente alguns elementos químicos Terras Raras, apresentam transições energéticas suficientemente baixas de seus elétrons, permitindo que a radiação NIR as promova ao interagir com uma amostra contendo estas espécies em sua composição.



**Figura 1.** (A) Características da radiação eletromagnética em uma parte do espectro eletromagnético que inclui o infravermelho próximo (NIR). (B) Modelo ondulatório da radiação eletromagnética mostrando a propagação ortogonal dos campos elétrico e magnético e a definição do comprimento de onda ( $\lambda$ ).

Em outro extremo, situam-se os fenômenos associados a transições energéticas de baixíssima energia, como as envolvidas na rotação de moléculas no estado gasoso. Neste caso, a energia transportada pela radiação NIR excede aquela dos níveis rotacionais das moléculas e, em função de restrições quânticas, impede também que esse tipo de radiação promova as moléculas a níveis rotacionais mais energéticos. As limitações impostas pela mecânica quântica, que explica o comportamento dos sistemas microscópicos em nível atômico e molecular, postulam que as espécies atômicas e moleculares apresentem níveis discretos de energia eletrônica, vibracional e rotacional (para gases) e que a transferência da energia da radiação eletromagnética para a matéria, em um processo de absorção, para aumentar sua energia total, requer que esta seja igual à diferença de energia entre os estados discretos de maior e menor energia  $\Delta E = (E_f - E_i)$ , onde  $E_f$  é a energia do estado mais energético (estado excitado) e  $E_i$  é a energia inicial do estado não excitado (estado fundamental). Ou seja,

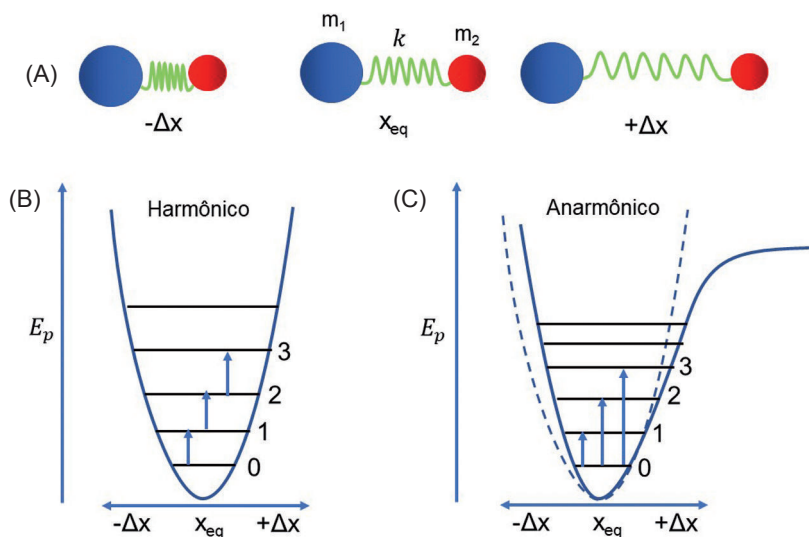
$$\Delta E = hv \quad (1)$$

Na escala de energia, as transições associadas à energia vibracional das moléculas também se mostram inferiores à energia da radiação NIR, quando as transições entre níveis quânticos energéticos adjacentes são consideradas. Estas transições resultam da interação e absorção da radiação na faixa do infravermelho médio (MIR), a qual se situa também vizinha ao NIR, do lado de menor energia. Então, quais são as formas de interação da radiação NIR com a matéria que possibilitam sua absorção? Que tipo de transição energética os fótons NIR podem promover? Para responder estas perguntas, é necessário compreender o comportamento vibracional das moléculas de forma geral para, a partir dele, compreender como a radiação NIR interage com a matéria e como esta interação permite a obtenção de informação associada à sua composição química. Esta informação constitui a base para o desenvolvimento dos métodos analíticos qualitativos e quantitativos baseados em espectroscopia NIR, apresentados neste livro.

## Espectroscopia vibracional

### Modelo harmônico clássico

Uma molécula diatômica simples pode ser representada classicamente por duas pequenas esferas, feitas, digamos, de aço inox, de massas diferentes e conectadas uma à outra por meio de uma mola, como mostrado na Figura 2A.



**Figura 2.** (A) Modelo de uma molécula diatômica representada por duas esferas com massas  $m_1$  e  $m_2$  unidas por uma mola com constante de força  $k$ . (B) Variação da energia potencial em um sistema vibracional harmônico mostrando o comportamento previsto pela mecânica quântica para os níveis vibracionais. (C) Variação da energia potencial em um sistema vibracional anarmônico e o comportamento dos níveis vibracionais. As transições possíveis entre níveis vibracionais são representadas pelas setas azuis.

Neste modelo, as massas e das esferas e a constante de força da mola estabelecem, de acordo com a mecânica clássica, uma frequência natural de vibração deste sistema dada por:

$$v = \frac{1}{2\pi} \sqrt{\frac{k}{m_r}} \quad (2)$$

onde  $v$  é a frequência de vibração do sistema, e  $m_r$  é a sua massa reduzida, dada por:

$$m_r = \frac{m_1 m_2}{m_1 + m_2} \quad (3)$$

A energia potencial  $E_p$  deste sistema pode ser alterada por compressão ou distensão da mola em relação à sua posição de equilíbrio (energia igual a zero), que define a amplitude da vibração  $x$  e é expressa pela equação:

$$E_p = \frac{1}{2} kx^2 \quad (4)$$

Um sistema que se comporta desta forma em relação à sua energia vibracional é denominado de sistema harmônico, representado na Figura 2B. Neste modelo clássico, o sistema pode vibrar com a mesma frequência em qualquer amplitude e transitar entre elas continuamente sem nenhuma restrição.

## Modelo harmônico quântico

Entretanto, o modelo baseado em mecânica clássica não se sustenta para sistemas moleculares, nos quais as esferas são, na verdade, átomos ligados entre si, e a mola é constituída pela ligação química que os une. Neste caso, o comportamento vibracional do sistema é melhor descrito pela mecânica quântica. Uma restrição importante imposta pela mecânica quântica aos sistemas vibracionais moleculares refere-se ao fato de que eles não podem vibrar em qualquer amplitude, ou seja, sua energia vibracional é quantizada e somente certos estados energéticos vibracionais são permitidos à molécula. Resultados quantitativos da abordagem mecânica-quântica dos sistemas moleculares harmônicos revelam também que os níveis de energia permitidos são igualmente espaçados, como representados na Figura 2B. A equação (2) ainda define a frequência de vibração do sistema molecular, mas as diferenças de energia entre os estados (0 e 1) e (1 e 2), representados na Figura 2B, são iguais. O nível de energia vibracional designado como zero é o nível que as moléculas ocupam em temperatura ambiente, na qual elas vibram com amplitudes e energia vibracional relativamente baixas e no qual quase todas as moléculas se encontram. A energia vibracional seria igual a zero somente em uma situação em que a movimentação dos átomos das moléculas cessasse completamente, assumindo suas posições de equilíbrio.

Considerando que a energia transportada pela radiação eletromagnética pode ser transferida para um sistema vibracional molecular harmônico e a restrição quântica, que impõe que esta energia deve ser igual à diferença de energia entre os estados energéticos vibracionais da molécula, concluímos, das equações (1) e (2), que a radiação capaz, por exemplo, de promover a energia vibracional do estado 0 para 1 equivale àquela da radiação eletromagnética de frequência  $\nu$  (igual à frequência natural de vibração do sistema harmônico molecular), que proporciona a energia igual a  $h\nu$ , correspondente à diferença de energia entre os níveis vibracionais. Se, por exemplo, em uma temperatura mais alta, tivéssemos muitas moléculas ocupando o nível 1, para promovê-las ao nível 2 teríamos que irradiá-las com a mesma frequência  $\nu$ , uma

vez que os níveis energéticos vibracionais são igualmente espaçados e transições saltando, por exemplo, de 0 para 2 diretamente são proibidas pelo modelo quântico harmônico. A faixa de frequência que corresponde à diferença de energia necessária para excitação vibracional de sistemas harmônicos moleculares situa-se na região do infravermelho médio (MIR), variando principalmente de acordo com a massa dos átomos ligados e da energia da ligação química entre eles (“o valor do  $k$  da mola”), como previsto pela equação (2). Sendo assim, a interação da radiação NIR com sistemas vibracionais moleculares não ocorreria, e a espectroscopia NIR não existiria!

## Modelo anarmônico quântico

No entanto, os sistemas moleculares vibracionais não se comportam de forma perfeitamente harmônica, como assumido anteriormente. A repulsão de átomos, ao se aproximarem em movimentos de compressão da ligação química, e o fato de as ligações químicas tenderem a se enfraquecer e se romper quando os átomos se afastam suficientemente em movimentos de distensão da ligação, geram um comportamento anarmônico da energia potencial vibracional em sistemas moleculares. A Figura 2C mostra uma função, denominada de função de Morse, que descreve o comportamento da energia potencial de sistemas vibracionais moleculares de forma mais próxima à realidade. A descrição mecânica-quântica destes sistemas anarmônicos produz resultados bem diferentes daqueles do sistema harmônico.

O primeiro é que nestes sistemas, os níveis energéticos vibracionais não são igualmente espaçados, mas diminuem a diferença entre si à medida que se consideram os níveis de energia vibracional mais altos. O fato de que a ruptura da ligação química pode ocorrer para grandes deslocamentos de afastamento dos átomos ligados é considerado no sistema anarmônico, pois, após a ruptura, o sistema não pode ter mais sua energia potencial aumentada pelo distanciamento dos átomos livres. A discrepância entre as diferenças de energia entre os níveis vibracionais adjacentes, sempre iguais para um sistema harmônico, e aquelas previstas entre dois estados adjacentes de um sistema anarmônico permite calcular o grau de anarmonicidade do sistema vibracional. Alguns modos vibracionais são bastante harmônicos e todos eles são assim quando se referem a transições denominadas fundamentais, que ocorrem entre os níveis vibracionais 0 e 1. Observe que a curva que descreve a energia potencial do sistema anarmônico se aproxima muito da parábola prevista para os sistemas harmônicos, considerando-se as transições fundamentais entre os níveis vibracionais 0 e 1.

O segundo resultado é bastante relevante para a espectroscopia NIR e refere-se ao fato de que, em função do grau de anarmonicidade do modo vibracional, a energia

vibracional de uma molécula pode transitar, por exemplo, diretamente do nível fundamental 0 para o nível 2, 3 ou mesmo 4. Estas transições são denominadas transições de sobretom. A transição direta entre o nível 0 (no qual se encontra a maioria das moléculas em temperatura ambiente) e o nível 2 é denominada de primeiro sobretom, do nível 0 para o 3 de segundo sobretom, e assim por diante. Agora, a restrição quântica de que a energia da radiação eletromagnética deve ser igual à diferença de energia entre os estados mais e menos energético nos leva a faixas de energia para a transição direta entre os níveis 0 e 2 próximas do dobro daquela referente às transições fundamentais entre os níveis 0 e 1. As transições que requerem energias da radiação MIR relativamente altas para excitar estas transições terão as energias dos seus sobretoms associadas à radiação mais energética localizada na faixa característica do infravermelho próximo (NIR). Assim, as transições referentes a sobretoms podem ser induzidas pela radiação NIR. A transferência de energia é medida pela intensidade de absorvância da radiação em uma determinada frequência (ou comprimento de onda). Assim, nos comprimentos de onda referentes às transições de sobretom observam-se bandas de absorção na região energética referente ao NIR. Deve ser observado que estas transições energéticas referentes aos sobretoms são permitidas pela distorção do modelo harmônico. Portanto, a intensidade com as quais elas são registradas nos espectros NIR são muito inferiores àquelas registradas nos espectros MIR, onde as transições fundamentais são observadas. A diferença típica é de 10, 100 e 10.000 vezes menor, à medida que passamos para o primeiro, segundo e terceiro sobretoms, respectivamente. As espectroscopias MIR e NIR estão bastante correlacionadas, como se pode concluir até aqui.

Outro resultado importante da anarmonicidade para a espectroscopia NIR está no fato de que, também em função da sua intensidade em uma ligação química, a combinação de modos vibracionais também se torna possível. No exemplo simples de uma molécula diatômica empregado até o momento, há somente um modo vibracional que consiste no estiramento e contração da ligação química entre os dois átomos ligados. Porém, para moléculas com mais de dois átomos em sua estrutura, diversos outros modos vibracionais podem existir. Considere, por exemplo, a molécula não linear da água. Nesta molécula é possível facilmente distinguir dois modos vibracionais associados ao estiramento e à contração das ligações O – H e outro associado à deformação do ângulo entre as ligações H – O – H. Com alguma restrição e sempre como consequência da distorção anarmônica, estes modos podem se combinar em um modo misto que irá requerer, para a sua excitação, a soma das energias necessárias para excitar cada um deles individualmente. Assim, a frequência da radiação eletromagnética necessária para excitar uma combinação

de modos vibracionais será igual à soma das frequências necessárias para excitar cada um dos modos. Isso resulta em que muitas combinações de modos vibracionais são excitadas por radiação de energia na região NIR do espectro eletromagnético. Novamente, como as combinações de modos vibracionais são permitidas apenas pela anarmonicidade, a intensidade com a qual elas ocorrem é menor que as transições individuais fundamentais registradas na região MIR.

Até o momento foram considerados somente os aspectos mecânicos associados às vibrações moleculares. Daí estabeleceu-se a igualdade das frequências (ressonância) entre as ondas eletromagnéticas e as frequências de vibração das ligações químicas nas moléculas para efetiva transferência de energia da radiação. Porém, há um outro aspecto relevante relacionado à interação da radiação eletromagnética com a matéria, associado à possibilidade ou não da transferência da energia por ela transportada para os sistemas vibracionais moleculares. Este aspecto considera que o acoplamento entre a energia transferida pela radiação eletromagnética e um modo vibracional mecânico dos átomos em uma ligação química somente é possível se o modo vibracional gerar uma variação do momento de dipolo local. Assim, um campo elétrico oscilante deve acompanhar o movimento vibracional dos átomos ligados, permitindo que o campo elétrico também oscilante da radiação eletromagnética acople sua energia ao sistema mecânico vibracional. Observe que, em razão deste fato, moléculas diatômicas como o  $N_2$  ou  $O_2$ , embora tenham seus níveis vibracionais bem estabelecidos, não podem ter suas energias vibracionais aumentadas por meio de excitação com radiação eletromagnética, pois o único modo vibracional nestas moléculas consiste no estiramento da ligação entre seus dois átomos iguais, que não gera nenhuma variação do momento de dipolo. Desta forma, estas moléculas não absorvem radiação na região do MIR e do NIR, sendo completamente transparentes nestas faixas espectrais. No exemplo que foi considerado anteriormente de uma molécula com dois átomos diferentes, a variação de eletronegatividade entre eles define um dipolo permanente na molécula, que se altera com a distância entre os átomos produzida pelo modo vibracional. Esta molécula certamente absorverá radiação eletromagnética nas regiões do MIR e muito provavelmente na região do NIR.

## **Informação química em espectros de absorção NIR**

Além dos sobretons e combinações de modos vibracionais, outros fenômenos relacionados com os modos vibracionais em moléculas podem ser observados na região

NIR. A ressonância de Fermi (acoplamento de 2ª ordem) é um fenômeno de acoplamento entre uma transição energética vibracional fundamental e um sobretom. Neste caso, ao redor do comprimento de onda esperado para a ocorrência do sobretom observam-se duas bandas oriundas do fenômeno de ressonância. A ressonância de Darling-Dennison é outro tipo de fenômeno que pode ser observado entre sobretoms de modos vibracionais na região NIR do espectro eletromagnético.

Sobretoms, combinações e ressonâncias são, em grau de relevância decrescente, os principais fenômenos responsáveis pelo aparecimento de bandas de absorção nos espectros NIR. Para serem ativos e absorverem a radiação NIR, estes fenômenos vibracionais precisam ser acompanhados por uma variação do momento de dipolo. A intensidade de absorção da radiação NIR por espécies químicas moleculares depende do grau de anarmonicidade do modo vibracional e da grandeza da variação do momento de dipolo que ele produz. Assim, por exemplo, moléculas de água, absorvem fortemente na região NIR por causa do grau de anarmonicidade das ligações O – H, que permitem a ocorrência de sobretoms e combinações dos seus modos vibracionais, e da alteração significativa do momento de dipolo gerado pelos seus modos vibracionais.

Como manifestado, o grau de anarmonicidade das ligações químicas envolvidas nos modos vibracionais moleculares é muito relevante para a espectroscopia NIR. Desta forma, ligações químicas mais anarmônicas têm maior probabilidade de gerarem bandas de absorção na região NIR. Além disso, para que uma banda de absorção, por causa de um sobretom ou combinação, seja observada na região NIR, a energia da transição fundamental, no caso dos sobretoms, ou das transições fundamentais nas combinações, deve ser alta. Assim, a energia destes modos vibracionais situa-se na região espectral NIR. Estes requisitos são atendidos por ligações envolvendo o átomo de hidrogênio ligado a um átomo mais pesado, como carbono, nitrogênio oxigênio e enxofre. Desta forma, espectros NIR sempre apresentam bandas de absorção referentes aos sobretoms e combinações de modos vibracionais associados às ligações C – H, N – H, O – H e S – H, denominadas de grupos funcionais. Outra possibilidade vem de ligações químicas de alta energia, como C = C, C ≡ C e C = O, cujas interações com a radiação eletromagnética podem ser observadas na região NIR.

Desta forma, as informações registradas por meio de espectros de absorção NIR apresentam característica universal para compostos orgânicos, pois todos eles contêm em sua estrutura os grupos de átomos mencionados acima. Por outro lado, a seletividade das absorções na região NIR é bastante comprometida, uma vez que, para



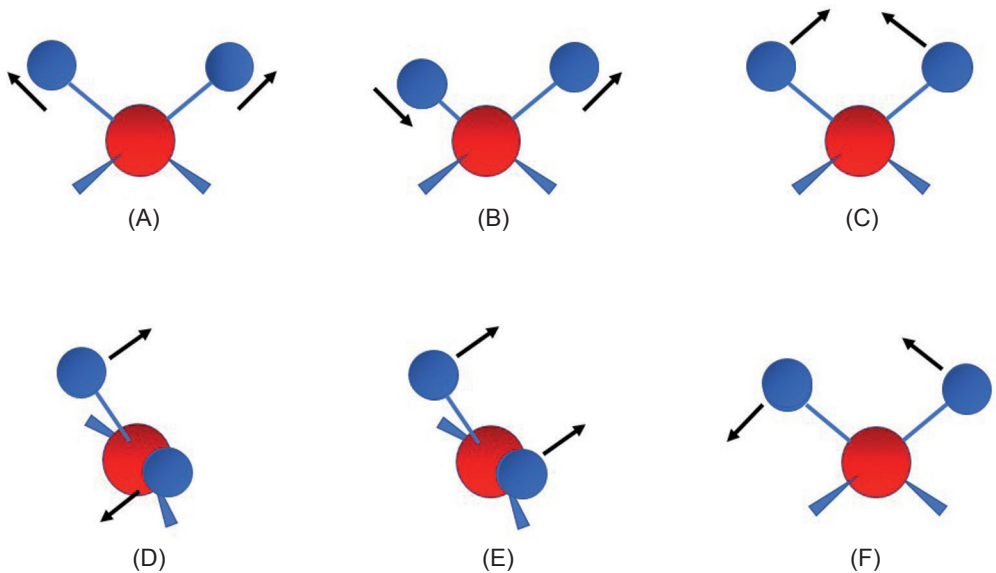
cada grupo funcional, as energias vibracionais se alteram muito pouco nas moléculas, dependendo principalmente da massa e energia da ligação entre os átomos ligados, de acordo com a equação (2). Neste ponto, caberia a questão: se os espectros NIR registram informações sobre a absorção basicamente atribuída somente aos grupos funcionais C – H, N – H, O – H e S – H, como a espectroscopia NIR poderia ser empregada como fonte de informação nos diversos métodos analíticos descritos neste livro e em outros milhares, descritos na literatura?

## Efeitos primários e secundários em espectros de absorção NIR

A possibilidade de uso dos espectros de absorção NIR como fonte de informação para o desenvolvimento de métodos analíticos qualitativos e quantitativos advém principalmente das absorções geradas pelos grupos funcionais mencionados acima e de uma série de efeitos sobre estes grupos e sobre suas interações intra e intermoleculares associadas à composição química global da amostra.

Efeitos primários são aqueles associados diretamente ao grupo funcional em nível atômico. Assim, o grupo C – H deve contribuir com bandas de absorção NIR em comprimentos de onda menores que o grupo funcional O – H, em virtude da diferença de massa entre o átomo de carbono e oxigênio e da energia da ligação entre eles e os átomos de hidrogênio, conforme previsto pela equação (2). Estes efeitos são aqueles capazes de causar maiores diferenças entre espectros de absorção NIR. Mesmo assim, as bandas de absorção de sobretons e combinações associadas diretamente aos grupos funcionais sobrepõem-se significativamente ao longo da faixa espectral NIR. A Figura 3 mostra os possíveis modos vibracionais associados ao grupo funcional – CH<sub>2</sub>. Seus sobretons e combinações geram diversas ocorrências espectrais na região do infravermelho próximo.

A Figura 4 mostra a localização aproximada das absorções atribuídas a três grupos funcionais comumente encontrados em amostras determinadas por meio de espectroscopia NIR: C – H, O – H e N – H.



**Figura 3.** Modos vibracionais normais do grupo – CH<sub>2</sub> em uma molécula: (A) estiramento simétrico, (B) estiramento assimétrico, (C) deformação angular no plano, (D) torção, (E) deformação angular fora do plano e (F) deformação no plano. Em (D) e (E) o plano dos átomos de hidrogênio do grupo funcional foi inclinado para possibilitar melhor visualização dos modos vibracionais. As esferas vermelhas e azuis representam os átomos de carbono e hidrogênio, respectivamente.

Comprimento de onda, nm	2.500	C – H	Combinação de vibrações
	2.200	O – H N – H	Combinação de vibrações
	1.800	C – H	Primeiro sobretom
	1.600	N – H O – H	Primeiro sobretom
	1.420	C – H	Sobretom de combinações
	1.300	C – H	Segundo sobretom
	900	N – H C – H	Terceiro sobretom
800			

**Figura 4.** Localização aproximada das ocorrências espectrais devido à absorção da radiação NIR por grupos funcionais C – H, O – H e N – H.

Os efeitos secundários referem-se às modificações induzidas nos dois grupos funcionais em nível atômico, microscópico e macroscópico, as quais podem induzir alterações nas energias de ligação química entre os átomos dos grupos funcionais, alterando as frequências de vibração, como previsto pela equação 2, e/ou do grau de anarmonicidade destas ligações, alterando as intensidades de absorção e também os comprimentos de onda nos quais elas são observadas. Por exemplo, em nível atômico, a ligação O – H da água envolvida na formação de uma ligação de hidrogênio em uma interação intermolecular diminui a energia de ligação do grupo O – H. Além disso, o átomo de hidrogênio envolvido na ligação de hidrogênio está confinado entre dois átomos de oxigênio, impedindo sua movimentação mais ampla, por exemplo, em um modo vibracional de estiramento. Assim, o grau de anarmonicidade da ligação diminui, com consequências sobre o espectro de absorção NIR, observadas, por exemplo, pela diminuição da intensidade de absorção associadas a este grupo funcional. Outro efeito secundário, agora em nível intramolecular, refere-se à simetria da molécula. Combinações de modos vibracionais são suscetíveis ao arranjo tridimensional dos átomos formando a molécula. As combinações de modos vibracionais, outra fonte de informação importante em espectroscopia NIR, são dependentes desta distribuição, e a sua modificação pode favorecer ou desfavorecer as combinações dos modos vibracionais. Finalmente, a substituição de um átomo pertencente a um dos grupos funcionais mencionados por um átomo com características muito diferentes pode influenciar as energias de ligação dos átomos nos grupos funcionais restantes na molécula. Assim, por exemplo, a substituição de um único átomo de hidrogênio, dentre os seis presentes em uma molécula de benzeno, por um átomo de cloro é facilmente perceptível por meio das alterações observadas entre os espectros de absorção do benzeno (molécula original) e do cloro-benzeno (molécula derivada). Observe-se que a ligação C – Cl do clorobenzeno não pode gerar absorções na região NIR, pois as massas envolvidas na ligação remetem a energia e, conseqüentemente, a frequência de vibração deste sistema para valores muito baixos, diminui a anarmonicidade da ligação e, conseqüentemente, as intensidades de absorção em razão dos primeiros sobretons, que em sua maioria, pela baixa energia da transição fundamental, são também observados na região MIR. Ou seja, os espectros do benzeno e do clorobenzeno continuam refletindo somente a presença dos grupos funcionais C – H em suas moléculas. Porém, a molécula substituída teve a energia das ligações C – H remanescentes alterada pela inclusão de um átomo pesado e muito eletronegativo no anel aromático. Em conseqüência, o seu espectro NIR modificou-se em relação ao da molécula original.

Em nível microscópico, destacam-se os efeitos associados ao arranjo cristalino das moléculas e aqueles associados à segregação de espécies em domínios microscópicos de acordo com seu potencial de interação. Assim, o mesmo material,

em sua forma amorfa ou cristalina, gera espectros de absorção que apresentam diferenças significativas em razão do arranjo ordenado ou desordenado das moléculas e seu efeito secundário sobre a energia e anarmonicidade das ligações dos grupos funcionais presentes na molécula.

Em nível macroscópico, os efeitos secundários são associados a alterações mecânicas e a efeitos da temperatura sobre os espectros NIR. O efeito da temperatura é mais relevante sob o ponto de vista analítico, pois chama a atenção sobre a necessidade de se considerar a temperatura da amostra no desenvolvimento de métodos analíticos qualitativos e, principalmente, quantitativos. Desta forma, por exemplo, o espectro de absorção da água é afetado pela temperatura, desviando, inclusive, a posição de máximos de absorção em consequência do seu efeito na dinâmica de formação das ligações de hidrogênio associadas à estrutura da água líquida. Efeitos mecânicos são observados comparando-se os espectros de absorção de polímeros obtidos antes e depois deles serem submetidos a estresse mecânico por estiramento. A alteração causada pelo estresse na disposição das cadeias poliméricas, e consequente alteração das forças de interação entre elas, modifica os espectros de absorção do polímero.

Os efeitos secundários descritos acima alteram os espectros de absorção NIR mesmo que, em sua essência, eles estejam associados a poucos grupos funcionais (C – H, N – H, O – H e S – H) presentes nas espécies químicas constituintes de uma amostra e, desta forma, os espectros NIR adquirem alguma seletividade. Porém, de forma diferente da espectroscopia analítica convencional, as informações seletivas presentes em espectros NIR não se encontram evidenciadas e dificilmente podem ser identificadas sem o uso de algoritmos computacionais desenvolvidos para extraí-las dos espectros NIR. Um número muito grande de métodos analíticos desenvolvidos com base na espectroscopia NIR está fundamentado nestes efeitos secundários.

## Considerações finais

Neste primeiro capítulo deste livro, a origem da informação analítica contida em espectros NIR foi situada. A espectroscopia NIR pode então ser compreendida como uma modalidade de espectroscopia vibracional de energia relativamente alta. A região NIR do espectro eletromagnético permite observar os fenômenos de transferência de energia eletromagnética da radiação NIR para os sistemas vibracionais mecânicos atômicos encontrados principalmente em moléculas orgânicas que apresentam os grupos funcionais C – H, N – H, O – H e S – H. Os sobretons, combinações e alguns

tipos de ressonância dos modos vibracionais destes grupos funcionais são responsáveis pela maioria absoluta das absorções observadas nos espectros NIR destas moléculas. O caráter destas informações é universal, pois todas as moléculas orgânicas possuem um ou mais destes grupos funcionais em sua estrutura.

Os métodos analíticos também necessitam de seletividade. Esta somente pode ser alcançada se as diferentes espécies químicas apresentarem, em algum grau, espectros NIR diferentes. Em espectroscopia NIR, estas diferenças são produzidas em sua maioria por efeitos indiretos (secundários) sobre os grupos funcionais mencionados anteriormente, alterando suas características relacionadas à energia de ligação, seu grau de anarmonicidade e as possibilidades de combinação entre modos vibracionais. A seletividade adquirida por meio destes efeitos não permite o uso das técnicas analíticas quantitativas convencionais e somente pode ser explorada com o auxílio das ferramentas matemáticas e estatísticas desenvolvidas no âmbito da área de estudo denominada Quimiometria, abordada no capítulo 2.

A sensibilidade da espectroscopia NIR pode ser considerada muito boa, principalmente porque a relação sinal/ruído dos espectrofotômetros NIR é muito alta ( $\geq 10.000$ ). Porém, a origem da informação analítica nas transições de sobretons e combinações limitam esta sensibilidade. Um método analítico desenvolvido e devidamente validado para determinar constituintes de uma amostra ou suas propriedades associadas a estes constituintes presentes em teores abaixo de 0,1% (m/m) precisa ter a fonte da informação analítica que permite a obtenção deste desempenho explicitada.

A espectroscopia NIR tem sido muito bem explorada para produzir métodos analíticos precisos, exatos, rápidos e que prescindem de tratamentos complexos e poluentes da amostra, durante mais de 50 anos. Ainda há muitas aplicações potenciais a serem exploradas.

## Agradecimentos

O autor agradece ao CNPq (proc. número 465768/2014-8) e à Fapesp (proc. número 2014/50951-4) pelo financiamento do Instituto Nacional de Ciências e Tecnologias Analíticas Avançadas (INCTAA), e ao CNPq (proc. número 30754/2014-2) pela bolsa de produtividade em pesquisa.

## Referências

AGELET, L. E.; HURBURGH, C. R. A tutorial on near infrared spectroscopy and its calibration. **Critical Reviews in Analytical Chemistry**, v. 40, n. 4, p. 246-260, 2010.

ASTM. **E1655 - 05**: standard practices for infrared multivariate quantitative analysis. West Conshohocken, 2012.

ASTM. **E1790 - 04e1**: standard practice for near infrared qualitative analysis. West Conshohocken, 2016.

BENGERA, I.; NORRIS, K. H. Direct spectrophotometric determination of fat and moisture in meat products. **Journal of Food Science**, v. 33, n. 1, p. 64-67, 1968.

CHU, X. L.; LU, W. Z. Research and application progress of near infrared spectroscopy analytical technology in China in the past five years. **Spectroscopy and Spectral Analysis**, v. 34, n. 10, p. 2595-2605, 2014.

ELLIS, J. W.; BATH, J. Alterations in the infrared absorption spectrum of water in gelatin. **Journal of Chemical Physics**, v. 6, n. 2, p. 108, 1938.

HART, J. H.; NORRIS, K. H. Direct spectrophotometric determination of moisture content of grain and seeds. In: WEXLER, A. (Ed.). **Humidity and moisture, principles and methods of measuring moisture in liquid and solids**. New York: Reinhold Publishing Co., 1965. v. 4, p. 19-25.

HERSCHEL, F. W. Experiments on the refrangibility of the invisible rays of the sun, philosophical. **Transactions of the Royal Academy**, v. 90, p. 284-292, 1800. Disponível em: <<https://archive.org/stream/philtrans08733349/08733349#page/n0/mode/2up>>. Acesso em: 10 nov. 2017.

LIMA, K. M. G.; RAIMUNDO, I. M.; SILVA, A. M. S.; PIMENTEL, M. F. Near infrared and mid infrared optical sensors. **Quimica Nova**, v. 32, n. 6, p. 1635-1643, 2009.

MCCLURE, W. F. 204 years of near infrared technology: 1800-2003. **Journal of Near Infrared Spectroscopy**, v. 11, n. 6, p. 487-518, 2003.

OZAKI, Y. Near-infrared spectroscopy - its versatility in analytical chemistry. **Analytical Sciences**, v. 28, n. 6, p. 545-563, 2012.

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.

- PASQUINI, C. Near infrared spectroscopy: A mature analytical technique with new perspectives: a review. **Analytica Chimica Acta**, v. 1026, p. 8-36, 2018.
- SHAO, X. G.; NING, Y.; LIU, F. X.; LI, J. H.; CAI, W. S. Application of near-infrared spectroscopy in micro inorganic analysis. **Acta Chimica Sinica**, v. 70, n. 20, p. 2109-2114, 2012.
- SHAPIRO, H. M. In Memoriam: Tomas Hirschfeld (1939-1986). **Cytometry**, v. 7, n. 5, p. 399, 1986.
- SPAKOWSKI, A. E.; EVANS, A.; HIBBARD, R. R. Determination of aromatics and olefins in wide petroleum fractions. **Analytical Chemistry**, v. 22, n. 11, p. 1419-1422, 1950.
- WILSON, R. H.; NADEAU, K. P.; JAWORSKI, F. B.; TROMBERG, B. J.; DURKIN, A. J. Review of short-wave infrared spectroscopy and imaging methods for biological tissue characterization. **Journal of Biomedical Optics**, v. 20, n. 3, p. 030901-10, 2015.
- WILLIAMS, P. How it all began. In: INTERNATIONAL CONFERENCE ON NEAR INFRARED SPECTROSCOPY, 17., 2015, Foz do Iguassu. **Proceedings...** Campinas: International Council for Near Infrared Spectroscopy, 2015. p. 128-132. Disponível em: <<https://proceedings.galao.com.br/NIR/papers/how-it-all-began>>. Acesso em: 20 ago. 2017.
- WILLIAMS, P.; NORRIS, K. **Near-infrared technology in the agricultural and food industries**. 2nd ed. St. Paul: American Association of Cereal Chemists, 2001. 296 p.
- ZHOU, W. H.; YING, Y. B.; XIE, L. J. Spectral database systems: a review. **Applied Spectroscopy Reviews**, v. 47, n. 8, p. 654-670, 2012.

# Quimiometria Aplicada aos Dados Espectrais no Infravermelho Próximo

Marcelo Martins de Sena  
Mariana Ramos de Almeida

## Introdução

A absorção de radiação na região do infravermelho próximo (NIR, *near infrared*) por moléculas dá origem a espectros bastante complexos, os quais no passado já foram considerados impossíveis de utilizar (Davies, 1998). Os sinais observados nessa região, referentes a bandas de combinações e sobretons de ligações C-H, O-H, N-H e S-H, são relativamente pouco intensos e bastante sobrepostos. Como tais ligações estão presentes na maior parte das moléculas, a utilização da espectroscopia NIR para comparações com padrões espectrais torna-se difícil, ao contrário do que acontece tradicionalmente com a espectroscopia no infravermelho médio. Além disso, a grande sobreposição espectral torna rara a possibilidade de se encontrar um comprimento de onda seletivo, mesmo para matrizes pouco complexas. Por isso, a utilização da calibração univariada em conjunto com espectros NIR é praticamente inviável. Em função destes motivos, a espectroscopia NIR ficou por muito tempo “adormecida” dentre as técnicas espectroscópicas (Wetzel, 1983). Na verdade, a espectroscopia NIR só começou a se difundir a partir dos anos 1970, com a chegada dos microprocessadores e microcomputadores aos laboratórios químicos, e o consequente aumento da capacidade de geração de dados. Com a necessidade de tratamento dos dados gerados pelos químicos, a aplicação de estatística multivariada em química ganhou maturidade e deu origem a uma nova disciplina, batizada de quimiometria (Bruns; Faigle, 1985).

A difusão do uso da espectroscopia NIR sempre esteve estritamente ligada ao desenvolvimento da quimiometria. De maneira recíproca, pode-se afirmar também que a evolução da quimiometria recebeu grande contribuição das demandas geradas pelas aplicações desta técnica. Na verdade, existe historicamente uma “simbiose” entre espectroscopia NIR e quimiometria, que tem contribuído sinergicamente para a evolução de ambas. Desde os primeiros trabalhos publicados pelo pioneiro Karl Norris nos anos 1960, já se notava a necessidade de analisar mais de uma variável para se



extrair informação útil de espectros NIR. A “maneira de pensar multivariada” já estava presente nos primeiros modelos NIR quantitativos desenvolvidos por Norris, o qual buscava empiricamente encontrar alguns comprimentos de onda seletivos que permitissem a utilização de regressão linear múltipla para, por exemplo, determinar o teor de umidade em grãos e sementes (Norris; Hart, 1996; Pasquini, 2003).

Enquanto num período de cinco décadas, entre 1930 e 1980, apenas 255 artigos foram publicados citando espectroscopia NIR (Pasquini, 2003), a década de 1990-1999 marcou uma verdadeira explosão do uso desta técnica (McClure, 1994), com mais de 13 mil artigos publicados. Foi nesta década que surgiu um periódico específico, o *Journal of Near Infrared Spectroscopy*, não por acaso, apenas alguns anos depois do surgimento das duas revistas especializadas em quimiometria fundadas na década anterior, *Chemometrics and Intelligent Laboratory Systems* e *Journal of Chemometrics*. Nos períodos seguintes, o número de publicações envolvendo NIR continuaria a crescer incrivelmente (mais de 35 mil artigos entre 2000-2009 e mais de 62 mil entre 2010-2017, de acordo com pesquisa no *Web of Science*). Seguramente, pode-se afirmar que este crescimento só foi possível graças à evolução simultânea da quimiometria. Atualmente, o desenvolvimento de modelos multivariados qualitativos e quantitativos encontra grande aplicação nas mais diversas áreas, tais como na análise de alimentos, produtos farmacêuticos, combustíveis e materiais. Uma das áreas de aplicação de maior relevância envolve a análise de produtos agrícolas, particularmente grãos e sementes.

A quimiometria trata da aplicação de métodos estatísticos multivariados em química e pode ser definida como a disciplina química que usa métodos matemáticos e estatísticos para planejar ou selecionar condições ótimas de medidas e experimentos e extrair o máximo de informações de dados químicos (Bruns; Faigle, 1985; Otto, 1999). Uma definição mais atual, recomendada pela International Union of Pure and Applied Chemistry - IUPAC, apresenta a quimiometria como a ciência que relaciona medidas feitas em um sistema químico ou processo com o estado do sistema através da aplicação de métodos matemáticos ou estatísticos (Hibbert, 2016).

A quimiometria, como disciplina e linha de pesquisa, foi formalizada no começo dos anos 1970, a partir da colaboração entre seus pioneiros, o químico sueco Svante Wold e o matemático e químico analítico estadunidense Bruce Kowalski, os quais fundaram a Sociedade Internacional de Quimiometria, em 1974 (Geladi; Esbensen, 1990). Desde o começo, as aplicações de quimiometria concentraram-se na área de química analítica. Com a evolução dos equipamentos relacionados às modernas técnicas de espectroscopia molecular, os quais são cada vez mais acessíveis a um menor custo, modelos multivariados tornaram-se mais frequentes e importantes na

literatura. Uma das características mais vantajosas dos métodos quimiométricos é a possibilidade de quantificação ou classificação sem a necessidade de resolução do sinal analítico. Desta forma, a separação física ou química dos interferentes pode ser substituída pela separação matemática dos seus sinais. Isto tem contribuído para o desenvolvimento de inúmeros métodos analíticos mais simples, rápidos e relativamente baratos. Esses métodos demandam um mínimo (ou mesmo nenhum) pré-tratamento das amostras, sendo ambientalmente amigáveis por não gerarem resíduos nem consumirem reagentes ou solventes.

Os métodos quimiométricos podem ser divididos em diferentes tipos de acordo com seus objetivos, tais como planejamento e otimização de experimentos, análise exploratória dos dados, classificação supervisionada, calibração multivariada e resolução de curvas. Com o objetivo de cobrir os principais aspectos teóricos relacionados à construção de modelos quimiométricos utilizando espectros NIR, serão apresentadas neste capítulo quatro seções focadas, respectivamente, no pré-processamento dos dados, na análise exploratória de dados, com destaque para a análise de componentes principais, na calibração multivariada e na classificação supervisionada.

Os dados gerados por espectroscopia NIR são considerados de primeira ordem, pois para cada amostra é gerado um vetor (espectro). Na linguagem da álgebra matricial, vetores são tensores de primeira ordem. Como esses vetores/espectros são coletados em uma matriz de dados, modelos bilineares são usados no seu tratamento. Embora modelos não lineares possam também ser utilizados, tais como redes neurais artificiais e máquina de vetores de suporte (SVM, *support vector machines*), eles não serão abordados aqui. O uso de métodos não lineares é reduzido, pois só se justifica quando métodos lineares não são capazes de descrever matematicamente os dados (princípio da parcimônia), o que normalmente está associado à modelagem de faixas analíticas muito extensas.

Os arranjos de dados a serem processados serão organizados em forma de matrizes, de modo que os espectros de cada amostra estarão dispostos em suas linhas, e os valores de absorvância, dispostos em suas colunas. Para a representação formal dos modelos matemáticos, este capítulo adota a notação algébrica padrão. Matrizes são representadas por letras maiúsculas em negrito, enquanto que vetores são representados por letras minúsculas em negrito e escalares por letras minúsculas normais. A letra T maiúscula sobrescrita será usada para indicar matrizes transpostas, e “-1” sobrescrito, para indicar matrizes inversas.

## Pré-processamento dos dados

Na análise multivariada é importante adotar uma sequência de etapas que, dentre outros objetivos, visa a detectar possíveis erros grosseiros e amostras anômalas, além de extrair o máximo de informação dos dados. Invariavelmente, o pré-processamento dos dados é a etapa inicial deste processo. O seu objetivo geral é eliminar ou reduzir a variância aleatória, além de fontes de variação sistemáticas não desejadas. Dessa maneira, a extração das informações poderá focar-se na variância que realmente interessa aos objetivos da análise.

Na análise de dados espectroscópicos, particularmente de espectros NIR, é comum a presença de variação sistemática indesejável, a qual pode mascarar a variância de interesse. Por exemplo, se o objetivo é determinar a composição química de grãos, as variações de densidade de amostra para amostra, ou as diferenças no tamanho das partículas no caso de amostras trituradas, gera variação física que pode prejudicar a habilidade preditiva do modelo. Essa variância deve, portanto, ser eliminada. Mas a escolha do pré-processamento deve ser feita com cautela, baseada no conhecimento sobre os princípios de cada método e sobre a natureza das fontes químicas e físicas de variação nas amostras. O uso de métodos de pré-processamento inadequados pode eliminar informação relevante, prejudicando a qualidade do modelo quimiométrico.

Os métodos de pré-processamento podem ser aplicados tanto nas linhas (amostras) quanto nas colunas (variáveis) da matriz de dados (Ferreira, 2015; Sena et al., 2017). O pré-processamento nas linhas é aplicado em uma amostra de cada vez, considerando todas as variáveis. Reciprocamente, o pré-processamento nas colunas é aplicado a cada variável, considerando todas as amostras. As fontes mais comuns de variância indesejável em espectros NIR são ruídos, espalhamentos e desvios de linha-base (Engel et al., 2013).

Espectros NIR sempre terão certo nível de ruído aleatório, o qual dependerá da amostra, das condições de medida e do detector do espectrofotômetro. Os métodos de alisamento (*smoothing*) são usados para reduzir matematicamente o ruído, aumentando a razão sinal/ruído dos espectros. Assume-se que os ruídos têm alta frequência em relação ao sinal de interesse. O método mais usado para esse fim é o alisamento Savitzky-Golay (Savitzky; Golay, 1964). Este método utiliza um filtro de média móvel, o qual sucessivamente ajusta um polinômio a cada parte/janela do espectro. A escolha do número de variáveis dessa janela é crucial. Um número muito pequeno de variáveis pode não filtrar o ruído de maneira satisfatória, enquanto um número muito grande pode eliminar informação relevante e distorcer o espectro. O usuário deverá, portanto, ajustar o tamanho da janela, que costuma variar entre 7 e 15 pontos, e esco-

lher o grau do polinômio, sendo que os de 2º ou 3º grau são os mais usados (Rinnan et al., 2009).

Outro artefato comum em espectros NIR é a presença de desvios de linha-base, os quais podem ser lineares, causados pela presença de um sinal de fundo (*offset*) que precisa ser corrigido por causa de variações instrumentais ou compensação inadequada do branco, ou não-lineares, causados por efeitos multiplicativos. Desvios de linha-base podem ser corrigidos pelo uso da simples compensação do *offset*, derivadas ou métodos específicos para a correção de espalhamentos multiplicativos. As derivadas são ferramentas matemáticas úteis para melhorar a resolução de sinais analíticos, embora apresentem, como efeito colateral, o aumento do ruído. Por isso, o uso de derivadas deve ser feito conjuntamente com métodos de alisamento (frequentemente, usam-se derivadas de Savitzky-Golay). A primeira derivada elimina desvios de linha-base aditivos, enquanto que a segunda derivada elimina também efeitos multiplicativos.

A eliminação de desvios causados pelo espalhamento de luz é particularmente relevante no processamento quimiométrico de espectros NIR obtidos para sólidos através de medidas de reflectância difusa. O uso de acessórios de reflectância foi muito importante na difusão da espectroscopia NIR, permitindo medidas diretas e aumentando a versatilidade desta técnica. Esse tipo de desvio de linha-base não linear (*drift*) é causado pelas diferenças de caminho ótico oriundas da falta de homogeneidade no tamanho das partículas ou das diferentes densidades dos grãos. Os métodos mais usados para mitigar esta fonte de variação são a correção do espalhamento multiplicativo (MSC, *multiplicative scatter correction*) e a variação normal padrão (SNV, *standard normal variate*), além da normalização vetorial, empregada menos frequentemente (Rinnan et al., 2009). O MSC corrige o espalhamento multiplicativo através de uma regressão em relação a um espectro de referência, normalmente o espectro médio do conjunto de calibração. Dessa forma, interceptos e inclinações são corrigidos, eliminando variação causada pelo espalhamento. O SNV é um tipo de normalização feito ao longo das linhas da matriz. Nele, o valor médio de cada linha é subtraído de todos os valores da respectiva linha, os quais são, em seguida, divididos pelos respectivos desvios-padrão. Portanto, o SNV corresponde ao autoescalamamento da matriz de dados transposta (Dhanoa et al., 1994). Pelo fato de não utilizar um espectro médio de referência, o SNV é considerado mais robusto à presença de amostras anômalas (*outliers*) do que o MSC.

Nos três parágrafos anteriores desta seção, foram descritos métodos de pré-processamento aplicados ao longo das linhas da matriz, cuja utilização é mais focada em dados espectrais. Por outro lado, métodos aplicados nas variáveis possuem um uso

mais geral. Dentre estes, os mais importantes são centrar os dados na média e no autoescalamento (Bro; Smilde, 2003). Centrar na média é o pré-processamento mais utilizado, aplicado a todo o tipo de dados, incluindo espectros de um modo geral. Neste método, a média de cada variável é calculada e subtraída do valor de cada elemento da respectiva coluna. A distribuição de cada variável passa a ter média zero, ou seja, o centro das coordenadas do sistema é movido da origem natural para a média multivariada dos dados (espectro médio). O autoescalamento é mais utilizado em dados obtidos para variáveis discretas (medidas físico-químicas, concentrações de elementos, etc.), em modelos de análise exploratória. Ele inclui a “centragem” na média, mas, após esta etapa, cada elemento é dividido pelo desvio-padrão da respectiva coluna. Como consequência, todas as variáveis terão média zero e variância um, o que dará a elas pesos iguais, independentes de suas escalas naturais. É importante frisar que espectros de um modo geral, e de NIR em particular, não devem ser autoescalados, pois os picos de maior intensidade normalmente contribuem com mais informação para o modelo. Além disso, o autoescalamento de espectros aumenta a contribuição de variáveis ruidosas, em regiões espectrais de pouca absorbância, para o modelo.

Após o pré-processamento, o próximo passo da análise quimiométrica é a construção dos modelos. Nas três seções seguintes, serão apresentados os principais tipos de modelos utilizados para a análise de espectros NIR.

## **Análise de componentes principais (PCA)**

Os métodos de análise exploratória de dados são também conhecidos como métodos de classificação não supervisionada. O termo “não supervisionada” indica que não serão fornecidas ao modelo quaisquer informações sobre a origem das amostras. Os objetivos desse tipo de método são verificar a presença de agrupamentos naturais de amostras, reduzir as dimensões dos dados e extrair os padrões latentes de informações mais relevantes. Por isso, às vezes usa-se também o termo “reconhecimento de padrões” para se referir a esses métodos. Modelos exploratórios extraem informação latente, que pode ser visualizada em gráficos de dispersão simples ou em dendrogramas. Dessa forma, o analista pode identificar mais facilmente as tendências e os padrões subjacentes em grandes conjuntos de dados, os quais não seriam perceptíveis observando-se os valores de uma variável de cada vez. Os mais importantes métodos quimiométricos exploratórios são a análise de componentes principais (PCA, *principal component analysis*) e a análise hierárquica de agrupamentos (HCA, *hierarchical cluster analysis*).

A HCA (Otto, 1999; Ferreira, 2015) busca observar agrupamentos de amostras (*clusters*) baseando-se em medidas de similaridade, normalmente na distância euclidiana, sem envolver uma etapa de redução da dimensionalidade dos dados. Como resultados, são obtidos dendrogramas que permitem analisar a formação de agrupamentos em função da distância entre as amostras no espaço multidimensional da variância dos dados. Por não envolver uma etapa de projeção dos dados e não permitir observar simultaneamente as relações entre amostras e variáveis, a HCA possui menor capacidade de extrair informações do sistema do que a PCA (Sena et al., 2017) e, por isso, não será discutida em profundidade aqui.

A PCA pode ser considerada o mais importante método quimiométrico. Além de ter ampla aplicação em diversas áreas da ciência, a sua compreensão teórica é um pré-requisito para o entendimento da maioria dos métodos quimiométricos de classificação e calibração multivariada. Historicamente, pode-se afirmar que a PCA foi reinventada por diversas vezes (Bro; Smilde, 2014), datando do início do século XX a sua formulação original (Pearson, 1901).

A PCA é um método que permite a redução da dimensionalidade de um conjunto de dados através de sua projeção em um subespaço vetorial, um novo sistema de eixos. Os novos eixos são estimados através de combinações lineares das variáveis originais e são denominados componentes principais (CPs). Os CPs condensam a informação dos dados em um número bem menor de fatores, que são mais facilmente visualizados. Por exemplo, espectros NIR típicos têm centenas de variáveis, que contêm muita variância redundante. Na verdade, um número muito menor de fatores, os componentes químicos das amostras, é responsável pela variação sistemática de interesse. Nesses casos, a PCA reduz centenas de variáveis a apenas uns poucos CPs, que retêm a maior parte da variância. A PCA pode ainda ser vista como um método que “filtra” as correlações presentes nos dados, o que é essencial na análise de espectros, que em geral contêm muitas variáveis correlacionadas. Na literatura, podem ser encontrados bons artigos tutoriais recentes descrevendo a aplicação da PCA tanto a dados discretos (variáveis físico-químicas) (Bro; Smilde, 2014) quanto contínuos (espectros) (Souza; Poppi, 2012).

Uma matriz de dados (espectros)  $\mathbf{X}_{(n,m)}$ , formada por  $n$  linhas (amostras) e  $m$  colunas (comprimentos de onda), é decomposta em um modelo PCA como o produto de uma matriz de escores  $\mathbf{T}_{(n,A)}$  e uma matriz de pesos (loadings)  $\mathbf{P}_{(m,A)}$ , de acordo com a seguinte equação:

$$\mathbf{X} = \mathbf{t}_1 \mathbf{p}_1^T + \mathbf{t}_2 \mathbf{p}_2^T + \dots + \mathbf{t}_A \mathbf{p}_A^T + \mathbf{E} = \mathbf{TP}^T + \mathbf{E} \quad (1)$$

sendo  $A$  o número de CPs no modelo PCA,  $\mathbf{t}_1$  e  $\mathbf{p}_1$  os vetores de escores e pesos do primeiro CP, respectivamente, e  $\mathbf{E}$  a matriz de resíduos contendo a variância não descrita pelos CPs incluídos no modelo. Desta forma, a estimativa da matriz de dados experimentais pelo modelo PCA é dada por  $\hat{\mathbf{X}} = \mathbf{TP}^T$ . Deve-se ressaltar que os CPs são definidos de modo a explicar a máxima variância dos dados. O vetor  $\mathbf{p}_1$  descreve a direção de maior variação no espaço multidimensional. Se os dados não estiverem centrados na média,  $\mathbf{p}_1$  descreverá o espectro médio das amostras. Na sequência, o vetor de pesos  $\mathbf{p}_2$  descreve a direção de maior variância restante nos dados, mas submetido à restrição de ser ortogonal ao primeiro CP, e assim por diante. Por isto, os CPs são sempre modelados em ordem decrescente da variância explicada. A restrição de ortogonalidade entre os CPs indica que eles são independentes, ou seja, totalmente não correlacionados.

Os escores e pesos são os mais importantes parâmetros de modelos PCA. Algebricamente, os elementos de cada vetor de pesos ( $\mathbf{p}$ ) representam os cossenos dos ângulos entre o respectivo CP e os eixos definidos pelas variáveis originais. Ou seja, os elementos de  $\mathbf{p}$  mostram o quanto cada variável original contribui para o respectivo CP. Por outro lado, os escores de cada vetor  $\mathbf{t}$  são as projeções ortogonais das amostras no eixo de cada CP. Para a interpretação de modelos PCA, é essencial entender que os escores fornecem a composição de cada CP relacionada aos objetos/amostras, enquanto os pesos relacionam essa mesma composição com as variáveis mais importantes (Sena et al., 2002). Após a escolha do número de CPs a ser incluído no modelo, os dados originais são projetados no novo espaço de dimensões reduzidos. As inter-relações e os possíveis agrupamentos de amostras podem ser observados através de gráficos de dispersão bi ou tridimensionais, nos quais os escores dos principais CPs são plotados uns contra os outros. Esta é a principal ferramenta de interpretação de modelos PCA. As projeções ortogonais dos escores em cada eixo separadamente indicam contrastes, que discriminam amostras mais negativas de amostras mais positivas. Amostras com escores próximos de zero em um CP indicam sua falta de importância na interpretação desta componente. É importante ressaltar que bons modelos PCA nem sempre mostram a formação de agrupamentos de amostras claramente distintos. Se o fenômeno que está sendo modelado dá origem, por exemplo, a amostras de composição contínua entre dois extremos, não é esperada a observação de agrupamentos nitidamente separados.

A observação dos escores deve sempre ser complementada pela observação simultânea dos pesos, buscando relacionar quais variáveis contribuem mais para caracterizar o contraste de amostras observado em cada CP. O analista deve sempre interpretar o significado de cada um dos CPs em termos químicos, biológicos, etc. Ele obrigatoriamente deve ser capaz de responder à pergunta “O que significa CP1?”.

A resposta pode ser, por exemplo, um contraste entre sementes de duas origens diferentes, ou entre grãos saudios e infestados por pragas. A observação conjunta de gráficos de escores e de pesos de CP1 x CP2 pode ser condensada em gráficos bivariados (*biplots*). Os *biplots* mostram escores e pesos, após normalização específica, no mesmo gráfico (Gabriel, 1971) e estão disponíveis como recurso visual em todos os softwares quimiométricos. No entanto, *biplots* e gráficos de dispersão de pesos de CP1 x CP2 são úteis apenas na análise de dados discretos, constituídos de teores de elementos e/ou variáveis físico-químicas. Quando os dados são contínuos, como no caso de espectros, os pesos devem ser observados individualmente em função das variáveis, tais como comprimento ou número de onda (Souza; Poppi, 2012). Nesse caso, *biplots* não terão utilidade.

Uma das etapas mais importantes na PCA é a escolha do número de CPs. Idealmente, o número de CPs significativos é igual ao número de fontes de variação linearmente independentes presentes nos dados. Por exemplo, na análise de espectros de misturas de compostos, esse número deve ser igual ao número de componentes químicos que contribuem de maneira não correlacionada para o sinal espectral. O número de CPs de um modelo depende da natureza dos dados em particular. Um dos modos mais intuitivos utilizados nessa estimativa é baseado na quantidade de variância capturada pelo modelo PCA ( $V_{exp}$ ). Essa quantidade pode ser estimada pela razão entre as somas dos quadrados dos elementos de  $\hat{\mathbf{X}}$  e de  $\mathbf{X}$ , expressa pela equação:

$$V_{exp} = \frac{\sum_{i=1}^n \sum_{j=1}^m (\hat{x}_{i,j})^2}{\sum_{i=1}^n \sum_{j=1}^m (x_{i,j})^2} 100 \quad (2)$$

na qual  $x_{i,j}$  e  $\hat{x}_{i,j}$  são os valores experimentais e estimados obtidos para a amostra  $i$  medida na variável  $j$  em  $\mathbf{X}$  e  $\hat{\mathbf{X}}$ , respectivamente. Com base nos valores de  $V_{exp}$  obtidos de 1 a  $A$  CPs, a quantidade de variância explicada para cada CP é calculada, em conjunto com a variância acumulada pelo modelo PCA. Na modelagem de espectros em geral e NIR em particular, os primeiros CPs costumam explicar a maior parte da variação nos dados. Então, uma regra útil sugere a rejeição do CP cuja variância for menor que uma porcentagem predefinida (por exemplo, 5%). Em muitos casos, apenas os primeiros dois ou três CPs são suficientes para modelar o fenômeno de interesse. Em outras situações, o interesse pode estar em alguma informação que representa uma pequena parte da variação de dados (por exemplo, 2%). Nestes casos, uma regra mais conservadora é inspecionar os vetores peso e rejeitar os CPs que descrevam apenas ruído, ou seja, que apresentem comportamento aleatório.



Um último aspecto importante a ser ressaltado é a detecção de amostras anômalas em modelos PCA. Esta detecção é feita através da observação de gráficos que mostram dois tipos de parâmetros, os resíduos  $Q$  e os valores de  $T^2$  de Hotelling. Estes parâmetros estão relacionados, respectivamente, à variância residual não modelada e à influência da variância incluída no modelo para cada amostra (Sena et al., 2014). Amostras que possuam simultaneamente altos valores para esses dois parâmetros são candidatas a serem detectadas como outliers.

## Calibração multivariada

Em um processo de calibração, a concentração de um componente ou outra propriedade de interesse é quantificada indiretamente através de medidas realizadas em um sistema químico. Quando existir uma variável seletiva que possa ser medida para esse sistema, a calibração univariada é usada. No entanto, espectros NIR raramente apresentam variáveis seletivas, mesmo para matrizes menos complexas (Pasquini, 2003). Nesse caso, a alternativa é o uso da calibração multivariada (Martens; Naes, 1996; Brereton, 2000), na qual muitas variáveis, centenas ou até milhares no caso de espectros de infravermelho em geral, são usadas para construir o modelo. Como na calibração multivariada predomina a calibração inversa, define-se um bloco de variáveis independentes (espectros), contidas na matriz  $\mathbf{X}$ , e uma (ou mais) variável dependente a ser prevista, contida no vetor  $\mathbf{y}$ . Assim, o modelo é definido matematicamente como  $\mathbf{y} = \mathbf{X}\mathbf{b}$ , sendo  $\mathbf{b}$  o vetor dos coeficientes de regressão, um para cada variável (comprimento de onda) em  $\mathbf{X}$ .

Dentre as vantagens da calibração multivariada, pode-se mencionar a possibilidade de determinações diretas, na ausência de resolução do sinal analítico, na presença de interferentes desde que presentes no conjunto de calibração, e também a possibilidade de quantificações simultâneas a partir do mesmo conjunto de espectros, quando houver mais de um analito ou propriedade a ser analisado. Pelo menos dois conjuntos de amostras são necessários para o desenvolvimento do modelo. O primeiro é o conjunto de calibração, para o qual os valores das concentrações ou propriedades a ser determinadas são conhecidos, seja pela preparação de amostras de composição planejada ou pelo uso de métodos de referência. Estes valores em conjunto com os espectros são utilizados para estimar os coeficientes de regressão. O outro conjunto de dados é o de validação, no qual o modelo desenvolvido é testado para prever a concentração de amostras independentes, não incluídas na calibração. As amostras costumam ser divididas em cerca de dois terços para o conjunto de calibração e um terço para o conjunto de validação. As amostras de calibração devem ser represen-

tativas de toda a variância sistemática a ser modelada e estar distribuídas homogeneamente. O número de amostras de validação deve ser suficiente para testar uma generalização representativa das previsões. Quando não for possível obter amostras de composição planejada, situação usual na análise de grãos e sementes, deve-se usar algum método para selecionar as amostras de calibração mais representativas a partir do conjunto total disponível inicialmente. Os métodos mais usados para isto são os algoritmos de Kennard-Stone e Duplex (Westad; Marini, 2015).

Enquanto que a regressão clássica ou direta (na qual a concentração é a variável independente/bloco X e a absorvância espectral a variável dependente/bloco Y) predomina na calibração univariada, o seu uso é muito restrito na calibração multivariada. O método dos mínimos quadrados clássicos (CLS, *classical least squares*) tem como limitação a exigência de que as concentrações de todas as espécies absorventes no sistema sejam previamente conhecidas. Obviamente, esta condição não está presente na imensa maioria das situações que envolvem análises espectrais de produtos agrícolas, tornando impossível o uso do CLS, o qual não fornece previsões aceitáveis na presença de interferentes. Por isso, o uso de calibração multivariada inversa é largamente predominante. As diferentes maneiras de estimar o vetor de regressão (**b**) dão origem aos principais métodos de calibração multivariada: regressão linear múltipla (RLM), regressão em componentes principais (PCR, *principal component regression*) e mínimos quadrados parciais (PLS, *partial least squares*).

A maneira mais simples e direta de estimar **b** na equação  $\mathbf{y} = \mathbf{Xb}$  é usando a RLM. A solução dos mínimos quadrados para esta equação é  $\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ . No entanto, a matriz  $\mathbf{X}^T\mathbf{X}$  só será inversível na ausência de colinearidade entre as variáveis independentes e se existirem mais linhas do que colunas na matriz **X**, ou seja, mais amostras que variáveis. Mesmo que o número de amostras disponível para construir o modelo seja maior do que o número de comprimentos de onda modelados, espectros NIR sempre apresentam alto grau de colinearidade. Como resultado, a solução dessa matriz inversa representará um sistema instável (mal condicionado). De outra maneira, pode-se dizer que a RLM é limitada por incluir toda a variância espectral no modelo. Com isso, quantidade significativa de informação não relevante será incorporada, tornando o modelo pouco robusto. Por isso, a RLM só funcionará satisfatoriamente para sistemas simples e bem-comportados, que apresentem respostas lineares, sem interferentes, com baixo ruído e nenhuma colinearidade. Uma alternativa é combinar a RLM com uma prévia seleção de variáveis, que encontre um pequeno número de comprimentos de onda não correlacionados.

Porém, outra alternativa de uso mais amplo foi imaginada. O que acontece se a RLM for combinada com a PCA? E se, ao invés de a regressão ser feita nos dados

espectrais brutos, ela fosse feita nos escores ( $\mathbf{T}$ ) de um modelo PCA? O resultado é exatamente a PCR. A estimativa de  $\mathbf{b}$  se torna igual a  $(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{y}$ , que não representa mais um sistema instável, pois os escores em  $\mathbf{T}$  não são correlacionados e estão em número bem mais reduzido. A variância redundante é descartada nos CPs não aproveitadas no modelo.

A decomposição dos dados espectrais num modelo PCR é feita de maneira idêntica ao modelo PCA, ignorando nesta etapa a informação contida nas variáveis dependentes. Isto deu origem a críticas à PCR e culminou na formulação de um novo modelo que simultaneamente decompõe ao blocos X e Y (espectros e concentrações), maximizando a correlação entre eles. Esse modelo é o PLS, proposto inicialmente pelo econométrista Herman Wold nos anos 1960 (Sanchez, 2015) e introduzido na literatura química na década de 1980 (Geladi; Kowalski, 1986). Nos anos seguintes, o PLS tornou-se o mais importante método de calibração multivariada. Assim como o PCR, o PLS é capaz de quantificar analitos na presença de interferentes, desde que eles tenham sido incluídos no conjunto de calibração. Entretanto, os fatores/componentes decompostos no PLS estão sob a restrição de explicar simultaneamente a variância dos espectros e das concentrações, o que provoca perda de ortogonalidade. Por isto, os fatores estimados em um modelo PLS não são chamados CPs como na PCR, mas variáveis latentes (VLs), um termo mais geral. Como consequência e de maneira diferente do PCR, as VLs em um modelo PLS não são modeladas necessariamente em ordem decrescente da variância explicada.

Uma das etapas mais importantes na construção de um modelo PLS/PCR é a escolha do número de VLs/CPs. Ela é feita através da validação cruzada, na qual uma amostra (ou grupo) do conjunto de calibração é retirada de cada vez e prevista pelo modelo construído com as demais. Este processo é repetido até que todas as amostras tenham sido previstas. A escolha do número de VLs corresponde ao menor erro de validação cruzada (RMSECV, *root mean square error of cross validation*) ou ao último decréscimo significativo desse valor. O RMSECV é calculado como:

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n_c}} \quad (3)$$

sendo  $n_c$  o número de amostras no conjunto de calibração,  $\hat{y}_i$  o valor previsto pelo modelo e  $y_i$  o valor de referência da  $i$ -ésima amostra. A aplicação do modelo construído às amostras de calibração e validação origina outros dois parâmetros importantes, RMSEC (*RMSE of calibration*) e RMSEP (*RMSE of prediction*). Existem vários tipos de validação cruzada, dependendo dos critérios utilizados na reamostragem. Um dos mais usados é o *leave-one-out* (LOO), no qual uma amostra é removida por vez. No entanto, o LOO é recomendado apenas para pequenos conjuntos de calibração, com

não muito mais do que 20 amostras. Quando o conjunto de calibração for maior, as alternativas são baseadas na remoção de grupos de amostras, tais como blocos contíguos, subconjuntos aleatórios ou venezianas (*venetian blinds*).

A determinação do número correto de VLs é fundamental. Se esse número for menor do que um valor ideal, variância relevante fica fora do modelo, que se torna subajustado e produz resultados inexatos. Por outro lado, se um número maior do que o ideal de VLs é usado, variância redundante é incluída no modelo, que se torna sobreajustado (Faber; Rajkó, 2007). Com isso, este modelo produzirá previsões exatas apenas para o conjunto de dados usado na sua própria construção. Muitas práticas inadequadas, incluindo o uso de softwares como “caixas pretas”, geram modelos com sobreajuste, tornando este tipo de erro um dos mais comuns na literatura quimiométrica. Um sinal típico de sobreajuste ocorre quando o RMSEP é muito superior ao RMSEC. Todos os modelos construídos devem ser rigorosamente validados. Práticas como a escolha tendenciosa das amostras de calibração, ou o aumento do conjunto de dados à custa do uso de replicatas como se fossem amostras verdadeiras, devem ser evitadas.

Um bom modelo de calibração multivariada deve explicar a maior parte da variância, geralmente mais de 90%, nos dois blocos, X e Y. A validação externa deve ser realizada através da previsão do conjunto de validação independente e da estimativa do RMSEP. Outro aspecto importante é a interpretação espectral dos modelos construídos, muitas vezes ausente em artigos publicados na literatura. Para isto, vetores informativos gerados no modelo PLS devem ser interpretados criticamente e associados às respectivas bandas espectrais. Os vetores informativos mais importantes são o vetor dos coeficientes de regressão e os VIP (*variable importance in projection*) escores.

Finalmente, dois últimos aspectos importantes devem ser mencionados. A detecção de *outliers* em calibração multivariada (Martens; Naes, 1996; Valderrama et al., 2007) é mais importante do que em análise exploratória, pois a presença deles prejudica a habilidade preditiva do modelo. Além da detecção de amostras com altos resíduos espectrais (resíduos Q) e alta influência (alto *leverage*), deve-se testar também a presença de *outliers* no bloco Y (valores previstos). Após a construção do modelo de calibração multivariada, é necessário ainda checar se o método possui um desempenho adequado para sua aplicação específica. Este processo de validação analítica é realizado através da estimativa de figuras de mérito, tais como linearidade, exatidão (veracidade + precisão), sensibilidade analítica, viés (*bias*) e relação de desempenho do desvio (RPD, *residual prediction deviation*). Atualmente, a grande maioria das normas regulatórias oficiais é concebida de acordo com um raciocínio univariado, o que gera a necessidade de harmonização para métodos multivariados. Nesse sentido, foi importante a criação de uma norma pioneira voltada para calibração multivariada

usando dados de infravermelho, no ano de 2000 (ASTM, 2012). Nos últimos anos, esse assunto tem sido bastante discutido na literatura (Valderrama et al., 2009; Botelho et al., 2013; Sena et al., 2017).

## Classificação supervisionada

O desenvolvimento de métodos qualitativos, de classificação supervisionada, na literatura quimiométrica, cresceu muito nos últimos anos, levando alguns autores a qualificar esse crescimento como “vibrante” (Szymanska et al., 2015). O objetivo desses métodos é atribuir amostras desconhecidas a classes previamente definidas. Conforme já mencionado, nos métodos não supervisionados as amostras podem ser atribuídas a grupos pela observação de gráficos de dispersão de escores. Porém, nesses métodos, informações sobre as classes não são fornecidas ao modelo. Por isso, a classificação supervisionada é mais objetiva, proporcionando uma maneira sistemática de classificar novas amostras. No entanto, recomenda-se sempre a construção de modelos PCA previamente aos modelos de classificação, pois eles fornecem um bom quadro dos padrões e diferenças naturalmente presentes nas amostras. De outra maneira, pode-se definir o objetivo da classificação supervisionada como a delimitação de regiões específicas do hiperespaço das variáveis associadas a cada classe, sendo que a definição de classe dependerá de cada problema em particular (Marini, 2010; Sena et al., 2017).

A combinação de classificação supervisionada com técnicas espectroscópicas é bastante apropriada para desenvolver métodos de triagem capazes de fornecer ferramentas rápidas para a solução de problemas que envolvam comprovação de autenticidade, atribuição de origem, detecção de contaminações, fraudes e falsificações, etc. Particularmente, publicações recentes têm combinado espectroscopia NIR e métodos de classificação na análise de grãos e sementes, em problemas que envolvem atribuição de origem geográfica ou genotípica, autenticidade e rastreabilidade de cereais, etc. (Vitale et al., 2013; Cozzolino, 2014; Marquetti et al., 2016).

As estratégias de classificação e a escolha do método mais apropriado podem variar bastante, dependendo de cada situação em particular. Em muitas situações, apenas duas classes são definidas, quando existem somente duas classificações possíveis. Exemplos disso são métodos que buscam detectar a ausência ou presença de um contaminante, se uma molécula é ativa ou não, se um produto está ou não adulterado, etc. Por outro lado, quando modelos para discriminar mais de duas classes são construídos, pode ou não fazer sentido que uma amostra seja atribuída a mais de uma

classe simultaneamente, ou pode-se permitir que o modelo rejeite a amostra como não pertencente a nenhuma das classes e seja classificada como um *outlier*. Assim como na calibração multivariada, os dados devem ser divididos em dois conjuntos e a validação cruzada deve ser usada. Em classificação supervisionada, os conjuntos usados para construir e validar o modelo são mais comumente denominados conjuntos de treinamento e teste, respectivamente. A divisão das amostras pode ser feita usando os já citados algoritmos de Kennard-Stone ou Duplex, mas em modelos de classificação é mais apropriado que eles sejam aplicados às amostras de cada classe em separado. Da mesma forma que muitos modelos de calibração sobre-ajustados são publicados em artigos, modelos de classificação enviesadamente otimistas também são encontrados. Dentre as principais práticas incorretas que devem ser evitadas, podem-se citar ausência de validação apropriada, uso de um número pequeno de amostras não representativas da variância que deveria ser associada a determinada classe, e o uso de repetições como amostras independentes, aumentando artificialmente o tamanho de uma ou mais classes.

Métodos de classificação podem ser subdivididos de várias maneiras, como lineares ou não lineares, paramétricos ou não paramétricos, etc. De particular relevância é a subdivisão desses métodos em discriminantes ou de modelagem de classe. Os primeiros definem um ou mais delimitadores entre duas ou mais classes, dividindo o hiperespaço das variáveis em um número correspondente de regiões. Os do segundo tipo modelam cada classe individualmente, sem levar em conta informações relativas às demais classes. Recentemente, alguns autores têm criticado o predomínio de publicações envolvendo métodos discriminantes quando aplicados a problemas de autenticação, pois frequentemente não é possível obter um conjunto de amostras representativo de todas as possíveis situações de não autenticidade (Rodionova et al., 2016).

O mais representativo método quimiométrico de modelagem de classes é o SIMCA (*soft independent modeling of class analogy*), que pode ser traduzido como “modelagem independente por analogia de classes”, desenvolvido por Wold nos anos 1970 (Wold, 1976). O SIMCA consiste numa coleção de modelos PCA, construídos individualmente para cada classe. Os números de CPs são selecionados por validação cruzada e o SIMCA pode ser usado tanto em modelagem multiclases quanto na modelagem de apenas uma classe, situação indicada em problemas que envolvem a comprovação de autenticidade ou a detecção de fraudes.

Atualmente, o modelo discriminante mais comum na literatura quimiométrica é o PLS-DA (*PLS discriminant analysis*), uma extensão do PLS para classificação (Brereton; Lloyd, 2014). Com tal, o PLS-DA partilha das propriedades do PLS, sendo que o bloco

Y é composto por variáveis categóricas, que assumem os valores 1 ou 0 dependendo se a amostra pertence ou não à respectiva classe. Outra diferença em relação ao PLS está na validação cruzada, na qual a escolha do número de VLs é feita não em função de um parâmetro quantitativo (RMSECV), mas do menor número de amostras incorretamente classificadas, um parâmetro qualitativo. As críticas ao frequente uso do PLS-DA em problemas de autenticação deram origem ao desenvolvimento recente de novos métodos de modelagem de classe, baseados no PLS ou no SIMCA, tais como PLS-DM (*PLS density modeling*) (Oliveri et al., 2014), OC-PLS (*one-class PLS*) (Xu et al., 2013) e DD-SIMCA (*data-driven SIMCA*) (Pomerantsev; Rodionova, 2014). Finalmente, um último tópico que merece destaque é a validação analítica de métodos qualitativos (Isabel López et al., 2015), que se baseia na estimativa de figuras de mérito específicas para a análise qualitativa, tais como a sensibilidade (taxa de verdadeiro positivo) e especificidade ou seletividade (taxa de verdadeiro negativo).

## Conclusão e perspectivas

Este capítulo descreveu os mais importantes tipos de métodos quimiométricos que podem ser combinados com espectros NIR para gerar modelos úteis na extração de informações de grandes conjuntos de dados. A difusão de métodos analíticos baseados em espectroscopia NIR propicia análises cada vez mais simples, rápidas e ambientalmente amigáveis. Essa difusão é constatada através do grande aumento de publicações científicas. Mas existe um potencial também para o crescimento do uso prático de tecnologias NIR em laboratórios de controle de qualidade de indústrias e órgãos reguladores. Espera-se que, no futuro, este tipo de tecnologia se torne mais acessível em toda a cadeia produtiva, do produtor ao consumidor. No setor agrícola, em particular, duas novas tendências são importantes. Primeiro, o desenvolvimento de espectrofotômetros NIR portáteis, cada vez menores, que permitem a obtenção de medidas no campo, tornando as análises mais simples e versáteis (Santos et al., 2013). O uso de equipamentos portáteis aumenta a demanda pelo desenvolvimento de modelos de transferência de calibração (Honorato et al., 2007), os quais permitem generalizar o uso de um método quantitativo também para espectrofotômetros de bancada, que ainda costumam fornecer resultados mais precisos. Outra tendência importante é o uso de imagens hiperespectrais, as quais permitem, por exemplo, obter mapas de concentração para analitos distribuídos ao longo da superfície de amostras heterogêneas (Burger; Gowen, 2011).

## Referências

- ASTM. E1655-05: standards practices for infrared multivariate quantitative analysis. West Conshohocken, 2012. Annual Book of ASTM Standards.
- BOTELHO, B. G.; MENDES, B. A. P.; SENA, M. M. Implementação de um método robusto para o controle fiscal de umidade em queijo Minas artesanal: abordagem metrológica multivariada. **Química Nova**, v. 36, n. 9, p. 1416-1422, 2013.
- BRERETON, R. G. Introduction to multivariate calibration in analytical chemistry. **Analyst**, v. 125, n. 11, p. 2125-2154, 2000.
- BRERETON, R. G.; LLOYD, G. R. Partial least squares discriminant analysis: taking the magic way. **Journal of Chemometrics**, v. 28, n. 4, p. 213-225, 2014.
- BRO, R.; SMILDE, A. K. Centering and scaling in component analysis. **Journal of Chemometrics**, v. 17, n. 1, p.16-33, 2003.
- BRO, R.; SMILDE, A. K. Principal component analysis. **Analytical Methods**, v. 6, n. 9, p. 2812-2831, 2014.
- BRUNS, R. E.; FAIGLE, J. F. G. Quimiometria. **Química Nova**, v. 8, n. 2, p. 84-99, 1985.
- BURGER, J.; GOWEN, A. Data handling in hyperspectral image analysis. **Chemometrics and Intelligent Laboratory Systems**, v. 108, n. 1, p. 13-22, 2011.
- COZZOLINO, D. An overview of the use of infrared spectroscopy and chemometrics in authenticity and traceability of cereals. **Food Research International**, v. 60, p. 262-265, June 2014.
- DAVIES, T. The history of near infrared spectroscopic analysis: past, present and future – “From sleeping technique to the morning star of spectroscopy”. **Anulysis**, v. 26, n. 4, p. M17-M19, 1998.
- DHANOVA, M. S.; LISTER, S. J.; SANDERSON, R.; BARNES, R. J. The link between Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) transformations of NIR spectra. **Journal of Near Infrared Spectroscopy**, v. 2, n. 1, p. 43-47, 1994.
- ENGEL, J.; GERRETZEN, J.; SZYMANSKA, E.; JANSEN, J. J.; DOWNEY, G.; BLANCHET, L.; BUYDENS, L. M. C. Breaking with trends in pre-processing? **TrAC – Trends in Analytical Chemistry**, v. 50, p. 96-106, Oct. 2013.



FABER, N. M.; RAJKÓ, R. How to avoid over-fitting in multivariate calibration – the conventional validation approach and an alternative. **Analytica Chimica Acta**, v. 595, n. 1/2, p. 98-106, 2007.

FERREIRA, M. M. C. **Quimiometria: conceitos, métodos e aplicações**. Campinas: Ed. UNICAMP, 2015. 833 p.

GABRIEL, K. R. The biplot display of matrices with application to principal component analysis. **Biometrika**, v. 58, n. 3, p. 453-467, 1971.

GELADI, P.; KOWALSKI, B. R. Partial least squares regression: a tutorial. **Analytica Chimica Acta**, v. 185, p. 1-17, 1986.

GELADI, P.; ESBENSEN, K. The start and early history of chemometrics: selected interviews. Part 1. **Journal of Chemometrics**, v. 4, n. 5, p. 337-354, 1990.

HIBBERT, D. B. Vocabulary of concepts and terms in chemometrics (IUPAC Recommendations 2016). **Pure and Applied Chemistry**, v. 88, n. 4, p. 407-443, 2016.

HONORATO, F. A.; BARROS NETO, B.; MARTINS, M. N.; GALVÃO, R. K. H.; PIMENTEL, M. F. Transferência de calibração em métodos multivariados, **Química Nova**, v. 30, n. 5, p. 1301-1312, 2007.

ISABEL LÓPEZ, M.; PILAR CALLAO, M.; RUISÁNCHEZ, I. A tutorial on the validation of qualitative methods: from the univariate to the multivariate approach. **Analytica Chimica Acta**, v. 891, p. 62-72, Sep. 2015.

MARINI, F. Classification methods in chemometrics. **Current Analytical Chemistry**, v. 6, n. 1, p. 72–79, 2010.

MARQUETTI, I.; LINK, J. D.; LEMES, A. L. G.; SCHOLZ, M. B. S.; VALDERRAMA, P.; BONA, E. Partial least square with discriminant analysis and near infrared spectroscopy for evaluation of geographic and genotypic origin of arabica coffee. **Computers and Electronics in Agriculture**, v. 121, p. 313-319, Feb. 2016.

MARTENS, H.; NAES, T. **Multivariate calibration**. Chichester: Wiley, 1989. 419 p.

MCCLURE, W. F. Near-infrared spectroscopy. The giant is running strong. **Analytical Chemistry**, v. 66, n. 1, p. 43A-53A, 1994.

NORRIS, K. H.; HART, J. H. Direct spectrophotometric determination of moisture content of grain and seeds. **Journal of Near Infrared Spectroscopy**, v. 4, n. 1, p. 23-30, 1996.

OLIVERI, P.; ISABEL LÓPEZ, M.; CASOLINO, M. C.; RUISÁNCHEZ, I.; PILAR CALLAO, M.; MEDINI, L.; LANTERI, S. Partial least squares density modeling (PLS-DM) - A new class-modeling strategy applied to the authentication of olives in brine by near-infrared spectroscopy. **Analytica Chimica Acta**, v. 851, p. 30-36, Dec. 2014.

OTTO, M. **Chemometrics**: statistics and computer application in analytical chemistry. Weinheim: Wiley-VCH, 1999. 343 p.

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.

PEARSON, K. On lines and planes of closest fit to systems of points in space. **Philosophical Magazine**, v. 2, n. 11, p. 559-572, 1901.

POMERANTSEV, A. L.; RODIONOVA, O. Y. On the type II error in SIMCA method. **Journal of Chemometrics**, v. 28, n. 6, p. 518-522, 2014.

RINNAN, A.; VAN DEN BERG, F.; ENGELSEN, S. B. Review of the most common pre-processing techniques for near-infrared spectra. **TrAC - Trends Analytical Chemistry**, v. 28, p. 1201-22, Nov. 2009

RODIONOVA, O. Y.; OLIVERI, P.; POMERANTSEV, A. L. Rigorous and compliant approaches to one-class classification. **Chemometrics and Intelligent Laboratory Systems**, v. 159, p. 89-96, Dec. 2016.

SANCHEZ, G. **The saga of PLS**. 2015. Disponível em: <<http://sagaofpls.github.io>>. Acesso em: 9 fev. 2018.

SANTOS, C. A. T.; LOPO, M.; PÁSCOA, R. N. M. J.; LOPES, J. A. A review on the applications of portable near-infrared spectrometers in the agro-food industry. **Applied Spectroscopy**, v. 67, n. 11, p. 1215-1233, 2013.

SAVITZKY, A.; GOLAY, M. Smoothing and differentiation of data by simplified least squares procedures. **Analytical Chemistry**, v. 36, n. 8, p. 1627-1639, 1964.

SENA, M. M.; FRIGHETTO, R. T. S.; VALARINI, P. J.; TOKESHI, H.; POPPI, R. J. Discrimination of management effects on soil parameters by using principal component analysis: a multivariate analysis case study. **Soil & Tillage Research**, v. 67, n. 2, p. 171-181, 2002.

SENA, M. M.; ALMEIDA, M. R.; BRAGA, J. W. B.; POPPI, R. J. Multivariate statistical analysis and chemometrics. In: FRANCA, A. S.; NOLLET, L. M. L. (Org.). **Spectroscopic methods in food analysis**. Boca Raton: CRC Press, 2017. p. 273-314.

SOUZA, A. M.; POPPI, R. J. Experimento didático de quimiometria para análise exploratória de óleos vegetais comestíveis por espectroscopia no infravermelho médio e análise de componentes principais: um tutorial, parte I. **Química Nova**, v. 35, n. 1, p. 223-229, 2012.

SZYMANSKA, E.; GERRETAEN, J.; ENGEL, J.; GEURTS, B.; BLANCHET, L.; BUYDENS, L. M. C. Chemometrics and qualitative analysis have a vibrant relationship. **TrAC – Trends in Analytical Chemistry**, v. 69, p. 34-51, June 2015.

VALDERRAMA, P.; BRAGA, J. W. B.; POPPI, R. J. Variable selection, outlier detection, and figures of merit estimation in a partial least-squares regression multivariate calibration model. A case study for the determination of quality parameters in the alcohol industry by near-infrared spectroscopy. **Journal of Agricultural and Food Chemistry**, v. 55, n. 21, p. 8331-8338, 2007.

VALDERRAMA, P.; BRAGA, J. W. B.; POPPI, R. J. Estado da arte de figuras de mérito em calibração multivariada. **Química Nova**, v. 32, n. 5, p. 1278-1287, 2009.

VITALE, R.; BEVILACQUA, M.; BUCCI, R.; MAGRÌ, A. D.; MARINI, F. A rapid and non-invasive method for authenticating the origin of pistachio samples by NIR spectroscopy and chemometric. **Chemometrics and Intelligent Laboratory Systems**, v. 121, p. 90-99, Feb. 2013.

WESTAD, F.; MARINI, F. Validation of chemometric models – a tutorial. **Analytica Chimica Acta**, v. 893, p. 14-24, Set. 2015.

WETZEL, D. L. Near-infrared reflectance analysis - sleeper among spectroscopic techniques. **Analytical Chemistry**, v. 55, n. 12, p. 1165A-1176A, 1983.

WOLD, S. Pattern recognition by means of disjoint principal component analysis. **Pattern Recognition**, v. 8, n. 3, p. 127-139, 1976.

XU, L.; YAN, S. M.; CAI, C. B.; YU, X. P. One-class partial least squares (OCPLS) classifier. **Chemometrics and Intelligent Laboratory Systems**, v. 126, p. 1-5, 2013.

# 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<sup>-1</sup> a 10.000 cm<sup>-1</sup>, com resolução de 4 cm<sup>-1</sup> 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.

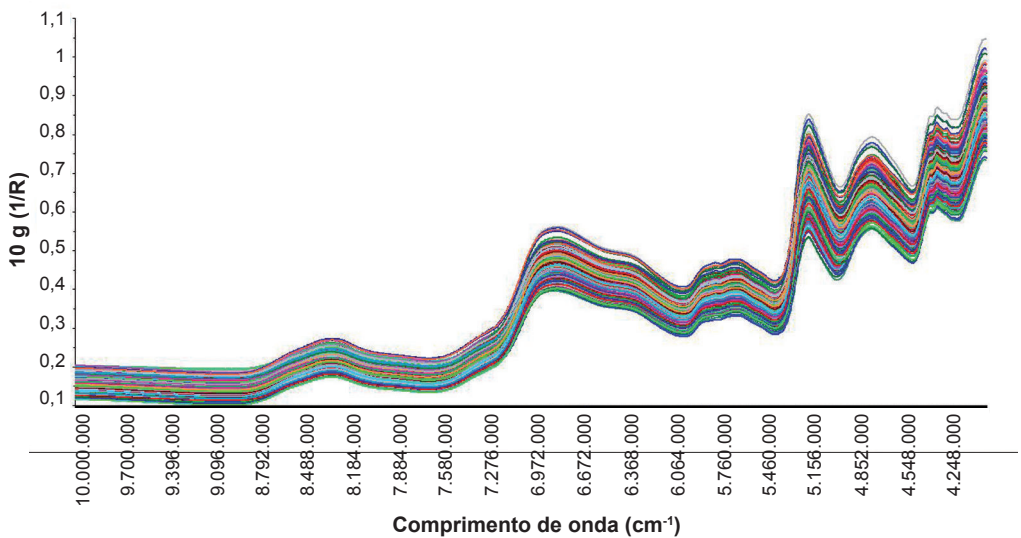
A escolha do número de variáveis latentes foi realizada por meio da resposta obtida pela soma do quadrado da diferença entre os erros residuais (PRESS) na validação cruzada, conforme Haaland e Thomas (1988). A situação ideal é aquela em que ocorre a minimização dos erros residuais.

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 absorbância, 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  $6.820 \text{ cm}^{-1}$ , referente ao primeiro sobreton OH, e outro em  $5.170 \text{ cm}^{-1}$ , referente a uma banda de combinação de OH, para todas as amostras analisadas.

Uma região rica em informação a partir de  $4.600 \text{ cm}^{-1}$  a  $4.000 \text{ cm}^{-1}$  pode ser atribuída a combinações de estiramento de OH/ligação de hidrogênio, C-C estiramento  $4.393 \text{ cm}^{-1}$  e combinações de C-H/C-C ( $4.382 \text{ cm}^{-1} - 4.063 \text{ cm}^{-1}$ ), 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  $4.624 \text{ cm}^{-1}$  a  $4.508 \text{ cm}^{-1}$  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  $4.878\text{ cm}^{-1}$  a  $4.852\text{ cm}^{-1}$  está associada a bandas de combinação de estiramento N-H; em  $4.988\text{ cm}^{-1}$  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 <sup>a</sup>	Val <sup>b</sup>	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 <sup>2</sup> 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 <sup>d</sup> (%)	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 <sup>e</sup>	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 <sup>f</sup>	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 <sup>g</sup>	31,68	27,85	50,98	46,41	33,62	29,86	29,34	28,13	31,31	20,60	16,98	14,61

<sup>a</sup> Cal: conjunto de amostras de calibração

<sup>b</sup> Val: conjunto de amostras de validação

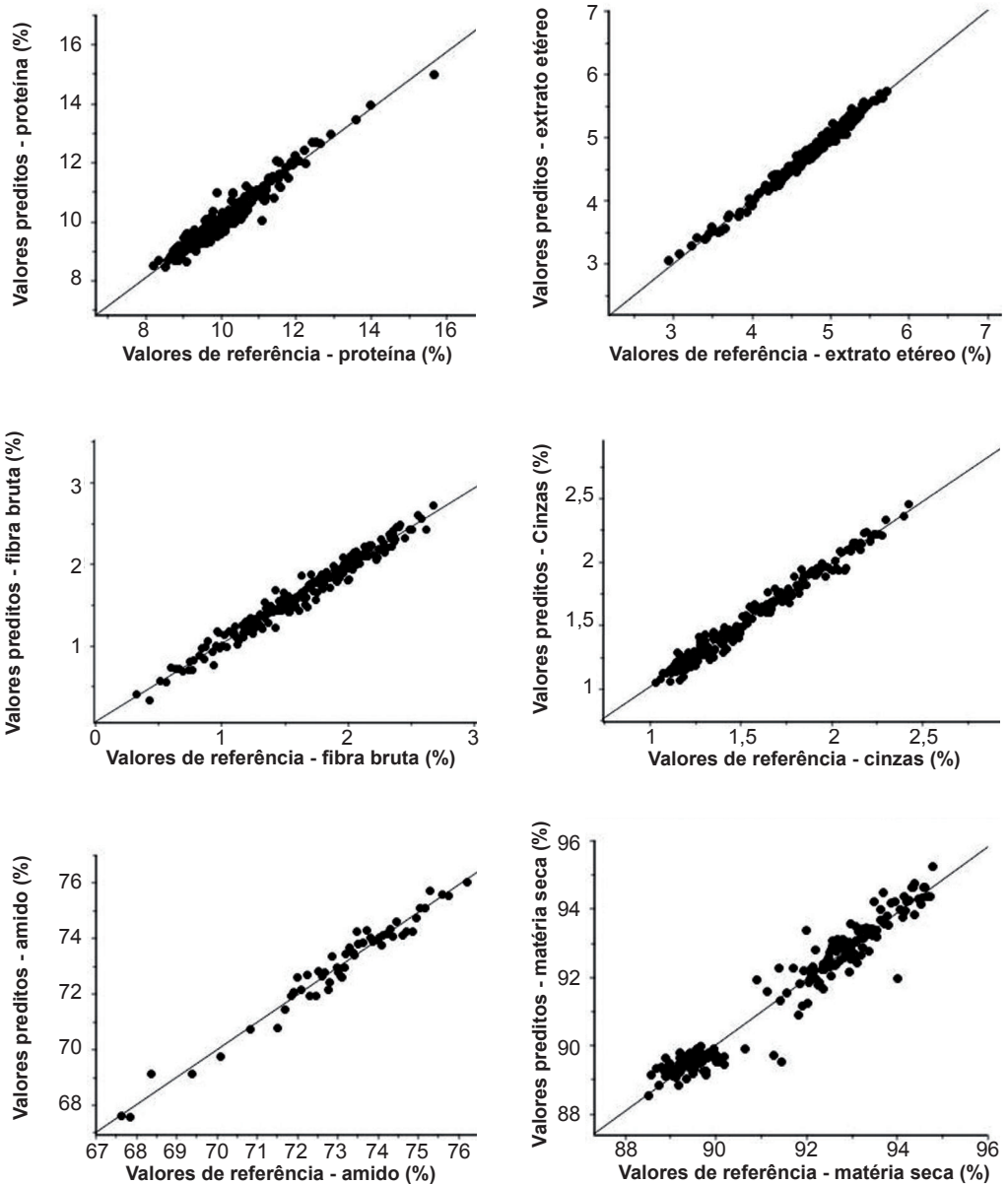
<sup>c</sup> R<sup>2</sup>: coeficiente de determinação

<sup>d</sup> RMSE: raiz quadrada do erro médio – Cal (RMSEC); Val (RMSEP).

<sup>e</sup> Bias: viés

<sup>f</sup> RPD: relação de desempenho do desvio

<sup>g</sup> 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.

# Avaliação de Indicadores de Qualidade Tecnológica da Soja por Espectroscopia no Infravermelho Próximo

Marcelo Alvares de Oliveira  
Rodrigo Santos Leite  
José Marcos Gontijo Mandarino

## Introdução

A análise da composição centesimal em alimentos visa a garantir a qualidade de produtos, atender às exigências de órgãos oficiais de fiscalização e regulamentação, além de contribuir para a proteção da saúde do consumidor. Estas análises incluem a determinação de umidade, proteínas, cinzas, lipídios e carboidratos, que, tradicionalmente, utilizam métodos químicos primários que consomem muito tempo e geram despesas com a compra de reagentes, além de muitos resíduos para o meio ambiente. Assim, novas metodologias têm sido desenvolvidas para reduzir o tempo de análise e minimizar o uso de reagentes químicos.

A determinação do teor de lipídios utiliza tradicionalmente a extração direta em aparelho de Soxhlet<sup>1</sup>. A determinação da concentração de nitrogênio (N) pelo método Kjeldhal é considerada a metodologia padrão para determinação de N total e, indiretamente, o teor de proteínas. Entretanto, esse método gera grande quantidade de resíduos, é moroso e passível de erros durante as etapas, tais como a digestão, a destilação e a titulação (Oliveira Junior et al., 2011).

A espectroscopia de infravermelho próximo (NIRS) vem demonstrando vantagens comparativas, como a rapidez do processo e a não utilização de reagentes químicos, sendo, portanto, um método que não gera resíduos. É um método rápido e não destrutivo, uma vez que pode ser realizado em amostras de sementes e grãos *in natura*. Entretanto, a calibração de modelos multivariados utilizando a espectroscopia no infravermelho próximo é o ponto fundamental na determinação das propriedades que se deseja determinar, como o teor de proteínas e lipídios em sementes e grãos de soja.

---

<sup>1</sup> Processo de extração da fase etérea segundo adaptações ao método oficial AOCS Bc 3-49.



A determinação indireta do teor de proteínas e lipídios em amostras de sementes e grãos de soja, por meio da espectroscopia no infravermelho próximo, vem sendo realizada com sucesso na Embrapa Soja, e tem sido aprimorada desde 2009, época em que a Unidade adquiriu o equipamento.

O programa de melhoramento de soja da Embrapa Soja tem utilizado o NIRS para quantificação dos teores de proteínas e lipídios em amostras de sementes de linhagens de soja, nos avanços de gerações como um *screening* do programa. Anualmente, mais de 2.000 amostras de sementes de soja são analisadas pela técnica de NIRS, visando à quantificação dos teores de proteínas e lipídios, e esses dados são enviados aos melhoristas da Unidade, que os utilizam nas tomadas de decisões quanto ao avanço de gerações do Programa de Melhoramento Genético.

Vários pesquisadores de diferentes equipes da Embrapa Soja têm realizado pesquisas em que a espectroscopia NIR foi a metodologia utilizada para a determinação de componentes da soja, como relatado nos trabalhos a seguir. Oliveira Junior et al. (2011) estudaram a viabilidade de utilização do NIRS em substituição ao método Kjeldahl para determinação da concentração de N total em grãos de soja. Estes autores concluíram que o método de análise por NIR pode substituir a análise convencional de N em grãos de soja, apresentando uma boa relação custo-benefício. No entanto, enfatizam que a comparação e a associação entre os resultados dos dois métodos devem ser realizadas a partir de um grande número de amostras, para refinar os ajustes e obter maior confiabilidade nas estimativas.

Alves et al. (2011) compararam as metodologias de referência com a metodologia utilizando a espectroscopia NIR para a determinação dos teores de proteínas e lipídios em oito cultivares comerciais de soja. Os modelos multivariados foram construídos utilizando o método PLS (partial least squares), com um pré-processamento (1ª derivada) dos espectros. Cada espectro obtido foi relacionado ao respectivo teor de lipídios e proteínas quantificados pelo método de referência. Na construção dos modelos foram selecionadas regiões espectrais distintas para cada componente a ser analisado. A metodologia utilizada não interferiu nos resultados dos teores de lipídios e proteínas das cultivares, porque a análise de variância mostrou que a variável Método e a interação Cultivar-Método não foram significativas a 5%, mostrando assim a eficiência do equipamento de NIR, quando foram utilizados modelos de calibração bem ajustados.

Oliveira et al. (2012) avaliaram a qualidade comercial em amostras de soja colhidas na safra 2011/2012, nos estados do Paraná e Mato Grosso do Sul, considerando a classificação comercial, os teores de lipídios e proteínas, e a presença de fungos contaminantes. Para a determinação dos teores de lipídios e proteínas foi utilizada a

técnica de NIRS. Os teores de lipídios e proteína não sofreram alterações significativas nessas amostras em função dos defeitos de campo.

Oliveira et al. (2013) avaliaram as variações nos teores de lipídios e proteínas em grãos de soja, cultivados sob diferentes condições de manejos de percevejo, desde o momento da colheita até 12 meses de armazenamento utilizando a técnica de NIRS. Estes autores concluíram que, em condições ideais de armazenamento (temperatura de 25 °C e umidade relativa de 60%), ocorreu um aumento nos teores de proteína e uma diminuição nos teores de lipídios dos grãos de soja ao longo do armazenamento. Os grãos de soja com os maiores danos por ataque de percevejos no campo apresentaram aumento nos teores de proteínas durante o armazenamento, sem, entretanto, alterar os teores de lipídios.

Oliveira et al. (2014) verificaram variações nos teores de lipídios e proteínas (utilizando a técnica de NIRS) em grãos de soja, com diferentes manejos para controle de percevejos, provenientes das safras 2011/2012 e 2012/2013, desde o momento da colheita até 12 meses de armazenamento. Em condições ideais de armazenamento, temperatura de armazenamento de 25 °C e umidade relativa de 60%, ocorreu um aumento na concentração nos teores de proteína dos grãos de soja ao longo dos 360 dias de armazenamento, nas duas safras. Os grãos de soja com os maiores danos por ataque de percevejos no campo apresentaram diminuição dos teores de lipídios.

Quirino et al. (2014) estudaram a influência comparativa dos sistemas de exaustão e respiro instalados nos silos, na evolução dos defeitos dos grãos de soja (avariados e fermentados), presença de fungos contaminantes, nos teores de proteínas e lipídios (utilizando a técnica de NIRS), e na acidez do óleo (através de análises químicas), durante 180 dias de armazenamento. Estes autores verificaram que os diferentes sistemas instalados em cada silo não promoveram diferenças significativas nos teores de lipídios, de proteínas e na acidez durante o período de armazenamento. Entretanto, independentemente do sistema instalado, os teores de acidez aumentaram ao longo dos 180 dias de armazenamento.

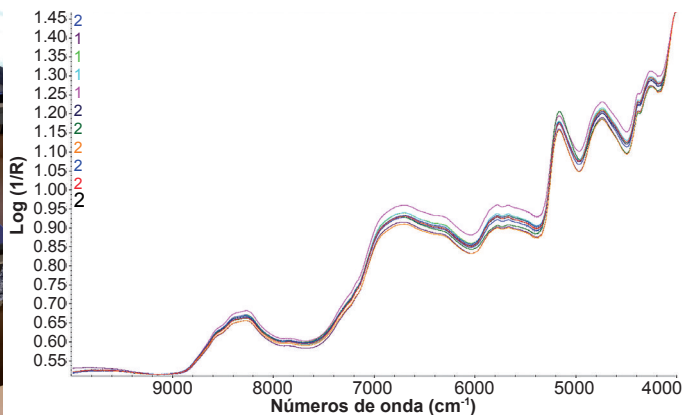
Os dados acima encontrados por Quirino et al. (2014) não foram diferentes dos teores de proteína e lipídios durante os 180 dias de armazenamento, quando comparados os dois diferentes sistemas de exaustão. A hipótese mais provável para explicar o que ocorreu é que o sistema de termometria quando bem utilizado, ou seja, utilização da aeração no momento adequado e em condições adequadas, pode minimizar essas diferenças. Entretanto, quando se perdeu o *time* de aeração, por diferentes motivos, ou a aeração não foi feita de forma adequada, o sistema de exaustão eólico pode fazer a diferença.

Assim sendo, o objetivo desse trabalho foi apresentar o histórico de calibração de NIRS para aplicação nas pesquisas realizadas pela Embrapa Soja desde o ano de 2010, quando se iniciou o trabalho com o NIRS na Unidade.

## Material e métodos

Para a construção dos modelos de predição para teor de lipídios e proteínas em grãos e sementes de soja, foram selecionadas amostras de diversas localidades de cultivo no Brasil, provenientes dos ensaios do Núcleo de Melhoramento Genético da Embrapa Soja, da safra 2011/2012. Cento e oitenta amostras de grãos íntegros e limpos (40 gramas cada) constituíram o conjunto de amostras para a calibração inicial. Os espectros destas amostras foram coletados no equipamento NIR da marca Thermo Scientific, modelo Antaris II, na região  $10.000\text{ cm}^{-1}$  a  $4.000\text{ cm}^{-1}$ , com resolução de  $4\text{ cm}^{-1}$  e com leituras de 32 scans para cada amostra do conjunto de calibração (Figura 1). Foi utilizado, como porta-amostras, um recipiente próprio do fabricante com superfície em quartzo para melhor passagem da luz, e o *background* entre leituras foi obtido com material de referência interna do equipamento, em Spectralon.

Foto: Rodrigo Santos Leite



**Figura 1.** Equipamento NIR utilizado na Embrapa Soja desde 2010 e espectros originais de amostras de soja em grão coletados para a construção dos modelos de predição para teor de óleo e proteína.

Após a coleta dos espectros, os grãos foram moídos em equipamento refrigerado da marca Tecnal, modelo TE-631, e as farinhas obtidas foram passadas através de tamis com malha de 20 mesh. Cada padrão foi então submetido às análises pelos métodos

de referência do Instituto Adolfo Lutz para determinação do teor de lipídios totais, proteína bruta e umidade, em quadruplicata. Para a determinação do teor de proteína bruta foi utilizada a metodologia de Kjeldahl, que compreende as seguintes etapas: digestão ácida, destilação e titulação em equipamento automático marca Foss Tecator, modelo Kjeltec 2400. O teor de lipídios foi determinado em aparelho de Soxhlet, utilizando-se como solvente orgânico o *n*-hexano, seguido pela secagem da amostra padrão desengordurada e a quantificação do percentual de lipídios por gravimetria. O teor de umidade foi determinado por gravimetria, após secagem em estufa a 105 °C, até peso constante.

Após avaliação de todos os resultados obtidos pelos métodos de referência, eles foram separados em faixas para os teores de lipídios e proteína. Através de um estudo de Análises de Componentes Principais - PCAs foram selecionadas as amostras representativas dentro de cada faixa, e que foram utilizadas para a coleta dos espectros e construção dos modelos matemáticos de predição do NIR utilizando o software TQ Analyst, da Thermo Scientific. O restante das amostras foi separado para validação.

Para a predição dos teores de lipídios, proteína bruta e umidade total, foram utilizados modelos desenvolvidos com a análise de regressão linear via PLS (*Partial Least Squares* ou Mínimos Quadrados Parciais). Foram construídos modelos distintos para cada um dos constituintes. Para a modelagem, foram aplicadas técnicas de suavização e uniformização dos espectros, como a 1ª derivada de Savitsky-Golay com 9 pontos à direita e à esquerda e ordem polinomial 2, além da centralização dos dados na média. A validação cruzada foi realizada com a seleção de 34 pontos para predição, pelo método randômico.

Além desses, foram desenvolvidos modelos de predição para o perfil dos ácidos graxos palmítico, esteárico, oleico, linoleico e linolênico em amostras de sementes de soja, a fim de atender à demanda da equipe de Melhoramento Genético, que necessitava de uma forma mais rápida de identificar materiais ainda em desenvolvimento no campo e com potencial para altos teores de ácido oleico. Para a modelagem, foram selecionados 155 materiais do Banco Ativo de Germoplasma (BAG) da Embrapa Soja com teores baixos, médios e altos de ácido oleico em sua composição. Destes, 130 foram aplicados na calibração e os demais 25, na validação externa. As sementes de soja íntegras destes materiais foram identificadas e submetidas à leitura no equipamento NIR, de forma individualizada, com a semente posicionada diretamente sobre a superfície de leitura e coberta por uma tampa plástica escura para evitar escape de luz durante a coleta dos espectros.

A obtenção do espectro de cada amostra de semente de soja foi realizada na faixa 10.000 cm<sup>-1</sup> a 4.000 cm<sup>-1</sup>, utilizando 16 scans com resolução de 4 cm<sup>-1</sup>. Após a coleta

dos espectros, cada amostra foi triturada e o perfil de ácidos graxos foi determinado por meio da metodologia de extração com metóxido de sódio e *n*-heptano seguida da análise por cromatografia gasosa em cromatógrafo a gás, da marca Thermo Scientific, modelo Trace GC Ultra, de acordo com as referências Bannon et al. (1982), Christie (1989), Rayford et al. (1994) e Abidi et al. (1999). Foram determinados os teores dos ácidos oleico, linoleico e linolênico, considerados os mais importantes presentes na fração oleosa da soja, através da análise das mesmas sementes que tiveram os espectros coletados previamente no NIR. Por causa da sua massa reduzida, cada semente foi extraída uma única vez, e as injeções foram realizadas em duplicata.

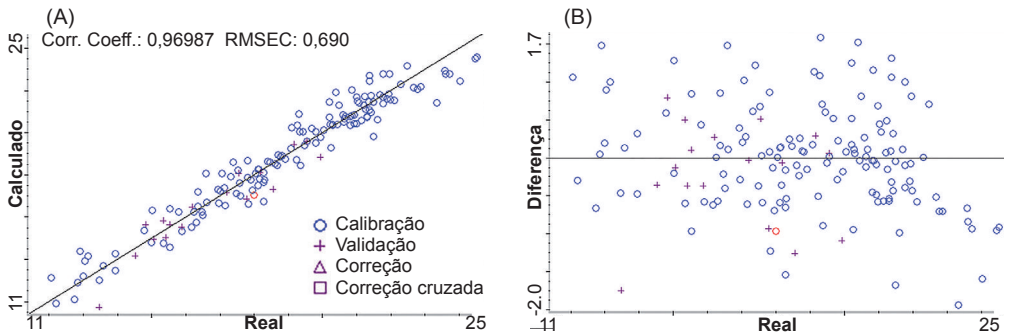
Assim como já havia sido feito para a predição dos teores de lipídios e proteína bruta, estes novos modelos foram desenvolvidos com a técnica de regressão linear PLS (*Partial Least Squares*). Foram construídos modelos distintos para cada um dos ácidos graxos a serem avaliados. Para a construção inicial dos modelos, os dados foram centrados na média e em seguida foram aplicados pré-processamentos para minimizar os efeitos de espalhamento da luz e deslocamentos de linha de base, como a 1ª derivada de Savitsky-Golay com 9 pontos à esquerda e à direita e ordem polinomial 2. Na construção dos modelos multivariados para os ácidos graxos oleico e linoleico, também foi utilizado o software Unscrambler versão 10.2 (Camo) para comparação dos modelos gerados.

Para todos os modelos desenvolvidos, foram avaliados os valores de R (correlação), RMSEC (Raiz Quadrada do Erro Quadrático Médio de Calibração), RMSEP (Raiz Quadrada do Erro Quadrático Médio de Previsão) e RMSECV (Raiz Quadrada do Erro Quadrático Médio de Validação Cruzada).

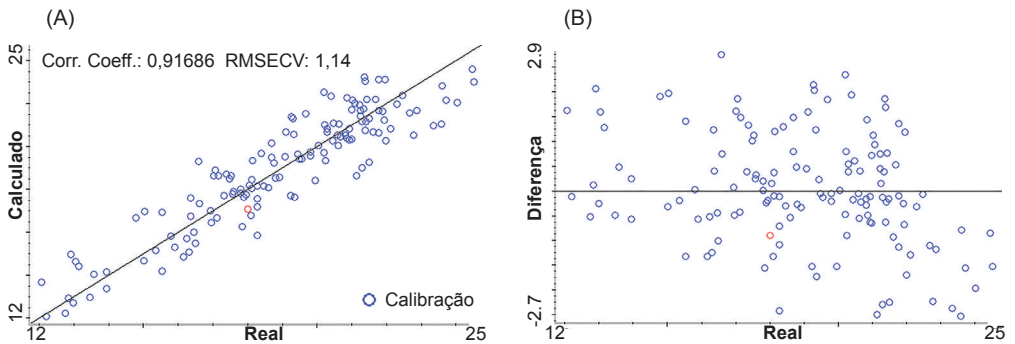
## Resultados e discussões

As amostras analisadas para o teor de lipídios apresentaram uma ampla faixa de resultados (entre 16,32% e 22,76%), com valor médio de 18,90% e desvio padrão 0,95. Na construção do modelo multivariado para a predição do teor de lipídios, o coeficiente de correlação alcançado ficou em 0,9698, com RMSEC 0,690 e RMSECV de 1,14, com correlação de 0,91 (Figuras 2 e 3).

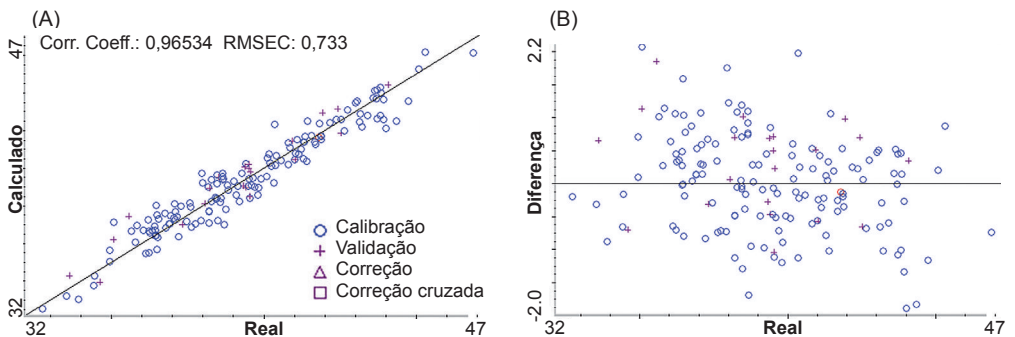
Para o modelo de predição do teor de proteína bruta, as amostras analisadas apresentaram uma faixa de resultados entre 32,94% e 45,90%, com valor médio de 38,96% e desvio padrão 0,83. O coeficiente de correlação alcançado foi de 0,9653, com RMSEC 0,733. Na validação cruzada, o RMSECV foi de 1,22, com correlação de 0,90 (Figuras 4 e 5).



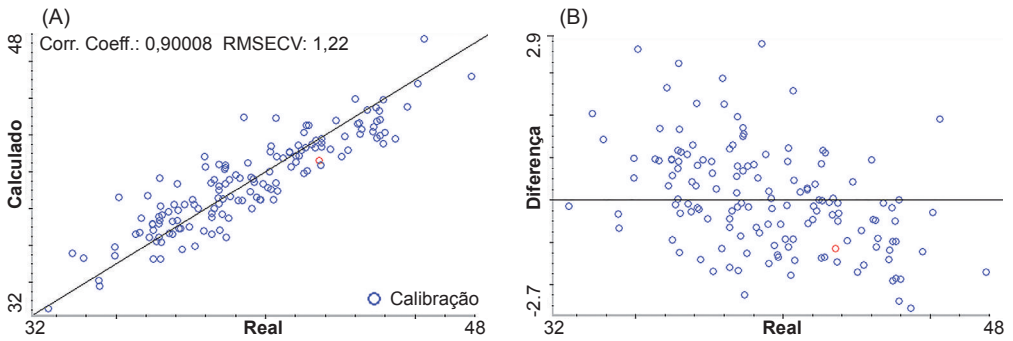
**Figura 2.** Valores de referência versus valores preditos para teor de lipídios em grãos de soja (A), juntamente com gráfico da distribuição dos erros relativos para a predição (B).



**Figura 3.** Validação cruzada e valores de referência versus valores preditos para teor de lipídios em grãos de soja (A), juntamente com gráfico da distribuição dos erros relativos para a predição (B).



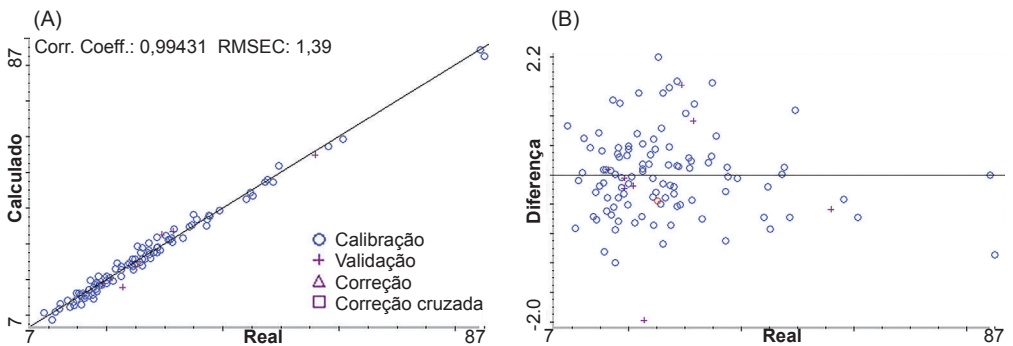
**Figura 4.** Valores de referência versus valores preditos para teor de proteína bruta em grãos de soja (A), juntamente com gráfico da distribuição dos erros relativos para a predição (B).



**Figura 5.** Validação cruzada e valores de referência versus valores preditos para teor de proteína bruta em grãos de soja (A), juntamente com gráfico da distribuição dos erros relativos para a predição (B).

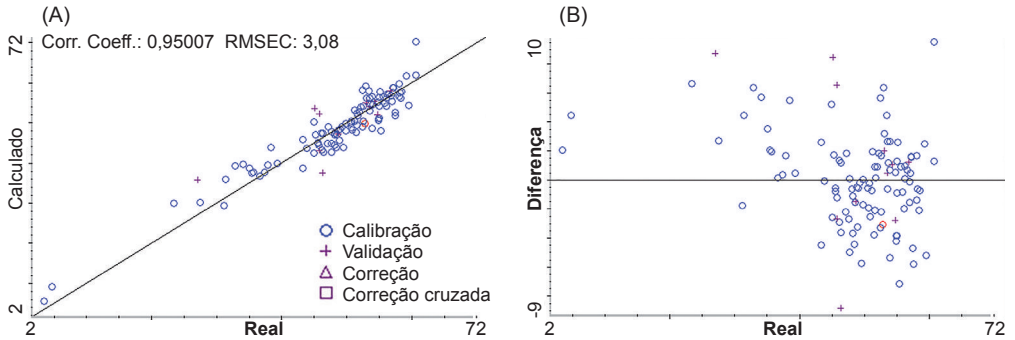
Para ambos os modelos (lipídios e proteína bruta), os valores alcançados na validação externa foram próximos aos da calibração, indicando que o modelo está bem ajustado e com as amostras bem distribuídas.

Na predição do perfil de ácidos graxos em soja, o modelo inicialmente desenvolvido para o ácido oleico apresentou correlação alta (0,9943), RMSEC igual a 1,39 e RMSEP de 2,20 (Figura 6). Na validação cruzada, os valores de correlação e RMSECV foram de 0,74 e 10,3, respectivamente.



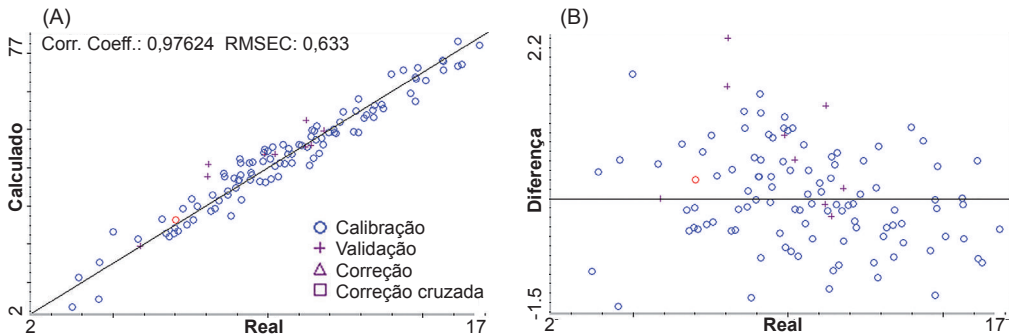
**Figura 6.** Valores de referência versus valores preditos para ácido oleico em grão de soja (A), juntamente com gráfico da distribuição dos erros relativos para a predição (B).

O modelo de predição para o ácido linoleico apresentou correlação de 0,95, porém com erros mais altos (RMSEC 3,08 e RMSEP 5,15) (Figura 7).



**Figura 7.** Valores de referência versus valores preditos para ácido linoleico em grão de soja (A), juntamente com gráfico da distribuição dos erros relativos para a predição (B).

O modelo de predição para o ácido linolênico também apresentou correlação satisfatória, de 0,976, e valores muito bons para RMSEC (0,633) e RMSEP (1,01) (Figura 8).



**Figura 8.** Valores de referência versus valores preditos linolênicos em grão de soja individual (A), juntamente com gráfico da distribuição dos erros relativos para a predição (B).

Por se tratar de uma predição mais complexa, a diferenciação entre as diferentes cadeias carbônicas (C18:1; C18:2 e C18:3) dos ácidos graxos diretamente nos grãos inteiros acarretou em modelos com valores de RMSEC e RMSEP superiores aos encontrados nos modelos para lipídios e proteínas, e por essa razão merecem uma constante atualização do modelo com novos materiais, a fim de melhorar a robustez da predição.



Os modelos de predição para os ácidos graxos oleico e linoleico apresentaram maior interesse pela equipe de pesquisadores da área de Melhoramento em Soja por serem estes dois os mais importantes na composição do óleo de soja e, após ajustes e inserção de novos materiais com valores de referência representativos dentro da faixa de calibração, foram submetidos à validação externa com materiais selecionados e distintos daqueles que haviam sido utilizados para a calibração (Tabela 1).

**Tabela 1.** Validação externa para os modelos de predição de frações de ácidos graxos oleico e linoleico em soja através de espectroscopia no infravermelho próximo por meio da determinação das concentrações desses ácidos graxos por cromatografia gasosa (CG). Embrapa Soja, Londrina, PR, 2018.

Código da amostra	%Oleico/%Total ácidos graxos		%Linoleico/%Total ácidos graxos	
	NIRS	Referência	NIRS	Referência
1	23,5	24,9	50,2	44,2
9	79,1	76,5	15,6	4,1
16	46,5	49,8	33,8	25,4
18	88,9	83,9	10,4	2,2
29	21,5	23,0	48,0	50,7
32	45,6	44,1	35,3	31,1
36	39,0	42,2	34,6	31,5
3	20,2	18,9	47,9	51,6
9	15,8	23,2	54,3	48,4
23	60,4	66,0	29,2	12,8
27	19,7	16,3	51,0	56,1
33	47,2	52,1	35,8	22,7
38	28,9	16,4	50,4	54,5
56	15,6	15,9	49,0	57,3
63	50,3	57,6	33,8	18,7
69	50,1	52,3	34,0	25,1
78	50,2	46,7	30,8	28,0
85	23,7	26,5	47,5	47,1
87	38,0	32,4	35,3	42,0

continua...

**Tabela 1.** Continuação.

Código da amostra	%Oleico/%Total ácidos graxos		%Linoleico/%Total ácidos graxos	
	NIRS	Referência	NIRS	Referência
94	43,0	41,8	33,0	33,2
102	17,3	21,2	53,0	53,6
108	47,1	41,7	29,9	30,5
120	20,8	18,4	52,0	54,6
131	63,0	64,0	26,0	13,9
138	21,3	17,1	50,3	56,2

Os resultados obtidos na validação externa para os teores de ácido oleico e linoleico reforçaram a necessidade de ampliar o número de amostras nos modelos, principalmente quando se analisam genótipos com teores intermediários de ácido linoleico (30% a 50% da fração total de ácidos graxos). Já para os genótipos com um intervalo maior entre os teores de oleico e linoleico, os resultados da validação apresentaram erros relativos menores que 2%.

## Tendências e perspectivas

Modelos para a predição dos teores de clorofila total e acidez titulável em grãos e farinha integral de soja estão sendo desenvolvidos, e para isto, ao longo das safras 2014/2015 e 2015/2016, foram avaliadas aproximadamente 1.500 amostras de grãos de soja para estes parâmetros, utilizando-se métodos primários químicos validados.

Visando à futura calibração destas propriedades, foram coletados os espectros tanto dos grãos inteiros de soja quanto da farinha integral obtida pela moagem, para a construção de modelos de calibração multivariados utilizando a espectroscopia no infravermelho próximo. Com isso, teremos a redução do tempo das análises, além de geração de economia com aquisição de grandes quantidades de reagentes de alto custo, nocivos à saúde humana e prejudiciais ao meio ambiente.

## Conclusões

Os modelos multivariados de calibração utilizando a espectroscopia no infravermelho próximo para a predição dos teores de lipídios e proteína em soja têm se mostrado satisfatórios, e desde a sua implantação já foram analisadas mais de 8.000 amostras do programa de Melhoramento Genético da Embrapa Soja. Os resultados encontrados com o uso dessa tecnologia são muito próximos aos obtidos por métodos químicos de referência. Com isso houve grande economia na aquisição de reagentes e materiais de consumo empregados nas análises químicas tradicionais, sem que houvesse prejuízo quanto aos resultados obtidos, demonstrando assim a importância do NIRS.

Em relação aos ácidos graxos, trata-se de uma demanda que não necessita de valores exatos, mas sim de valores aproximados, que facilitem a seleção de genótipos para o Programa de Melhoramento Genético da Embrapa Soja. Os modelos iniciais podem ser considerados satisfatórios e, atualmente, estão sendo feitos ajustes nos modelos de predição em soja para melhorar a robustez das predições.

Assim, a experiência com a utilização do equipamento NIR nesses últimos oito anos na Embrapa Soja tem demonstrado que a acuidade e a robustez dos modelos de calibração são de vital importância para o sucesso da tecnologia. Pelo fato de a Embrapa Soja ter ensaios em todas as regiões produtoras de soja do Brasil, podem-se utilizar amostras representativas de todo o País nas calibrações do equipamento de NIR, o que tem sido um aspecto importantíssimo na obtenção de modelos de predição cada vez mais ajustados.

## Referências

ABIDI, S. L.; LIST, G. R.; RENNICK, K. A. Effect of genetic modification on the distribution of minor constituents in canola oil. **Journal of American Oil Chemistry Society**, v. 76, n. 4, p. 463 – 467, 1999.

ALVES, F. P.; OLIVEIRA, M. A.; LEITE, R. S.; OLIVEIRA JUNIOR, A. Comparação de diferentes métodos para análise do teor de lipídio e proteína em cultivares de soja. In: SEMANA DE CIÊNCIA E TECNOLOGIA DE ALIMENTOS, 3., 2011, Londrina. **Resumos...** Londrina: UTFPR, 2011. Poster.

BANNON, C. D.; BREEN, G. J.; CRASKE, J. D.; HAI, N. T.; HARPER, N. L.; CZONYIC, C. **Journal of Chromatography**, v. 247, p. 71, 1982.

CHRISTIE, W. W. **Gas chromatography and lipids: a practical guide**. Ayr, Scotland: The Oil Press, 1989. 307 p.

OLIVEIRA JUNIOR, A.; CASTRO, C. de; GERMANO, M. G.; LEITE, R. S.; KLEINERT, J. J.; NARDO, A. E.; FARINE, E. L.; MOREIRA, D. A.; CAMARGO, I. S. P. de. Avaliação da espectroscopia de infravermelho próximo (NIR) para determinação da concentração de Nitrogênio em amostras de grãos de soja. In: CONGRESSO BRASILEIRO DE CIÊNCIA DO SOLO, 33., 2010, Uberlândia. **Anais...** Uberlândia: UFU, 2011.

OLIVEIRA, M. A.; LORINI, I.; HENNING, A. A.; MANDARINO, J. M. G.; CHIODELLI, A. Caracterização da qualidade comercial da soja colhida na safra 2011-2012 no Estado do Paraná e Mato Grosso do Sul. In: CONGRESSO BRASILEIRO DE SOJA, 6., 2012, Cuiabá. **Soja: integração nacional e desenvolvimento sustentável: resumos**. Brasília, DF: Embrapa, 2012. p. 20.

OLIVEIRA, M. A.; LORINI, I.; MANDARINO, J. M. G.; LEITE, R. S.; CORREA-FERREIRA, B. S.; QUIRINO, J. R.; VILAS BOAS, R. L. P.; DELAFRONTA, B. Teores de óleo, proteína e acidez em grãos de soja, com diferentes manejos de perceiving, da colheita ao armazenamento, provenientes das safras 2011/2012 e 2012/2013. In: CONFERÊNCIA BRASILEIRA DE PÓS-COLHEITA, 6.; SIMPÓSIO PARANAENSE DE PÓS-COLHEITA DE GRÃOS, 8., 2014, Maringá. **Anais...** Londrina: ABRAPOS, 2014. p. 579-585.

OLIVEIRA, M. A.; LORINI, I.; MANDARINO, J. M. G.; LEITE, R. S.; QUIRINO, J. R.; CORRÊA-FERREIRA, B. S.; VILAS BOAS, R. L. P.; DELAFRONTA, B. Teores de óleo e proteína em grãos de soja, com diferentes manejos de perceiving, da colheita ao armazenamento, utilizando a espectroscopia no infravermelho próximo (NIR). In: AMERICAS: INTERNATIONAL CONFERENCE ON SOYBEAN UTILIZATION, 2013, Bento Gonçalves. **Proceedings...** Brasília, DF: Embrapa, 2013. 1 CD-ROM. 5 p.

QUIRINO, J. R.; OLIVEIRA, M. A.; LORINI, I.; ROSA, E. E.; HENNING, A. A.; FRANCA-NETO, J. B.; KRZYZANOWSKI, F. C.; LEITE, R. S. Efeito comparativo dos sistemas de exaustão e respiro do silo na qualidade da soja armazenada. In: CONFERÊNCIA BRASILEIRA DE PÓS-COLHEITA, 6.; SIMPÓSIO PARANAENSE DE PÓS-COLHEITA DE GRÃOS, 8., 2014, Maringá. **Anais...** Londrina: ABRAPOS, 2014. p. 230-236.

RAYFORD, W. E.; THOMAS, D. I.; ELAM, L. M.; WALKER, S. M. **Analytical chemical support soybean uniform test analysis**. Peoria: USDA, Agricultural Research Service, Midwest Area, National Center for Agricultural Utilization Research, 1994. 26 p.

# Avaliação de Parâmetros de Qualidade de Trigo por Espectroscopia no Infravermelho Próximo

Casiane Salete Tibola  
Martha Zavariz de Miranda  
Eliana Maria Guarienti

## Introdução

O mercado de grãos diferenciados está em expansão, e as indústrias apresentam novas demandas para segregação e rastreabilidade de produtos, visando a atender ao consumidor cada vez mais exigente em qualidade. Entretanto, análises de qualidade tecnológica e de contaminantes, que orientem a segregação e a rastreabilidade, são limitadas pela disponibilidade de laboratórios, de analistas e de equipamentos, além do elevado custo de reagentes e de materiais de laboratório. Isto é especialmente relevante para trigo, que demanda análises de multiparâmetros para caracterizar a aptidão tecnológica e a inocuidade.

O trigo é a principal fonte de nutrientes para a população mundial. No Brasil, o consumo diário em 2013 foi de 56,3 kg/habitante/ano (De Mori, 2016). A definição do padrão oficial de classificação do trigo no Brasil foi estabelecida no Regulamento Técnico do Trigo (Brasil, 2010). Além de exigências de normativas oficiais, cada produto derivado de trigo necessita ter parâmetros físico-químicos e reológicos específicos, para que seja obtido o máximo de rendimento e o produto final tenha características de qualidade desejadas pelo consumidor.

Na pós-colheita de trigo, um dos fatores limitantes na escolha do método de análise é a duração do teste, pois quando esse é demorado (como a alveografia, a farinografia e o teor de glúten), dificulta a segregação por qualidade. Assim, uma das grandes demandas em pós-colheita de grãos é a seleção de método rápido e eficaz para segregação e rastreabilidade de lotes comerciais, que possa ser empregado na recepção em unidades armazenadoras. Atualmente, a segregação é feita em função da classificação comercial da cultivar. Entretanto, este indicativo de qualidade, muitas vezes, não se confirma, em função de fatores de cultivo como adubação, controle fitossanitário, manejo do solo, condições climáticas, manejos pré e pós-colheita, entre

outros. Portanto, somente após a adequada avaliação da qualidade na pós-colheita pode-se definir a segregação.

Neste contexto, a espectroscopia no infravermelho próximo (NIRS) tem sido utilizada como método para prever a qualidade de diferentes produtos agrícolas, por causa da rapidez de análise, da mínima preparação de amostra e do baixo custo. Esta metodologia fundamenta-se na absorção de luz infravermelha por compostos orgânicos, que apresentam características de absorção específicas, causando vibrações nas ligações de hidrogênio induzidas pelo calor. Essas reações permitem que os componentes sejam quantificados por meio de equações de calibração. Na região do infravermelho, com comprimento de onda ( $\lambda$ ) entre 800 nm e 2.500 nm, radicais como -OH, -NH e -CH vibram fortemente. Como estes grupos são componentes típicos de alimentos (água, proteína, lipídios e carboidratos), o método NIRS é particularmente adequado para análises nestes produtos (Kersting; Zwinoelberg, 2006). Os métodos baseados em espectroscopia têm sido popularizados, especialmente por possibilitarem a análise simultânea de vários parâmetros, por constituírem-se em uma técnica não destrutiva (permitindo a reutilização da amostra), por possuírem rapidez de processamento das informações e de fornecimento de resultados quantitativos, por não consumirem reagentes químicos nocivos ao meio ambiente e por serem menos laboriosos e de custo relativamente baixo, quando comparados aos métodos tradicionais (Williams, 2007).

As indústrias de moagem e de panificação foram precursoras na adoção da tecnologia NIRS. A quantificação de proteína no trigo moído e na farinha é a aplicação mais antiga e mais bem estabelecida (Osborne, 2007). A tecnologia NIRS foi usada com sucesso para a classificação de trigo de acordo com o teor de proteínas (Delwiche; Norris, 1993), na predição da dureza do grão (Maghirang; Dowell, 2003), no acompanhamento da maturação pós-colheita de trigo e de cevada (Cassels et al., 2007) e no estudo das mudanças do pão durante o armazenamento (Xie et al., 2003). Dowell et al. (2006) avaliaram vários parâmetros de qualidade de trigo com o uso do NIRS e obtiveram valores de coeficiente de determinação ( $R^2$ ) superiores ou iguais a 0,97, para umidade, proteína e parâmetro de cor de farinha  $b^*$  (coordenada de cromaticidade  $b^*$ , cor amarela). Estes coeficientes são considerados preditivos, sendo adequados para controle de processo (Dowell et al., 2006). Valores de  $R^2 \geq 0,70$  de NIRS são considerados adequados para *screening*, ou seja, para triagem ou seleção de lotes de acordo com características de interesse. Estes coeficientes foram obtidos para peso do hectolitro, diâmetro de grão, parâmetro de cor  $a^*$  (coordenada de cromaticidade  $a^*$ , cor branca), absorção de água, farinografia, extensibilidade (L), índice de intumescimento (G) e força de glúten (W), alveografia, absorção de água na panificação e volume do pão (Dowell et al., 2006). Outras aplicações do NIRS na avaliação de parâmetros de

qualidade em trigo incluíram extração de farinha, sedimentação e teor de proteína, com alta precisão para quantificação de proteínas (Jirsa et al., 2008).

As principais aplicações analíticas do NIRS em cereais e derivados foram descritas por Jespersen e Munck (2009). Pojić e Mastilović (2013) apresentaram revisão detalhada quanto às aplicações do NIRS para avaliar a composição e as propriedades funcionais de trigo, para ser utilizado em melhoramento genético, processamento e comercialização. Mais recentemente, Porep et al. (2015) revisaram as aplicações *on line* do NIRS na produção de alimentos, destacando o uso de NIRS conectado à colhedora para avaliar níveis de proteína durante a colheita de trigo. A correlação da predição de proteína na colhedora (*in line*) com os resultados obtidos no método de referência foi alta ( $R^2=0,94$ ), demonstrando a viabilidade de emprego da espectroscopia sob condições de campo (Long et al., 2008). Ziegler et al. (2016) utilizaram NIRS para a rápida discriminação de grãos e de farinhas de diferentes espécies de trigo, demonstrando a aplicabilidade da técnica para identificar a autenticidade de produtos.

Avanços na capacidade de predição de parâmetros de qualidade tecnológica de trigo e de derivados com uso do NIRS possibilitarão novas oportunidades de segregação na cadeia produtiva. O objetivo deste capítulo foi descrever as funcionalidades da espectroscopia NIR para avaliar parâmetros de qualidade tecnológica no trigo. Os parâmetros analisados através dos espectros do trigo moído foram peso do hectolitro, teor de umidade e de proteínas totais, índice de dureza do grão, número de queda, força de glúten (alveografia), estabilidade (farinografia) e glúten úmido.

## Material e métodos

### Seleção de amostras, equipamento e obtenção de espectros no NIRS

Amostras de trigo de diferentes cultivares foram obtidas de experimentos e de lavouras comerciais, no período de 2009 a 2016, das principais regiões produtoras de trigo, especialmente do Paraná e do Rio Grande do Sul. No presente estudo, somente a avaliação de safra 2015 foi realizada em duplicata. Todas as amostras foram moídas com moinho de laboratório (Modelo 3100® Perten, Suíça), obtendo-se partículas menores que 0,8 mm de diâmetro. Após a homogeneização, alíquotas de, aproximadamente 4,5 g, acondicionadas em cubeta cilíndrica com 3,8 cm de diâmetro, com janela de quartzo, foram escaneadas no NIR para a coleta de espectros, em modo estacionário. O modelo utilizado foi XDS RCA (*Rapid Content Analyzer*), da

FOSS NIRsystems, Hillerød, Dinamarca. Os dados espectrais foram registrados no modo reflectância, a cada 2 nm, para compor espectro com 1.050 pontos, na faixa de 400 nm – 2.500 nm, com o software ISIScan™ (Infrasoft International LLC, 2008). Os dados espectrais foram armazenados como log 1/R.

A metodologia NIRS demanda o desenvolvimento de equações de calibração, utilizando-se os resultados obtidos nos métodos de referência, a partir da mesma amostra. Assim, utilizou-se a Quimiometria, técnica que aplica a estatística e a matemática com o objetivo de encontrar relações quantitativas ou qualitativas entre os espectros obtidos no NIRS e os resultados obtidos no método de referência.

## **Análises de referência**

Foram realizadas as seguintes análises, com os respectivos métodos:

Peso do hectolitro (PH) - é a massa de 100 litros de trigo, expressa em quilograma por hectolitro (kg/hL). Os resultados são influenciados por uniformidade, forma, densidade, tamanho e presença de matérias estranhas e grãos quebrados na amostra (Guarienti, 1996). De acordo com a Instrução Normativa (IN) nº 38 (Brasil, 2010), além de outros atributos, o peso do hectolitro faz parte das análises para tipificação de trigo: Tipo 1 (PH  $\geq$  78), Tipo 2 (PH  $\geq$  75 e  $<$  78), Tipo 3 (PH  $\geq$  72 e  $<$  75) e Fora de Tipo (PH  $<$  72). O PH foi obtido em balança Dalle Molle, pelo método 55-10.01(AACC International, 1999c).

Umidade do grão (UG) - de acordo com a IN 38 (Brasil, 2010), a umidade é o percentual de água encontrado na amostra. O teor de umidade tecnicamente recomendável para trigo, no Brasil, é de 13%, valor máximo permitido em razão do clima. A umidade foi obtida pelo método AACC 44-15.02 (AACC International, 2010), com secagem em estufa a 130 °C, por 1 hora, em duplicata.

Proteínas totais do grão (PTG) - as proteínas do trigo estão divididas em proteínas não formadoras de glúten (globulinas e albuminas) e em formadoras de glúten (gliadinas e gluteninas). No Brasil, foram estabelecidos teores mínimos de proteínas totais para diversos usos da farinha: panificação artesanal e panificação industrial (12%), farinha doméstica (10%), massas (14%), massas frescas/instantâneas (12%), biscoitos fermentados (9% a 12%), biscoitos moldados doces e biscoitos laminados doces (8% a 9%) e bolos (8%) (Tabela 21 ..., 2016). O teor de proteínas foi quantificado pelo método 46-13.01, da AACC International (1999a), em duplicata.



Índice de dureza do grão (IDG) - definido como a dificuldade de desintegração dos grãos quando sobre eles é exercida uma pressão (Simmonds, 1974): Extraduro (acima de 91); Muito duro (81 a 90); Duro (65 a 80); Semiduro (45 a 64); Semimole (35 a 44); Mole (25 a 34); Muito mole (10 a 24), e Extramole (abaixo de 10). Foi determinado em equipamento SKCS – Single Kernel Characterization system, marca Perten, pelo método de análise 55-31.01, da AACC International (1999d).

Número de queda (NQ) - conhecido como *Falling Number*, mede a intensidade da atividade da enzima alfa-amilase no grão, expresso em segundos. Baixos valores indicam elevada atividade de alfa-amilase, situação que comumente é resultado do processo de germinação pré-colheita. A IN 38 (Brasil, 2010) utiliza o número de queda como um dos parâmetros para a classificação do trigo por Classe: Melhorador ( $\geq 250$  s), Pão ( $\geq 220$  s), Doméstico ( $\geq 220$  s), Básico ( $\geq 200$  s) e Outros usos (qualquer valor). Foi determinado em equipamento Falling number marca Perten, segundo o método 56-81.03, da AACC International (1999e).

Força de glúten (W) - representa o trabalho de deformação da massa e indica a qualidade panificativa da farinha (força da farinha). O teste simula o comportamento da massa na fermentação, imitando a formação de alvéolos originados na massa pelo gás carbônico. A IN 38 (Brasil, 2010) utiliza a força de glúten (expressa em  $\times 10^{-4}$  J), além de outros parâmetros, para classificar os trigos por Classes: Melhorador ( $\geq 300$ ), Pão ( $\geq 220$  e  $< 300$ ), Doméstico ( $\geq 160$  e  $< 220$ ), Básico ( $\geq 100$  e  $< 160$ ) e Outros usos ( $< 100$ ). Obtida em alveoconsistógrafo marca Chopin, modelo NG, pelo método 54-30.02, da AACC International (1999b).

Estabilidade (EST) - fornece a medida de tolerância da massa ao amassamento. A IN 38 (Brasil, 2010) utiliza a estabilidade (em minutos), além de outros parâmetros, para classificar o trigo por Classes: Melhorador ( $\geq 14$ ), Pão ( $\geq 10$  e  $< 14$ ), Doméstico ( $\geq 6$  e  $< 10$ ), Básico ( $\geq 3$  e  $< 6$ ) e Outros usos ( $< 3$ ). Obtida em farinógrafo/resistógrafo marca Brabender, modelo 821000, com cuba/masseira de 50 g, pelo método 54-60.01 da AACC International (2011).

Glúten úmido - o glúten é uma rede formada pelas proteínas insolúveis do trigo (gliadinas e gluteninas) quando se adiciona água à farinha. O teste de glúten úmido fornece a medida quantitativa dessas proteínas, sendo superior a 28% ideal para panificação. Determinado pelo Sistema Glutomatic, marca Perten, modelo 2200, pelo método 38-12.02, da AACC International (2000).

Na Tabela 1 são apresentados os números de amostras avaliadas, eliminadas e utilizadas para elaboração dos modelos de calibração, além dos valores mínimos e máximos obtidos em cada parâmetro de qualidade tecnológica de trigo analisado.

**Tabela 1.** Caracterização do conjunto de amostras utilizadas para construção dos modelos de calibração no NIRS, de parâmetros de qualidade tecnológica de trigo, de acordo com resultados obtidos nos métodos de referência. Embrapa Trigo, Passo Fundo, RS, 2018.

Parâmetro	Amostra avaliada (nº)	Amostra eliminada (nº)	Amostra usada na calibração (nº)	Valor mínimo da variável	Valor máximo da variável
Peso do hectolitro (kg/hL) <sup>1</sup>	6.878	293	6.585	59,2	84,3
Teor de umidade (%) <sup>2</sup>	568	30	538	10,0	17,1
Teor de proteínas totais (%) <sup>2</sup>	588	28	560	8,3	17,7
Índice de dureza dos grãos <sup>2</sup>	6.845	340	6.505	7	100
Número de queda (s) <sup>2</sup>	6.891	291	6.600	62	800
Força de glúten (x 10 <sup>-4</sup> J) <sup>2</sup>	6.701	214	6.487	13	601
Estabilidade (min) <sup>2</sup>	1.708	158	1.550	0,6	45,9
Glúten úmido (%) <sup>2</sup>	6.580	365	6.215	12,0	69,5

<sup>1</sup>Amostras obtidas nos anos 2012 a 2016. <sup>2</sup>Amostras obtidas nos anos 2009 a 2015.

## Pré-processamento de dados

Os modelos de calibração e de validação foram desenvolvidos utilizando-se o software WinISITM 4 (Infrasoft International LLC, 2007).

O pré-processamento antes da calibração foi utilizado para eliminar ruídos, interferentes e fontes indesejadas de variação dos espectros originais (difusão de luz e reflexão especular), visando a obter modelos acurados, robustos e estáveis. O principal objetivo do pré-processamento é transformar os dados (sinais multivariados/múltiplos) para aderir à Lei de Beer, que determina que a absorvância e a concentração do constituinte de interesse (analito) apresentem correlação linear (Rinnan et al., 2009).

Dois métodos de pré-processamento espectrais, variação padrão normal – Detrend (SNV-D) ou correção de difusão multiplicativa (MSC), foram utilizados para caracterizar a qualidade tecnológica do trigo. SNV-D compõe-se de dois algoritmos usualmente aplicados juntos, sendo que o primeiro corrige tamanho de partícula e efeito de interferentes na difusão, e o segundo remove tendência linear e curvilínea de cada espectro (Haba et al., 2006). O conceito básico de MSC é remover as não linearidades nos dados causadas pela dispersão de partículas nas amostras (Rinnan et al., 2009).

Duas a três derivações matemáticas foram aplicadas aos dados para reduzir a variação de base e potencializar as variações ao longo do espectro. Dentre as derivadas testadas, foram selecionadas: 1.4.4.1; 2.8.4.1 e 2.8.6.1, significando que o primeiro dígito é a ordem da derivada, o segundo é o intervalo sobre o qual a derivada é calculada, o terceiro é o número de pontos de dados considerados no cálculo da média, ou alisamento, e o quarto é o segundo alisamento (Fernández-Ibañez et al., 2009).

## Modelos de calibração e de validação

A construção de modelos de calibração envolve as etapas de seleção de amostras representativas de cada parâmetro a ser analisado, de quantificação do parâmetro de interesse em método de referência (padrão), de obtenção/coleta do espectro no NIRS, de construção do modelo de calibração e de validação do modelo desenvolvido para a predição do parâmetro em amostras desconhecidas (novas). Esses procedimentos foram adotados para todos os parâmetros de qualidade tecnológica avaliados nos métodos baseados em espectroscopia.

Na validação cruzada, o conjunto de amostras foi dividido em quatro grupos; cada grupo é então validado usando calibração desenvolvida com as amostras restantes (Infrasoft International, 2007). Este procedimento foi utilizado para selecionar o número ótimo de termos para quadrados mínimos parciais (PLS). Para a validação externa foram separadas, aleatoriamente, 20 amostras por ano, durante cinco anos para compor o conjunto de validação, mantendo a representatividade e as faixas de concentração dos parâmetros analisados.

As principais técnicas de calibração foram descritas em Porep et al. (2015). As regressões de componentes principais (PCR) e (PLS) são frequentemente aplicadas na compactação de dados para avaliar novas variáveis. A regressão PLS contém todas as informações espectrais e está mais diretamente relacionada com a variabilidade espectral do que quando se consideram somente os componentes principais (regressão PCR). As combinações lineares resultantes das variáveis utilizadas na calibração são nominadas como variáveis latentes, fatores ou componentes. O objetivo é assegurar que as primeiras variáveis latentes contenham tanta informação quanto possível do valor preditivo (Blanco Romía; Alcalá Bernárdez, 2009). Na avaliação de parâmetros de qualidade tecnológica de trigo, a regressão por quadrados mínimos parciais modificada (MPLS) foi utilizada, para correlacionar os resultados obtidos no método de referência com os dados espectrais do conjunto de calibração no NIRS. A regressão MPLS implementada no programa WinISI trata-se de uma versão melhora-

da do PLS, por causa da eficácia na identificação de interferentes alheios ao analito de interesse. A regressão MPLS foi detalhada em Zhang e Littlejohn (1996).

A análise de componentes principais (PCA) foi aplicada no conjunto de amostras para calcular a distância de Mahalanobis (H) de cada espectro, em comparação com a média, para avaliar presença de valores atípicos (*outliers*) (Waes et al., 2005). De acordo com Shenk e Westerhaus (1991), as amostras com valor de H superior a 3 foram consideradas *outliers* e eliminadas do conjunto de calibração (Tabela 1).

O número ideal de fatores e a acurácia do modelo de calibração foram testados pela validação cruzada, selecionando-se o menor erro padrão de validação cruzada (SECV) e o maior coeficiente de determinação na validação cruzada (1-VR).

O desempenho do modelo de calibração foi mensurado em termos de precisão e acurácia, de acordo com as seguintes estatísticas: desvio padrão (SD), erro padrão de calibração (SEC), coeficiente de determinação da calibração ( $R^2$ ), erro padrão da validação cruzada (SECV) e coeficiente de determinação da validação cruzada (1-VR).

A melhor curva de calibração foi validada utilizando-se um conjunto independente (externo) de amostras de trigo, do qual foram calculados o erro padrão de predição (SEP), o teste para detecção de erros sistemáticos (*bias*) e o coeficiente de determinação ( $R^2$ ).

A utilidade prática dos modelos de calibração foi ainda avaliada utilizando-se relação de desempenho do desvio (RPD), que é definido como a razão entre o desvio padrão (SD) e o erro padrão de predição (SEP) (Shi; Yu, 2017). Chang et al. (2001) classificaram os modelos de calibração obtidos no NIRS em três categorias, utilizando RPD e  $R^2$ : categoria A ( $RPD > 2,0$ ;  $R^2 > 0,80$ ) - satisfatório; categoria B ( $1,4 < RPD < 2,0$ ;  $0,50 < R^2 < 0,80$ ) - modelo poderá discriminar baixo e altos valores da variável de resposta *screening*; e categoria C ( $RPD < 1,4$ ;  $R^2 < 0,50$ ) - sem habilidade de predição.

## Resultados e discussão

Os dados estatísticos obtidos na calibração, como médias, desvio padrão, erro padrão de calibração e coeficientes de determinação da calibração e da validação cruzada para os parâmetros de qualidade tecnológica de trigo, estão apresentados na Tabela 2.

**Tabela 2.** Dados de calibração obtidos no NIRS para peso do hectolitro, teores de umidade e de proteínas totais, índice de dureza, número de queda, força de glúten, estabilidade e glúten úmido em trigo moído. Embrapa Trigo, Passo Fundo, RS, 2018.

Parâmetro	Modelo de regressão	Pré-procesamento	Amostra n <sup>o(1)</sup>	Média do parâmetro	SD <sup>(2)</sup>	SEC <sup>(3)</sup>	R <sup>2(4)</sup>	SECv <sup>(5)</sup>	1-VR <sup>(6)</sup>	Termos (n <sup>o</sup> ) <sup>(7)</sup>	Bias <sup>(8)</sup>
Peso do hectolitro (kg/hL)	MPLS	SNV-D	6.585	79,6	3,32	1,82	0,70	1,84	0,69	12	-1,10
Teor de umidade (%)	MPLS	MSC	538	12,8	0,95	0,11	0,99	0,12	0,98	10	-0,07
Teor de proteínas totais (%)	MPLS	MSC	560	12,4	1,70	0,17	0,99	0,18	0,99	10	-0,11
Índice de dureza do grão	MPLS	MSC	6.505	72	15,03	4,78	0,90	4,84	0,90	14	-2,91
Número de queda do grão (s)	MPLS	Nenhum	6.600	342	109,25	68,10	0,61	68,62	0,61	14	-41,17
Força de glúten (x 10 <sup>-4</sup> J)	MPLS	SNV-D	6.487	220	76,85	54,31	0,50	54,78	0,49	13	-32,9
Estabilidade (min)	MPLS	SNV-D	1.550	7,2	4,58	3,30	0,48	3,43	0,44	12	-2,05
Glúten úmido (%)	MPLS	MSC	6.215	32,0	5,90	2,55	0,81	2,57	0,81	11	-1,54

<sup>(1)</sup> Número de amostras; <sup>(2)</sup> Desvio padrão; <sup>(3)</sup> Erro padrão de calibração; <sup>(4)</sup> Coeficiente de determinação da calibração; <sup>(5)</sup> Erro padrão da validação cruzada;

<sup>(6)</sup> Coeficiente de determinação da validação cruzada; <sup>(7)</sup> Número de termos; <sup>(8)</sup> Erro sistemático.

Na Tabela 3 são apresentados os dados estatísticos obtidos na validação externa: erro padrão de predição, erro sistemático, coeficiente de determinação e erro de desvio de desempenho.

**Tabela 3.** Dados de validação externa obtidos no NIRS para peso do hectolitro, teores de umidade e de proteínas totais, índice de dureza dos grãos, número de queda, força de glúten, estabilidade e glúten úmido em trigo moído. Embrapa Trigo, Passo Fundo, RS, 2018.

Parâmetro	Amostra (nº)	SEP <sup>(1)</sup>	Bias <sup>(2)</sup>	R <sup>2(3)</sup>	RPD <sup>(4)</sup>
Peso do hectolitro (kg/hL)	100	2,10	-0,21	0,73	1,58
Teor de umidade (%)	58	0,15	0,02	0,92	6,33
Teor de proteínas totais (%)	60	0,25	0,06	0,97	6,80
Índice de dureza do grão	100	5,35	-1,62	0,91	2,81
Número de queda do grão (s)	100	78,60	-26,16	0,72	1,39
Força de glúten ( $\times 10^{-4}$ J)	100	54,05	-1,58	0,67	1,42
Estabilidade (min)	100	3,40	-0,47	0,31	1,35
Glúten úmido (%)	100	2,69	0,054	0,81	2,19

<sup>(1)</sup> Erro padrão da predição; <sup>(2)</sup> Erro sistemático; <sup>(3)</sup> Coeficiente de determinação; <sup>(4)</sup> Erro de desvio de desempenho (RPD = SD/SEP).

Na análise do peso do hectolitro, foi utilizado o modelo MPLS 2.8.6.1 com pré-processamento SNV-D, resultando no coeficiente de determinação da calibração ( $R^2$ ) de 0,70 (Tabela 2). Na validação externa (Tabela 3), o coeficiente de determinação obtido foi de 0,73. De acordo com Dowell et al. (2006), esse resultado qualificou a equação de calibração para ser utilizada em *screening* de novas amostras.

O coeficiente de determinação da calibração ( $R^2$ ) para os teores de umidade e de proteínas totais foi de 0,99 (Tabela 2), indicando alta correlação linear entre resultados obtidos no NIRS e nos métodos de referência. Valores similares foram obtidos nos coeficientes de determinação na validação externa, com 0,92 e 0,97 (Tabela 3) para os teores de umidade e de proteínas totais, respectivamente. Nestas aplicações, o NIRS pode ser utilizado em análises de rotina, substituindo os métodos químicos de referência. Os baixos valores de SEC, SEP e alto  $R^2$  (tabelas 2 e 3), além da pouca diferença entre estes parâmetros, indicam que a predição de proteína e de umidade em trigo moído é estável e acurada (Jha, 2010). Shi e Yu (2017) reportaram que modelo de calibração desenvolvido no NIRS (comprimento de onda 1.400 nm – 2.500 nm) para proteína bruta em trigo apresentou excelente desempenho de predição ( $R^2 =$

0,97); em relação à umidade, o melhor modelo obtido no NIRS (comprimento de onda 1.100 nm – 2.500 nm) apresentou menor desempenho de predição ( $R^2 = 0,86$ ).

O NIRS mostrou adequada capacidade de predição para o parâmetro “índice de dureza”. Na calibração e na validação externa, os coeficientes de determinação  $R^2$  obtidos foram de 0,90 e 0,91, respectivamente. Os erros registrados na calibração SEC (4,78) e na validação cruzada SECV (4,84) são considerados baixos, do ponto de vista prático na determinação de dureza. Maghirang e Dowell (2003) utilizaram NIRS (550 nm - 1.690 nm) para avaliar índice de dureza em grãos de trigo e obtiveram coeficiente de determinação de 0,91 e erro na validação cruzada de 7,70. De acordo com os autores, na validação, o NIRS apresentou desempenho similar ao método de referência ( $R^2 = 0,88$ ), confirmando o potencial de aplicação na avaliação rápida e não destrutiva da dureza em trigo moído (Maghirang; Dowell, 2003).

Para o parâmetro “número de queda”, os coeficientes de determinação  $R^2$  obtidos na calibração e validação foram 0,61 e 0,72, respectivamente, indicando adequada correlação entre dados espectrais e químicos na construção do modelo, com possibilidade de viabilizar o uso para *screening* de lotes de trigo. Similar aos resultados obtidos neste trabalho, Risius et al. (2015) analisaram número de queda por meio de NIR *in line*, acoplado em colhedora, e obtiveram coeficiente de determinação de  $R^2 = 0,70$  e erro padrão da predição de 37 s. Os autores indicaram a viabilidade desta informação para otimizar a comercialização de trigo. Os resultados obtidos na validação externa para número de queda apresentaram discrepâncias quando comparados com a equação de calibração (que obteve menor  $R^2$ ). Assim, novas estratégias matemáticas e estatísticas serão aplicadas na avaliação do número de queda no NIRS. Com esta alternativa pretende-se, também, reduzir o erro padrão de predição (78,6 s) (Tabela 3), que foi elevado quando comparado com outros trabalhos.

Para os parâmetros “força de glúten” (W) e “estabilidade”, os valores de  $R^2$  na calibração foram 0,50 e 0,48, respectivamente (Tabela 2), indicando que o modelo de calibração não apresentou habilidade de predição destes parâmetros em trigo. A avaliação destes parâmetros qualitativos depende da formação de rede de glúten, limitando a análise acurada através do NIRS. Os coeficientes de determinação ( $R^2$ ) obtidos no presente estudo, para força de glúten (W), foram similares aos reportados por Arazuri et al. (2012). Na validação externa, os autores propuseram um indicador de erro menor que 15% na predição pelo NIRS, definindo assim a correta classificação das amostras. No presente estudo, os erros padrão de validação cruzada e de predição foram maiores. Assim, a redução do erro inerente ao método de referência e outras técnicas de estatísticas multivariada deverão ser testadas para aprimorar o desempenho de predição de força de glúten e estabilidade no NIRS.

Neste estudo, foram incluídas amostras comerciais e de experimentos com várias procedências, diversas cultivares de trigo, diferentes condições de armazenamento e com grande amplitude de resultados para “força de glúten” e “estabilidade” (Tabela 1). A redução na precisão é influenciada pela heterogeneidade das amostras (Williams, 2007). Outra dificuldade na predição de “força de glúten” e “estabilidade” são os métodos de referência, que não são suficientemente acurados para permitir a construção de modelos de calibração robustos (Osborne, 2007). Uma das principais limitações do NIRS está relacionada com o fato de ser um método secundário. A modelagem dos dados espectrais no NIRS depende da acurácia e da reprodutibilidade dos resultados obtidos no método de referência (Pasquini, 2003). A seleção de amostras, o tamanho da amostragem, a acurácia do método de referência e o tratamento estatístico são fatores decisivos na construção de modelos baseado em NIRS (Hell et al., 2016).

O NIRS forneceu adequada estimativa quanto aos teores de glúten úmido. O coeficiente de determinação  $R^2$  obtido na calibração e na validação foi de 0,81 e o erro padrão de predição foi de 2,69 (tabelas 2 e 3). Chen et al. (2017) utilizaram quimiometria, pré-tratamento dos espectros e intervalos de sinergia para melhorar o desempenho de predição do NIRS para os parâmetros glúten úmido e proteína total em trigo, obtendo coeficiente de determinação para glúten úmido de 0,85.

O erro de desvio de desempenho (RPD) é comumente utilizado para interpretar as calibrações, sendo que valores altos indicam melhor habilidade de predição (Porep et al., 2015). Com base neste indicador, as equações de calibração desenvolvidas poderão ser utilizadas para análise de novas amostras de trigo quanto ao teor de umidade (6,33), de proteína total (6,80), de índice de dureza dos grãos (2,81) e de glúten úmido (2,19). Os valores de RPD obtidos para os parâmetros peso do hectolitro (1,58) e número de queda (1,39) habilitam a equação de calibração para *screening* de novas amostras. Finalmente, de acordo com RPD, os parâmetros força de glúten e estabilidade não podem ser preditos pelo NIRS (Tabela 3).

Outra estatística útil é o coeficiente de determinação ( $R^2$ ). Representa, essencialmente, a proporção da variância explicada da variável de resposta no conjunto de calibração ou de validação (Cozzolino et al., 2011). Na proposta de Chang et al. (2001) classificaram-se os modelos de calibração em três categorias (satisfatório, moderado e sem habilidade de predição), de acordo com RPD e  $R^2$ . Assim, podem ser considerados satisfatórios os resultados preditos no NIRS para os parâmetros teor de umidade, de proteína total, de índice de dureza dos grãos e de glúten úmido. Os parâmetros peso do hectolitro, número de queda e força de glúten são classificados como moderados, adequados para *screening*. E, no caso da estabilidade, o NIRS não apresenta habilidade de predição (Tabela 3).



## Tendências

A espectroscopia NIR apresenta potencial para ser explorada na análise de qualidade de trigo, especialmente para os parâmetros teor de umidade, teor de proteínas, índice de dureza e glúten úmido, podendo ser utilizada em análises de rotina, substituindo os métodos químicos de referência. O NIRS pode, também, ser utilizado para *screening* em novas amostras, no caso de parâmetros peso do hectolitro e número de queda. Entretanto, apresenta limitações na predição de alguns parâmetros, como força de glúten e estabilidade, que são qualitativos (precisam que a rede de glúten seja formada antes de serem avaliados). As moléculas devem absorver radiação e apresentar vibrações características da sua composição, para serem identificadas por técnicas de infravermelho. Os parâmetros reológicos não são representados diretamente no espectro NIRS, pois resultam da composição proteica da amostra (Shiroma; Rodriguez-Saona, 2009; Arazuri et al., 2012). Quanto ao método de referência, uma prática fundamental na calibração do NIRS é a quantificação do erro intrínseco (Williams, 2007). O método de referência deve ser o mais acurado e preciso possível. Entretanto, em muitas análises de qualidade tecnológica de trigo, o erro associado ao método de referência não é conhecido. Os erros intrínsecos aos métodos de referência interferem diretamente no desempenho do NIRS. Para minimizar este efeito, poderão ser incluídos resultados de amostras analisadas em duplicata/triplicata, além de realizar a validação intralaboratorial e interlaboratorial para verificação dos resultados.

Adicionalmente, Williams (2007) identificou mais de 50 fatores que afetam as aplicações do NIRS em grãos, especialmente em trigo e derivados. Dentre os principais fatores que influenciam o espectro NIRS e limitam a utilização do modelo de calibração, destacam-se: região produtora, ano de produção, cultivar, composição e técnicas quimiométricas empregadas. Assim, refinamentos na prática de calibração deverão ser constantemente testados. De acordo com Shi e Yu (2017), tanto o pré-processamento dos espectros como a seleção do comprimento de onda afetam o desempenho dos modelos de calibração no NIRS. Desta forma, os modelos propostos poderão ser melhorados utilizando-se diferentes ferramentas estatísticas, ampliando as possibilidades de utilização do NIRS.

Outra tendência é o uso do equipamento NIR hiperespectral, que associa as funcionalidades do infravermelho próximo com imagens obtidas em número superior a 100 canais ou comprimentos de onda, gerando a definição de imagens hiperespectrais. Estas imagens representam a composição química a partir de informações espectrais. Com o conjunto de informações distribuídas em uma matriz tridimensional (imagem em x e y; comprimento de onda em z), é possível avaliar parâmetros físicos e químicos, considerando a distribuição espacial global ou por pixel de uma espécie ou

propriedade em diferentes produtos, em nível qualitativo, como coloração, formato, dimensão e impurezas, além de quantitativo em função da concentração ou teor dos analitos (Barbedo et al., 2015). Verdú et al. (2015) utilizaram NIR hiperespectral e evidenciaram que a quantidade de glúten e de suas frações proteicas influenciaram a resposta espectral, podendo ser diferenciada por meio de análise multivariada (PCA). O NIR associado a imagens hiperespectrais pode viabilizar análises *on line* do comportamento da farinha de trigo durante a elaboração de produtos finais. Outras informações podem ser obtidas nos Capítulos 7 e 8 desta publicação.

A disponibilidade de equipamentos NIRS portáteis também representa avanço na aplicação da técnica na avaliação de parâmetros de qualidade. Estas funcionalidades são exploradas no Capítulo 9 desta publicação. Um espectrômetro de fibra óptica em miniatura, associado com análise multivariada, foi eficiente no monitoramento, *in situ* e em tempo real, do processo enzimático de glúten em trigo (Zhang et al., 2017).

Outra aplicação da espectroscopia combinada com quimiometria é na classificação de lotes de grãos visando à rastreabilidade. O perfil espectral (*fingerprint*) de amostras é diferenciado de acordo com a região produtora, a genética, o ano de produção e os parâmetros de qualidade. Quando estes resultados são associados a um sistema de rastreabilidade, pode diferenciar a produção de acordo com diferentes características de interesse.

## Conclusões

Nesse capítulo, foi apresentada a aplicação da espectroscopia NIR como método rápido e não invasivo, para avaliação de parâmetros de qualidade tecnológica em trigo. A espectroscopia NIR apresenta potencial para ser utilizada em seleção de genótipos de trigo em programas de melhoramento genético e para orientar a segregação de trigo no recebimento de unidades armazenadoras, de acordo com os resultados dos principais componentes de qualidade tecnológica. Níveis de umidade, teores de proteína, índice de dureza e glúten úmido foram avaliados com adequada acurácia pela espectroscopia NIR. Para os parâmetros peso do hectolitro e número de queda, o NIRS pode ser utilizado para *screening*. O NIRS não apresentou habilidade de predição para os parâmetros força de glúten e estabilidade.

## Agradecimentos

Agradecemos a Rogerio Delanora, empregado da Embrapa Trigo, pela condução de análises no NIRS e pela colaboração nas análises estatísticas (calibração e validação). Também agradecemos a contribuição de Antônio Sergio Brisola de Oliveira, Helena de Araújo Andrade, Ellen Rogoski, Paulo Rocha Albuquerque e Pihetra Oliveira Tatsch, empregados da Embrapa Trigo, nas análises de referência de qualidade tecnológica.

## Referências

AACC INTERNATIONAL. **Method 38-12.02**: wet gluten, dry gluten, water-binding capacity, and gluten index. 11th ed. St. Paul, 2000. Approved methods of analysis, Nov. 8, 2000.

AACC INTERNATIONAL. **Method 44-15.02**: moisture-air-oven methods. 11th ed. St. Paul, 2010. Approved methods of analysis, May 28, 2010.

AACC INTERNATIONAL. **Method 46-13.01**: crude protein-Kjeldahl method, boric acid modification. 11th ed. St. Paul, 1999a. Approved methods of analysis, Nov. 3, 1999.

AACC INTERNATIONAL. **Method 54-30.02**: alveograph method for soft and hard wheat flour. 11th ed. St. Paul, 1999b. Approved methods of analysis, Nov. 3, 1999.

AACC INTERNATIONAL. **Method 54-60.01**: rheological behavior of flour by farinograph: constant flour weight procedure. 11th ed. St. Paul, 2011. Approved methods of analysis, Jan. 6, 2011.

AACC INTERNATIONAL. **Method 55-10.01**: test weight per bushel. 11th ed. St. Paul, 1999c. Approved methods of analysis, Nov. 3, 1999.

AACC INTERNATIONAL. **Method 55-31.01**: single-kernel characterization system for wheat kernel texture. 11th ed. St. Paul, 1999d. Approved methods of analysis, Nov. 3, 1999.

AACC INTERNATIONAL. **Method 56-81.03**: determination of falling number. 11th ed. St. Paul, 1999e. Approved methods of analysis, Nov 3, 1999.

ARAZURI, S.; ARANA, J. I.; ARIAS, N.; ARREGUI, L. M.; GONZALEZ-TORRALBA, J.; JAREN, C. Rheological parameters determination using near infrared technology in whole wheat grain. **Journal of Food Engineering**, v. 111, n. 1, p. 115, 121, 2012.

BARBEDO, J. G. A.; TIBOLA, C. S.; FERNANDES, J. M. C. Detecting Fusarium head blight in wheat kernels using hyperspectral imaging. **Biosystems Engineering**, v. 131, p. 65-76, Mar. 2015.

BLANCO ROMÍA, M.; ALCALÀ BERNARDEZ, M. Multivariate calibration for quantitative analysis. In: SUN, D.-W. (Ed.). **Infrared spectroscopy for food quality analysis and control**. Amsterdam: Academic Press, 2009. p. 51-82.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Instrução Normativa nº 38, de 30 de novembro de 2010. Regulamento técnico do trigo. **Diário Oficial da União**, 1 dez. 2010. Seção 1, p. 2.

CASSELS, J. A.; REUSS, R.; OSBORNE, B. G.; WESLEY, I. J. Near infrared spectroscopic studies of changes in stored grain. **Journal Near Infrared Spectroscopy**, v. 15, n. 3, p. 161-167, 2007.

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 America Journal**, v. 65, n. 2, p. 480-490, 2001.

CHEN, J.; ZHU, S.; ZHAO, G. Rapid determination of total protein and wet gluten in commercial wheat flour using siSVR-NIR. **Food Chemistry**, v. 221, p. 1939-1946, Apr. 2017.

COZZOLINO, D.; CYNKAR, W. U.; SHAH, N.; SMITH, P. Multivariate data analysis applied to spectroscopy: Potential application to juice and fruit quality. **Food Research International**, v. 44, n. 7, p. 1888-1896, 2011.

DE MORI, C. Aspectos comerciais e econômicos do trigo. In: DE MORI, C.; ANTUNES, J. M.; FAE, G. S.; ACOSTA, A. da S. (Ed.). **Trigo: o produtor pergunta, a Embrapa responde**. Brasília, DF: Embrapa, 2016. cap. 16, p. 297-309. (Coleção 500 perguntas, 500 respostas, n. 309).

DELWICHE, S. R.; NORRIS, K. H. Classification of hard red wheat by near-infrared diffuse reflectance spectroscopy. **Cereal Chemistry**, v. 70, n. 1, p. 29-35, 1993.

DOWELL, F. E.; MAGHIRANG, E. B.; XIE, F.; LOOKHART, G. L.; PIERCE, R. O.; SEABOURN, B. W.; BEAN, S. R.; WILSON, J. D.; CHUNG, O. K. Predicting wheat quality characteristics and functionality using near-infrared spectroscopy. **Cereal Chemistry**, v. 83, n. 5, p. 529-536, 2006.

FERNÁNDEZ-IBAÑEZ, V.; SOLDADO, A.; MARTÍNEZ-FERNÁNDEZ, A.; DE LA ROZA-DELGADO, B. Application of near infrared spectroscopy for rapid detection of aflatoxin B1 in

maize and barley as analytical quality assessment. **Food Chemistry**, v. 113, n. 2, p. 629-634, 2009.

GUARIENTI, E. M. **Qualidade industrial de trigo**. 2. ed. Passo Fundo: EMBRAPA-CNPQ, 1996. 36.p. (EMBRAPA-CNPQ. Documentos, 27).

HABA, M. J. DE LA; GARRIDO-VARO, A.; GUERRERO-GINEL, J. E.; PEREZ-MARÍN, D. C. Near-infrared reflectance spectroscopy for predicting amino acids content in intact processed animal proteins. **Journal of Agriculture and Food Chemistry**. v. 54, n. 20, p. 7703-7709, 2006.

HELL, J.; PRÜCKLER, M.; DANNER, L.; HENNIGES, U.; APPRICH, S.; ROSENAU, T.; BÖHMDORFER, S. A comparison between near-infrared (NIR) and mid-infrared (ATR-FTIR) spectroscopy for the multivariate determination of compositional properties in wheat bran samples. **Food Control**, v. 60, p. 365-369, Feb. 2016.

INFRAISOFT INTERNATIONAL LLC. **ISIScan™**: routine analysis software for FOSS near infrared instruments. Version 3.10. State College, PA, USA, 2008.

INFRAISOFT INTERNATIONAL LLC. **WinISI™ 4**: routine calibration software for FOSS near infrared instruments. Version 4.0. State College, PA, USA, 2007.

JESPERSEN, B. M.; MUNCK, L. Cereals and cereal products. In: SUN, D.-W. (Ed.). **Infrared spectroscopy for food quality analysis and control**. Burlington: Academic Press, 2009. p. 275-319.

JHA, S. N. Near infrared spectroscopy. In: JHA, S. (Ed.). **Nondestructive evaluation of food quality: theory and practice**. New York: Springer, 2010. p. 141-212.

JIRSA, O.; HRUŠKOVÁ, M.; ŠVEC, I. Near-infrared prediction of milling and baking parameters of wheat varieties. **Journal of Food Engineering**, v. 87, n. 1, p. 21-25, 2008.

KERSTING, H. J.; ZWINOELBERG, H. Spectroscopy: electromagnetic radiation for identifying and measuring the ingredients of food. In: POPPER, L.; SCHAFER, W.; FREUD, W. (Ed.). **Future of flour: a compendium of flour improvement**. Bergen-Dumme: Agrimedia, 2006. p. 160-168.

LONG, D. S.; ENGEL, R. E.; SIEMENS, M. C. Measuring grain protein concentration with in-line near infrared reflectance spectroscopy. **Agronomy Journal**, v. 100, n. 2, p. 247-252, 2008.

MAGHIRANG, E. B.; DOWELL, F. E. Hardness measurement of bulk wheat by single-kernel visible and near-infrared reflectance spectroscopy. **Cereal Chemistry**, v. 80, n. 3, p. 316-322, 2003.

OSBORNE, B. G. Flours and breads. In: OZAKI, Y.; MCCLURE, W. F.; CHRISTY, A. A. (Ed.). **Near-infrared spectroscopy in food science and technology**. Hoboken: John Wiley & Sons Inc., 2007. p. 281-296.

PASQUINI, C. Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. **Journal Brazilian Chemical Society**, v. 14, n. 2, p. 198-219, 2003.

POJIĆ, M. M.; MASTILOVIĆ, J. S. Near infrared spectroscopy-advanced analytical tool in wheat breeding, trade, and processing. **Food and Bioprocess Technology**, v. 6, n. 2, p. 330-352, 2013.

POREP, J. U.; KAMMERER, D. R.; CARLE, R. On-line application of near infrared (NIR) spectroscopy in food production. **Trends in Food Science & Technology**, v. 46, n. 2, p. 211-230, 2015.

RINNAN, A.; NORGAARD, L.; VAN DEN BERG, F.; THYGESEN, J.; BRO, R.; ENGELSEN, S. B. Data pre-processing. In: SUN, D.-W. (Ed.). **Infrared spectroscopy for food quality analysis and control**. Amsterdam: Academic Press, 2009. p. 29-48.

RISIUS, H.; HAHN, J.; HUTH, M.; TÖLLE, R.; KORTE, H. In-line estimation of falling number using near-infrared diffuse reflectance spectroscopy on a combine harvest. **Precision Agriculture**, v. 16, n. 3, p. 261-274, 2015.

SHENK, J. S.; WESTERHAUS, M. O. Population definition, sample selection, and calibration procedures for near-infrared reflectance spectroscopy. **Crop Science**, v. 31, n. 2, p. 469-474, 1991.

SHI, H.; YU, P. Comparison of grating-based near-infrared (NIR) and Fourier transform mid-infrared (ATR-FT/MIR) spectroscopy based on spectral preprocessing and wavelength selection for the determination of crude protein and moisture content in wheat. **Food Control**, v. 82 p. 57-65, Dec. 2017.

SHIROMA, C.; RODRIGUEZ-SAONA, L. Application of NIR and MIR spectroscopy in quality control of potato chips. **Journal of Food Composition and Analysis**, v. 22, n. 6, p. 596-605, 2009.

SIMMONDS, D. H. Chemical basis of hardness and vitreosity in the wheat kernel. **Bakers Digest**, v. 48, n. 5, p. 16-29, 1974.

TABELA 21. Indicações de características de qualidade por produto à base de trigo. In: REUNIÃO DA COMISSÃO BRASILEIRA DE PESQUISA DE TRIGO E TRITICALE, 10., 2016, Londrina. **Informações técnicas para trigo e triticale - safra 2017**. Brasília, DF: Embrapa, 2017. p. 43.

VERDÚ, S.; IVORRA, E.; SÁNCHEZ, A. J.; BARAT, J. M.; GRAU, R. Study of high strength wheat flours considering their physicochemical and rheological characterisation as well as fermentation capacity using SW-NIR imaging. **Journal of Cereal Science**, v. 62, p. 31-37, Mar. 2015.

WAES, C. V.; MESTDAGH, I.; LOOTENS, P.; CARLIER, L. Possibilities of near infrared reflectance spectroscopy for the prediction of organic carbon concentrations in grassland soils. **Journal of Agricultural Science**, v. 143, n. 6, p. 487-492, 2005.

WILLIAMS, P. Grains and seeds. In: OZAKI, Y.; MCCLURE, W. F.; CHRISTY, A. A. (Ed.). **Near-infrared spectroscopy in food science and technology**. Hoboken: John Wiley & Sons Inc., 2007. p. 165-217.

XIE, F.; DOWELL, F. E.; SUN, X. S. Comparison of near-infrared reflectance spectroscopy and texture analyzer for measuring wheat bread changes in storage. **Cereal Chemistry**, v. 80, n. 1, p. 25-29, 2003.

ZHANG, Y.; LUO, L.; LI, J.; LI, S.; QU, W.; MA, H.; YE, X. In-situ and real-time monitoring of enzymatic process of wheat gluten by miniature fiber NIR spectrometer. **Food Research International**, v. 99, p. 147-154, Sept. 2017.

ZHANG, P.; LITTLEJOHN, D. Interference assessment and correction in the partial least squares regression method for multicomponent determination by UV spectrophotometry. **Chemometrics and Intelligent Laboratory Systems**, v. 34, n. 2, p. 203-215, Sep. 1996.

ZIEGLER, J. U.; LEITENBERGER, M.; LONGIN, C. F. H.; WÜRSCHUM, T.; CARLE, R.; SCHWEIGGERT, R. M. Near-infrared reflectance spectroscopy for the rapid discrimination of kernels and flours of different wheat species. **Journal of Food Composition and Analysis**, v. 51, p. 30-36, Aug. 2016.

# Controle de Qualidade para Espectroscopia no Infravermelho Próximo

Gilberto Batista de Souza  
Ana Rita de Araujo Nogueira  
Igor Renato Bertoni Olivares

## Introdução

O conhecimento da composição química de uma matriz, para o estabelecimento de suas condições de estocagem ou para sua utilização nos mais variados sistemas de produção, está diretamente relacionado com suas características físico-químicas. Neste contexto, é importante que a caracterização química, tanto a qualitativa quanto a quantitativa, seja realizada com alta velocidade analítica e produza resultados confiáveis e com rastreabilidade metrológica. Para atender aos requisitos descritos, a técnica analítica de espectroscopia no infravermelho próximo (NIRS) tem se mostrado altamente eficiente, principalmente em termos de precisão e frequência analítica, em substituição aos métodos clássicos de análises laboratoriais, nos diversos setores, como em análises de alimentos e produtos agropecuários, dentre outros, viabilizando a avaliação de vários parâmetros de forma simultânea, em diversos tipos de matrizes (Pasquini, 2018).

Para que a técnica analítica de NIRS seja implantada em um laboratório, é necessário o desenvolvimento de modelos de calibração multivariada, os quais requerem procedimentos iniciais que envolvem estratégias de preparo de amostras, em que sejam previstos cuidados nas etapas de secagem e moagem, prevendo a temperatura de secagem adequada para cada tipo de matriz e o perfil do tamanho de partículas. Na seleção das amostras do conjunto de calibração também deve estar prevista a inclusão de amostras que abranjam a maior variabilidade possível na composição físico-química, considerando a similaridade dentro de uma determinada espécie ou família botânica.

O número de amostras utilizadas para a calibração é considerado de grande importância e, talvez, ainda mais importante do que a própria técnica quimiométrica empregada para o desenvolvimento do modelo. As normas ASTM E1655-05 para o desenvolvimento de análises quantitativas, usando espectrometria NIR e análise multivariada, tratam de recomendações relativas à elaboração das amostras dos conjun-



tos de calibração e validação e a definição dos parâmetros estatísticos para avaliar, otimizar e recalibrar modelos de calibração, entre outras (ASTM Standards, 2000).

Na ASTM E1655-05 é citado que, se o modelo multivariado é desenvolvido usando três ou menos variáveis ( $k$ ), então o conjunto de calibração deve conter um mínimo de 24 amostras, depois da eliminação das amostras *outliers*. No caso de o modelo multivariado ter sido desenvolvido usando  $k(>3)$  variáveis, então o conjunto de calibração deve conter um mínimo de  $6k$  espectros depois da eliminação das amostras *outliers*. Por fim, se o modelo é centrado na média, um mínimo de  $6(k + 1)$  espectros deve restar após a eliminação das amostras *outliers* (ASTM Standards, 2000).

No entanto, esse ponto é relativo, pois a quantidade de amostras a ser analisada para que o método seja validado depende da complexidade das amostras e das propriedades que estão sendo utilizadas. A recomendação do número de amostras para a construção dos modelos de calibração, para amostras naturais com muitos atributos, geralmente está na faixa de 50 a 100 amostras, dependendo da complexidade e variabilidade da matriz que acompanha as amostras reais. Por exemplo, na construção de um modelo para avaliar o teor de proteína de grãos de trigo em geral, todas as cultivares de trigo devem estar representadas no conjunto de calibração, sendo que, para cada classe de trigo, a concentração de proteínas deve abranger os valores esperados (Pasquini, 2003).

Após os procedimentos iniciais, os quais estão relacionados às etapas de preparo e seleção de amostras, é dado seguimento à construção dos modelos de calibração. Essa próxima etapa é conhecida como validação analítica do modelo e prevê avaliar algumas figuras de mérito multivariadas. As figuras de méritos previstas na validação de métodos analíticos baseadas em espectrometria NIR estão previstas na ASTM E1655-05 (ASTM Standards, 2000), sendo que, além das figuras de mérito, da linearidade e da faixa de trabalho, outras também devem ser avaliadas, considerando-se como as principais:

- **exatidão**, que é avaliada por meio da raiz quadrada do erro médio de predição (RMSEP – *Root Mean Square Error of Calibration*) (Equação 6) (Valderrama et al., 2009);
- **precisão**, que é a medida do erro aleatório associado ao método, em termos do desvio padrão absoluto ou relativo (são avaliados três níveis de precisão: repetibilidade, precisão intermediária e reprodutibilidade) (Valderrama et al., 2009);
- **robustez**, que avalia a sensibilidade do método às variações nas condições analíticas;

- **viés** (*bias*), indica a presença de erros sistemáticos, que são calculados pela diferença entre a média da população e o valor verdadeiro. Seus componentes são procedentes de efeitos não aleatórios. Para verificar efeitos significativos no valor do viés, a ASTM E1655-05 preconiza que seja aplicado um teste t Student ( $t_{viés}$ ) e, caso o valor de  $t_{viés}$  apresentar valor inferior ao valor de  $t$  tabelado (crítico), para  $P-1$  graus de liberdade com 95% de confiança, é possível afirmar a inexistência de efeito sistemático significativo no modelo de calibração (equações 1 e 2) (ASTM Standards, 2000; Valderrama et al., 2009);

$$viés = \frac{\sum_{P=1}^P (y_P - \hat{y}_P)}{P} \quad (1)$$

$$t_{viés} = \frac{|viés|\sqrt{P}}{SEP} \quad (2)$$

Onde:

P: número de amostras do conjunto teste;

$y_P$  e  $\hat{y}_P$ : valores de referência e valores previstos pelo modelo, respectivamente;

$t_{viés}$ : distribuição t Student;

SEP: Erro Padrão de Previsão (*Standard Error of Prediction or of Performance*)

- **sensibilidade**, que corresponde à fração do sinal responsável pelo acréscimo de uma unidade de concentração à propriedade de interesse, ou seja, definida como o sinal analítico líquido (NAS - *Net Analytical Signal*) gerado por uma concentração do analito igual à unidade (Valderrama et al., 2009);
- **sensibilidade analítica**, é a sensibilidade em termos da unidade de concentração que é utilizada, sendo definida como a razão entre a sensibilidade do método e o desvio padrão da resposta instrumental do branco (Valderrama et al., 2009; Ferreira, 2015);
- **seletividade**, é uma medida do quanto o sinal da espécie de interesse se encontra sobreposto ao sinal dos interferentes, fornecendo assim uma descrição da parte do sinal medido que é útil para previsão;
- **limite de detecção** (LD), é a menor quantidade do analito que pode ser detectada quantitativamente, sendo que, para dados centrados na média, pode ser calculado considerando o desvio padrão do sinal do branco ( $\sigma_{br}$ ) e o módulo dos vetores de regressão ( $||b||$ ) (equação 3) (Ferreira, 2015);

$$LD = 3,29 \times \sigma_{br} \times \|b\| \quad (3)$$

- **limite de quantificação (LQ)**, é usado para definir o limite inferior de uma medida quantitativa precisa e para métodos multivariados, centrados na média; pode ser obtido por meio da razão entre o desvio padrão do sinal do branco ( $\sigma_{br}$ ) e o módulo dos vetores de regressão ( $\|b\|$ ) (equação 4) (Ferreira, 2015).

$$LQ = \frac{10 \times \sigma_{br}}{\|b\|} \quad (4)$$

Outras figuras de mérito também devem ser utilizadas para avaliar a qualidade e a confiabilidade dos modelos de regressão multivariada, os quais são parâmetros estatísticos. Dessa forma, são estimados com base na análise dos resíduos entre o valor previsto ( $y_p$ ) e o valor referência ( $\hat{y}_p$ ), sendo sintetizados através da soma quadrática desses resíduos (PRESS - *Predicted Residual Error Sum of Square*) (equação 5). A partir da raiz quadrática do PRESS dividida pelo número de amostras do conjunto de calibração ( $P$ ), obtém-se o erro médio de previsão (RMSEP - *Root Mean Square Error of Prediction*) (equação 6).

$$PRESS = \sum_{P=1}^P (y_p - \hat{y}_p)^2 \quad (5)$$

$$RMSEP = \sqrt{\frac{PRESS}{P}} \quad (6)$$

Considerando-se o erro padrão de previsão (SEP), este pode ser estimado a partir da raiz quadrática da soma quadrática dos resíduos de previsão menos o viés, dividido pelo número de amostras do conjunto de calibração menos um ( $P-1$ ) (equação 7). A partir dos resíduos do conjunto de amostras de calibração, para dados centrados na média, estima-se o erro padrão de calibração (SEC - *Standard Error of Calibration*), o qual é calculado por meio da raiz quadrática da somatória quadrática dos resíduos do conjunto de amostras de calibração ( $y_i - \hat{y}_i$ ), onde o número de graus de liberdade no denominador é o número de amostras ( $I$ ) subtraído do número de parâmetros no modelo ( $A$ ), menos 1 (equação 8).

$$SEP = \sqrt{\frac{\sum_{P=1}^P (y_P - \hat{y}_P - viés)^2}{P - 1}} \quad (7)$$

$$SEC = \sqrt{\frac{\sum_{i=1}^I (y_i - \hat{y}_i)^2}{I - A - 1}} \quad (8)$$

No modelo de regressão, cuja opção foi por meio da validação interna cruzada, são consideradas a soma quadrática dos resíduos ( $PRESS_{val}$ ) (equação 9), raiz quadrada do erro médio quadrático de validação cruzada (RMSECV) calculada a partir da raiz quadrada do  $PRESS_{val}$  dividido pelo número de amostras usadas na validação interna cruzada ( $I$ ) (equação 10).

$$PRESS_{val} = \sum_{i=1}^I (y_i - \hat{y}_i)^2 \quad (9)$$

$$RMSECV = \sqrt{\frac{PRESS_{val}}{I}} \quad (10)$$

Para que a técnica de NIRS seja implantada em laboratórios de rotina, se faz necessário o monitoramento contínuo, para verificar a estabilidade da calibração por meio de métodos de avaliação da precisão quanto à repetibilidade, à precisão intermediária e à verificação da presença de efeitos sistemáticos significativos, os quais prejudicam a exatidão dos resultados.

Conforme previsto na ABNT NBR ISO/IEC 17025:2017, item 7.7 – garantia da validade dos resultados, é mandatório que o laboratório tenha um procedimento para monitorar a validade dos resultados (item 7.7.1) e deve também monitorar o seu desempenho por meio de comparação com resultados de outros laboratórios (item 7.7.2). Dessa forma, para atender aos requisitos citados, o laboratório deve participar em rodadas de programas de ensaios de proficiência por comparação interlaboratorial, utilizar materiais de referência ou materiais para controle de qualidade (amostras padrão) e monitorar a rotina com o uso de cartas de controle do processo analítico, sendo fundamental ter ações preventivas e corretivas para garantir a qualidade dos resultados

gerados no laboratório (Associação Brasileira de Normas Técnicas, 2017a). Dessa forma, esse capítulo abordará algumas ferramentas aplicadas na garantia dos resultados para que os usuários da técnica de NIRS possam utilizá-las em suas rotinas analíticas.

## Controle de qualidade por meio de comparação intralaboratorial

A inerente reprodutibilidade ou imprecisão de um método deve ser determinada como parte de um procedimento de validação. Esta informação pode, então, ser aplicada em um programa de controle de qualidade intralaboratorial (IQC – *Internal Quality Control*), que é projetado para verificação de variação ou viés (*bias*) e/ou na alteração da reprodutibilidade de um método. O IQC garante que fatores determinados pela incerteza permaneçam constantes durante o uso em rotina de um método analítico durante longos períodos de tempo. Junto com a validação, o IQC forma a base da prática de qualidade em uma análise química. A validação compreende a estimativa de incerteza de um resultado a partir do emprego do método analítico em uma dada condição. Os analistas são os responsáveis pela avaliação a respeito da adequação do método, em razão do conhecimento da incerteza, se esta é adequada ao propósito pretendido. O IQC constitui o procedimento que avalia se a incerteza obtida a partir da validação mantém-se constante durante o uso em rotina (Thompson, 2010).

Um programa de controle de qualidade intralaboratorial pode envolver vários elementos, mas o mais comum é o uso de um analito (ou analitos) com concentração conhecida em cada lote de amostra. Os materiais controle são tratados da mesma forma que as amostras e os resultados são colocados em cartas de controle.

**Importância/razão da necessidade da rastreabilidade metrológica:** desde que a primeira versão da ISO/IEC 17025 foi introduzida em 1999, houve aumento na atenção dada a medidas de incerteza e rastreabilidade a uma referência padrão, tanto em métodos químicos quanto microbiológicos. É cada vez maior o número de laboratórios acreditados. Dentre os requisitos de acreditação está a documentação dos procedimentos, sendo o IQC parte importante desta documentação. Quando os laboratórios estimam as medidas de incerteza, os resultados do IQC são essenciais e a rastreabilidade e a determinação de incertezas são parte integrante da validação dos métodos (Christensen, 1996).

A validação dos métodos é obrigatória, de acordo com a ISO/IEC 17025:2017 (Associação Brasileira de Normas Técnicas, 2017). Apesar de esta validação ser aplica-

da para métodos não normalizados, a verificação/confirmação de desempenho dos métodos normalizados também é obrigatória, pois, conforme descrito pela norma de acreditação, é necessária a confirmação de que o laboratório tem condição de operar adequadamente métodos normalizados antes de implantar os ensaios ou as calibrações (Manual, 2014). Para validação de métodos, existem protocolos, para diferentes áreas de aplicação e tipos de análises. Dentre esses protocolos podem ser citados o ICH (International Conference on Harmonization of Technical Requirements, 2005) e a European Union (2007), além de documento orientativo elaborado pelo Instituto Nacional de Metrologia-Inmetro (Inmetro, 2016a). Para a estimativa de incerteza de medição, dois guias podem ser destacados, o EURACHEM/CITAC (Ellison; Williams, 2012) e a ISO GUM (International Organization for Standardization, 2008), de grande aplicação nos laboratórios de ensaio e calibração.

**Uso de materiais de referência:** é importante que as figuras de mérito de um procedimento sejam sistematicamente determinadas, o que implica que o usuário tenha confiança em relação a qualquer resultado obtido a partir do método utilizado. De uma maneira geral, a verificação do trabalho em laboratório e a avaliação das necessidades do cliente são essenciais, assim como é claro o papel da documentação baseada nos resultados obtidos com materiais de referência e a qualidade analítica (International Organization for Standardization, 2015a).

Quando se realiza uma medição química de uma amostra real, deve-se comparar o resultado obtido com um valor designado de um material conhecido, ou seja, de um material de referência certificado (CRM – *Certified Reference Material*). Os valores das propriedades (geralmente a concentração do analito de interesse) nos CRMs são suficientemente homogêneos e bem estabelecidos. Desta forma, torna-se possível a identificação e a quantificação da substância de interesse. Os CRMs também são utilizados na etapa de validação dos métodos e em outras etapas dos processos necessárias para garantir o IQC, tais como a estimativa da incerteza de medição, treinamento de pessoal e no controle estatístico de processo, com o emprego de cartas de controle (International Organization for Standardization, 2017).

Um material de referência (RM – *Reference Material*) é um material, suficientemente homogêneo e estável em relação a uma ou mais propriedades específicas, que é estabelecido como apto para o seu uso em um processo de medição. Quando os valores das propriedades do RM são rastreáveis a um valor exato da unidade pelo Sistema Internacional de Unidades (SI), têm valores de incerteza para um nível de confiança estabelecido e estão acompanhados por documentação emitida por uma entidade reconhecida, o RM é denominado “material de referência certificado”, CRM (Joint Committee for Guides in Metrology, 2012).

Segundo descrição do Ministério da Agricultura, Pecuária e Abastecimento-Mapa (Manual, 2014), na validação de métodos analíticos devem ser utilizados materiais/padrões de referência certificados, produzidos por produtores acreditados na ABNT NBR ISO 17034:2017 (Associação Brasileira de Normas Técnicas, 2017b). Caso os materiais/padrões certificados não estejam disponíveis, podem ser empregados materiais de referência internos, amostras fortificadas ou ainda amostras com valores conhecidos. A utilização de material/padrão de trabalho é permitida, desde que o material/padrão de trabalho seja, devidamente, rastreável a um material/padrão de referência, que por sua vez tenha sido calibrado com um CRM. O RM deve ser o ponto final da cadeia de rastreabilidade nos laboratórios usuários. Todo RM deve ser rastreável a uma referência estabelecida, independentemente do uso pretendido do material, mas é o usuário quem define qual é o nível de incerteza apropriado para cada processo analítico em particular e, portanto, qual é o material adequado para o seu processo de medição (Zschunke, 2000).

Os CRMs são utilizados para fornecer resposta imediata a respeito de um resultado que está sendo apresentado após uma corrida analítica ou um ensaio intralaboratorial. Em termos objetivos, possibilitam o acesso à qualidade dos resultados. Dentro deste contexto de garantia da qualidade, os materiais de referência são utilizados para a caracterização dos métodos, do controle intralaboratorial e do controle interlaboratorial. A seguir, serão apresentados os detalhes referentes ao controle de qualidade intralaboratorial (IQC), que envolve a caracterização dos métodos e o controle de qualidade interno.

**Caracterização dos Métodos:** o desenvolvimento de um novo método analítico deve representar avanço em relação aos métodos existentes e ser efetivo para o propósito pretendido, com a análise de espécies com concentrações conhecidas, isto é, materiais de referência certificados (CRMs), cujos resultados são apresentados para o escrutínio da comunidade científica. Fica então demonstrado, para aqueles que tenham outros interesses, a possibilidade de utilizar este novo procedimento para suas próprias aplicações. Este tipo de trabalho representa um dos mais visíveis usos para os CRMs (Cave et al. 1999; Olivares et al., 2018).

Outra aplicação na qual os CRMs são essenciais refere-se à introdução no laboratório de um novo método ou a transferência de um método pré-estabelecido ou, ainda, de uma nova instrumentação. Mesmo quando as condições analíticas são padronizadas, como no caso específico da espectrometria de infravermelho próximo (NIRS), que muitas vezes possui curvas de calibração fornecidas pelo fabricante, é essencial a validação das curvas com amostras semelhantes às que estão sendo analisadas, as quais devem ter resultados conhecidos dos analitos que estão sendo analisados. O

trabalho de validação é necessário, pois os dados devem ser documentados para o propósito de garantia de qualidade, por exemplo, acreditação com base no IQC. Isso é importante como referência futura, caso sejam necessárias investigações relacionadas a problemas com equipamento, reagentes ou mão de obra.

O trabalho de validação inclui a determinação dentro do lote e entre diferentes lotes de amostra, os desvios padrão, a linearidade, a sensibilidade e os limites de detecção e quantificação. Para demonstrar a exatidão do método, para investigar possíveis interferências e estabelecer faixas de trabalho, é necessário analisar-se um CRM com concentrações conhecidas, sendo ideal de matriz o mais próximo possível da amostra que se pretende analisar.

A resposta se um método está adequado para uso na rotina do laboratório no ensaio de CRM pode ser realizada considerando o resultado obtido mais o seu limite de confiança ou incerteza de medição, comparando com os valores declarados no certificado do CRM (devendo haver uma intersecção entre as faixas de resultados), conforme descrito na Equação 11:

$$\bar{x} \pm \frac{t_{n-1} \times s}{\sqrt{n}} \quad (11)$$

Avaliação do valor do CRM (Miller; Miller, 2005)

Onde:

$\bar{x}$ : média das repetições referentes à análise do CRM;

$t$ :  $t$  de Student para  $n$  números de repetições;

$n$ : número de repetições (no mínimo 5);

$s$ : desvio padrão para número  $n$  de repetições.

Face ao exposto, fica evidente a enorme importância dos CRMs na garantia da qualidade dos resultados analíticos, porém seu uso ainda está longe de ser difundido nos laboratórios de ensaios rotineiros, visto que muitas vezes não é possível dispor de um RM de matriz semelhante à das amostras em estudo ou com níveis de concentração similares.

Para a execução de um controle de qualidade intralaboratorial, os laboratórios podem preparar seu próprio material de referência. Com planejamento cuidadoso e atenção a detalhes, materiais de referência extremamente úteis podem ser preparados como descrito pela norma ISO 80 (International Organization for Standardization, 2014). A introdução de práticas referentes ao preparo de RMs apropriados para o uso



pretendido seguindo os protocolos estabelecidos pela ISO 80 representa um ganho de qualidade nos resultados. De acordo com esta norma, o material de referência preparado para o controle de qualidade intralaboratorial é denominado “Material para Controle de Qualidade” (QCM – *Quality Control Material*). O preparo do QCM segue todo o protocolo descrito para o preparo de um CRM, que está detalhado e pode ser consultado nas normas ABNT NBR ISO 17034:2017 (Associação Brasileira de Normas Técnicas, 2017b) e ISO GUIDE 35:2017 (International Organization for Standardization, 2017). Muitos laboratórios utilizam amostras reais, semelhantes às utilizadas nas análises de rotina. Exemplos de como o material e os dados são processados são apresentados no anexo da ISO Guia 80:2014 (International Organization for Standardization, 2014).

Na hora da escolha de um material de referência, há vários aspectos a ser levados em conta. Deve ser o mais semelhante possível (tipo de matriz e nível de concentração) à amostra a ser analisada, para que abarque todos os problemas analíticos que possam causar erros nas medições. Também deve-se considerar a homogeneidade, a estabilidade e a incerteza dos valores das propriedades de interesse. A utilidade de um RM vai depender de cada processo analítico em particular e o usuário deve decidir que aspectos são os mais importantes a levar em conta. Por exemplo, para alguns casos em que o efeito matriz seja muito significativo, pode ser mais importante utilizar um material da mesma matriz que as amostras, mas com uma incerteza mais alta nos valores das propriedades. Como o material será utilizado no próprio laboratório, requisitos exigidos para o preparo de CRMs não são tão necessários, como a avaliação da estabilidade referente ao transporte do material.

O emprego do QCM inclui, dentre outras aplicações:

- elaboração de cartas de controle: para demonstrar o comportamento de uma medição interna realizada pelo laboratório ou para confirmar a efetividade do IQC, ou para demonstrar a variação de uma medida em certo período de tempo;
- comparação dos resultados: por exemplo, de uma ou duas séries de amostras relacionadas tanto em períodos curtos quanto mais longos, quando é sabido que os resultados podem variar;
- desenvolvimento de métodos: para o estabelecimento da consistência das medidas (para validar um material certificado de referência que deve ser utilizado);
- para checar o desempenho de um instrumento;

- em estudos de repetibilidade e reprodutibilidade: que devem ser repetidos a cada período de tempo, para verificar a reprodutibilidade das medidas, a robustez das medidas e o desempenho dos instrumentos e analistas, etc.

A consequência do sucesso de um programa de Controle de Qualidade Intralaboratorial (IQC) é uma variação mínima entre uma medida e outra. Com a estabilidade analítica, variações temporais como período do dia ou sazonais, podem ser facilmente identificadas. Da mesma forma, o controle ou a faixa de variação pode ser confirmado. Assim, mesmo pequenas variações de concentração serão identificadas e mudanças cruciais, a exemplo de falha de algum processo ou efeito de uma doença em uma cultura ou em um paciente, serão diagnosticadas (Cotlove et al. 1970; Thompson; Wood, 1995).

Quando confirmado que um procedimento de medida está estatisticamente controlado (International Organization for Standardization, 1993, 2012, 2013, 2014), o desempenho do laboratório é conhecido e pode ser obtido comparando tanto o desvio padrão quanto a faixa aceitável dos resultados individuais obtidos pelo MQC em relação a um critério pré-estabelecido. Se uma variação fora de controle for identificada, o laboratório deve tomar ação para a correção. Nos casos mais simples, pode haver a necessidade de repetição da medida “suspeita”, talvez seguida de uma recalibração dos instrumentos.

A reprodutibilidade de um método ou o intervalo de confiança atribuído a um QCM permite ao analista conhecer a faixa aceitável de trabalho. Uma faixa aceitável de  $\pm 2$  desvios padrão é normalmente aplicada e um procedimento prático é o uso de cartas de controle, inicialmente propostas por Walter Andrew Shewhart (1891 - 1967). Uma série de regras de utilização para o uso das cartas de controle de Shewhart foi desenvolvido por Westgard e colaboradores (Westgard et al., 1981). Nessas regras estão compreendidos o número e a concentração dos materiais de referência a serem incluídos, a frequência com que deverão ser incluídos em uma corrida e como os resultados deverão ser aplicados para a tomada de decisão, ou seja, se uma corrida deverá ser aceita ou rejeitada. No entanto, apesar de formalmente adotar as regras de Westgard, muitos laboratórios desenvolvem seus próprios critérios, principalmente quando não têm acesso a um razoável número de materiais certificados.

De maneira resumida, o IQC envolve práticas que certifiquem que os erros nos dados analíticos gerados estejam dentro de uma variação aceitável. Na prática, o IQC depende do uso de duas estratégias, a análise do QCM ou CRM para monitorar a veracidade dos resultados com o controle estatístico e o número de repetições independentes, como forma de se conhecer a precisão das medidas (Thompson; Wood, 1995).

Para mais detalhes a respeito dos procedimentos, pode ser consultado o documento produzido pelo Mapa - Manual de Garantia da Qualidade Analítica - Áreas de Identidade e Qualidade de Alimentos e de Insumos (Manual, 2015), que apresenta detalhadamente as definições e os cálculos empregados em um programa de controle de qualidade intralaboratorial.

## Controle estatístico de processo analítico – Cartas de controle

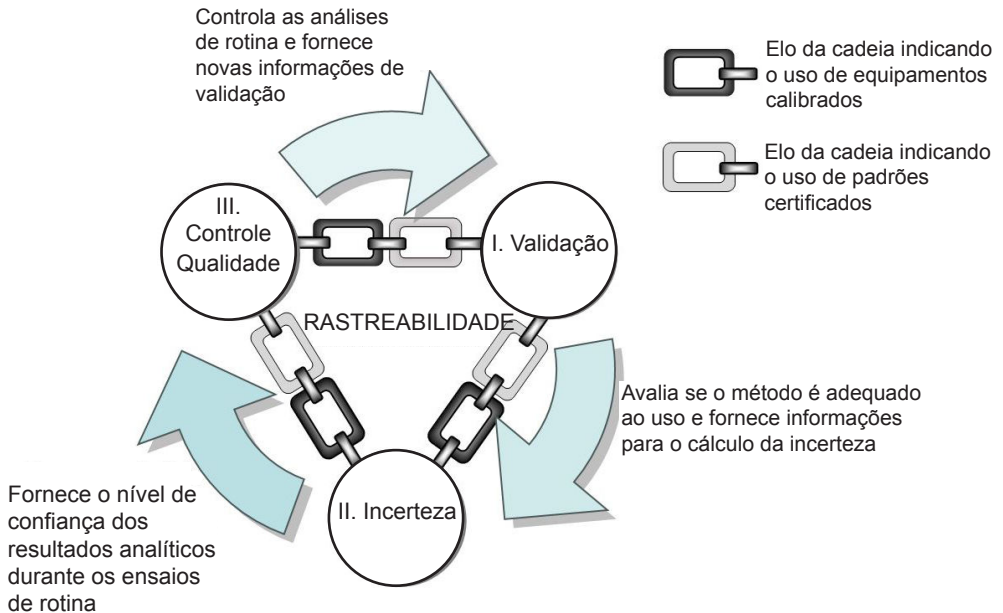
Uma carta de controle é definida como um gráfico com limites de controle em que uma medida estatística de uma série de amostras é representada em uma ordem particular para orientar o processo em relação a essa medida (International Organization for Standardization, 2014). Também pode ser definida como Carta de Controle de Shewhart: gráfico de controle com limites de controle Shewhart destinados principalmente a distinguir entre a variação na medida plotada em razão de causas aleatórias e aquelas por causas especiais (International Organization for Standardization, 2013).

As cartas de controle atuam no controle de qualquer tipo de processo, sendo incorporadas à realidade dos laboratórios principalmente após a publicação da norma ISO/IEC 17025 em 1999. Entre os requisitos desta norma, o item 5.9 estabelece que controles de qualidade devem ser aplicados pelo laboratório, sendo os resultados destes controles apresentados de maneira que as tendências sejam detectadas, desta maneira, as cartas de controle sempre foram uma das melhores opções para avaliação desta tendência. No ano de 2017, a norma ISO/IEC 17025, em sua 3ª edição, incorporou o termo carta de controle no texto da própria norma, indicando, no requisito 7.7.1 d), quanto ao uso de “padrões de checagem ou padrões de trabalho com cartas de controle, quando aplicável” (Associação Brasileira de Normas Técnicas, 2017).

A aplicação das cartas de controle alinha-se diretamente com a ideologia do Ciclo de Qualidade Analítica (Olivares; Lopes, 2012), conforme apresentado na Figura 1.

Neste ciclo, destaca-se inicialmente que um método de ensaio deve ser avaliado para verificar se ele é adequado ao uso e fornece resultados confiáveis, sendo aplicado à validação do método (aplicado a métodos novos ou normalizados utilizados fora de seu escopo), ou à verificação de desempenho de um método normalizado, para verificar se o método normalizado fornece resultados confiáveis nas condições operacionais do laboratório (a verificação de desempenho pode ser entendida como uma validação simplificada). Em seguida, com base nos resultados da validação ou da verificação de desempenho, é necessário estimar a incerteza de medição, visando

a avaliar os limites de confiança dos resultados analíticos. Finalmente, na rotina de ensaios do laboratório, é necessário avaliar se o método continua produzindo resultados adequados e se a incerteza estimada é pertinente. Desta maneira, amostras de controle (que apresentam características conhecidas como a concentração de um determinado analito) são analisadas durante a rotina de ensaios, e os resultados são representados nas cartas de controle visando a avaliar as tendências dos resultados.



**Figura 1.** Ciclo de Qualidade Analítica.

Fonte: Olivares e Lopes (2012).

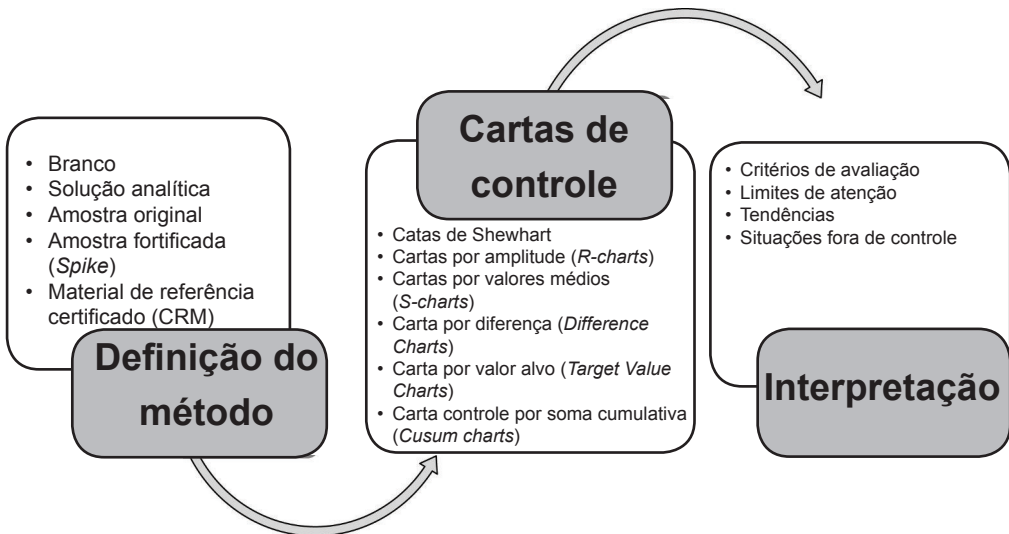
De acordo com a ISO 7870-1:2014 (International Organization for Standardization, 2014), as cartas de controle são um instrumento fundamental de controle estatístico de processo (CEP) que fornecem um método gráfico simples que pode ser usado para:

- a) indicar se o processo é estável, operando dentro de um sistema estável de causas aleatórias;
- b) estimar a magnitude da variabilidade inerente do processo;

- c) comparar a informação a partir de amostras que representam o estado atual de um processo em relação aos limites de controle que refletem essa variabilidade, com o objetivo de determinar se a variabilidade do processo manteve-se estável ou se está reduzindo ou aumentado;
- d) identificar, investigar e, possivelmente, reduzir/eliminar o efeito de causas especiais de variação, que podem levar o processo a um nível inaceitável de desempenho;
- e) ajudar na regulação de um processo através da identificação de padrões de variabilidade, como tendências, ciclos, etc.

Uma grande vantagem das cartas de controle é a sua facilidade de construção e utilização. No entanto, para que uma carta de controle possa ser um indicador confiável e eficiente, deve-se ter atenção especial na fase de planejamento quanto às questões de como selecionar o tipo adequado de carta para o processo em estudo e como determinar esquema de amostragem adequada, bem como a escolha da amostra controle (Turuta, 2015).

Existem diferentes tipos de cartas de controle aplicáveis na garantia da qualidade analítica para monitoramento de alguns dos parâmetros, como valor único, valor médio, taxa de recuperação, desvio padrão e alcance. Essas ferramentas são utilizadas em concordância com a amostra controle pré-estabelecida para uma metodologia específica (Figura 2).



**Figura 2.** Tipos de amostras de controle, cartas de controle e interpretação.

Fonte: Funk et al., (2007)

Considerando que as cartas de controle de Shewhart são as mais aplicadas em laboratórios de ensaio, um guia prático para sua aplicação pode ser encontrado no Manual de Garantia da Qualidade Analítica (Manual, 2015), conforme apresentado a seguir.

### **Elaboração de Cartas de Controle de Shewhart:**

- analisar entre 10 e 20 amostras (homogêneas e de características similares às comumente ensaiadas pelo laboratório);
- eliminar os *outliers* por testes estatísticos apropriados (por exemplo, teste de Grubbs);
- calcular a média e o desvio padrão dos resultados;
- realizar um teste de hipótese sobre a diferença estatística insignificante entre o valor obtido e o esperado utilizando um teste  $t$  (equação 12);

$$t = \frac{|x_m - \mu|}{sd} \sqrt{n} \quad (12)$$

Onde:

$x_m$ : média dos resultados

$\mu$ : valor predito

$sd$ : desvio padrão

$n$ : número de repetições

$t$ : resultado a ser comparado com valor crítico tabelado

- caso  $t$  calculado seja menor que o tabelado, elaborar a carta;
- marcar o número consecutivo de resultados no eixo  $x$  da carta e os valores das características observadas no eixo  $y$ ;
- marcar uma linha central correspondendo ao valor de referência (CL – *Central Line*);
- traçar duas linhas paralelas, sendo uma e outra abaixo da linha central, correspondendo a  $\pm 2$  x desvios padrão para estabelecer o Limite Superior de Alerta (UWL - *Upper Warning Limit*) e o Limite Inferior de Alerta (LWL - *Lower Warning Limit*);
- traçar duas linhas paralelas, sendo uma e outra abaixo da linha central, correspondendo a  $\pm 3$  x desvios padrão para estabelecer o Limite Superior de Ação (UAL - *Upper Action Limit*) e o Limite Inferior de Ação (LAL - *Lower Action Limit*);

- com a carta preparada, continuar a inserir os dados de controle e avaliar conforme a Tabela 1.

**Tabela 1.** Condições para avaliação da carta de Shewhart.

Condição	Avaliação
1	Caso um resultado esteja dentro da faixa WL este valor é considerado aprovado;
2	Valores obtidos entre as faixas são aceitos; no entanto não podem ser mais que 2 em 20 determinações;
3	Caso um resultado seja encontrado fora do limite de AL, ou caso existam 7 valores consecutivos criando uma tendência (acima ou abaixo), deve ser realizada uma avaliação do método;
4	3 valores consecutivos entre AL e WL podem indicar problemas no ensaio;
5	2 valores consecutivos entre AL e WL, do mesmo lado da média podem indicar problemas no ensaio;
6	10 valores consecutivos do mesmo lado da média indicam erros sistemáticos.

AL – limite de atenção (*Attention Limit*); WL – limite de alerta (*Warning Limit*).

## Controle de processo estatístico multivariado

Apesar de um grande número de engenheiros e químicos ainda utilizar medidas univariadas, o Controle de Processo Estatístico Multivariado (MSPC – *Multivariate Statistical Process Control*) tem se tornado cada vez mais presente. Um problema que surge com o emprego da análise univariada convencional é a quantidade de parâmetros que deve ser medida no tempo, em intervalos regulares. Desta forma, em um dia centenas ou mesmo milhares de dados e informações são disponibilizados para serem processados. Como muitas dessas medidas são relacionadas, faz mais sentido o emprego de um modelo multivariado do que de diversos modelos univariados. Além disso, alguns instrumentos, que são cada vez mais utilizados nas indústrias, utilizam monitoramento espectroscópico, tais como o NIR, o qual sozinho é capaz de gerar um novo espectro em poucos segundos. Incluir essa informação na base de dados torna-se inimaginável, podendo gerar centenas de possibilidades de cartas de controle (Brereton, 2007).

Neste enfoque, métodos tais como Análise de Componentes Principais (PCA – *Principal Components Analysis*) e a Regressão Parcial por Mínimos Quadrados (PLS – *Partial Least Square*) são alternativas bastante úteis, pois possibilitam o registro dos Componentes Principais de diversos parâmetros, em vez do registro de um único parâmetro. Em lugar de se utilizar uma única medida física, deve-se obter a concentração em tempo real por meio de uma calibração multivariada de um ou mais componentes do produto.

Existem diferentes medidas multivariadas que possibilitam o controle dos resultados espectroscópicos. Os objetivos do MSPC são observar amostras cuja característica (normalmente espectroscópica) apresenta desvio significativo das características previamente estabelecidas como normais e calcular estatisticamente este desvio. Se a amostra estiver fora desses limites, indica que provavelmente há dificuldade no processo. Especialmente em termos de aplicação para o controle dos resultados e processos, essa é a área que se apresenta com maior potencial de crescimento, sendo o controle multivariado a principal tendência na área de quimiometria (Brereton, 2007).

## Controle de qualidade por meio de comparação interlaboratorial

Os métodos empregados pelos diferentes programas de ensaio de proficiência (PT – *Proficiency Test*) variam dependendo da natureza dos itens ensaiados (amostras), do método de ensaio utilizado e do número de laboratórios participantes. As normas ABNT NBR ISO/IEC 17043:2011 descrevem as características comuns dos programas de ensaio de proficiência que abrangem os três tipos básicos de ensaios de laboratório (quantitativos, qualitativos e interpretativos): programa de participação sequencial; programas de participação simultâneos (programas de ensaio de amostras divididas e programas de processos parciais) e programas de avaliação externa de qualidade (Associação Brasileira de Normas Técnicas, 2011).

Geralmente, independentemente do tipo, são atribuições do provedor do PT: a) produzir os itens de ensaio; b) estabelecer o protocolo de funcionamento do programa; c) definir o modelo estatístico para a avaliação do desempenho; d) determinar os valores designados e as incertezas; e) distribuir os itens de ensaio aos laboratórios participantes; f) receber os resultados dos participantes e as informações sobre os métodos; g) comparar os resultados dos participantes com a faixa aceitável; e h) elaborar relatórios com comentários orientadores e educativos.



Para selecionar um provedor de ensaio de proficiência adequado ao escopo de análises do laboratório, sugere-se pesquisar na base de dados EPTIS (*European Proficiency Testing Information System*), que contém cerca de 4.000 PTs cadastrados, procedentes de aproximadamente 40 países (*European Proficiency Testing Information System*, 2018).

Em um contexto geral, a participação em programas de ensaios de proficiência traz como benefícios aos participantes: a) avaliação do desempenho e monitoramento contínuo; b) evidência de obtenção de resultados confiáveis; c) identificação de problemas relacionados com a sistemática de ensaios ou o sistema analítico; d) possibilidade de tomada de ações corretivas e/ou preventivas; e) avaliação da eficiência de controles internos de qualidade; f) determinação das características de desempenho e validação de métodos e tecnologias; e g) padronização das atividades frente ao mercado, sendo um dos requisitos da norma ISO/IEC 17025:2017 para a acreditação ou habilitação do laboratório de ensaio (*International Organization for Standardization*, 2017).

Os benefícios citados, na prática, minimizam os riscos da ocorrência de inexatidão e imprecisão significativas no conceito estatístico, o que pode gerar resultados analíticos contaminados com erros sistemáticos ou aleatórios, tendências ou diferenças significativas quando comparados a laboratórios similares, tornando imprescindível e imperativo que os laboratórios participem de ao menos uma rodada de ensaios de proficiência por ano.

De acordo com orientação descrita na norma nº NIT-DICLA-026 do Inmetro, cabe ao laboratório definir o intervalo do seu plano de participação em rodadas de PT (por exemplo: anual, bienal, quadrienal, etc.). Os resultados dessa participação fornecem aos gestores do laboratório uma confirmação de que o desempenho do laboratório é satisfatório ou alerta para a necessidade de investigação de potenciais problemas (*Inmetro*, 2016b).

O objetivo maior dos ensaios de proficiência (PT) é estimular o bom desempenho dos participantes, disponibilizando meios objetivos para que o responsável pelo laboratório possa avaliar e demonstrar a confiabilidade dos dados que produz.

O projeto estatístico a ser empregado em programas de ensaios de proficiência deve fornecer resultados simples e transparentes, para que os participantes e outros interessados, por exemplo, os clientes do laboratório, possam avaliar facilmente as informações contidas nos relatórios fornecidos pelo provedor do PT.

Dessa forma, a hipótese estatística fundamental para o escopo do PT deve ser formulada da seguinte maneira: a hipótese nula ( $H_0$ ) é que não haja diferença estatística significativa do resultado do laboratório participante em relação aos resultados do grupo de participantes, ou seja, que o laboratório atenda aos requisitos de qualidade. Por outro lado, se a hipótese nula  $H_0$  for rejeitada, a hipótese alternativa ( $H_1$ ) torna-se válida, isto é, o resultado do laboratório participante apresenta diferença estatística significativa em um determinado nível de confiança (por exemplo: 95%), quando comparado com os resultados do grupo e, portanto, não atende aos requisitos de qualidade (Uhlig; Lischer, 1998).

Um dos pontos críticos para todos os provedores de PT consiste na interpretação dos resultados do programa e, conseqüentemente, a avaliação do desempenho dos participantes. Na literatura, são citadas diversas técnicas estatísticas empregadas para avaliar os resultados de programas de comparação interlaboratorial. No entanto, essas técnicas devem ser apropriadas para cada situação, sendo que na ISO 13528 são apresentadas orientações sobre delineamentos estatísticos e análises visuais de dados (Associação Brasileira de Normas Técnicas, 2011; International Organization for Standardization, 2015b).

Geralmente, duas etapas são comuns para todos os PTs: 1) estimar o valor designado da concentração do analito e a incerteza associada; e 2) efetuar a análise estatística para avaliar o desempenho do laboratório e determinar a homogeneidade e a estabilidade do item do ensaio de proficiência. Não existe um procedimento padronizado que descreva em detalhes as estratégias a ser utilizadas. Dessa forma, dependendo do propósito, diferentes critérios estatísticos empregados podem afetar a determinação do valor designado e o intervalo de aceitação, e conseqüentemente comprometer a avaliação do desempenho do laboratório (Maio et al., 2006; Visser, 2006).

## Como calcular o valor designado

Por definição, o valor designado é a melhor estimativa do valor verdadeiro, sendo este utilizado com o propósito de calcular os índices de desempenho dos participantes dos PTs. Esse valor deve ser definido de forma criteriosa para avaliar de maneira justa os participantes e incentivar a concordância entre métodos e laboratórios (Inmetro, 2016b; Associação Brasileira de Normas Técnicas, 2011).

Na literatura, são sugeridos diferentes métodos para determinar o valor designado, sendo os mais comuns listados a seguir: valor designado fornecido por um laboratório referência; valor certificado por um material de referência certificado (CRM) utili-

zado como material de ensaio; valor designado fornecido por laboratório especialista; valor designado obtido através de materiais formulados; valores de consenso que são derivados diretamente dos resultados relatados pelos participantes.

## **Valores de consenso que são derivados diretamente dos resultados relatados pelos participantes**

Este é o procedimento mais utilizado para determinar o valor designado em ensaios de proficiência, sendo utilizada normalmente a média dos resultados após a exclusão de valores dispersos (estatística clássica) ou a mediana dos resultados de todos os participantes na rodada do ensaio (estatística robusta).

Dentre as principais vantagens do uso deste método, estão a facilidade de aplicação e o baixo custo. A principal desvantagem é a existência de subpopulações, ou seja, em PT onde os participantes utilizam métodos distintos, o valor alvo poderá sofrer tendência no valor proveniente dos resultados de um grupo majoritário de laboratórios.

## **Procedimentos para a determinação do desvio padrão alvo**

Esse parâmetro é definido como sendo a incerteza associada ao valor designado que caracteriza a dispersão ou a variabilidade dos valores atribuídos ao mensurando, ou seja, determina com alta probabilidade o intervalo no qual o valor verdadeiro provavelmente está contido (Ellison; Willians, 2012).

Vários testes estatísticos são empregados para estimar o intervalo de confiança em PTs. Na ISO/IEC 13528:2015 são descritos procedimentos que podem ser utilizados para a definição do desvio padrão alvo (International Organization for Standardization, 2015b). Geralmente, o procedimento para determinar a incerteza deve ser adequado para satisfazer o objetivo comum de todos os laboratórios e, dessa forma, precisa ser definido pelo provedor do PT e deverá ser divulgado para os participantes antes do início de cada rodada (Thompson; Ellison, 2006).

Dentre os procedimentos mais utilizados, estão: a) métodos baseados nos resultados dos participantes, como na estatística clássica que utiliza a estimativa do desvio padrão e na estatística robusta, calculados a partir do intervalo quartílico normalizado; b) métodos baseados na raiz quadrática da soma da incerteza-padrão combinada dos resultados dos participantes e a incerteza-padrão do valor designado; c) método baseado na raiz quadrática da soma da incerteza-padrão expandida do resultado do

participante; e d) incerteza expandida do valor designado do laboratório de referência ou do material de referência utilizado como item de ensaio.

## Procedimentos para a avaliação do desempenho dos laboratórios participantes

Independentemente do modelo estatístico utilizado pelo provedor para a avaliação do desempenho dos participantes, é necessário verificar a presença de comportamento multimodal nos resultados, a qual poderia causar impactos significativos quanto à validade da avaliação de desempenho. Dessa forma, na verificação visual da normalidade dos dados, podem ser utilizados gráficos tipo histogramas.

Dentre os procedimentos estatísticos previstos na ABNT NBR ISO/IEC 17043:2011 (Associação Brasileira de Normas Técnicas, 2011), empregados por programas interlaboratoriais para avaliar o desempenho dos laboratórios, estão:

a) **Índice z (z-escore)**: o índice z é procedimento estatístico mais comum empregado por programas interlaboratoriais para avaliar o desempenho dos laboratórios. Esse índice é recomendado pelo Protocolo Internacional Harmonizado para Ensaio de Proficiência (IUPAC, 2006) e é obtido conforme a equação 13, onde  $x_i$  é o resultado informado pelo participante,  $\bar{X}$ , o valor designado (melhor estimativa do valor verdadeiro) e  $\hat{\sigma}$ , o desvio padrão alvo para o propósito conveniente do PT (Thompson; Ellison, 2006).

$$z = \frac{(x_i - \bar{X})}{\hat{\sigma}} \quad (13)$$

O índice z tem, como vantagem, permitir a comparação direta dos resultados de diferentes amostras e de diferentes unidades, porque o valor desse índice não é expresso na unidade original da medida, ou seja, é normalizado e descrito como sendo a distância entre  $x_i$  e  $\bar{X}$  em unidades de desvio padrão. Dessa forma, a maioria dos PTs por comparação interlaboratorial, que utilizam o índice z para avaliar o desempenho dos ensaios, interpreta que, quanto mais próximo de 0 (zero) for o valor de z, mais exato o resultado, melhor o desempenho do laboratório na realização do ensaio (Thompson; Ellison, 2006; Analytical Methods Committee, 1989).

Para distribuição normal do conjunto de dados, a probabilidade é que aproximadamente 95% dos resultados estejam dentro do intervalo de  $|z| \leq 2,0$ . Desse modo, o desempenho do laboratório é considerado aceitável ou satisfatório para realizar o ensaio. Por outro lado, valores de  $|z| \geq 3,0$  são atípicos e, dessa forma, devem ser considerados inaceitáveis ou insatisfatórios, uma vez que estão acima de três desvios padrão da média e fora do intervalo de confiança de 99,7%. Nesse caso, é preciso abrir uma não conformidade, investigar a causa e tomar medidas corretivas para eliminar as fontes de erros no processo analítico.

Valores entre  $2,0 < |z| < 3,0$  são considerados resultados com desempenho questionável. Espera-se que os resultados dentro dessa faixa ocorram aproximadamente em 5% das vezes, caracterizando-se como um evento isolado que não requer ações corretivas, mas apenas monitoramento.

- b) **Índice zeta ( $\zeta$ )**: atualmente, esse índice vem sendo amplamente utilizado para a avaliação do desempenho de laboratórios e PTs. É semelhante ao erro normalizado ( $E_n$ ), no entanto, utiliza a incerteza-padrão ao invés da incerteza expandida, permitindo interpretar os intervalos de confiança da mesma maneira que o índice  $z$  (z-escore). O cálculo desse índice está descrito na Equação 14, onde  $u_{lab}$  é a incerteza-padrão combinada do resultado do participante e  $u_{av}$  é a incerteza-padrão do valor designado (International Organization for Standardization, 2015b).

$$\zeta = \frac{x_i - \bar{X}}{\sqrt{u_{lab}^2 + u_{av}^2}} \quad (14)$$

- c) **erro normalizado ( $E_n$ )**: o cálculo do erro normalizado é baseado na Equação 15, onde  $x_i$  é o resultado do laboratório participante;  $X_{ref}$  é o valor designado fornecido pelo laboratório referência;  $U_{lab}^2$  é a incerteza-padrão expandida do resultado do participante e  $U_{lab}^2$  é a incerteza expandida do valor designado pelo laboratório referência. Nesse caso, os participantes que apresentarem  $|E_n| \leq 1,0$  obtêm desempenho satisfatório e os participantes que apresentarem  $|E_n| > 1,0$  obtêm desempenho insatisfatório, havendo a necessidade de gerar uma não conformidade para dar início à ação preventiva ou corretiva (Associação Brasileira de Normas Técnicas, 2011).

$$E_n = \frac{x_i - X_{ref}}{\sqrt{U_{lab}^2 + U_{ref}^2}} \quad (15)$$

## Aplicação, uso e interpretação de resultados dos Programas de Ensaio de Proficiência

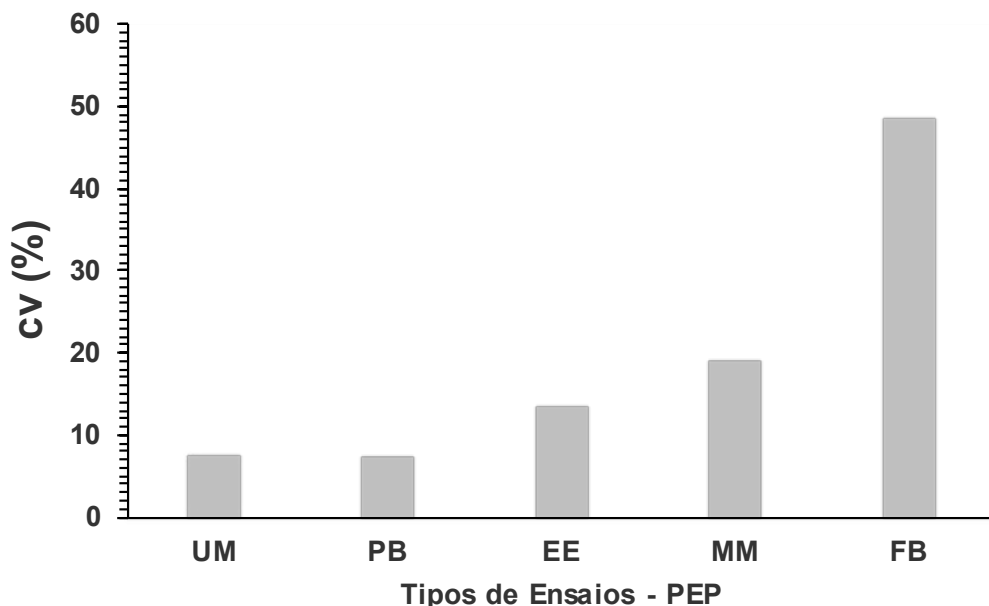
Para que o laboratório participante do PT possa usufruir dos vários benefícios anteriormente citados, é preciso que este avalie suas próprias conclusões sobre o seu desempenho, analise de forma criteriosa os resultados apresentados pelo provedor, por meio dos relatórios analíticos de cada rodada e, em caso de resultados insatisfatórios, seja registrada uma não conformidade para que seja tomada a ação corretiva apropriada. A seguir, estão descritas algumas aplicações que o participante de PT pode utilizar a partir da interpretação dos resultados apresentados no relatório do provedor:

a) **comparação de desvio analítico interlaboratorial:** para expressar a relação percentual do desvio padrão com a média, utiliza-se o coeficiente de variação (CV), também conhecido como desvio padrão relativo (DPR), o qual é expresso em porcentagem e está relacionado à dispersão dos resultados entre os laboratórios.

A informação desse coeficiente proporciona uma visão geral da precisão de um método analítico em termos de reprodutibilidade, independentemente da grandeza dos valores. Quanto maior for o CV, menos uniforme é o conjunto de dados, ou seja, maior a variabilidade interlaboratorial.

Em programas de ensaio de proficiência por comparação interlaboratorial, geralmente o CV é calculado após a exclusão dos resultados de laboratórios que são considerados discrepantes (*outliers*). Para identificar esses resultados discrepantes, é preciso aplicar testes estatísticos, como teste de Grubbs, teste de Dixon ou teste de Hampel (Souza et al., 2009). Na Figura 3 são apresentados valores dos coeficientes de variação de alguns tipos de ensaios realizados em uma amostra de milho em grão. Essa amostra foi utilizada como item de ensaio no Programa de Proficiência para Laboratório de Nutrição Animal, coordenado pela Embrapa Pecuária Sudeste, no qual foram avaliados os teores de umidade (UM), proteína bruta (PB), extrato etéreo (EE), matéria mineral (MM) e fibra bruta (FB). Todos esses ensaios foram realizados por NIRS, considerando modelos de calibração e marcas independentes, próprias de cada laboratório participante.

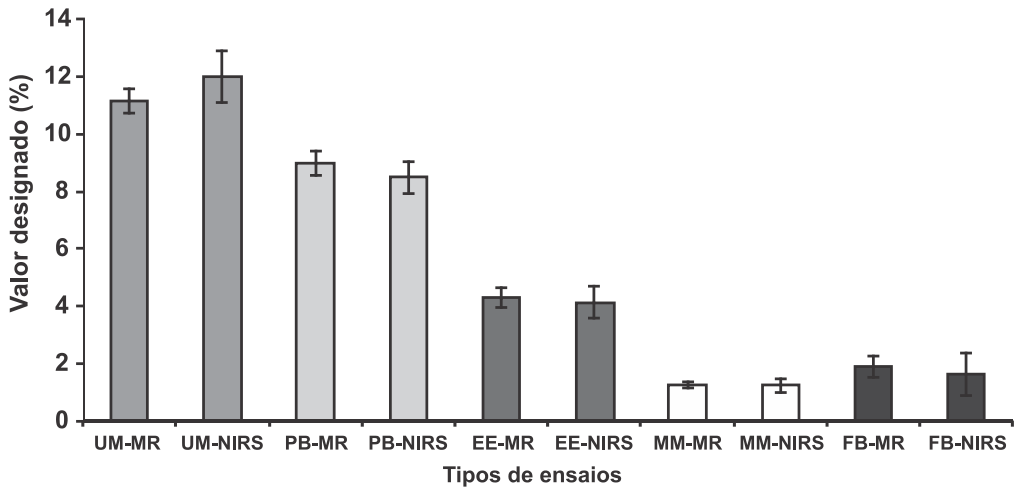
Por meio do PT, foi possível observar que o ensaio que apresentou maior variabilidade interlaboratorial foi a FB, com CV igual a 48,6%. Sendo assim, a orientação é que os laboratórios avaliem os seus modelos de calibração e o método referência e verifiquem possíveis problemas em relação à precisão analítica e à exatidão.



**Figura 3.** Valores dos coeficientes de variação (CV) para os ensaios de proficiência em amostra de milho grão moído, analisado por NIRS. UM - umidade, PB - proteína bruta, EE - extrato etéreo, MM - matéria mineral, FB - fibra bruta.

b) **comparação de métodos analíticos:** utilizando a mesma amostra de milho do item anterior, os laboratórios informaram os resultados dos ensaios obtidos pelo método referência (MR) e os resultados previstos nos modelos de calibração de NIRS. Por meio desses resultados, foram definidos, para cada tipo de ensaio, os valores designados e o desvio padrão alvo, ambos para fins de avaliação de desempenho no PT.

Por meio da Figura 4, é possível observar que os resultados obtidos através dos dois procedimentos analíticos foram comparáveis e não apresentaram diferença estatística significativa no nível de 95%.

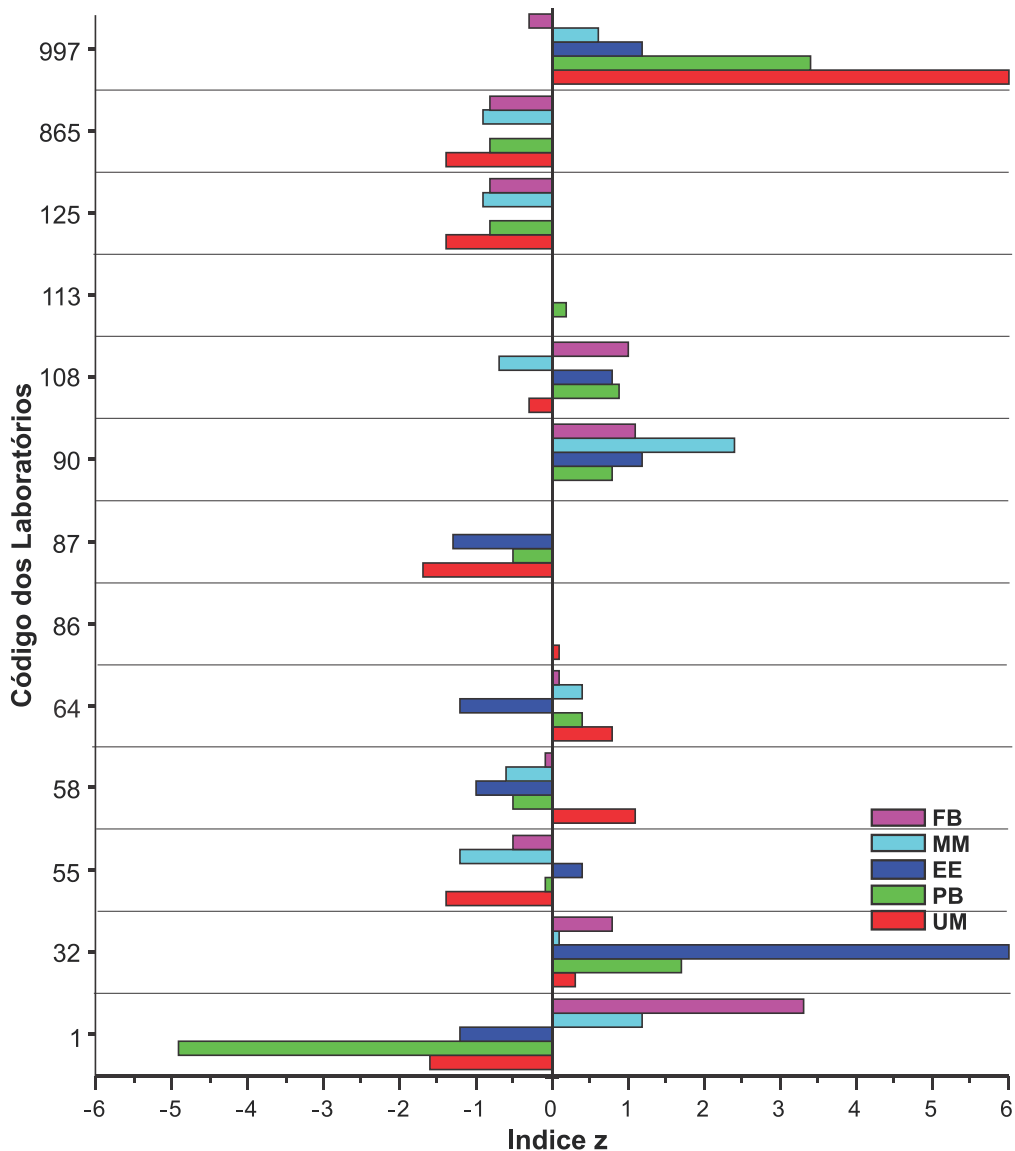


**Figura 4.** Valores do coeficiente de variação (CV) para os ensaios de uma rodada de ensaio de proficiência em amostra de milho grão moído analisado por NIRS e por método referência (MR). UM - umidade, PB - proteína bruta, EE - extrato etéreo, MM - matéria mineral, FB - fibra bruta.

c) **avaliação de desempenho interlaboratorial:** por meio do gráfico do índice z, os laboratórios puderam avaliar quais ensaios apresentaram desempenho satisfatório, questionável ou insatisfatório. Na Figura 5 é possível observar que o laboratório código 997 apresentou resultados com desempenho insatisfatório em dois tipos de ensaio, o que demonstra a ocorrência de problemas na realização dos ensaios UM e PB. Nesse caso, o responsável técnico deverá abrir uma não conformidade e avaliar qual é a causa raiz do problema e corrigi-la antes de continuar o processo analítico no laboratório.

Outra observação feita é a evidência de efeitos sistemáticos nos resultados; por exemplo, o laboratório código 90 apresentou todos os resultados dos ensaios com valores positivos para o índice z. Nesse caso, o responsável técnico deverá observar com atenção se esse efeito sistemático é frequente e, caso necessário o modelo deverá passar por revalidação para verificar, principalmente, a presença de viés significativo.





**Figura 5.** Avaliação do desempenho dos laboratórios com o uso do índice z. UM - umidade, PB - proteína bruta, EE - extrato etéreo, MM - matéria mineral, FB - fibra bruta.

d) **monitoramento do desempenho do laboratório ao longo do tempo:** o desempenho individual ou coletivo dos laboratórios participantes do PT também pode ser monitorado por avaliação dos índices de desempenho, índice z ou índice zeta. Por meio desses índices, os participantes podem visualizar a variabilidade tempo-

ral do seu desempenho e verificar em quais ensaios existem tendências sistemáticas aleatórias. Esse monitoramento poderá ser conduzido com o uso de cartas de controle de Shewhart, por exemplo.

## Conclusões

Qualidade não é um atributo passível de ser observado apenas com uma visão superficial. Deve ser medida e avaliada para ser expressa em termos como boa-ruim ou aceitável-recusada. A qualidade pode também melhorar ou deteriorar, dependendo do esforço, entusiasmo e experiência dos analistas e do ambiente no qual o experimento é realizado.

O emprego de materiais de referência possibilita a avaliação de fatores inerentes ao laboratório (equipamentos, pessoal, etc.), sendo importante seu uso nas etapas cruciais do método analítico, desde a validação, aplicação e trabalho de rotina, para a obtenção de bons desempenhos em programas de qualidade e certificação.

O controle de qualidade intralaboratorial é um aspecto essencial para certificar que os dados gerados por um laboratório estão de acordo com o resultado esperado, ou seja, dentro dos limites aceitáveis. Quando executadas de forma apropriada, as ferramentas de controle de qualidade (ensaios intra e interlaboratoriais, cartas de controle, etc.) podem monitorar os vários aspectos relacionados à qualidade dos dados. Em rodadas nas quais o desempenho esteja fora dos limites aceitáveis, os dados produzidos devem ser rejeitados e, após a tomada de ações corretivas, as análises devem ser repetidas.

Finalmente, deve-se salientar que apenas a execução de sistemas controle de qualidade não garante a confiabilidade dos dados gerados. Os procedimentos corretos para ações de remediação e motivação dos empregados envolvidos devem ser documentados e colocados em ação. Somente a participação e a conscientização de todos os envolvidos no processo garantem o sucesso de um programa de qualidade.

## Referências

ANALYTICAL METHODS COMMITTEE. Robust statistics – how not to reject outliers. Part 1. Basic concepts. **Analyst**, v. 114, n. 12, p. 1693-1697, 1989.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR ISO/IEC 17025:2017**: requisitos gerais para a competência de laboratórios. Rio de Janeiro, 2017a.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR ISO 17034:2017**: requisitos gerais para a competência de produtores de materiais de referência. Rio de Janeiro, 2017b.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR ISO/IEC 17043**: avaliação de conformidade – requisitos gerais para ensaios de proficiência. Rio de Janeiro, 2011.

ASTM STANDARDS. **E1655, vol 03.06**: standards practices for infrared, multivariate, quantitative analysis, West Conshohocken: ASTM International, 2000. Annual Book.

BRERETON, R. G. **Applied chemometrics for scientists**. West Sussex: John Wiley & Sons, Ltd, 2007. 379 p.

CAVE, R.; BUTLER, O.; COOK, J. M.; CRESSEM, R. S.; GARDEL, N. M.; HOLDEN, A. J.; MILES, D. L. Atomic spectrometry update: environmental analysis. **Journal of Analytical Atomic Spectrometry**, v. 14, n. 2, p. 279-353, 1999.

CHRISTENSEN, J. M. Correcting measurement errors using reference materials in method validation. **Mikrochimica Acta**, v. 123, n. 1/4, p. 231-240, 1996.

COTLOVE, E.; HARRIS, E. K.; WILLIAM, G. Z. Biological and analytic components of variation in long-term studies of serum constituents in normal subjects. 3. Physiological and medical implications. **Clinical Chemistry**, v. 16, n. 12, p. 1028-1032, 1970.

ELLISON, S. L. R.; WILLIAMS, A. (Ed.). **EURACHEM/CITAC guide quantifying uncertainty in analytical measurement**. 3rd ed. [S.I.]: EURACHEM; CITAC, 2012. EURACHEM/CITAC Guide CG 4. 133 p. Disponível em: <[https://www.eurachem.org/images/stories/Guides/pdf/QUAM2012\\_P1.pdf](https://www.eurachem.org/images/stories/Guides/pdf/QUAM2012_P1.pdf)>. Acesso em: 5 março 2018

EUROPEAN PROFICIENCY TESTING INFORMATION SYSTEM. EPTIS. Disponível em: <<http://www.eptis.bam.de>>. Acesso em: 12 maio 2018.

EUROPEAN UNION. Commission Regulation (EC) No 333/2007 of 28 march 2007. Laying down the methods of sampling and analysis for the official control of the levels of lead, cadmium, mercury, inorganic tin, 3-MCPD and benzo(a)pyrene in foodstuffs. **Official Journal of the European Union**, p. L.88/29-L.88/38, 2007.

FERREIRA, M. M. C. **Quimiometria**: conceitos, métodos e aplicações. Campinas: Ed. da Unicamp, 2015. 493 p.

FUNK, W.; DAMMANN, V.; DONNEVERT, G. **Quality assurance in analytical chemistry**. Weinheim: Wiley -VCH Verlag GmbH & Co. KGaA, 2007.

INTERNATIONAL CONFERENCE ON HARMONISATION OF TECHNICAL REQUIREMENTS FOR REGISTRATION OF PHARMACEUTICALS FOR HUMAN USE (ICH). **Validation of analytical procedures**: text and methodology Q2(R1): step 4 version. London: ICH, 2005.

INMETRO. **Doq-Cgcre-008-Rev.05**: orientação sobre validação de métodos analíticos. Rio de Janeiro, 2016a.

INMETRO. **NIT-DICLA-26**: requisitos para a participação de laboratórios em ensaios de proficiência. Rio de Janeiro, 2016b.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 7870-1**: control charts. Part 1: general guidelines. Geneva, 2014.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 7870-2**: control charts. Part 2: shewhart control charts. Geneva, 2013.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 7870-3**: control charts. Part 3: acceptance control charts. Geneva, 2012.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 7873**: control charts for arithmetic average with warning limits. Geneva, 1993.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO GUIDE 33**: reference materials - good practice in using reference materials. Geneva, 2015a.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO GUIDE 35**: reference materials - guidance for characterization and assessment of homogeneity and stability. Geneva, 2017.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO GUIDE 80:2014(E)**: guidance for the in-house preparation of quality control materials (QCMs). Geneva, 2014.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO GUM**: avaliação de dados de medição: guia para a expressão de incerteza de medição. 1. ed. brasileira. Rio de Janeiro, 2008.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 13528**: statistical methods for use in proficiency testing by interlaboratory comparison. Geneva, 2015b.

JOINT COMMITTEE FOR GUIDES IN METROLOGY. **JCGM 200:2012**: international vocabulary of metrology - basic and general concepts and associated terms (VIM). 3rd ed. Paris, 2012.

MAIO, F. D.; ZENEBON, O.; TIGLEA, P.; SAKUMA, A. M. Avaliação de critérios estatísticos utilizados em programas interlaboratoriais para pesquisa de chumbo em sangue. **Revista do Instituto Adolfo Lutz**, v. 65, n. 2, p. 89-93, 2006.

MANUAL de garantia da qualidade analítica: áreas de identidade e qualidade de alimentos e de insumos. Brasília, DF: Ministério da Agricultura, Pecuária e Abastecimento, 2015. 51 p.

MANUAL de validação, verificação/confirmação de desempenho, estimativa da incerteza de medição e controle de qualidade intralaboratorial. Brasília, DF: Ministério da Agricultura, Pecuária e Abastecimento, 2014.

MILLER, J. M.; MILLER, J. C. **Statistics and chemometrics for analytical chemistry**. 5th ed. Harlow, Essex: Pearson Education Limited Edinburgh Gate, 2005. 285 p.

OLIVARES, I. R. B.; LOPES, F. A. L. Essential steps to providing reliable results using the analytical quality assurance cycle. **Trends in Analytical Chemistry**, v. 35, p. 109-121, May 2012.

OLIVARES, I. R. B.; SOUZA, G. B.; NOGUEIRA, A. R. A.; TOLEDO, G. T. K.; MARCKI, D. C. Trends in developments of certified reference materials for chemical analysis - focus on food, water, soil, and sediment matrices. **Trends in Analytical Chemistry**, v. 100, p. 53-64, Mar. 2018.

PASQUINI, C. Near infrared spectroscopy: a mature analytical technique with new perspectives – a review. **Analytica Chimica Acta**, v. 1026, p. 8-36, Oct. 2018.

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.

SOUZA, G. B.; NOGUEIRA, A. R. A.; SANTO, V. R. D.; PICCHI, C. M. C.; GUIMARAES, E. S.; JUNIOR, W. B. Proficiency testing of animal nutrition laboratories. **Accreditation and Quality Assurance**, v. 14, n. 8/9, p. 455-460, 2009.

THOMPSON, M. Internal quality control in routine analysis. **AMC Technical Briefs**, n. 46, Feb 2010. 2 p.

THOMPSON, M.; ELLISON, L. R. Fitness for purpose – the integrating theme of the revised harmonised protocol for proficiency testing in analytical chemistry laboratories. **Accreditation and Quality Assurance**, v. 11, n. 8/9, p. 373-378, 2006.

THOMPSON, M.; WOOD, R. Harmonized guidelines for internal quality control in analytical chemistry laboratories (technical report). **Pure and Applied Chemistry**, v. 67, n. 4, p. 649-666, 1995.

TURUTA, T. B. **Aplicação de cartas de controle como ferramenta de melhoria frente às dificuldades operacionais de laboratórios acreditados na ABNT NBR ISO/IEC 17025**. 2015. 67 f. Dissertação (Mestrado em Química Analítica) – Instituto de Química, Universidade de São Paulo, São Carlos.

UHLIG, S.; LISCHER, P. Statistically-based performance characteristics in laboratory performance studies. **Analyst**, v. 123, n. 6, p. 167-172, 1998.

VALDERRAMA, P.; BRAGA, J. W. B.; POPPI, R. J. Estado da arte de figuras de mérito em calibração multivariada, **Química Nova**, v. 32, n. 5, p. 1278-1287, 2009.

VISSER, R. G. Interpretation of interlaboratory comparison results to evaluate laboratory proficiency. **Accreditation and Quality Assurance**, v. 10, n. 10, p. 521-526, 2006.

WESTGARD, J. O.; BARRY, P. L.; HUNT, M. R.; GROTH, T. A multi-rule Shewhart chart for quality control in clinical chemistry. **Clinical Chemistry**, v. 27, n. 3, p. 493-501, 1981.

ZSCHUNKE, A. The role of reference materials. **Accreditation and Quality Assurance**, v. 5, n. 10, p. 441-445, 2000.

# Princípios da Espectroscopia no Infravermelho Próximo Associada a Imagens Hiperespectrais

Francisco Fernandes Gambarra Neto  
Joabson Borges de Araújo  
Everaldo Paulo de Medeiros

## Introdução

Herschel observou que existia, além do espectro visível, radiação que poderia enegrecer uma placa fotográfica ou aumentar a temperatura de um termômetro (Herschel, 1800). O primeiro espectro infravermelho de um composto orgânico foi medido fotograficamente por Abney e Festing (1881), observando que “o fundamento de toda absorção nesses corpos é o hidrogênio”. Brackett (1928) estudou moléculas orgânicas sob alta resolução no infravermelho e mostrou que os átomos de hidrogênio primários, secundários e terciários ligados ao carbono resultam em bandas com comprimentos de onda diferentes. A absorção de luz de um comprimento de onda definido por uma molécula resulta em um aumento na frequência de vibração de uma ligação entre um par de átomos. A vibração na ligação comporta-se como um simples movimento harmônico. Todas as mudanças de energia dentro das moléculas são quantificadas e há uma frequência definida de vibração associada à mudança quantizada na energia vibracional (Herzberg, 1945; Wheeler, 1959).

Ao longo das últimas três décadas, a espectroscopia de reflectância no infravermelho próximo (NIR) provou ser uma das ferramentas mais eficientes e avançadas para a estimativa de atributos de qualidade e quantidade em amostras de diferentes matrizes.

A região espectral NIR é geralmente definida como o intervalo de comprimento de onda de 900 nm a 2.500 nm. As bandas de absorção nesta região são devidas a sobretons (frequências harmônicas) e combinações das bandas de vibração molecular em nível fundamental. As transições de energia estão entre o estado fundamental e o segundo ou terceiro estado vibratório excitados. Como as transições de energia mais altas são sucessivamente menos propensas a ocorrer, cada sobretom é sucessivamente mais fraco em intensidade (Weyer, 1985).

A espectroscopia NIR ganhou ampla aceitação em diferentes campos devido à sua capacidade de registrar espectros para amostras envolvendo uma mínima manipula-

ção prévia. Tornando-se uma ferramenta analítica de resposta rápida de maneira não destrutiva (Blanco; Villarroya, 2002). Os pontos críticos da precisão das medidas NIR e da quimiometria de quantificação são a decomposição do erro padrão de predição em componentes de *bias* (exatidão) e de variância (Bellon-Maurel; Fernandez-Ahumada, 2010). Novas técnicas como a espectroscopia espacial resolvida no tempo para estimativa de propriedade de absorção e dispersão da luz, assim como técnicas de imagens multi e hiperespectrais no infravermelho próximo (HSI-NIR), foram revisadas por Nicolaï, sendo dada atenção especial ao desenvolvimento de sistemas portáteis (Nicolaï et al., 2007).

A imagem hiperespectral (HSI) é uma combinação de duas tecnologias consolidadas que incluem espectroscopia e imagem, em que uma imagem é adquirida ao longo dos comprimentos de onda na região do NIR para especificar o espectro completo de comprimento de onda de cada pixel do plano de imagem. A grande vantagem da imagem hiperespectral é a capacidade de caracterizar as propriedades químicas inerentes de uma amostra observando sua distribuição no espaço (Ngadi; Liu, 2010).

As análises com HSI-NIR geram um grande conjunto de dados em uma superfície de amostra, conhecida como hipercubo. Esta estrutura pode ser compreendida como duas dimensões espaciais e uma dimensão espectral, contendo informações químicas e físicas da matriz. Portanto, experimentos que geram dados usando a técnica HSI-NIR tendem a sobrecarregar a capacidade de memória dos sistemas computacionais. Tais arquivos precisam ser analisados com uma abordagem computacional quimiométrica para reduzir a dimensionalidade dos dados, mantendo a informação espectral (Dorrepaal et al., 2016). Para a análise de HSI-NIR devem-se observar alguns procedimentos básicos, como aquisição de imagens hiperespectrais, pré-processamento, análise exploratória multivariada, resolução de imagem hiperespectral, classificação e processamento final de imagem digital (Amigo et al., 2015).

## **Aplicações em produtos agrícolas e agroindustriais usando NIR**

Os avanços nos campos da espectroscopia NIR e quimiometria aumentaram o potencial da tecnologia como uma ferramenta de monitoramento confiável em relação à qualidade do produto e à autenticidade dos alimentos, tornando os sistemas de controle de qualidade rápidos e eficazes (Dixit et al., 2017). Como amostras de mesma composição têm muitas fontes de variabilidade que afetam os espectros NIR, uma grande população de amostras é necessária para projetar modelos robustos de calibração, uma vez que o pré-processamento espectral e a complexidade do modelo são simplificados em relação à abordagem do espectro total (Ruckebusch et al., 2006).



A espectroscopia é a principal técnica experimental de física atômica e molecular que envolve a determinação dos estados de energia de átomos ou moléculas, observando a radiação eletromagnética absorvida ou emitida quando eles mudam de estado. Medir a frequência de luz absorvida ou emitida, determinada pela diferença de energia entre os dois estados, pode fornecer uma indicação sensível às interações que perturbam esses estados de energia (Aenugu et al., 2011).

Processos nos quais se utilizam a espectroscopia NIR e a quimiometria para quantificação e classificação são: identificação de fontes de licopeno e carotenoides, como o betacaroteno (Baranska et al., 2006), classificação de azeite extravirgem (Sinelli et al., 2010), determinação de parâmetros de qualidade em produtos lácteos (Ruzicková; Sustová, 2006), caracterização de azeitona de mesa (Casale et al., 2010), determinação de açúcar em uva (Jarén et al., 2001), quantificação do teor de proteína em produtos de leite em pó (Ingle et al., 2016), controle de qualidade de extratos de frutos silvestres durante o armazenamento (Georgieva et al., 2014).

A tecnologia NIR pode prever composição química, valor do pH, capacidade de retenção, força de cisalhamento da fatia e atributos sensoriais como a cor, forma, marmoreio, odor, sabor, suculência, sensibilidade e firmeza (Prieto et al., 2009). Produtos alimentares que passam por transformações químicas, bioquímicas e físicas, possuem em sua composição proteínas e amidas, permitindo assim que esses produtos sejam caracterizados por métodos usando medidas espectrais no infravermelho próximo e médio (MIR) (Aït Kaddour et al., 2008).

A espectroscopia pode ser usada para analisar pequenas amostras de qualquer composição orgânica e organometálica. Realizar análises precisas, de curto tempo, baixo custo e de forma não destrutiva são demandas pela indústria moderna. Portanto, incorpora os últimos avanços em instrumentação, informatização, calibração, transferência de calibração, desenvolvimento de métodos na espectroscopia NIR, preparação de amostras, controle de processos e análise de dados. Contudo, os conhecimentos de espectroscopia e de química proporcionam uma combinação incomparável de fundamentos teóricos, experiência prática e aplicações avançadas (Workman Jr, 2007).

## **Espectroscopia NIR associada à imagem hiperespectral**

Durante as últimas décadas, os investimentos em pesquisa e tecnologia de agricultura de precisão aumentaram significativamente em todo o mundo. Um desses avanços tecnológicos é na espectroscopia NIR, que aprimora as técnicas de sensoriamento

remoto e a análise de imagens de maneira mais rápida, eficiente e precisa do que com os métodos tradicionais de monitoramento (Kenaston; Crockett, 2018).

A NIR-CI (*Near-Infrared Chemical Imaging*) é a fusão da espectroscopia NIR e da análise de imagem, ou de imagens hiperespectrais com o NIR (HSI-NIR), e pode ser usada para visualizar a distribuição espacial dos compostos químicos em uma amostra, fornecendo imagem química. Cada medida da amostra gera um cubo de dados em mais de 100 canais espectrais ou comprimentos de onda, contendo milhares de espectros. Uma parte importante de uma análise NIR-CI é o processamento dos dados a partir dos hipercubos (Ravn et al., 2008).

A aplicação de imagem hiperespectral foi abordada pela primeira vez em satélites e, mais recentemente, tem sido uma técnica analítica emergente em laboratório. As HSI-NIR geram arquivos de dados que exigem novos pensamentos computacionais tanto em questões técnicas quanto conceituais. Configurações instrumentais típicas incluem uma câmera InGaAs (composto químico incluindo os metais Índio, Gálio e o semimetal Arsênio) com filtro sintonizável de cristal líquido que gera uma imagem de 256 pixels x 320 pixels com 81 bandas de comprimento de onda na região do NIR (900 nm-1.700 nm) (Geladi et al., 2004).

A técnica HSI-NIR é utilizada para obtenção de dados muito complexos, como na caracterização de arrozais por um sistema de sensor hiperespectral em miniatura, montado em um veículo autônomo não tripulado (VANT) (Uto et al., 2013), em projetos que medem índice de respostas ao estresse das plantas (Smith et al., 2004), e no estudo comparativo sobre a utilização do perfil de antocianina usando HSI-NIR como ferramenta para discriminar cultivares autóctones de uva vermelha (Nogales-Bueno et al., 2015).

Aplicação destinada à determinação de danos em frutas é um dos fatores de qualidade interna mais importante, que precisam ser detectados em processos de classificação de qualidade em pós-colheita. As técnicas de HSI-NIR têm potenciais efetivos para identificar e detectar injúrias, uma vez que os danos resultam na ruptura das paredes celulares internas. Um exemplo é o algoritmo desenvolvido para classificação de peras utilizando imagens hiperespectrais, que possibilitou a discriminação de frutas normais e com danos, com 92% de precisão (Lee et al., 2014).

Desenvolvimentos recentes em instrumentação analítica e de métodos de processamento de dados permitiram maior uso de técnicas espectroscópicas, sendo propostos para estabelecer métodos alternativos que substituam as técnicas de referência. Essas melhorias incluíram o desenvolvimento de métodos HSI-NIR combinados com ferramentas quimiométricas apropriadas (Fernández Pierna et al., 2012).

## Resolução de imagem hiperespectral NIR

A principal questão para uma análise de imagem é “Quais são as propriedades químicas e físicas da medida de um hiper-cubo?”. A resposta à pergunta delimitará os métodos que podem ser potencialmente utilizados e ajudará na concepção de novos algoritmos que possam levar em consideração as especificidades ligadas à natureza da imagem (De Juan et al., 2014).

Variáveis puras de imagens espectroscópicas Raman de partículas de poeira industrial foram selecionadas usando SIMPLISMA to use Interactive Self-Modeling Mixture Analysis (SIMPLISMA), que é uma técnica que seleciona variáveis puras do conjunto de dados espectrais. Os valores das variáveis puras são proporcionais às concentrações e podem, portanto, ser usados para resolver as análises dos espectros no NIR obtidos para as amostras. O SIMPLISMA é uma das metodologias pioneiras quando se tratam de variáveis puras de imagens espectroscópicas, podendo-se encontrar os mais puros pixels (Equação 1), ou os mais puros canais espectrais, ou a resolução espacial de uma imagem, que é representado pela soma ponderada das concentrações espectrais dos constituintes da imagem (Windig et al., 2007).

$$\rho_i = \frac{S_i}{m_i + f_i} \quad (1)$$

Em que  $S_i$  é o desvio padrão dos elementos nos espectros de pixel,  $m_i$  é a média e  $f_i$  é um fator *offset* que representa a porcentagem de ruído nos dados, prevenindo a seleção de *background* ou *pixels* de ruído. O primeiro espectro selecionado é o de maior pureza. Então, o resto dos espectros são normalizados e um índice de pureza,  $\rho_i$ , (Equação 2) é recalculado para cada um deles (De Juan et al., 2014):

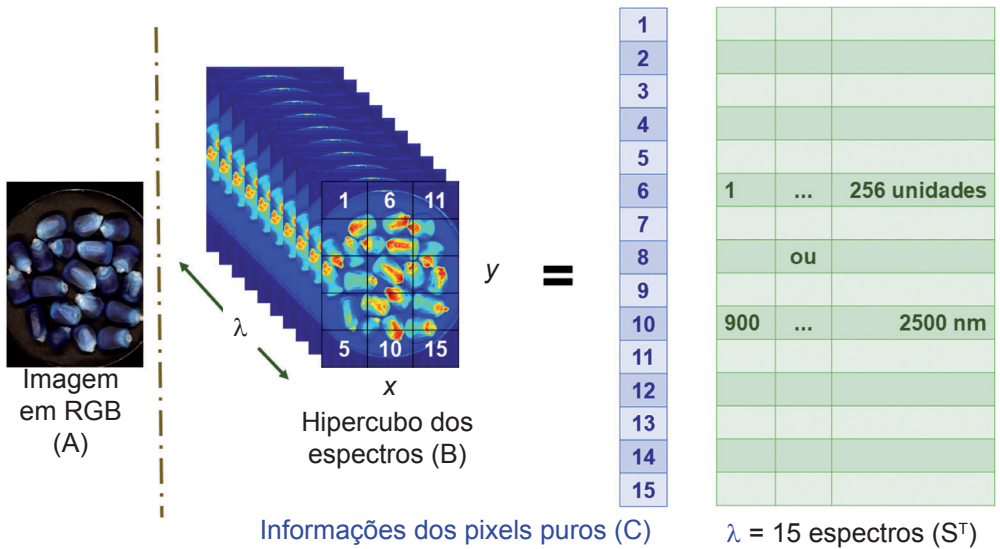
$$\rho'_i = w_i \times \rho_i \quad (2)$$

Em que  $w_i$  é um fator de ponderação que leva em consideração a dissimilaridade do *pixel* em análise com o mais puro *pixel* (*background*) anteriormente selecionado (quanto mais dissimilar o espectro, maior o peso) (De Juan et al., 2014).

O segundo *pixel* selecionado será aquele com maior pureza recalculada,  $\rho_i$ . A sequência de recálculo de pureza (de acordo com novos pesos baseados na dissimilaridade com todos os *pixels* selecionados anteriormente) e nova seleção de *pixels* continuará até que um número de espectros de *pixels* igual ao número de constituintes da imagem tenha sido obtido (De Juan et al., 2014).

Estes pesos de concentração variam de *pixel* a *pixel*, dependendo da composição do *pixel*. O hiper-cubo de espectros NIR ( $\lambda$ ) pode ser expresso matematicamente por um

modelo bilinear. Primeiro, devemos desdobrar o cubo da imagem em uma tabela de dados que contém os espectros de todos os *pixels* em linhas (Figura 1) (De Juan et al., 2014).



**Figura 1.** HSI-NIR de grãos de milho nas imagens espacial em Red, Green and Blue (RGB) (A) e espectral (hiperespectros ou hiper-cubo de espectros (B)). A separação do hiper-cubo (B) em matrizes bidimensionais de informação dos pixels puros (C) e dos espectros NIR ( $\lambda$ ).

A tabela de dados da Figura 1 contém a medição da imagem de grãos de milho sem tratamento. Ela pode ser expressa como o produto de uma matriz de espectros puros pelos pesos de concentração desses componentes puros em cada *pixel*. Na forma matricial, a expressão lembra a equação linear (Equação 3) (De Juan et al., 2014).

$$D = CS^T + E \quad (3)$$

As linhas da matriz  $S^T$  contém os espectros puros dos constituintes da imagem e as linhas da matriz  $C$ , os valores das concentrações destes constituintes em cada *pixel*. E contém os erros experimentais por causa da vibração de sinal não associada com a informação físico-química (De Juan et al., 2014).

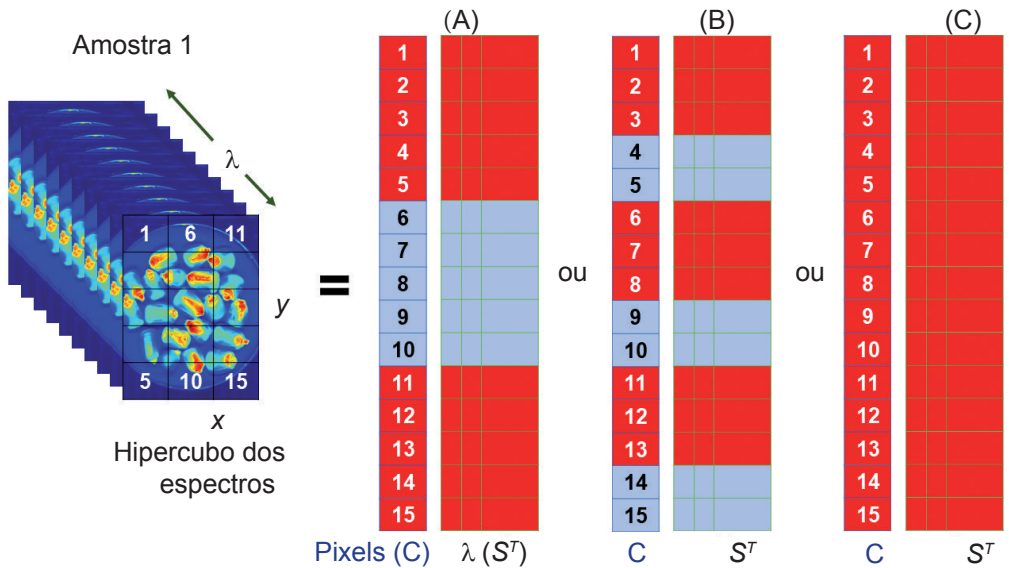
Comparando com as câmeras RGB atuais, as câmeras hiperespectrais existentes são limitadas em resolução espacial. Pode-se conseguir resoluções mais elevadas pela combinação de uma HSI-NIR de baixa resolução e uma imagem RGB de alta resolução. Aplica-se um *unmixing algorithm* para busca de uma fatoração da entrada hiperespectral em um conjunto de coeficientes de máxima dispersão, para estimar

uma base que representa os espectros de reflectância. Utiliza-se essa representação em conjunto com a entrada RGB para produzir o resultado desejado (Kawakami et al., 2011).

O HSI-NIR apresenta um grande potencial para obter informações precisas e confiáveis sobre a qualidade de produtos finais e também é uma excelente ferramenta para controle de processos. Análises quantitativas e estudos de heterogeneidade em imagens NIR são, muitas vezes, realizados por técnicas de calibração multivariada, projetadas neste contexto como regressão de imagem multivariada. Os mínimos quadrados alternados em uma resolução de curva multivariada (MCR-ALS - *Multivariate Curve Resolution – Alternating Least Square*) é outro método de análise de dados, voltado principalmente para recuperar os espectros puros e os mapas de distribuição de imagens (Piqueras et al., 2012).

Há diferentes possibilidades em relação aos componentes usados para calibração e validação do algoritmo MCR-ALS (Figura 2). Existem três opções: (A) A totalidade das imagens é analisada dentro do mesmo multiconjunto em uma única análise de resolução. Os mapas de distribuição (perfis de concentração) de cada imagem são divididos por colunas e o valor médio de uma coluna (Figura 2A, vermelho) é usado para calibração, e a média da outra coluna (Figura 2A, azul) é usada para validação. (B) Todas as imagens são divididas em duas partes antes da resolução. Quando se colocam as três colunas (1 a 5, 6 a 10 e 11 a 15) em paralelo voltando-se ao hiper-cubo, separam-se as amostras em três quintos e dois quintos para a calibração e validação (Figura 2B). A análise de resolução múltipla é executada em uma parte do hiper-cubo e os modelos de calibração são construídos. (C) O conjunto de matrizes dos hiper-cubos é dividido em um conjunto de imagens de calibração e um conjunto de imagens de validação (Figura 2C, apenas a imagem de calibração). A análise de resolução é feita no conjunto de ajuste de imagem e as linhas de calibração são construídas. As três estratégias fornecem resultados muito semelhantes com as figuras de mérito relacionadas (Piqueras et al., 2012).

A resolução do HSI-NIR é mantida a mesma utilizando os pré-processamentos da variação de linha de base e dos efeitos de dispersão multiplicativos (Wold et al., 1998). Para isso, existem dois algoritmos: a variação normal padrão padrão (SNV – Standard Normal Variation) e a correção de dispersão multiplicativa (MSC - Multiplicative Scatter Correction) que acarretam em resultados de resolução muito semelhantes e são amplamente utilizados em HSI-NIR (Fearn et al., 2009).



**Figura 2.** Hiper cubo dos grãos de milho (Amostra 1) com os esquemas de separação dos espectros NIR (A, B e C) de calibração (vermelho) e validação (azul) para análise quantitativa usando informações MCR-ALS.

O SNV opera por espectro, transformando o espectro  $x$  ( $x_1, x_2, \dots, x_p$ ) medido em comprimentos de onda  $p$  para  $z$  ( $z_1, z_2, \dots, z_p$ ) (Equação 4).

$$z_i = \frac{(x_i - m)}{s} \tag{4}$$

Em que  $m$  é a média e  $s$  é o desvio padrão dos valores  $x_1, x_2, \dots, x_p$ .

Após a transformação, o espectro  $z$  tem o comprimento quadrado  $z^T z = p - 1$  ou  $p$ , normalizado pelo divisor  $s$ , produzindo uma estrutura espectral curvada ou elíptica no espaço dimensional (Fearn et al., 2009).

O MSC transforma o espectro  $x$  em  $z$  de acordo com a Equação 5.

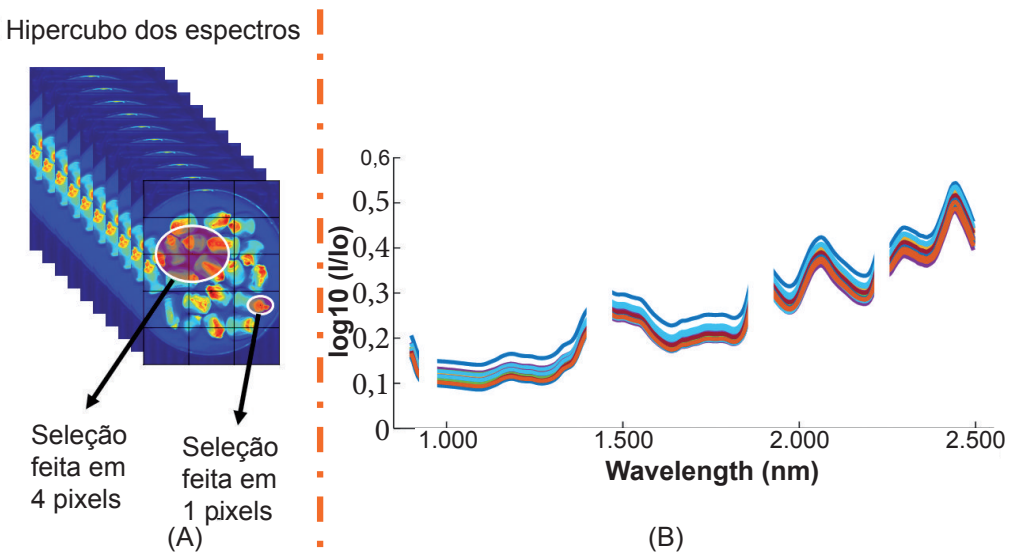
$$z_i = \frac{(x_i - a)}{b} \tag{5}$$

Em que  $a$  é a interseção e  $b$  é a inclinação da reta da regressão dos mínimos quadrados dos valores  $x_1, x_2, \dots, x_p$ , nos valores correspondentes aos coeficientes de correlação linear  $r_1, r_2, \dots, r_p$  para um espectro médio de todos os espectros disponíveis, por exemplo, o espectro médio de um conjunto de calibração. Portanto, os espectros pré-processados por MSC não são normalizados, e a geometria no espaço espectral do SNV não é a mesma do MSC.

## Aplicações de imagens hiperespectrais NIR

Os espectros medidos correspondentes aos *pixels* com o mesmo tipo de superfície exibem uma variabilidade espectral inerente que impede a caracterização de superfícies homogêneas por assinaturas espectrais únicas (Figura 3B). Isso ocorre porque o sensor integra a irradiância de toda a informação dentro da superfície de um *pixel* de imagem (Figura 3A). A variabilidade espectral e a interferência de *pixels* mistos são os principais obstáculos que precisam ser resolvidos e superados pelos algoritmos para as aplicações envolvendo HSI (Manolakis et al., 2003).

Na Figura 3(B), observa-se que as lacunas em branco nos espectros são correspondentes aos comprimentos de onda próximo das bandas de absorção de água, os quais foram descartados por causa de sua baixa relação sinal/ruído (Manolakis et al., 2003).



**Figura 3.** Ilustração de variabilidade hiperespectral e a interferência dos pixels mistos das amostras de grãos de milho. (A) A irradiância de todos os materiais dentro dos pixels ou células de resolução 3 ( $x$ ) x 5 ( $y$ ) é reduzida pelo sensor por um único pixel de imagem. (B) Os espectros medidos ( $\lambda$ ) correspondentes aos pixels com o mesmo tipo de superfície exibem uma variabilidade espectral que impede a caracterização de superfícies heterogêneas e com o *background*.

Por exemplo, o dano causado por insetos em trigo afeta a qualidade do produto e é considerado um ponto crítico na pós-colheita desta cultura. O potencial de HSI-NIR para a detecção de grãos de trigo danificados por insetos foi investigado. As sementes de trigo saudáveis e visivelmente danificadas por *Sitophilus oryzae*, *Rhizopertha dominica*, *Cryptolestes ferrugineus* e *Tribolium castaneum* foram analisadas na faixa de comprimento de onda 1.000 nm a 1.600 nm. A dimensão dos dados hiperespectrais adquiridos foi reduzida com a análise de imagens multivariadas. Seis características de imagem estatística como máximo, mínimo, médio, mediano, desvio padrão e variância, e 10 recursos de histogramas foram extraídos nas imagens reduzidas à informação nas variáveis de 1.101,69 nm e 1.305,05 nm. Usou-se, para essa aplicação, recursos quimiométricos baseados em função linear, função quadrática e equação do modelo de Mahalanobis em estratégia de classificação (Singh et al., 2009).

A HSI-NIR foi utilizada para estudar três cepas de cada um de três *Fusarium* spp. (*Fusarium subglutinans*, *Fusarium proliferatum* e *Fusarium verticillioides*) inoculados em meio de cultura batata-dextrose-ágar em placas de Petri após 72 h ou 96 h de incubação. A análise de imagens multivariadas foi utilizada para compor modelos de reconhecimento de padrão empregando PCA e PLS-DA (*Partial Least Square – Discriminant Analysis*). As análises, incluindo todas as cepas, evidenciaram quão diferentes as cepas foram umas das outras. Usando gradientes de classificação, foi possível observar a mudança no crescimento do micélio ao longo do tempo. A primeira e segunda componentes principais (PC1 e PC2) explicaram as variações entre os diferentes *Fusarium* spp. como dispersão e diferenças na produção de proteínas, respectivamente. Os resultados de previsão de PLS-DA discriminaram as espécies de *F. verticillioides*. Para *F. subglutinans*, 78%-100% de *pixels* foram corretamente previstos, dependendo dos conjuntos de treinamento e teste usados, e a porcentagem de valores preditos corretamente de *F. proliferatum* foi de 60%-80%. A visualização do crescimento radial do micélio nas imagens da pontuação na construção de PCA foi possível em razão do uso da HSI-NIR, o que seria impossível com espectroscopia nas regiões do visível ou NIR convencionais (Williams et al., 2012).

Yang et al. (2015) conseguiram classificar sementes de milho ceroso de acordo com o teor de amilopctina, através da análise simultânea de imagens hiperespectrais com as características morfológicas dos grãos. Auxiliada com ferramentas quimiométricas para classificação, como SNV, foram obtidos índices de classificação em torno de 98,2% e de 96,3%, em comparação com o modelo PLS-DA.

Outros estudos realizados por Ambrose (2016), nos quais grãos de milho foram analisados de forma não destrutiva, utilizando a técnica de HSI-NIR, associada ao modelo de classificação de Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA), apresentaram exatidão de 97,6% para calibração e 95,6% para predição de semen-



tes viáveis e não viáveis. Assim também Monteiro et al. (2007) utilizaram a HSI para analisar os teores de sacarose, glicose, frutose e de aminoácidos em soja, projetando um modelo de predição baseado em algoritmos de redes neurais artificiais, e comparando-o com outros métodos de pré-processamento como reflectância bruta, segunda derivada e PCA.

## Referências

- ABNEY, W. de W.; FESTING, E. R. On the influence of the atomic grouping in the molecules of organic bodies on their absorption in the infra-red region of the spectrum. **Philosophical Transactions of the Royal Society of London**, v. 172, p. 887-918, jan. 1881. Disponível em: <<http://rstl.royalsocietypublishing.org/cgi/doi/10.1098/rstl.1881.0020>>. Acesso em: 24 mar. 2018.
- AENUGU, H. P. R.; KUMAR, D. S.; PARTHIBAN, S. N.; GHOSH, S. S.; BANJI, D. Near infrared spectroscopy - an overview. **International Journal of ChemTech Research**, v. 3, n. 2, p. 825-836, 2011.
- AÏT KADDOUR, A.; MONDET, M.; CUQ, B. Application of two-dimensional cross-correlation spectroscopy to analyse infrared (MIR and NIR) spectra recorded during bread dough mixing. **Journal of Cereal Science**, v. 48, n. 3, p. 678-685, 2008.
- AMBROSE, A. High speed measurement of corn seed viability using hyperspectral imaging. **Infrared Physics & Technology**, v. 75, p. 173-179, 2016. Disponível em: <<https://doi.org/10.1016/j.infrared.2015.12.008>>. Acesso em: 24 mar. 2018.
- AMIGO, J. M.; BABAMORADI, H.; ELCOROARISTIZABAL, S. Hyperspectral image analysis. A tutorial. **Analytica Chimica Acta**, v. 896, p. 34-51, 2015.
- BARANSKA, M.; SCHÜTZE, W.; SCHULZ, H. Determination of lycopene and  $\beta$ -carotene content in tomato fruits and related products: comparison of FT-raman, ATR-IR, and NIR spectroscopy. **Analytical Chemistry**, v. 78, n. 24, p. 8456-8461, 2006.
- BELLON-MAUREL, V.; FERNANDEZ-AHUMADA, E.; PALAGOS, B.; ROGER, J.-M.; MCBRATNEY, A. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. **TrAC Trends in Analytical Chemistry**, v. 29, n. 9, p. 1073-1081, 2010.
- BLANCO, M.; VILLARROYA, I. NIR spectroscopy: a rapid-response analytical tool. **TrAC Trends in Analytical Chemistry**, v. 21, n. 4, p. 240-250, 2002.

BRACKETT, F. S. Characteristic differentiation in the spectra of saturated hydrocarbons. **Proceedings of the National Academy of Sciences of the United States of America**, v. 14, n. 11, p. 857-864, nov. 1928. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/16577147>>. Acesso em: 24 mar. 2018.

CASALE, M.; ZUNIN, P.; COSULICH, M. E.; PISTARINO, E.; PEREGO, P.; LANTERI, S. Characterization of table olive cultivar by NIR spectroscopy. **Food Chemistry**, v. 122, n. 4, p. 1261-1265, 2010.

DE JUAN, A.; MAEDER, M.; HANCEWICZ, T.; DUPONCHEL, L.; TAULER, R. Chemometric tools for image analysis. In: SALZER, R.; SIESLER, H. W. (Org.). **Infrared and raman spectroscopic imaging**. 2. ed. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 2014. p. 57-110.

DIXIT, Y.; CASADO-GAVALDA, M. P.; CAMA-MONCUNILL, R.; CAMA-MONCUNILL, X.; MARKIEWICZ-KESZYCKA, M.; CULLEN, P. J.; SULLIVAN, C. Developments and challenges in online NIR spectroscopy for meat processing. **Comprehensive Reviews in Food Science and Food Safety**, v. 16, p. 1172-1187, 2017.

DORREPAAL, R.; MALEGORI, C.; GOWEN, A. Tutorial: time series hyperspectral image analysis. **Journal of Near Infrared Spectroscopy**, v. 24, n. 2, p. 89-107, 2016.

FEARN, T.; RICCIOLI, C.; GARRIDO-VARO, A.; GUERRERO-GINEL, J. E. On the geometry of SNV and MSC. **Chemometrics and Intelligent Laboratory Systems**, v. 96, n. 1, p. 22-26, 2009.

FERNÁNDEZ PIERNA, J. A.; VERMEULEN, P.; AMAND, O.; TOSSENS, A.; DARDENNE, P.; BAETEN, V. NIR hyperspectral imaging spectroscopy and chemometrics for the detection of undesirable substances in food and feed. **Chemometrics and Intelligent Laboratory Systems**, v. 117, p. 233-239, 2012.

GELADI, P.; BURGER, J.; LESTANDER, T. Hyperspectral imaging: calibration problems and solutions. **Chemometrics and Intelligent Laboratory Systems**, v. 72, n. 2, p. 209-217, 2004

GEORGIEVA, M.; NEBOJAN, I.; MIHALEV, K.; YONCHEVA, N. Application of NIR spectroscopy and chemometrics in quality control of wild berry fruit extracts during storage. **Croatian Journal of Food Technology**, v. 8, n. 3/4, p. 67-73, 2014. Disponível em: <[http://hrcak.srce.hr/index.php?show=clanak&id\\_clanak\\_jezik=171380](http://hrcak.srce.hr/index.php?show=clanak&id_clanak_jezik=171380)>. Acesso em: 24 mar. 2018.

HERSCHEL, W. XIII. Investigation of the powers of the prismatic colours to heat and illuminate objects; with remarks, that prove the different refrangibility of radiant heat. To which is added, an inquiry into the method of viewing the sun advantageously, with telescopes of large apertures and high magnifying powers. **Philosophical Transactions of the Royal Society**

of London, v. 90, p. 255-283, 1800. Disponível em: <<http://rstl.royalsocietypublishing.org/cgi/doi/10.1098/rstl.1800.0014>>. Acesso em: 24 mar. 2018.

HERZBERG, G. **Molecular spectra and molecular structure: Infrared and Raman Spectra of Polyatomic molecules**. New York: D. Van Nostrand Company, 1945. v. 2, 644 p.

INGLE, P. D.; CHRISTIAN, R.; PUROHIT, P.; ZARRAGA, V.; HANDLEY, E.; FREEL, K.; ABDO, S. Determination of protein content by NIR spectroscopy in protein powder mix products. **Journal of AOAC International**, v. 99, n. 2, p. 360-363, 2016.

JARÉN, C.; ORTUÑO, J. C.; ARAZURI, S.; ARANA, J. I.; SALVADORES, M. C. Sugar determination in grapes using NIR technology. **International Journal of Infrared and Millimeter Waves**, v. 22, n. 10, p. 1521-1530, 2001.

KAWAKAMI, R.; WRIGHT, J.; TAI, Y.-W.; MATSUSHITA, Y.; BEN-EZRA, M.; IKEUCHI, K. High-resolution hyperspectral imaging via matrix factorization. In: CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2011, Colorado Springs. **Proceedings...** Colorado Springs: IEEE Computer Society: University of Colorado, 2011. p. 2329-2336.

KENASTON, B.; CROCKETT, C. ASD - Highland Precision Agriculture case study: improving precision agriculture with visual and near-infrared spectroscopy. **NIR news** v. 29, n. 2, p. 7-9, 2018. Disponível em: <<http://journals.sagepub.com/doi/10.1177/0960336017751299>>. Acesso em: 24 mar. 2018.

LEE, W.-H.; KIM, M. S.; LEE, H.; DELWICHE, S. R.; BAE, H.; KIM, D.-Y.; CHO, B.-K. Hyperspectral near-infrared imaging for the detection of physical damages of pear. **Journal of Food Engineering**, v. 130, p. 1-7, 2014.

MANOLAKIS, D.; MARDEN, D.; SHAW, G. hyperspectral image processing for automatic target detection applications. **Lincoln Laboratory Journal**, v. 14, n. 1, p. 79-116, 2003.

MONTEIRO, S. T.; MINEKAWA, Y.; KOSUGI, Y.; AKAZAWA, T.; ODA, K. Prediction of sweetness and amino acid content on soybean crops from hyperspectral imagery. **Journal of Photogrammetry & Remote Sensing**, v. 62, p. 2-12. 2007.

NGADI, M. O.; LIU, L. Hyperspectral image processing techniques. In: SUIN, D.-W. (Ed.). **Hyperspectral imaging for food quality analysis and control**. London: Academic Press, 2010. Chap. 4, p. 99-127.

NICOLAÏ, B. M.; BEULLENS, K.; BOBELYN, E.; PEIRS, A.; SAEYS, W.; THERON, K. I.; LAMMERTYN, J. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. **Postharvest Biology and Technology**, v. 46, n. 2, p. 99-118, 2007.

NOGALES-BUENO, J.; RODRÍGUEZ-PULIDO, F. J.; HEREDIA, F. J.; HERNÁNDEZ-HIERRO, J. M. Comparative study on the use of anthocyanin profile, color image analysis and near-

infrared hyperspectral imaging as tools to discriminate between four autochthonous red grape cultivars from la Rioja (Spain). **Talanta**, v. 131, p. 412-416, 2015.

PIQUERAS, S.; BURGER, J.; TAULER, R.; JUAN, A. Relevant aspects of quantification and sample heterogeneity in hyperspectral image resolution. **Chemometrics and Intelligent Laboratory Systems**, v. 117, p. 169-182, ago. 2012. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/S0169743911002577>>. Acesso em: 28 mar. 2018.

PRIETO, N.; ROEHE, R.; LAVÍN, P.; BATTEN, G.; ANDRÉS, S. Application of near infrared reflectance spectroscopy to predict meat and meat products quality: a review. **Meat Science**, v. 83, n. 2, p. 175-186, 2009.

RAVN, C.; SKIBSTED, E.; BRO, R. Near-infrared chemical imaging (NIR-CI) on pharmaceutical solid dosage forms-Comparing common calibration approaches. **Journal of Pharmaceutical and Biomedical Analysis**, v. 48, n. 3, p. 554-561, 2008.

RUCKEBUSCH, C.; ORHAN, F.; DURAND, A.; BOUBELLOUTA, T.; HUVENNE, J. P. Quantitative analysis of cotton-polyester textile blends from near-infrared spectra. **Applied Spectroscopy**, v. 60, n. 5, p. 539-544, 2006.

RUZICKOVÁ, J.; SUSTOVÁ, K. Determination of selected parameters of quality of the dairy products by NIR spectroscopy. **Czech Journal of Food Sciences**, v. 24, n. 6, p. 255-260, 2006. Disponível em: <<http://www.agriculturejournals.cz/publicFiles/50302.pdf>>. Acesso em: 24 mar. 2018.

SINELLI, N.; CERRETANI, L.; DI EGIDIO, V.; BENDINI, A.; CASIRAGHI, E. Application of near (NIR) infrared and mid (MIR) infrared spectroscopy as a rapid tool to classify extra virgin olive oil on the basis of fruity attribute intensity. **Food Research International**, v. 43, n. 1, p. 369-375, 2010.

SINGH, C. B.; JAYAS, D. S.; PALIWAL, J.; WHITE, N. D. G. Detection of insect-damaged wheat kernels using near-infrared hyperspectral imaging. **Journal of Stored Products Research**, v. 45, n. 3, p. 151-158, 2009.

MITH, K. L.; STEVEN, M. D.; COLLS, J. J. Use of hyperspectral derivative ratios in the red-edge region to identify plant stress responses to gas leaks. **Remote Sensing of Environment**, v. 92, n. 2, p. 207-217, 2004.

UTO, K.; SEKI, H.; SAITO, G.; KOSUGI, Y. Characterization of rice paddies by a UAV-mounted miniature hyperspectral sensor system. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 6, n. 2, p. 851-860, 2013.

WEYER, L. G. Near-infrared spectroscopy of organic substances. **Applied Spectroscopy Reviews**, v. 21, n. 1/2, p. 1-43, 1985.

WHEELER, O. H. Near infrared spectra of organic compounds. **Chemical Reviews**, v. 59, n. 4, p. 629-666, 1959.

WILLIAMS, P. J.; GELADI, P.; BRITZ, T. J.; MANLEY, M. Near-infrared (NIR) hyperspectral imaging and multivariate image analysis to study growth characteristics and differences between species and strains of members of the genus *Fusarium*. **Analytical and Bioanalytical Chemistry**, v. 404, n. 6/7, p. 1759-1769, 2012.

WINDIG, W.; MARKEL, S.; THOMPSON, P. M. Self-modeling image analysis with SIMPLISMA. **Techniques and applications of hyperspectral image analysis**. Chichester, UK: John Wiley & Sons, 2007. p. 69-87. Disponível em: <<http://doi.wiley.com/10.1002/9780470010884.ch4>>. Acesso em: 30 maio 2018.

WOLD, S. Orthogonal signal correction of near-infrared spectra. **Chemometrics and Intelligent Laboratory Systems**, v. 44, n. 1/2, p. 175-185, Dec. 1998.

WORKMAN JR., J. J. NIR spectroscopy calibration basics. In: BURNS, D. A.; CIURCZAK, E. W. (Ed.). **Handbook of near-infrared analysis**. 3. ed. Boca Raton: CRC Press, 2007. p. 123-149. (Practical spectroscopy, 35). Disponível em: <<http://books.google.lu/books?id=6EEed1a0uka0C>>. Acesso em: 30 maio 2018.

YANG, X.; HONG, H.; YOU, Z.; CHENG, F. Spectral and image integrated analysis of hyperspectral data for waxy corn seed variety classification. **Sensors**, v. 15, n. 7, p. 15578-15594, July 2015. Disponível em: <[https://www.researchgate.net/publication/280030104\\_Spectral\\_and\\_Image\\_Integrated\\_Analysis\\_of\\_Hyperspectral\\_Data\\_for\\_Waxy\\_Corn\\_Seed\\_Variety\\_Classification](https://www.researchgate.net/publication/280030104_Spectral_and_Image_Integrated_Analysis_of_Hyperspectral_Data_for_Waxy_Corn_Seed_Variety_Classification)>. Acesso em: 19 jul. 2018.

# Aplicação de Imagens Hiperespectrais na Detecção de Fungos e na Predição de Micotoxinas em Grãos de Trigo

Jayme Garcia Arnal Barbedo

Casiane Salete Tibola

Maria Imaculada Pontes Moreira Lima

## Introdução

Grãos são a base da cadeia alimentar, servindo tanto como matéria-prima para a indústria alimentícia quanto para formulação de rações nas cadeias de produção animal. Assim, grande importância é atribuída à análise em termos de qualidade e de contaminantes, fatores estes que podem reduzir o valor comercial dos lotes e causar danos à saúde de pessoas e animais. Historicamente, essas análises são realizadas manualmente, seja por inspeção visual, seja através da seleção de amostras a serem submetidas a análises laboratoriais. Essa abordagem manual, apesar de acurada, tende a ser demorada e cara. Além disso, se a amostragem não é realizada corretamente, lotes contaminados podem não ser detectados.

Por esse motivo, técnicas alternativas capazes de superar algumas das limitações das técnicas tradicionais vêm sendo investigadas cada vez com mais frequência. Dentre essas, a aplicação de imagens hiperespectrais vem ganhando destaque à medida que o custo diminui e a resolução espectral e espacial dos equipamentos aumenta. Algumas aplicações têm recebido considerável atenção nos últimos anos em razão da sua importância comercial e das características do problema a ser resolvido. Alguns dos trabalhos de maior destaque nessas aplicações são brevemente descritos a seguir.

Insetos podem causar sérios prejuízos à produção de grãos, tanto pela indução de brotamento (Singh et al., 2010a) quanto pela deterioração da qualidade física e nutricional (Kaliramesh et al., 2013). Singh et al. (2009, 2010a, 2010b) focaram seus esforços na detecção de danos causados por insetos em grãos de trigo. Em 2009, os autores utilizaram análise multivariada de imagens, atributos estatísticos e de histograma e classificadores estatísticos para detectar danos causados por quatro espécies de insetos. Os trabalhos seguintes, publicados em 2010, empregaram abordagens similares, variando o número de atributos e, em alguns casos, utilizando redes neurais do tipo “backpropagation” para classificação. A detecção de grãos de feijão-mungo danificados por insetos foi o foco de trabalho de Kaliramesh et al. (2013), os

quais extraíram, a partir das bandas mais relevantes (1.100 nm, 1.290 nm e 1.450 nm), os mesmos atributos sugeridos por Singh et al. (2009). Chelladurai et al. (2014) combinaram imagens hiperespectrais e de raio-x para detectar grãos de soja danificados, a partir das quais 81 atributos foram extraídos e usados para alimentar classificadores de análise discriminante linear e quadrática, alcançando taxa de acerto próxima a 80%.

A classificação é importante na determinação da qualidade de lotes de sementes. Essa classificação pode buscar a identificação de diferentes cultivares de um mesmo lote (Choudhary et al., 2009) ou de sementes pertencentes a diferentes espécies (Serranti et al., 2013b). O método proposto por Mahesh et al. (2008) fez a detecção de oito diferentes cultivares de trigo utilizando um classificador estatístico e redes neurais. Uma abordagem similar foi usada por Choudhary et al. (2009) na classificação das mesmas oito cultivares de trigo. A detecção de diferentes cultivares de milho foi o objetivo de Nansen et al. (2008), os quais utilizaram análise discriminante e análise de regressão, Zhang et al. (2012), os quais utilizaram redes neurais, e Wang et al. (2016), os quais utilizaram máquinas de suporte vetorial. A detecção de cultivares de outras culturas, como arroz (Wang et al., 2015a) e aveia (Serranti et al., 2013b), tem também recebido atenção da comunidade científica.

A detecção de danos causados por brotamento é outra tarefa que vem se beneficiando da informação contida em imagens hiperespectrais. Xing et al. (2010) utilizaram análise de componentes principais, combinada com a extração de atributos morfológicos, para detectar brotamentos em grãos de trigo. O método proposto por McGoverin et al. (2011) utilizou análise de componentes principais e análise discriminante para detectar brotamento em trigo, cevada e sorgo. Arngren et al. (2011) utilizaram classificação supervisionada baseada em um conjunto de atributos para detectar germinação em grãos de cevada.

Pierna et al. (2012) combinaram imagens hiperespectrais com quimiometria para detectar uma variedade de impurezas em trigo e cevada. A detecção de impurezas em grãos de trigo foi também o objetivo do método proposto por Ravikanth et al. (2015), os quais empregaram diversas técnicas (Máquinas de Suporte Vetorial, Naïve Bayes, k-vizinhos mais próximos) para classificar os contaminantes.

A detecção de sintomas de doenças causadas por fungos e de seus danos é especialmente a principal aplicação de imagens hiperespectrais na análise de grãos, especialmente no caso do trigo. Outro uso comum desse tipo de imagem é na quantificação de micotoxinas associadas a esses fungos, aplicação esta que tem sido mais comum no caso do milho (Wang et al., 2014; Wang et al., 2015b; Kandpal et al., 2015; Wang et al., 2015c). Essas aplicações (detecção de fungos e predição de micotoxinas), no contexto da análise de grãos de trigo, são o principal objeto deste capítulo e o foco da próxima seção.

## Estado da arte da detecção de doenças e micotoxinas em grãos de trigo utilizando imagens hiperespectrais

A giberela, doença fúngica que ocorre em cereais, destaca-se pela importância que representa em áreas de produção de trigo no mundo. O agente causal da doença no Brasil é *Gibberella zeae* Petch (Schwein.) (anamorfo: complexo de espécies em *Fusarium graminearum* Schwabe) (Del Ponte et al., 2013). Além dos danos diretos no rendimento de grãos e na redução da qualidade da farinha, a giberela pode ocasionar a contaminação de grãos com metabólitos secundários tóxicos, conhecidos como micotoxinas. No trigo, prevalecem as micotoxinas deoxinivalenol (DON) e zearalenona (ZEA). Para proteger os consumidores das micotoxicoses, muitos países estabeleceram níveis máximos permitidos para as micotoxinas prevalentes em cereais e derivados. No Brasil, o limite máximo permitido de DON em trigo não processado é de 3.000 ppb, na farinha é de 1.000 ppb e nos produtos destinados à alimentação infantil é de 200 ppb (Brasil, 2017).

A detecção de doenças e de seus danos em grãos de trigo têm recebido considerável atenção da comunidade acadêmica, dada a importância econômica e social desta atividade. À medida em que os equipamentos para captura de imagens hiperespectrais se tornam mais baratos e robustos, cresce a probabilidade de que técnicas baseadas neste tipo de informação venham a ser utilizadas na prática.

Um dos primeiros métodos a aplicar imagens hiperespectrais na detecção de doenças em trigo foi proposto por Polder et al. (2005). Os autores combinaram espectrografia e processamento de imagens para detectar giberela em grãos de trigo, utilizando regressão por mínimos quadrados parciais e agrupamento *fuzzy* do tipo *C-means* para relacionar a giberela com os espectros. Os resultados mostraram que a faixa do infravermelho próximo é mais adequada que o espectro visível para esse fim.

Outro estudo em trigo, envolvendo o uso de imagens hiperespectrais, foi feito por Berman et al. (2007) e teve como objetivo detectar três tipos de manchas, sendo duas causadas por fungos, que reduzem o valor comercial de grãos de trigo. Os autores aplicaram análise discriminante penalizada ao formato do espectro para classificar os pixels, e uma regra simples para classificação da qualidade dos grãos, alcançando índices de acerto acima de 95%.

Os fungos *Penicillium* spp., *Aspergillus glaucus* e *Aspergillus niger* foram o alvo do trabalho de Singh et al. (2007), os quais utilizaram análise multivariada de imagens baseada em análise de componentes principais (PCA) para reduzir a dimensionalidade dos dados, e *k-means* e análise discriminante (linear, quadrática e de Mahalanobis)



para realizar a classificação, também alcançando acertos acima de 95%. Os mesmos autores publicaram a evolução de seu trabalho cinco anos depois (Singh et al., 2012), a qual incluiu o cálculo de diversos atributos de cor e textura, além do uso das imagens hiperespectrais.

Shahin e Symons (2011) propuseram método para classificar grãos de trigo saudáveis, levemente danificados ou muito danificados pela giberela através da combinação de análise de componentes principais com análise discriminante linear. Os autores conseguiram chegar a índices de acerto de 92% utilizando apenas seis bandas espectrais, resultado este similar àquele obtido utilizando todas as bandas. Os mesmos autores publicaram outro trabalho em que a análise discriminante foi a de mínimos quadrados parciais, ao invés da linear (Shahin; Symons, 2012).

O problema da detecção de danos causados por giberela foi abordado por Delwiche et al. (2011), tendo como base a análise discriminante linear. Os autores alcançaram índices de acerto acima de 95%, argumentando que a região do infravermelho próximo, especialmente ao redor da banda de 1.200 nm (faixa de absorção do ergosterol), foi a que mais contribuiu para os resultados.

O objetivo do trabalho de Vermeulen et al. (2012) foi o de detectar ergotismo em grãos de trigo, combinando um escâner de imagens hiperespectrais e uma esteira de transporte de grãos. Os autores reportaram correlações acima de 99% entre os valores estimados utilizando análise discriminante e máquinas de vetores suporte e os valores de referência, destacando que foi possível identificar grãos com níveis de contaminação tão baixos quanto 0,01%. Os mesmos autores publicaram outro trabalho em que a técnica foi estendida para grãos de centeio (Vermeulen et al., 2013).

O método proposto por Serranti et al. (2013a) visou a identificar três tipos de problemas que acometem grãos de trigo (giberela, grãos vítreos e *yellow berry*). Este método também se baseou em análise de componentes principais para exploração dos dados e análise discriminante para classificação dos grãos. Bons resultados foram obtidos utilizando apenas três faixas estreitas do espectro.

Bauriegel e Herppich (2014) compararam técnicas de imageamento hiperespectral e fluorescência de clorofila para detectar giberela em grãos de trigo. Foi observado que as imagens hiperespectrais detectaram as modificações causadas por giberela nas assinaturas espectrais dos grãos, enquanto que a fluorescência detectou a diminuição da atividade fisiológica causada pelo fungo. Os autores concluíram que o imageamento é mais apropriado quando os sintomas se tornaram visíveis, enquanto que a fluorescência, por depender de medições ao longo do tempo, pode restringir a

velocidade da detecção. Afirmaram que, se ambas as técnicas pudessem ser combinadas, a taxa de detecção da giberela aumentaria consideravelmente.

Mais recentemente, pesquisadores da Empresa Brasileira de Pesquisa Agropecuária (Embrapa) desenvolveram um método para detectar giberela em grãos de trigo, o qual obteve bons resultados utilizando apenas algumas bandas espectrais sem depender de técnicas tradicionalmente utilizadas, como análise de componentes principais e análise discriminante (Barbedo et al., 2015). Esse método foi posteriormente estendido para a predição dos níveis de deoxinivalenol em lotes de trigo (Barbedo et al., 2017), algo ainda inédito para esta cultura, considerando que outros trabalhos reportados focaram no milho.

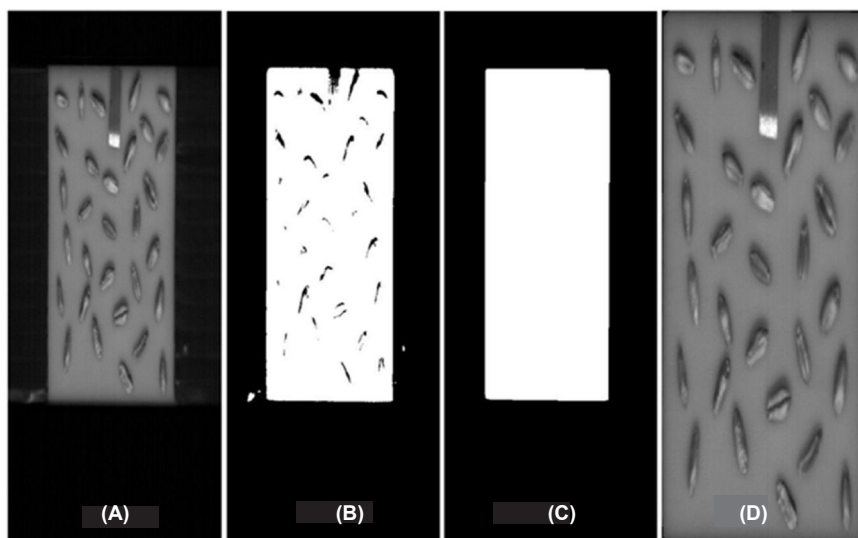
## **Métodos rápidos para detecção de giberela e predição do nível de DON**

Apesar dos avanços alcançados até o momento na detecção de giberela em grãos de trigo, ainda há muitas demandas para melhorar a acurácia e o tempo de análise. Com a finalidade de superar alguns dos problemas reportados na literatura até então, pesquisadores da Embrapa desenvolveram um novo método cuidadosamente projetado para incorporar o conhecimento existente a respeito das respostas espectrais de grãos de trigo saudáveis e contaminados (Barbedo et al., 2015).

Apesar dos princípios explorados aqui serem, basicamente, os mesmos usados em outros trabalhos encontrados na literatura, o algoritmo proposto possui algumas vantagens que nem sempre são encontradas em seus predecessores, como simplicidade de implementação, alta velocidade de execução, ausência de ajustes manuais, capacidade de separação de grãos aglomerados e existência de interface gráfica. O algoritmo foi calibrado e testado utilizando imagens capturadas pelo espectrômetro EV/NIR Hyperspec Model 1003B-10151 (Headwall Photonics Inc., Fitchburg, MA, USA), o qual empregou um sensor InGaAs (resolução de 320 pontos x 256 bandas) e uma câmera XENICS (Model XEVA-1246 XC 134, Leuven, Belgium). As reflectâncias foram medidas na faixa de 528 nm-1.785 nm, com resolução espectral entre 5 nm e 7 nm, resultando em 256 bandas. Um total de 800 linhas foi capturado, sendo a região de interesse delimitada por cerca de 500 dessas linhas. Mais detalhes sobre como as imagens foram capturadas podem ser encontrados em Barbedo et al. (2015).

O primeiro passo do algoritmo consiste na determinação da região de interesse (ROI), a qual deve incluir somente a superfície contendo os grãos de trigo, descartando todo o restante. Testes preliminares revelaram que a banda 25 (647 nm, Figura 1A) fornece o melhor contraste entre a ROI e a área a ser descartada. Todos os pixels

dessa banda são normalizados, e a imagem resultante é limiarizada de maneira que todos os pixels com valor maior que certo limite assumem valor 1, e todos os demais pixels assumem valor zero. Por causa do ruído inerente às imagens capturadas, a imagem binária resultante possui alguns elementos espúrios fora da ROI e algumas falhas dentro da ROI (Figura 1B). Para corrigir isso, algumas regras são aplicadas: a) apenas o maior componente conectado (grupo de pixels brancos) é mantido; b) a imagem é submetida a uma abertura morfológica utilizando um disco de 5 pixels de diâmetro como elemento estruturante, eliminando elementos espúrios nas bordas da ROI; c) todas as falhas na imagem resultante são preenchidas; d) a casca convexa do maior objeto é tomada como a região de interesse de fato. Como resultado, tem-se a máscara mostrada na Figura 1C, a qual, quando aplicada à imagem original, gera o resultado esperado (Figura 1D).

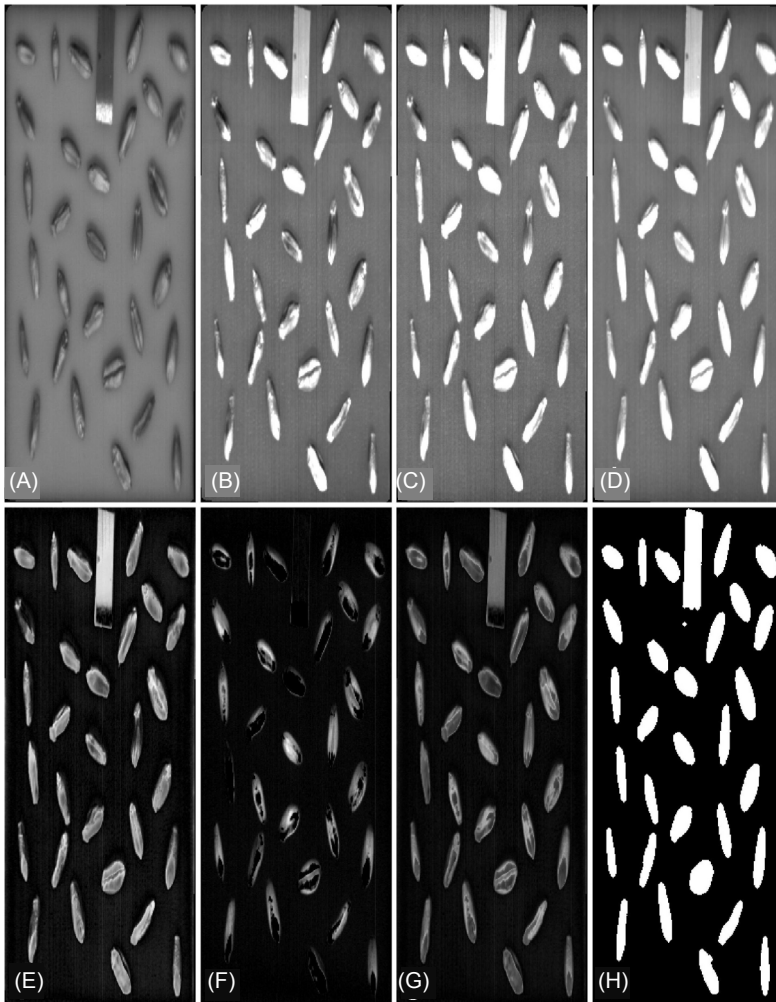


**Figura 1.** Processo de delineamento da região de interesse. A) A banda de 647 nm é selecionada. B) Imagem binária é gerada. C) Falhas são preenchidas e elementos espúrios são eliminados. D) Região de interesse é isolada.

Fonte: Barbedo et al., 2015.

Após a determinação da ROI, os grãos devem ser segmentados, isto é, separados do restante da imagem. Em razão da complexidade do processo, quatro bandas espectrais foram selecionadas para a tarefa: 672 nm (Figura 2A), 1.361 nm (Figura 2B), 1.509 nm (Figura 2C), e 1.657 nm (Figura 2D). Após uma normalização, essas quatro bandas são aritmeticamente combinadas duas a duas (2A com 2D, e 2B com 2C), e as duas imagens resultantes (Figuras 2E e 2F) são novamente combinadas, gerando a imagem mostrada na Figura 2G. A imagem resultante é limiarizada, gerando a má-

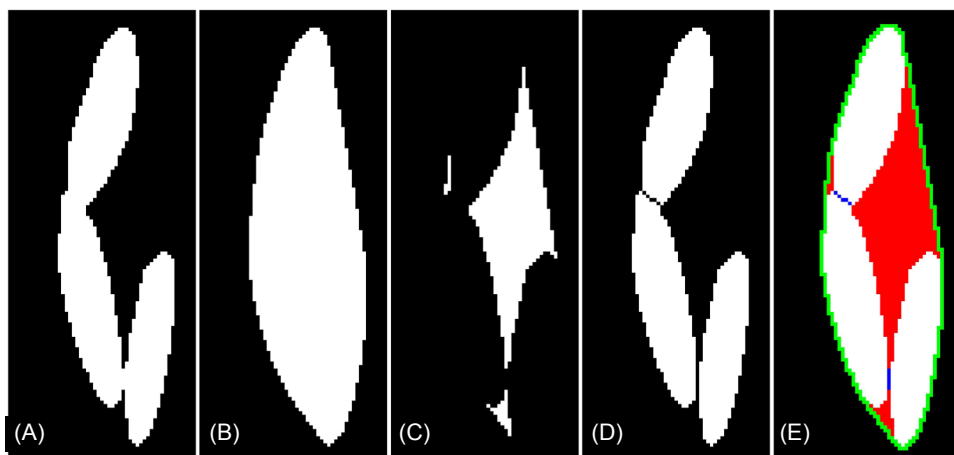
cara com o delineamento dos grãos (Figura 2H). Pode-se notar que alguns elementos espúrios permanecem após as operações, os quais são eliminados através da aplicação de algumas regras relacionadas ao tamanho dos objetos (Barbedo et al., 2015).



**Figura 2.** Processo de determinação da máscara para os grãos de trigo. Bandas de 672 nm (A), 1.361 nm (B), 1.509 nm (C) e 1.657 nm (D) são selecionadas e combinadas em duas novas imagens (E e F). Estas duas novas imagens são então combinadas (G) e binarizadas (H), gerando a máscara onde os grãos estão localizados.

Fonte: Barbedo et al., 2015.

Dependendo da posição dos grãos, alguns deles podem aparecer aglomerados nas imagens, sendo, portanto, tratados como entidades individuais. A fim de corrigir possíveis erros associados a esse fato, o método proposto inclui um conjunto de regras e procedimentos para separar os grãos aglomerados sem causar distorções no seu formato. Essas regras baseiam-se no fato de que os grãos possuem um formato quase perfeitamente convexo, propriedade esta que é perdida em aglomerados de grãos. Em linhas gerais, o procedimento adotado identifica o número de regiões côncavas que surgem ao se gerar uma casca convexa ao redor do aglomerado, e a partir dessa informação ele infere o número provável de grãos e as linhas mais apropriadas para separá-los de maneira adequada. Os passos utilizados na separação dos grãos podem ser visualizados na Figura 3, e mais detalhes podem ser encontrados em Barbedo et al. (2015).



**Figura 3.** Processo de separação de aglomerados de grãos de trigo, pela técnica: (A) grãos aglomerados; (B) casca convexa envolvendo os grãos; (C) regiões côncavas dentro da casca convexa, obtidas subtraindo (A) de (B); (D) separação das sementes dada pelas linhas mais curtas que conectam as regiões encontradas em (C); (E) representação de todos os elementos usados no processo.

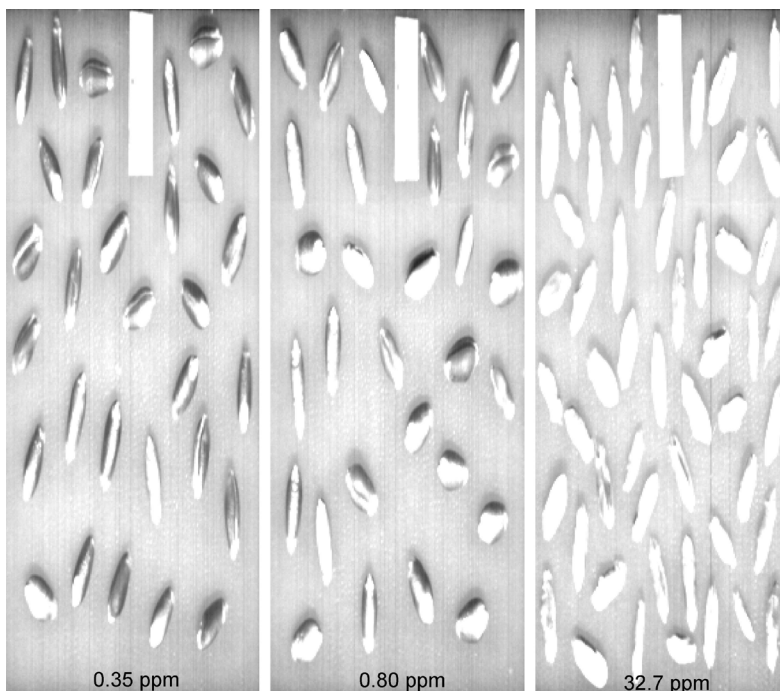
Fonte: Barbedo et al., 2015.

O passo final do algoritmo visa à identificação de grãos doentes e saudáveis. A banda espectral normalizada de 1.411 nm foi escolhida para esse fim. A partir dessa banda, é calculado o Índice Giberela (IG), o qual é dado pela proporção de pixels em um grão com valores maiores que 0,58. Quanto maior o valor de IG, maior a probabilidade de a giberela estar presente. Se uma abordagem absoluta é considerada, valores de IG acima de 0,5 indicam a presença de giberela, enquanto que valores menores que 0,5 indicam grãos saudáveis. Porém, uma estratégia mais apropriada é considerar esses valores como uma estimativa de que a giberela esteja de fato presente.

O algoritmo para identificação de grãos com sintomas de giberela está disponível em <https://www.digipathos.cnptia.embrapa.br>.

Tendo como base o algoritmo descrito acima, em 2017 foi proposta uma estratégia para estimar os níveis da micotoxina deoxinivalenol (DON) em lotes de grãos de trigo (Barbedo et al., 2017). Ambos os algoritmos foram idênticos até a separação de aglomerados de grãos. A primeira diferença entre eles foi o fato de que, enquanto o cálculo de IG faz uso apenas da banda espectral de 1.411 nm, o cálculo do nível de DON utiliza a diferença entre as bandas de 623 nm e 1.411 nm. A imagem de referência (IR) resultante evidenciou os efeitos espectrais causados pelo DON e atenuou as estruturas que não foram afetadas pela micotoxina, fornecendo assim melhor contraste entre grãos saudáveis e doentes. Após os grãos serem isolados, a proporção K de pixels com valor acima de 0,48 é calculada, e os resultados de todos os grãos são elevados à quarta potência e somados, resultando no chamado índice preliminar de DON (IPD).

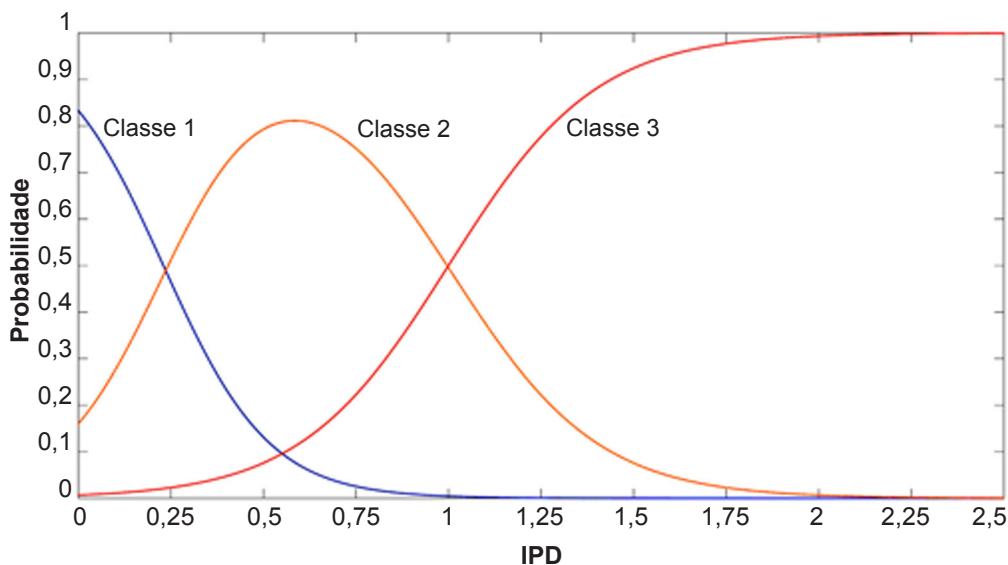
Três classes de concentração de DON foram definidas no desenvolvimento do algoritmo: classe 1 (<0,5 mg/kg – nível baixo), classe 2 (>0,5 mg/kg e <1,5 mg/kg – nível médio) e classe 3 (>1,5 mg/kg – nível alto). À medida que os níveis de DON aumentam, os grãos aparecem mais brilhantes (Figura 4).



**Figura 4.** Reflectâncias típicas para diferentes concentrações de DON.

Fonte: Barbedo et al., 2017.

A partir dos resultados para todas as imagens utilizadas na calibração do algoritmo, foram geradas curvas para as distribuições de probabilidade de cada uma das classes (Figura 5). As expressões para cada uma dessas curvas são apresentadas em Barbedo et al. (2017). Idealmente, a sobreposição entre as curvas seria mínima, porém, na prática, há uma superposição moderada entre as curvas, o que explica a maior parte dos erros observados nos testes realizados. Os valores de IPD nos quais as curvas das distribuições de probabilidade se cruzam foram adotados como os limites entre as diferentes classes. Assim, as classes ficaram definidas da seguinte maneira: classe 1:  $IPD \leq 0,26$ , classe 2:  $0,26 < IPD \leq 0,96$ , classe 3:  $IPD > 0,96$ . Uma vantagem de utilizarem-se distribuições de probabilidade é o fato de estas poderem ser substituídas no caso de ser consideradas outras cultivares de trigo.



**Figura 5.** Distribuições de probabilidade para cada uma das classes de concentração de DON consideradas, com respeito ao índice preliminar de DON (IPD). Classe 1: concentração de DON  $> 0,5$  mg/kg; classe 2: concentração de DON de 0,5 mg/kg a 1,5 mg/kg; classe 3: concentração de DON  $< 1,5$  mg/kg.

Fonte: Barbedo et al., 2017.

Pode-se observar que o desempenho do método é ao menos tão bom quanto o de outras estratégias, porém utilizando um algoritmo mais simples e rápido (Tabela 1). É importante destacar que o uso de PCA não melhorou os resultados, porque os componentes principais gerados por essa análise tendem a convergir para as bandas com melhor contraste, as quais foram selecionadas manualmente no algoritmo proposto.

**Tabela 1.** Comparação do algoritmo com alguns de seus predecessores na avaliação de grãos com sintomas de giberela em trigo. Todos os métodos basearam-se em imagens hiperespectrais.

Algoritmo	Erro
Proposto	9%
Delwiche e Kim (2000)	2% – 17%
Menesatti et al. (2009)	9% – 18%
Shahin e Symons (2011)	8%
Delwiche et al. (2011)	4% – 10%

Fonte: Barbedo et al., 2015.

Na matriz de confusão obtida nos testes com o algoritmo para predição dos níveis de DON, as colunas representam as classes estimadas e as linhas representam as classes de referência (Tabela 2). Os valores na diagonal principal (em cinza) representam a porcentagem de classificações corretas. O índice geral de acerto foi de 72%.

**Tabela 2.** Matriz de confusão obtida para o algoritmo para predição dos níveis de DON em grãos de trigo, com valores dados em porcentagens. Classe 1: concentração de DON >0,5 mg/kg; classe 2: concentração de DON de 0,5 mg/kg a 1,5 mg/kg; classe 3: concentração de DON <1,5 mg/kg.

	Classe 1	Classe 2	Classe 3
Classe 1	58	31	11
Classe 2	21	71	8
Classe 3	4	21	75

Fonte: Barbedo et al., 2017.

## Discussão, tendências e perspectivas

As pesquisas voltadas ao uso de imagens hiperespectrais na detecção de sintomas de doenças e na predição dos níveis de micotoxinas vêm crescendo consideravelmente, mas ainda há alguns desafios que precisam ser vencidos antes que esse tipo de técnica possa ser adotado em análises de rotina. Esta seção visa a discutir alguns desses desafios e possíveis soluções para os problemas associados.

A eficiência dos métodos para detecção de sintomas de doenças, e em particular da giberela, é prejudicada pela falta de amostras de referência rotuladas de maneira objetiva e consistente. Em geral, as amostras de referência são rotuladas de acordo com as observações visuais realizadas por pessoas com experiência na detecção de sintomas de doenças. Essa estratégia embute certo grau de subjetividade, o que faz com que o próprio conjunto de referência esteja sujeito a erros. Incerteza adicional



é introduzida quando a infecção está nos estágios iniciais, quando a manifestação visual da doença é praticamente imperceptível tanto para humanos quanto para computadores. Nesse contexto, é muito difícil ter certeza do diagnóstico, qualquer que seja o método utilizado.

Um fenômeno observado nos testes com o algoritmo foi o fato de as alterações espectrais causadas pela giberela aparecerem nas imagens menos pronunciadas para grãos alongados que para grãos arredondados. Hipóteses podem explicar isso. Primeiro, é possível que algumas das cultivares que têm formato alongado sejam mais resistentes à giberela. Segundo, como grãos alongados têm uma relação perímetro/área maior, efeitos de borda como sombras e distorções de pixels tendem a ter maior impacto. Adicionalmente, esse formato pode tornar mais difícil a detecção visual dos sintomas, prejudicando o processo de seleção dos grãos. Porém, não há dados concretos que suportem qualquer uma dessas possibilidades, indicando que mais estudos são necessários para uma explicação conclusiva.

O problema oposto é observado para grãos redondos, para os quais é relativamente comum que grãos saudáveis tenham altos valores de reflectância, o que seria um indicador de grãos doentes. Na maior parte das vezes, isso é causado pela presença de reflexões especulares produzidas pela superfície relativamente mais plana deste tipo de grão, artificialmente aumentando o valor da reflectância em todas as bandas espectrais. Uma possível solução para isso seria considerar apenas as porções dos grãos menos sujeitas à reflexão especular, porém há indicações de que, mesmo com a presença de distorções, considerar o grão inteiro normalmente leva a melhores resultados (Delwiche et al., 2011).

Dependendo do procedimento de segmentação adotado, sombras causadas pelos grãos podem ser tomadas como parte dos grãos ao invés de serem removidas, artificialmente reduzindo a reflectância média daquele grão. O efeito das sombras poderia ser reduzido pela utilização de outra fonte de luz no lado oposto da primeira, porém, isso poderia aumentar os problemas com reflexões especulares, além de poder tornar indiscerníveis detalhes importantes para a detecção da doença.

Algoritmos frequentemente empregam regras baseadas no tamanho dos objetos para eliminar elementos espúrios nas imagens. Em princípio, essas regras são dependentes da resolução espacial das imagens, valor este que determina, por exemplo, o número de pixels contidos em cada grão. Para que as regras de tamanho possam ser aplicadas de maneira geral, elas devem estar diretamente associadas às dimensões (em pixels) das imagens.

As operações de segmentação normalmente incluem a realização de limiarizações, as quais definem as faixas de valores de pixels que melhor representam as regiões

que se deseja separar (por exemplo, grãos e fundo). Por esse motivo, é sempre importante aplicar uma normalização aos valores de reflectância, compensando assim a maior parte das possíveis diferenças em iluminação. Contudo, é possível que as relações entre as reflectâncias dos grãos, bandeja e fundo possam ser diferentes, dependendo das condições. Em geral, os elementos nas imagens são suficientemente contrastantes para que isso não tenha um grande impacto, mas pode haver certas circunstâncias que requeiram uma redefinição dos valores de limiar.

A tarefa de estimar os níveis de micotoxinas em grãos tem também alguns desafios específicos. O maior deles é, provavelmente, o fato dos escâneres hiperespectrais não terem sensibilidade suficiente para detectar as micotoxinas diretamente, a menos que as concentrações sejam muito elevadas (Sun, 2009). Isso faz com que estimativas precisas sejam muito difíceis de ser obtidas. Portanto, o mais provável é que a resposta espectral de grãos contaminados com DON venha dos danos causados pela giberela ao invés de vir da própria micotoxina. Porém, apesar da relação entre giberela e DON ser inequívoca, ela não é perfeita, o que significa que grãos assintomáticos podem conter DON, e vice-versa. Isso sempre causa erros de predição, e será inevitável até que a sensibilidade dos equipamentos seja suficiente para detectar DON diretamente.

Nos testes realizados, as duas bandas mais efetivas na detecção de DON foram as de 623 nm e de 1.411 nm. A segunda banda está localizada exatamente na região espectral de absorção da água, o que indica que algoritmos utilizando esta banda devem se basear, ao menos parcialmente, nas diferenças de umidade entre os diferentes grãos. Isso pode causar problemas se diferentes lotes possuírem diferentes níveis de umidade, o que pode exigir a definição de diferentes parâmetros para diferentes lotes. Por outro lado, a banda de 1.408 nm está associada ao DON (Peiris et al., 2009), o que significa que esta banda também deve detectar DON diretamente quando as concentrações forem elevadas. Além disso, Dowell et al. (1999) indicaram que a giberela afeta o conteúdo de amido e proteína, causando absorção na faixa de 1.400 nm, contribuindo para a detecção indireta de DON. Portanto, apesar de a umidade ser um problema, diversos outros fatores contribuem para viabilizar a predição dos níveis de DON. Esse fato permitiu o desenvolvimento de métodos como os descritos em Barbedo et al. (2017) e em Peiris et al. (2017). Neste último trabalho, os autores reportaram que FT-NIR (*Fourier Transform - Near Infrared*) pode ser utilizado para segregar o trigo de acordo com níveis de DON, separando os lotes em baixo (<2.000 ppb) e alto (>2.000 ppb) níveis.

Como no caso da detecção de giberela, o desenvolvimento de algoritmos para estimar os níveis de DON é limitado pela falta de referências completamente confiáveis. Um dos métodos mais utilizados na avaliação do conteúdo de DON é o ELISA (*Enzyme-linked immunosorbent assay*), que possui limitações de sensibilidade e especificidade (reação

cruzada), consequentemente, aumentando a incerteza associada aos resultados fornecidos pelos algoritmos. Para contornar estas limitações, a amostragem deve ser representativa e os métodos laboratoriais devem ser acurados na determinação dos níveis de micotoxinas. As técnicas cromatográficas geralmente são empregadas como métodos de referência, por causa da precisão e da reprodutibilidade dos resultados, entretanto são procedimentos longos e caros, demandam estrutura (laboratório) e utilizam reagentes tóxicos, limitando sua utilização em análises de rotina de grande volume de amostras (Xu et al., 2010; Pierna et al., 2012). Dadas as muitas incertezas relacionadas à predição dos níveis de micotoxinas, é mais apropriado adotar uma estratégia de classificação associada com o processamento dos grãos em lotes (Barbedo et al., 2017). Dessa maneira, os erros associados a grãos individuais acabam compensando-se mutuamente, e desde que os erros remanescentes não causem uma mudança de classe, o resultado obtido será correto.

Esta classificação de lotes de trigo quanto aos níveis de micotoxinas, através do NIR associado a imagens hiperspectrais, possui ampla demanda na cadeia produtiva, principalmente para orientar a segregação de lotes. Considerando que estes contaminantes são imperceptíveis visualmente no produto final, a predição dos níveis de micotoxinas por NIR hiperespectral previne a mistura de lotes de trigo com altos níveis de contaminação com lotes sadios, que podem ser destinados à produção de alimentos para humanos e animais, contribuindo, assim, para a garantia de qualidade dos alimentos comercializados e para o atendimento da legislação vigente.

## Conclusões

Vários trabalhos encontrados na literatura corroboram a afirmação de que a banda do infravermelho próximo é apropriada para detectar giberela em grãos de trigo. A maior parte das inconsistências está associada à reflexão especular ou à incerteza inerente ao processo de rotulagem visual dos grãos. Como não há uma referência absoluta disponível, as estimativas geradas pelos algoritmos normalmente fornecem apenas um indicador da probabilidade de um determinado grão estar infectado, ao invés de fornecer uma avaliação categórica.

No caso da estimativa de DON, embora os resultados alcançados sejam relativamente acurados, eles não podem ser usados isoladamente como resposta definitiva na triagem de lotes de trigo, especialmente considerando os riscos à saúde representados por essa toxina. Por outro lado, eles podem ser utilizados como parte de um sistema de análise rápida, com o papel de selecionar lotes que mereçam uma análise mais detalhada. Contudo, esse tipo de uso tem algumas implicações importantes. Por

exemplo, falsos negativos são muito mais problemáticos que falsos positivos, porque o envio de um lote saudável para análise adicional vai ter apenas um impacto limitado nos custos, enquanto que a aprovação de lotes contaminados pode ter repercussões na saúde pública. Assim, é conveniente que algoritmos que venham a ser utilizados em sistemas de triagem sejam calibrados para evitar, tanto quanto possível, falsos negativos, ainda que isso prejudique o desempenho global.

## Referências

- ARNGREN, M.; HANSEN, P. W.; ERIKSEN, B.; LARSEN, J.; LARSEN, R. Analysis of pregerminated barley using hyperspectral image analysis. **Journal of Agricultural and Food Chemistry**, v. 59, n. 21, p. 11385-11394, Nov. 2011.
- BARBEDO, J. G. A.; TIBOLA, C. S.; FERNANDES, J. M. C. Detecting head blight in wheat kernels using hyperspectral imaging. **Biosystems Engineering**, v. 131, p. 65-76, Mar. 2015.
- BARBEDO, J. G. A.; TIBOLA, C. S.; LIMA, M. I. P. Deoxynivalenol screening in wheat kernels using hyperspectral imaging. **Biosystems Engineering**, v. 155, p. 24-32, Mar. 2017.
- BAURIEGEL, E.; HERPPICH, W. B. Hyperspectral and chlorophyll fluorescence imaging for early detection of plant diseases, with special reference to *Fusarium* spec. infections on wheat. **Agriculture**, v. 4, n. 1, p. 32-57, 2014.
- BERMAN, M.; CONNOR, P. M.; WHITBOURN, L. B.; COWARD, D. A.; OSBORNE, B. G.; SOUTHAN, M. D. Classification of sound and stained wheat grains using visible and near infrared hyperspectral image analysis. **Journal of Near Infrared Spectroscopy**, v. 15, n. 6, p. 351-358, 2007.
- BRASIL. Resolução nº. 138, de 08 de fevereiro de 2017. Regulamenta os limites máximos tolerados (LMT) de deoxinivalenol em alimentos. Diário Oficial da União, Brasília, DF, 9 fev. 2017. Seção 1, n. 29, p. 45.
- CHELLADURAI, V.; KARUPPIAH, K.; JAYAS, D. S.; FIELDS, P. G.; WHITE, N. D. G. Detection of (F.) infestation in soybean using soft X-ray and NIR hyperspectral imaging techniques. **Journal of Stored Products Research**, v. 57, p. 43-48, Apr. 2014.
- CHOUDHARY, R.; MAHESH, S.; PALIWAL, J.; JAYAS, D. S. Identification of wheat classes using wavelet features from near infrared hyperspectral images of bulk samples. **Biosystems Engineering**, v. 102, n. 2, p. 115-127, 2009.
- DEL PONTE, E. M.; TESSMANN, D. J.; SPOLTI, P.; KUHNEM, P. R.; SILVA, C. N. Species identification, genetic diversity and phenotypic variation studies on the *Fusarium graminearum* complex populations from Brazil. In: ALCONADA MAGLIANO, T. M.; CHULZE, S. N. (Ed.). **Fusarium Head Blight in Latin America**. Netherlands: Springer, 2013. p. 15-29.

- DELWICHE, S. R.; KIM, M. S. Hyperspectral imaging for detection of scab in wheat. In: DESHAZER, J. A.; MEYER, G. E. **Biological quality and precision agriculture II**. Bellingham, WA: SPIE, 2000. p. 13-20, (Proceedings of SPIE 4203).
- DELWICHE, S. R.; KIM, M. S.; DONG, Y. *Fusarium* damage assessment in wheat kernels by vis/NIR hyperspectral imaging. **Sensing and Instrumentation for Food Quality and Safety**, v. 5, n. 2, p. 63-71, 2011.
- DOWELL, F. E.; RAM, M. S.; SEITZ, L. M. Predicting scab, vomitoxin, and ergosterol in single wheat kernels using near infrared spectroscopy. **Cereal Chemistry**, v. 76, n. 4, p. 573-576, 1999.
- KALIRAMESH, S.; CHELLADURAI, V.; JAYAS, D. S.; ALAGUSUNDARAM, K.; WHITE, N. D. G.; FIELDS, P. G. Detection of infestation by *Callosobruchus maculatus* in mung bean using near-infrared hyperspectral imaging. **Journal of Stored Products Research**, v. 52, p. 107-111, Jan. 2013.
- KANDPAL, L. M.; LEE, S.; KIM, M. S.; BAE, H.; CHO, B.-K. Short wave infrared (SWIR) hyperspectral imaging technique for examination of aflatoxin B (AFB) on corn kernels. **Food Control**, v. 51, p. 171-176, May 2015.
- MAHESH, S.; MANICKAVASAGAN, A.; JAYAS, D. S.; PALIWAL, J.; WHITE, N. D. G. Feasibility of near-infrared hyperspectral imaging to differentiate Canadian wheat classes. **Biosystems Engineering**, v. 101, n. 1, p. 50-57, 2008.
- MCGOVERIN, C. M.; ENGELBRECHT, P.; GELADI, P.; MANLEY, M. Characterisation of non-viable whole barley, wheat and sorghum grains using near-infrared hyperspectral data and chemometrics. **Analytical and Bioanalytical Chemistry**, v. 401, n. 7, p. 2283-2289, Oct. 2011.
- MENESATTI, P.; ANTONUCCI, F.; COSTA, C.; SANTORI, A.; NICIARELLI, I.; INFANTINO, A. Application of morphometric image analysis system to evaluate the incidence of *Fusarium* head blight wheat infected kernels. In: BORNIMER AGRARTECHNISCHE BERICHTE, 69., 2009, Potsdam. **Proceedings...** Potsdam: Leibniz-Institut für Agrartechnik, 2009. p. 157-161.
- NANSEN, C.; KOLOMIETS, M.; GAO, X. Considerations regarding the use of hyperspectral imaging data in classifications of food products, exemplified by analysis of maize kernels. **Journal of Agricultural and Food Chemistry**, v. 56, n. 9, p. 2933-2938, May 2008.
- PEIRIS, K. H. S.; DONG, Y.; DAVIS, M. A.; BOCKUS, W. W.; DOWELL, F. E. Estimation of the deoxynivalenol and moisture contents of bulk wheat grain samples by FT-NIR spectroscopy. **Cereal Chemistry Journal**, v. 94, n. 4, p. 677-682, 2017.
- PEIRIS, K. H. S.; PUMPHREY, M. O.; DOWELL, F. E. NIR absorbance characteristics of deoxynivalenol and of sound and *Fusarium*-damaged wheat kernels. **Journal of Near Infrared Spectroscopy**, v. 17, n. 4, p. 213-221, 2009.

PIERNA, J. A. F.; VERMEULEN, P.; AMAND, O.; TOSSENS, A.; DARDENNE, P.; BAETEN, V. NIR hyperspectral imaging spectroscopy and chemometrics for the detection of undesirable substances in food and feed. **Chemometrics and Intelligent Laboratory Systems**, v. 117, p. 233-239, Aug. 2012.

POLDER, G.; VAN DER HEIJDEN, G. W. A. M.; WAALWIJK, C.; YOUNG, I. T. Detection of Fusarium in single wheat kernels using spectral imaging. **Seed Science and Technology**, v. 33, n. 3, p. 655-668, 2005.

RAVIKANTH, L.; SINGH, C. B.; JAYAS, D. S.; WHITE, N. D. G. Classification of contaminants from wheat using near-infrared hyperspectral imaging. **Biosystems Engineering**, v. 135, p. 73-86, July 2015.

SERRANTI, S.; CESARE, D.; BONIFAZI, G. The development of a hyperspectral imaging method for the detection of -damaged, yellow berry and vitreous Italian durum wheat kernels. **Biosystems Engineering**, v. 115, n. 1, p. 20-30, 2013a.

SERRANTI, S.; CESARE, D.; MARINI, F.; BONIFAZI, G. Classification of oat and groat kernels using NIR hyperspectral imaging. **Talanta**, v. 103, p. 276-284, Jan. 2013b.

SHAHIN, M. A.; SYMONS, S. J. Detection of Fusarium damage in Canadian wheat using visible/near-infrared hyperspectral imaging. **Journal of Food Measurement and Characterization**, v. 6, n. 1/4, p. 3-11, 2012.

SHAHIN, M. A.; SYMONS, S. J. Detection of Fusarium damaged kernels in Canada Western Red Spring wheat using visible/near-infrared hyperspectral imaging and principal component analysis. **Computers and Electronics in Agriculture**, v. 75, n. 1, p. 107-112, 2011.

SINGH, C. B.; JAYAS, D. S.; PALIWAL, J.; WHITE, N. D. G. Detection of insect-damaged wheat kernels using near-infrared hyperspectral imaging. **Journal of Stored Products Research**, v. 45, n. 3, p. 151-158, 2009.

SINGH, C. B.; JAYAS, D. S.; PALIWAL, J.; WHITE, N. D. G. Detection of midge-damaged wheat kernels using short-wave near-infrared hyperspectral and digital colour imaging. **Biosystems Engineering**, v. 105, n. 3, p. 380-387, 2010a.

SINGH, C. B.; JAYAS, D. S.; PALIWAL, J.; WHITE, N. D. G. Fungal damage detection in wheat using short-wave near-infrared hyperspectral and digital colour imaging. **International Journal of Food Properties**, v. 15, n. 1, p. 11-24, 2012.

SINGH, C. B.; JAYAS, D. S.; PALIWAL, J.; WHITE, N. D. G. Fungal detection in wheat using near-infrared hyperspectral imaging. **Transactions of the ASABE**, v. 50, n. 6, p. 2171-2176, 2007.

SINGH, C. B.; JAYAS, D. S.; PALIWAL, J.; WHITE, N. D. G. Identification of insect-damaged wheat kernels using short-wave near-infrared hyperspectral and digital colour imaging. **Computers and Electronics in Agriculture**, v. 73, n. 2, p. 118-125, 2010b.

SUN, S.-W. (Ed.). **Hyperspectral imaging for food quality analysis and control**. San Diego: Academic Press, 2009. 477 p.

VERMEULEN, P.; FERNÁNDEZ PIERNA, J. A.; VAN EGMOND, H. P.; DARDENNE, P.; BAETEN, V. On-line detection and quantification of ergot bodies in cereals using near infrared hyperspectral imaging. **Food Additives and Contaminants: Part A**, v. 29, n. 2, p. 232-240, 2012.

VERMEULEN, P.; FERNÁNDEZ PIERNA, J. A.; VAN EGMOND, H. P.; ZEGERS, J.; DARDENNE, P.; BAETEN, V. Validation and transferability study of a method based on near-infrared hyperspectral imaging for the detection and quantification of ergot bodies in cereals. **Analytical and Bioanalytical Chemistry**, v. 405, n. 24, p. 7765-7772, Sep. 2013.

WANG, L.; SUN, D.-W.; PU, H.; ZHU, Z. Application of hyperspectral imaging to discriminate the variety of maize seeds. **Food Analytical Methods**, v. 9, n. 1, p. 225-234, Jan. 2016.

WANG, W.; HEITSCHMIDT, G. W.; NI, X.; WINDHAM, W. R.; HAWKINS, S.; CHU, X. Identification of aflatoxin B1 on maize kernel surfaces using hyperspectral imaging. **Food Control**, v. 42, p. 78-86, Aug. 2014.

WANG, L.; LIU, D.; PU, H.; SUN, D.-W.; GAO, W.; XIONG, Z. Use of hyperspectral imaging to discriminate the variety and quality of rice. **Food Analytical Methods**, v. 8, n. 2, p. 515-523, 2015a.

WANG, W.; HEITSCHMIDT, G. W.; WINDHAM, W. R.; FELDNER, P.; NI, X.; CHU, X. Feasibility of detecting aflatoxin B1 on inoculated maize kernels surface using Vis/NIR hyperspectral imaging. **Journal of Food Science**, v. 80, n. 1, p. M116-M122, Jan. 2015b.

WANG, W.; NI, X.; LAWRENCE, K. C.; YOON, S.-C.; HEITSCHMIDT, G. W.; FELDNER, P. Feasibility of detecting Aflatoxin B1 in single maize kernels using hyperspectral imaging. **Journal of Food Engineering**, v. 166, p. 182-192, Dec. 2015c.

XING, J.; SYMONS, S.; SHAHIN, M.; HATCHER, D. Detection of sprout damage in Canada Western Red Spring wheat with multiple wavebands using visible/near-infrared hyperspectral imaging. **Biosystems Engineering**, v. 106, n. 2, p. 188-194, 2010.

XU, Y.; HUANG, Z.-B.; HE, Q.-H.; DENG, S.-Z.; LI, L.-S.; LI, Y.-P. Development of an immunochromatographic strip test for the rapid detection of deoxynivalenol in wheat and maize. **Food Chemistry**, v. 119, n. 2, p. 834-839, 2010.

ZHANG, X.; LIU, F.; HE, Y.; LI, X. Application of hyperspectral imaging and chemometric calibrations for variety discrimination of maize seeds. **Sensors**, v. 12, n. 12, p. 17234-17246, 2012.

# Equipamentos Portáteis de Espectrometria no Infravermelho Próximo

Jez Willian Batista Braga  
Laiz de Oliveira Magalhães

## Introdução

A espectroscopia no infravermelho próximo (NIR) desenvolveu-se juntamente com algumas linhas de pesquisa da quimiometria e atualmente é aplicada às mais diversas áreas do conhecimento (Williams; Norris, 2001). Ao longo desse desenvolvimento, por causa do avanço da instrumentação, da capacidade de processamento e do armazenamento de dados, a miniaturização de dispositivos espectroscópicos tornou-se uma realidade, possibilitando o surgimento de espectrômetros NIR portáteis.

Entre os primeiros trabalhos relatados na literatura empregando espectrômetros NIR portáteis, pode-se destacar o de Mitchell et al. (1990), que utilizou combinações de medidas de reflectância em comprimentos de onda na região do infravermelho próximo e do visível para avaliar os teores de fitomassa em alfafa, com a intenção de acompanhar o ganho de peso de ovelhas. Também no início da década de 1990, Lysaght et al. (1991) descreveram o desenvolvimento de um espectrômetro acoplado a um computador portátil, o qual utilizava como fonte uma lâmpada de tungstênio, um sistema de dispersão por grade de difração e detector por arranjo de diodos de silício capaz de realizar medidas entre 680 nm e 1.050 nm. De acordo com os autores, este aparelho apresentava dimensões de 15,2 cm x 35,6 cm x 45,7 cm de altura, largura e comprimento, respectivamente, e pesava aproximadamente 9 kg. Atualmente, existem mais de uma dezena de equipamentos comerciais disponíveis, os quais apresentam peso que pode variar de 60 g a menos de 3 kg, dependendo se o dispositivo representa apenas o espectrômetro ou se este é acoplado a um processador e visor digital que possibilitam a visualização do espectro medido e resultado da análise. As dimensões também possibilitam o fácil transporte e realização de medidas, geralmente pelos modos de transmitância, reflectância ou transreflectância.

O uso desses equipamentos tem despertado um grande interesse tanto da área acadêmica quanto de diversos setores produtivos nos últimos anos, pois além de apresentar todas as vantagens geralmente obtidas pela espectroscopia NIR (mínimo ou



nenhum preparo de amostra e geração de resíduos químicos, rápidas leituras, menor custo, etc.), também elimina a necessidade de transportar as amostras para análise em laboratório, eliminando custos e tempo com transporte ou perda das amostras destinadas à análise. Este último aspecto tem possibilitado a realização de análises em campo (*in loco*), o que pode representar ampliação significativa da aplicação da técnica para análises em locais distantes de laboratórios ou de difícil acesso, na análise direta de grãos, folhas, pastagens, frutos e uma infinidade de outros produtos de interesse agropecuários (Santos et al., 2013). Além disso, do ponto de vista de ações de fiscalização ou apreensão de produtos ilícitos, a realização de análise química *in loco* é capaz de munir agentes de fiscalização ou policiais de informações objetivas que podem embasar atuações com maior convicção e minimizar a ocorrência de erros.

Apesar das grandes possibilidades abertas pela chegada desses dispositivos, o desenvolvimento de métodos em diversas áreas ainda vem sendo um desafio, por causa principalmente das variações espectrais observadas em razão das diferentes condições encontradas nas medidas *in loco*. Fatores de difícil controle, como a umidade presente nas amostras, ou a variação sazonal observada em diversos tipos de grãos, frutas ou diferentes culturas, têm dificultado a obtenção de modelos de regressão e classificação que atendam aos requisitos mínimos de parâmetros de exatidão, precisão ou outras figuras de mérito.

Como forma de elucidar alguns aspectos referentes ao advento de espectrômetros NIR portáteis, esse capítulo apresentará uma breve descrição de alguns espectrômetros atualmente disponíveis no mercado, como se encontra o desempenho desses equipamentos quando comparados a espectrômetros de bancada e considerações sobre a realização de transferência de bases de dados espectrais e modelos entre equipamentos de bancada e portáteis. Além disso, uma revisão sucinta de recentes aplicações desenvolvidas com dispositivos NIR portáteis para a determinação de propriedades de interesse para o cultivo de grãos ou de seu controle de qualidade será apresentada, destacando o potencial e as principais dificuldades enfrentadas nesse setor.

## Espectrômetros NIR portáteis

Na última década, houve grande ampliação na diversidade de equipamentos NIR portáteis disponíveis comercialmente, os quais vêm se aprimorando com o passar dos anos. De forma geral, os equipamentos portáteis são aqueles que dispõem de uma fonte de energia independente, possuem baixo peso e estabilidade óptica que permitem seu transporte e realização de medidas em condições fora do ambiente de

laboratório. A fonte de radiação destes equipamentos é frequentemente composta por lâmpadas halógenas ou lâmpadas de tungstênio, que apresentam maior intensidade de radiação na faixa do NIR (Santos et al., 2013). Alguns dispositivos permitem a troca da lâmpada, mas, de forma geral, o tempo de vida útil das fontes é relativamente longo e pode aproximar-se do tempo de vida útil de alguns desses espectrômetros. O detector normalmente é baseado em um arranjo de diodos de arseneto de gálio e índio (InGaAs) ou arranjo de diodos de silício (Guillemain et al., 2017). Na Tabela 1 são apresentadas algumas das especificações de alguns espectrômetros NIR portáteis que podem ser encontrados no mercado. É possível perceber que alguns abrangem a faixa do visível, juntamente com parte ou toda a região do NIR. Além disso, alguns desses equipamentos podem apresentar-se como espectrômetros de tamanho e peso reduzidos, possibilitando fácil transporte, instalação e utilização, fazendo com que possam ser considerados portáteis. Contudo, alguns instrumentos, além de terem essas vantagens, apresentam sistema de alimentação acoplado ao dispositivo por meio de baterias, formato anatômico e peso menor que 2 kg, fazendo com que possam ser considerados portáteis de mão *handheld*, sendo os que mais têm atraído interesse para aplicações *in loco*.

**Tabela 1.** Especificações principais de alguns espectrômetros NIR portáteis, considerando as informações disponíveis nos websites de seus fabricantes no ano de 2018.

Equipamento	Fabricante	Fonte	Intervalo espectral (nm)	Dimensões (mm)*	Peso (kg)
QualitySpec Trek	ASD	H	350 a 2.500	310×100×300	2,5
USB 4000	OceanOptics	T	350 a 1.000	891×633×344	0,190
i-Spec 17	B&W Tek	H / T	350 a 2.200	342×170×347	4,9
i-Spec 25	B&W Tek	H / T	350 a 2.500	NI	NI
i-Spec Plus	B&W Tek	H / T	900 a 2.200	400×260×250	8
i-Spec Nano	B&W Tek	H / T	1.300 a 2.200	NI	NI
Sol 1.7	B&W Tek	H / T	900 a 1.700	NI	NI
Sol 2.2A	B&W Tek	H / T	900 a 2.200	NI	NI
Sol 2.6	B&W Tek	H / T	1.550 a 2.550	197×109×68	NI
Phazir	Polychromix	T	900 a 1.690	254×292×152	1,7
MicroPhazir	Thermo Scientific	T	1.600 a 2.400	266×109×251	1,8
MicroNIR 1700 Pro	JDSU	T	950 a 1.650	45×50	0,064
MicroNIR OnSite	JDSU	T	950 a 1.650	146×44,5	0,2
Scio	Consumer Physics Inc.	NI	700 a 1.100	68×41×19	0,035
DLP NIRscan Nano EVM	Texas Instruments	T	700 a 2.500	36×62×58	NI
Corona Plus 45 NIR	Zeiss	T	950 a 1.670	280×230×140	7

H – lâmpada halógena; T – lâmpada de tungstênio; NI – não informado.

\* Altura × largura × comprimento.

As empresas que produzem esses equipamentos também apostam na variabilidade de modelos, de modo a atender diferentes demandas. Há espectrômetros dotados de fibras óticas para leituras pontuais e que necessitam de mínimo contato do dispositivo com a amostra. Por outro lado, também existem espectrômetros de mão com formato que se assemelha a uma pistola, amplamente utilizados na indústria farmacêutica e aplicados em casos nos quais as amostras são pouco móveis e o dispositivo deve se adaptar a elas. Adicionalmente, há dois meios principais de exportação dos dados obtidos: via USB, quando conectado a um computador ou *tablet*, ou por meio de ferramentas como *bluetooth*, conexão *wi-fi* e armazenamento em servidores *online* (nuvem).

Os espectrômetros portáteis geralmente realizam leitura espectral no modo de reflectância difusa, e estes são amplamente utilizados na análise de vários tipos de amostras no estado sólido, como no caso da análise de grãos. Muitos desses equipamentos, como o USB 4000 que realiza medidas na região visível e NIR (VIS-NIR), o MicroNIR1700 Pro, MicroNIR OnSite e MicroPhazir, podem realizar a leitura em outros modos além da reflectância difusa, como por transmitância e transfletância, fato que confere maior versatilidade de aplicação destes equipamentos nos diferentes campos de pesquisa e industriais.

Em relação ao modo de exportar e armazenar os dados, muitas empresas que produzem esses espectrômetros portáteis valem-se da utilização de softwares próprios que realizam a aquisição dos espectros, que depois são transferidos para um computador via interface USB. Após essa etapa, os espectros podem ser analisados com outros softwares para a realização da análise multivariada, o que ocorre para espectrômetros como o MicroNIR e USB 4000 VIS-NIR. Há equipamentos, por outro lado, que já são dotados de uma interface interativa e espaço de armazenamento que, além de realizar a leitura espectral, também salvam os dados obtidos no próprio equipamento e são capazes de apresentar o resultado da análise de um modelo salvo no equipamento, como no caso do MicroPhazir. Outros modelos estão apostando na utilização do armazenamento online de dados ou em “nuvem de dados”, que são hospedados por um servidor do fabricante do equipamento. Contudo, se esta estratégia pode ser considerada uma vantagem para regiões com fácil acesso à internet, também pode ser uma desvantagem por vender um serviço que, de certo modo, deixa o usuário dependente da empresa e requer o pagamento de uma licença para ter acesso aos dados obtidos para eventual análise em outros softwares.

Como abordado anteriormente, por causa das suas características inerentes, a interpretação analítica do espectro NIR necessita de ferramentas de análise multivariada, de modo a pré-processar os dados, eliminar ou modelar os sinais de interferentes

para realizar a diferenciação entre espectros de amostras distintas e quantificar propriedades de interesse. Considerando este fator, alguns espectrômetros NIR portáteis, como MicroPhazir e SCIO, utilizam softwares próprios que realizam diferentes pré-processamento de dados e geram modelos de calibração e classificação multivariada. Estas implementações também justificam a disseminação destes dispositivos nos diferentes setores produtivos, principalmente no campo alimentício, uma vez que estes atributos contribuem para facilitar o desenvolvimento dos modelos para a realização da análise e para apresentar uma interface mais amigável ao usuário.

Considerando a variabilidade de modelos de espectrômetros NIR portáteis disponíveis no mercado, o analista depara-se com uma gama de preços, a depender dos objetivos de uso, do método de armazenamento dos dados, do que é oferecido pela empresa para a análise dos dados, além de outras características do equipamento. Dispositivos de menor resolução, arranjo ótico e eletrônico mais simplificado e isento de interface interativa ou softwares de análise costumam apresentar preços menores. Também apresentam baixos preços os dispositivos portáteis dependentes de login ou armazenamento de dados na nuvem, uma vez que o custo relacionado a esta fase da análise é considerado algo à parte do equipamento pelas empresas produtoras deste tipo de espectrômetro.

No geral, comparando-se com equipamentos de bancada, os espectrômetros NIR portáteis apresentam um custo menor, o qual pode variar de cerca de U\$ 1.000 até chegar a U\$ 25.000. Por essa razão, muitos desses dispositivos têm viabilizado a disseminação da espectroscopia NIR nas análises em campo. Para o propósito de análise de grãos, é necessário escolher um dispositivo portátil que se enquadre no objetivo da análise. Além dos aspectos que envolvem a portabilidade do instrumento, idealmente, deve-se considerar a forma em que os espectros serão medidos (refletância, transmitância ou transfletância), a região espectral medida, a área da amostra que é amostrada durante a análise, duração da bateria, estabilidade de sinal, etc. Se o objetivo for, por exemplo, quantificar proteínas ou gordura no grão intacto, espectrômetros com região espectral limitada podem não conter a informação ideal para análise. Uma região espectral mais abrangente (até cerca de 2.000 nm) pode fornecer melhores limites de detecção, uma vez que a região que se limita aos segundo e terceiro sobretons (até 1.600 nm) possui menor intensidade por apresentar menor probabilidade de ocorrência das transições vibracionais (Pasquini, 2003).

Tendo em vista que na análise dos grãos a amostra é geralmente analisada por refletância, diferentes dispositivos portáteis, com características distintas, podem ser utilizados, pois esse é o modo de leitura mais comum entre esses espectrômetros. A adaptação dos dispositivos portáteis para uma análise específica também já é uma

alternativa praticada em alguns centros de pesquisa ou por empresas dispostas a apostar na customização desses instrumentos para finalidades específicas, sendo esta uma opção que pode ser conduzir a melhores resultados.

## Comparação entre equipamentos de bancada, portáteis e aspectos a considerar na transferência de modelos

Um aspecto relevante que sempre é questionado quando se utilizam espectrômetros portáteis é se estes apresentam aproximadamente o mesmo desempenho dos de bancada empregados em laboratório, sob condições controladas. Diversos trabalhos, estudando diferentes aplicações, já foram relatados na literatura descrevendo a comparação de diferentes marcas e modelos (Herberholz et al., 2010; Sorak et al., 2012; Plans et al., 2013; Cozzolino, 2014; Cabassi et al., 2015; Malegori et al., 2017, Marchesini et al., 2018). Contudo, em razão da diversidade de equipamentos de configurações distintas disponíveis atualmente, tanto nas versões de bancada como nos portáteis, comparações de desempenho devem ser feitas sempre com muito critério e considerando a aplicação desejada.

Um trabalho que ilustra essa diferença de configurações foi realizado para a análise de esterco misto (*slurry*) por Cabassi et al. (2015). Foram comparados quatro espectrômetros NIR para determinação dos seguintes teores de parâmetros de qualidade: matéria seca (MS), cinzas (CI), nitrogênio total obtido pelo método de Kjeldahl (NT), amônio total (AT), nitrogênio orgânico (NO), carbono total (CT) e fósforo total (PT). Foram utilizados os espectrômetros de bancada NIR System 5000 (FOSS, Denmark), que possui óptica dispersiva e intervalo espectral de 1.100 nm a 2.498 nm com 2 nm de resolução e NIR Flex N-500 (Buchi, Italy) dotado de transformada de Fourier, intervalo espectral de 1.000 nm a 2.500 nm com resolução média de 12 nm. Os portáteis avaliados consistiram do CORONA 45 (Carl Zeiss, Germany), que possui dispersão por monocromador e detecção entre 960 nm e 1.690 nm com resolução de 10 nm; e o espectrômetro portátil de mão LAB POD™ Polycromix (Bedford, MA, USA), o qual utiliza um sistema mecânico microeletrônico baseado em transformada de Hadamard e detecção de 1.000 nm a 1.800 nm com 12 nm de resolução. Além disso, os espectrômetros de bancada foram avaliados em sua resolução nativa ou com resolução reduzida matematicamente, para simular a aquisição com resolução similar a um espectrômetro portátil. Para permitir a comparação dos instrumentos, todas as medidas foram realizadas por meio de reflectância difusa com todos os equipamentos em ambiente laboratório. A Tabela 2 resume as principais conclusões obtidas por Cabassi et al. (2015), no que diz respeito à comparação do desempenho

dos equipamentos. Observou-se que a equivalência ou não dos espectrômetros dependeu da propriedade de interesse analisada, o que pode estar relacionada a vários fatores, como a sensibilidade e a amplitude do intervalo de concentrações, pois quanto menor for a sensibilidade e a amplitude, maior tendem a ser as diferenças observadas entre espectrômetros de bancada e portáteis. Outro aspecto relevante é referente à resolução espectral, ao contrário do relatado por outros autores (Kolomiets; Siesler, 2004; Herberholz et al., 2010); de forma geral, a redução da resolução dos espectrômetros de bancada levou a uma piora nos resultados de predição, aproximando os resultados daqueles obtidos pelos equipamentos portáteis.

**Tabela 2.** Comparação do desempenho de espectrômetros NIR portáteis e de bancada para estimar propriedades de interesse em esterco misto (*slurry*) (Cabassi e al., 2015).

Espectrômetros	Propriedades de interesse <sup>(1)</sup>						
	MS	CI	NT	AT	NO	CT	PT
NIR System 5000 <sup>(2)</sup>	B	B	B	M	B	B	M
NIR Flex N-500 <sup>(2)</sup>	B	M	M	M	B	B	M
NIR System 5000 (RR) <sup>(1, 2)</sup>	M	M	M	I	M	B	M
NIR Flex N-500 (RR) <sup>(1, 2)</sup>	M	M	M	I	I	B	M
CORONA 45 <sup>(3)</sup>	B	M	M	I	B	B	M
LAB POD <sup>TM(4)</sup>	M	M	M	I	B	M	I

<sup>(1)</sup> RR: resolução reduzida; MS: matéria seca; CI: cinzas; NT: nitrogênio total obtido pelo método de Kjeldahl; AT: amônio total (AT); NO: nitrogênio orgânico; CT: carbono total; PT: fósforo total; B: bom desempenho, apropriado para realização de análises; M: desempenho mediano, apropriado para propósitos de seleção; I: inapropriado para estimar a propriedade de interesse. <sup>(2)</sup> Bancada. <sup>(3)</sup> Portátil. <sup>(4)</sup> Portátil de mão.

Outro trabalho que apresenta interessante comparação de instrumentos foi o realizado por Plans et al. (2013), que analisaram os teores de proteína, amido e amilose total em grãos moídos de feijão de diferentes genótipos cultivados na Espanha. Foram comparados os resultados de dois equipamentos NIR de bancada, sendo um deles dispersivo (NIRFOSS 5000, FOSS), outro baseado em transformada de Fourier (Varian Excalibur 3100, Agilent Technologies) e, um NIR portátil de mão (MicroPhazir RX, Thermo Scientific). Além disso, dois espectrômetros trabalhando na região do infravermelho médio (MIR, *Middle infrared*) também foram utilizados, sendo um de bancada (Varian Excalibur FT-IR 3100, Agilent Technologies) e outro portátil (Cary 630, Agilent Technologies), ambos utilizando transformada de Fourier. Com relação às condições de medida, para a região MIR foi utilizada a resolução de 4 cm<sup>-1</sup> e acessório

de reflectância total atenuada com cristal de diamante em ambos os equipamentos. Por outro lado, no NIR foram utilizadas resoluções e acessórios de reflectância difusa com configurações distintas nos três equipamentos. Os autores utilizaram análise de componentes principais (PCA, *Principal Component Analysis*) para definir as classes das amostras e a regressão por mínimos quadrados parciais (PLSR, *Partial Least Squares Regression*) para as análises quantitativas. Os resultados mostraram que, entre os espectrômetros MIR comparados, o portátil apresentou erros relativos cerca de 10% menores que o equipamento de bancada, sendo isso justificado pelos autores em razão de um caminho óptico interno muito curto no interferômetro, proporcionando níveis de desempenho comparáveis aos de espectrômetros de alto desempenho. Com relação aos espectrômetros de NIR, o de bancada baseado em transformada de Fourier apresentou o melhor desempenho, sendo melhor que os outros dois equipamentos em cerca de 10% em termos de erros relativos. Contudo, ao final do trabalho, os autores concluíram que os resultados foram comparáveis para a determinação dos três analitos e que os espectrômetros portáteis são alternativas atraentes aos espectrômetros de bancada por suas diversas vantagens, mas que pesquisas ainda necessitam ser feitas para testar os instrumentos portáteis (Plans et al., 2013).

Os dois estudos destacados anteriormente ressaltam que a viabilidade do uso de um espectrômetro NIR portátil para resolver determinado problema é que a diferença existente nos resultados por ele obtidos em relação a um instrumento de bancada dependerão muito dos dois instrumentos comparados e do problema específico. A qualidade e o custo dos espectrômetros atualmente são muito variáveis, assim como é variável a complexidade dos sistemas que podem ser abordados com esses instrumentos. Em sistemas simples, espectrômetros com diferenças marcantes podem apresentar resultados semelhantes que possibilitam a aplicação de ambos, o que pode não ser verdadeiro para problemas de média ou alta complexidade.

Já foi destacada anteriormente a necessidade do desenvolvimento de modelos multivariados na grande maioria das aplicações utilizando a espectroscopia NIR. Dependendo da complexidade do problema, o desenvolvimento do método pode requerer um considerável conhecimento de Quimiometria e um número relativamente elevado de amostras, o que torna o desenvolvimento do método oneroso do ponto de vista de tempo e de custo com profissionais especializados. Por essa razão, a manutenção desses modelos e a sua transferência entre diferentes espectrômetros é de extrema importância. Não são raras as situações em que um método é desenvolvido e atualizado por anos em um mesmo espectrômetro, gerando um banco de dados de milhares de espectros, e em seguida pode tornar-se incompatível com novas leituras realizadas após o equipamento ter passado por manutenção ou pela troca do equi-

pamento. Diversos trabalhos já foram realizados propondo diferentes estratégias de transferência de modelos ou banco de dados espectrais entre equipamentos distintos ou em um mesmo equipamento em situações de desempenho diferentes (Wang et al., 1991; Wang et al., 1992; Lima; Borges, 2002; Honorato et al., 2007; Silva et al., 2017; Marchesini et al., 2018). Detalhes quanto à parte matemática dessas diferentes estratégias estão além dos objetivos desse capítulo, e os leitores interessados nesse detalhamento podem recorrer aos diversos trabalhos publicados na área de Quimiometria focando esse tema específico (Wang et al., 1991; Wang et al., 1992; Honorato et al., 2007).

Entre os diversos métodos de transferência propostos na literatura, a padronização de medidas espectrais é a atualmente a mais aplicada. Dentro dessa abordagem, diferentes procedimentos de transferência podem ser empregados, tais como: padronização direta, padronização direta por partes, padronização reversa e padronização reversa por partes (Wang et al., 1991; Lima; Borges, 2002). Considerando que a transferência envolva equipamentos diferentes, o equipamento onde o modelo foi originalmente construído é comumente chamado de primário ou mestre e o equipamento para o qual o modelo é transferido é chamado de secundário ou escravo.

Nos métodos diretos tem-se, como objetivo, fazer com que o espectro registrado no equipamento secundário se torne semelhante ao que teria sido medido no equipamento primário. Após essa padronização, os espectros do equipamento secundário corrigidos são analisados no modelo multivariado desenvolvido no equipamento primário. Por outro lado, nos métodos reversos, os espectros do equipamento primário são padronizados para que se assemelhem aos do equipamento secundário. Deste modo, com a padronização reversa, todo o banco de dados espectrais de um instrumento inicial (primário) pode ser padronizado para que se torne equivalente ao de um novo instrumento (secundário). Entretanto, um requisito para a aplicação desses métodos é que se disponha de um pequeno grupo de amostras, geralmente de 5 a 20, que foi medido nos dois equipamentos envolvidos são conhecidas como amostras de transferência.

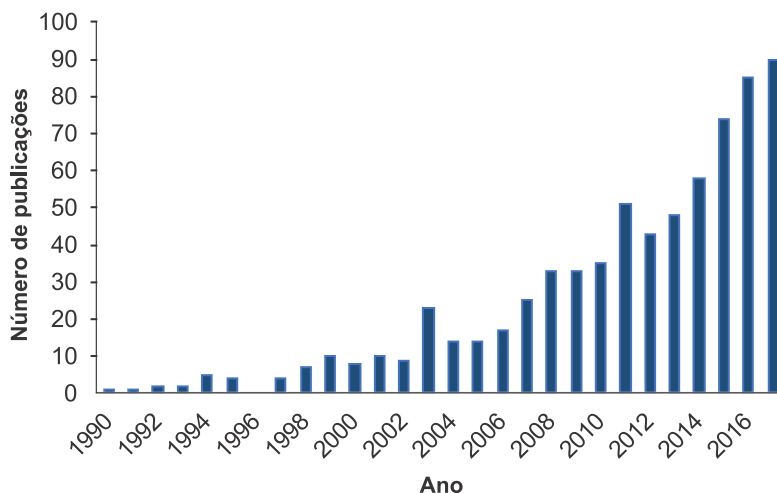
Trabalhos como os de Silva et al. (2017) já demonstraram que é possível a realização da transferência de modelos que foram desenvolvidos em equipamentos de bancada para portáteis com o objetivo de determinar parâmetros de qualidade de gasolina e misturas de biodiesel. Marchesini et al. (2018) também tiveram sucesso na transferência de modelos desenvolvidos de um espectrômetro de bancada para portáteis visando à determinação dos teores de matéria seca, proteína crua, fibra detergente neutra e ácida, para avaliar a qualidade de silagem de milho. Esses exemplos ilustram a potencialidade das estratégias de transferência entre equipamentos distintos, mes-



mo quando a diferença nas configurações dos instrumentos é considerável. Contudo, a facilidade ou o sucesso da realização da transferência de um modelo ou banco de dados é sempre maior para equipamentos similares. Um exemplo é o desenvolvimento de modelos em um espectrômetro portátil e a transferência desse para outras unidades do mesmo equipamento que serão utilizadas em diferentes localidades. O fato de diversos desses espectrômetros não possuírem partes móveis em sua óptica é outro fator que facilita a padronização de suas medidas espectrais.

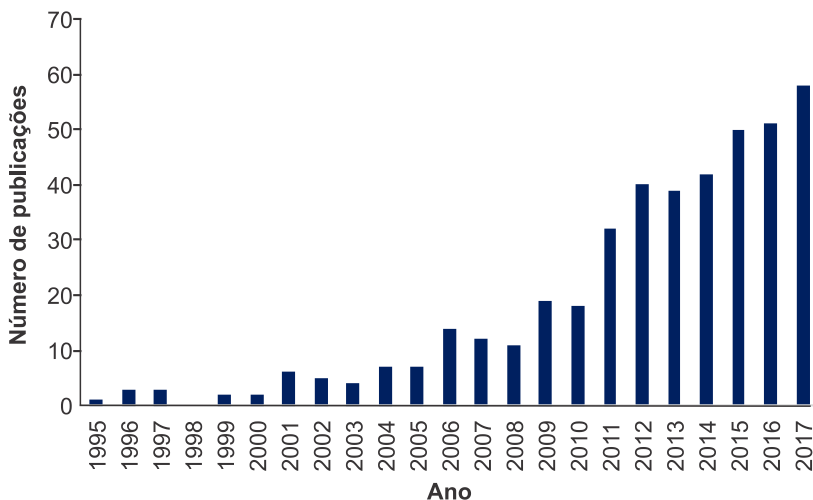
## Aplicações na análise de grãos

O número de artigos publicados utilizando espectrômetros NIR portáteis tem crescido expressivamente a cada ano desde 1990, como é possível observar no gráfico da Figura 1. Dentre estes trabalhos, aqueles que se dedicam à análise de grãos utilizando os equipamentos portáteis seguem um padrão semelhante quanto às publicações anuais. No ano de 2017, cerca de 60 trabalhos foram publicados, conforme é apresentado na Figura 2. Esses dados comprovam o interesse e a funcionalidade da aplicação de espectrômetros NIR portáteis para a análise de grãos e também indicam boas perspectivas para o desenvolvimento de novas técnicas de análise utilizando estes equipamentos.



**Figura 1.** Número de publicações utilizando espectrômetros NIR portáteis em diferentes análises no decorrer das últimas três décadas.

Fonte: Dados obtidos pela base de dados Web of Science em abril de 2018.



**Figura 2.** Número de publicações utilizando espectrômetros NIR portáteis para análise de grãos no decorrer das últimas três décadas.

Fonte: Dados obtidos pela base de dados Web of Science em abril de 2018.

Muitos desses trabalhos dedicam-se à quantificação de nutrientes, gordura, determinação do teor de umidade e classificação dos grãos quanto à qualidade, contaminação e deterioração. Kumagai et al. (2003) desenvolveram método de triagem que realizou a discriminação e a determinação do teor de cádmio em grãos não polidos de arroz utilizando um espectrômetro NIR portátil (PlaScan SH, OPT Research, Inc., Japão) e a análise canônica discriminante (CDA, *Canonical Discriminant Analysis*). Foram medidos espectros em triplicata na faixa de 1.100 nm a 2.500 nm de 106 amostras empacotadas dos grãos, as quais foram divididas em três níveis de concentração de cádmio, determinados anteriormente por espectroscopia de absorção atômica: baixo (< 0,4 ppm), médio (entre 0,4 ppm e 1,0 ppm) e alto (> 1,0 ppm). Os resultados obtidos comprovaram a eficiência na utilização do espectrômetro, com resultados corretos de discriminação de 86,7% para as amostras de baixa concentração de cádmio, de 83,9% para as de média e de 87,5% para as amostras com alto teor da espécie (Kumagai et al., 2003).

Outros autores dedicaram-se à avaliação da qualidade de castanhas portuguesas empregando equipamento NIR portátil. Moschetti et al. (2014a) desenvolveram método de discriminação baseado em análise discriminante linear (LDA, *Linear Discriminant Analysis*) para castanhas infestadas e não infestadas por larvas. O método de referência baseou-se na imersão das castanhas em uma solução salina e diferenciação de densidade por flutuação das castanhas danificadas. Dependendo

da forma de cultivo e de características intrínsecas das castanhas, este procedimento pode ocasionar muitos falsos negativos. A aquisição dos espectros NIR foi realizada em diferentes posições das castanhas, utilizando o Luminar 5030 (Brimrose Corp.), que emprega um filtro acústico-óptico sintonizável (AOTF, *Acousto-optic Tunable Filters*). Antes da análise por LDA, os autores realizaram uma seleção de comprimentos de onda empregando algoritmo genético. Comparando com o método por flutuação, o método NIR com o equipamento portátil apresentou uma melhora de cerca de 55% na classificação das castanhas, com erros totais em torno de 8,4%, empregando uma análise sem preparo de amostra, rápida e não destrutiva.

Em um trabalho seguinte dos mesmos autores, foi avaliado o dano causado por fungos às castanhas, utilizando-se o mesmo equipamento e três modelos de discriminação (LDA, QDA, *Quadratic Discriminant Analysis*, e *k*-NN, *k*-Nearest Neighbor). Avaliaram-se os melhores resultados em relação à posição na qual foram adquiridos os espectros e concluiu-se que a maior eficiência de discriminação de castanhas insalubres era obtida com espectros do lado plano das castanhas. Os resultados de eficiência na discriminação com diferentes pré-processamentos e escolha de variáveis foram maiores utilizando QDA (maior que 97% de eficiência), tornando o método portátil promissor como estratégia de triagem da qualidade de castanhas portuguesas (Moscetti et al., 2014b).

Grãos de milho são matéria-prima de grande importância econômica, uma vez que são parte principal da alimentação em muitas culturas. Muitos alimentos industrializados têm, em sua composição, flocos de milho e o estudo da qualidade desse material muitas vezes é necessário. Ayvaz et al. (2015) utilizaram dois espectrômetros portáteis, um portátil de mão na região do infravermelho próximo e outro portátil no infravermelho médio (MicroPhazir e Cary 630, respectivamente). Além disso, foi utilizado um espectrômetro NIR de bancada (Excalibur 3500, Varian) para realizar a classificação de flocos de milho de duas origens diferentes, orgânicos e convencionais. Para tanto, os autores utilizaram a modelagem independente por analogia de classe (SIMCA, *Soft Independent Modeling of Class Analogy*) com a avaliação e exclusão de amostras anômalas. Adicionalmente, realizaram a quantificação do teor de umidade e cinzas, bem como a determinação do tamanho de partícula dos flocos e de suas propriedades de liga por PLSR. Os melhores resultados de classificação obtidos seguindo a seguinte ordem: MIR portátil, NIR de bancada e NIR portátil. O melhor desempenho do MID foi justificado por causa das vibrações fundamentais de grupos funcionais específicos, que indicaram as diferenças entre as duas origens do milho. Contudo, os três instrumentos apresentaram elevadas eficiências, de forma que todos foram considerados aptos para aplicação em rotina. Em relação às quantificações por PLSR, o espectrômetro NIR de bancada obteve os menores erros de predição, mas novamente os três

espectrômetros atingiram resultados aceitáveis, com valores de razão entre o desvio padrão dos valores de referência e o erro padrão da validação variando entre 3,9 e 1,7 (Ayvaz et al., 2015).

O dano oculto em amêndoas foi avaliado por meio de espectroscopia NIR por Rogel-Castillo et al. (2016) utilizando o espectrômetro portátil de mão MicroNIR 2200 (JDSU) adaptado. Esta avaliação foi desenvolvida como alternativa de análise, uma vez que o dano não pode ser observado visualmente, a não ser que o grão seja aberto, isto é, destruído no procedimento analítico. Tendo em vista o estudo da absorção de óleos, proteínas e carboidratos, cujas taxas são mais baixas para amêndoas danificadas, as amostras foram tratadas para gerar os grupos de controle com diferentes teores de umidade, e os espectros foram obtidos por meio de um simples orifício até o centro do grão. Diferentes pré-processamentos foram testados e diferentes faixas espectrais foram selecionadas para realizar a discriminação por mínimos quadrados parciais para análise discriminante (PLS-DA, *Partial Least Squares for Discriminant Analysis*). Os modelos de discriminação desenvolvidos geraram taxas de falsos positivos/negativos de até 18%, sendo que o modelo desenvolvido com processamento de segunda derivada e utilizando quatro faixas espectrais apresentou melhor eficiência e robustez, com erros de até 9,2%.

Um dos mais recentes trabalhos acerca da análise de grãos utilizando um espectrômetro NIR portátil de mão foi desenvolvido por Correia et al. (2018) para a avaliação da qualidade de grãos de café (café arábica) no Brasil. Neste trabalho, foi utilizando o MicroNIR 1700 Pro para quantificação de adulterantes no café (café robusta, milho, cascas e paus). Para tanto, 125 amostras foram preparadas, com diferentes concentrações de adulterantes e em diferentes estágios de torra (leve, média e alta) e utilizadas para o desenvolvimento do modelo. Outras 16 amostras comerciais formaram o conjunto de teste do modelo desenvolvido. PCA e PLSR foram utilizados para o desenvolvimento dos modelos, os quais foram considerados eficientes e capazes de realizar a predição da quantidade de adulterantes no café, com limite de quantificação entre 5% m/m e 8% m/m. A simplicidade de utilização do MicroNIR, a velocidade de análise, a conservação da amostra e a portabilidade propiciaram avaliação em tempo real da qualidade do café.

Todos os trabalhos apresentados mostram a viabilidade de aplicação da espectroscopia NIR com equipamentos portáteis na caracterização e análise de diferentes grãos, principalmente gerando métodos de triagem de qualidade e métodos de quantificação de nutrientes ou adulterantes. A ampla aplicação destes dispositivos deve-se às vantagens já relatadas, e o cenário atual da literatura indica uma perspectiva de disseminação no uso dos espectrômetros portáteis no campo alimentício. Técnicas

rápidas, robustas, portáteis e que necessitam de mínimo preparo de amostra têm ganhado o seu espaço em diferentes campos de pesquisa e, muitas vezes, são preferência de aplicação, quando comparadas com técnicas onerosas e complexas.

## Conclusão e perspectivas

Entre os principais desafios na utilização de espectrômetros NIR portáteis ainda se pode citar a necessidade de profissionais especializados que tenham conhecimento sobre a técnica e, principalmente, sobre o desenvolvimento de métodos multivariados aplicando Quimiometria, a qual é essencial na análise e na obtenção das informações contidas nos dados espectrais NIR (Pasquini, 2003; Luo et al., 2013; Fernández Pierna et al., 2015). O objetivo da análise, escolha do espectrômetro, da região espectral, condições de medida, bem como número e forma de apresentação das amostras também são informações essenciais para a escolha do melhor modelo quimiométrico a ser aplicado na análise e sua possível transferência para outros instrumentos. Dessa forma, a popularização e o aumento na utilização de espectroscopia NIR tem que ser acompanhada pela formação de profissionais capacitados para desenvolvimento e validação de métodos. Uma vez vencidas essas etapas, pela existência de softwares com interface gráfica amigável e intuitiva, através de um treinamento relativamente rápido no equipamento, analistas com pouco conhecimento em Quimiometria podem realizar análises de forma rápida e eficiente.

Quando se comparam os dispositivos NIR portáteis com os equipamentos de bancada, algumas limitações são observadas em razão das diferenças instrumentais. Normalmente, a resolução dos equipamentos portáteis é menor, uma vez que a miniaturização dos equipamentos implica a retirada de módulos fundamentais para o aumento da resolução e da razão sinal/ruído. Além disso, muitos possibilitam a medida em apenas parte da região espectral do NIR, o que muitas vezes reduz a sensibilidade. Apesar destas e de outras limitações, para muitas situações o desempenho desses instrumentos pode ser comparável ao obtido com equipamentos de bancada ou ao menos suficiente para alcançar o objetivo pretendido (Sorak et al., 2012; Plans et al., 2013; Cabassi et al., 2015). Contudo, deve-se destacar que a grande maioria dos estudos relatados na literatura que utilizam instrumentos portáteis ou portáteis de mão ainda foi realizada nos ambientes controlados de um laboratório e não em condições de campo. Poucos são os trabalhos que apresentam uma tentativa de transferir um protocolo experimental de um método desenvolvido em laboratório para as situações de campo, para a detecção de estresses em plantas por espectroscopia VIS-NIR (Behrens et al., 2007). Portanto, ainda são necessários estudos práticos que demons-

trem de forma mais fidedigna o desempenho desses instrumentos em condições de campo, onde podem ser observadas variações de temperatura, umidade, presença de luz/radiação ambiente, variações sazonais entre as amostras, etc.

Por todas as características e vantagens mencionadas anteriormente, pode-se concluir que a utilização desses instrumentos tende a se intensificar. Em razão da tendência de redução dos custos de produção, da alta competitividade entre as diversas empresas que fabricam esses instrumentos e dos ótimos resultados já demonstrados em um grande número de aplicações, a espectroscopia NIR é uma das técnicas com maior potencial para o desenvolvimento de aplicações em campo, não só em indústrias e setor agropecuário para controle de grãos, pastagem, solo, etc., mas também para aplicações muito próximas da população em geral, como a verificação da qualidade de frutas e grãos em um mercado ou o monitoramento da qualidade do combustível diretamente nas bombas de postos de combustível.

## Referências

AYVAZ, H.; PLANS, M.; TOWERS, B. N.; AUER, A.; RODRIGUEZ-SAONA, L. E. The use of infrared spectrometers to predict quality parameters of cornmeal (corn grits) and differentiate between organic and conventional practices. **Journal of Cereal Science**, v. 62, p. 22-30, mar. 2015. Disponível em: <<http://dx.doi.org/10.1016/j.jcs.2014.12.004>>. Acesso em: 5 abr. 2018.

BEHRENS, T.; MULLER, J.; DIEPENBROCK, W. Optimizing a Diode Array VIS/NIR Spectrometer System to Detect Plant Stress in the Field. **J. Agronomy & Crop Science**, v. 193, n. 4, p. 292-304, 2007. Disponível em: <<https://doi.org/10.1111/j.1439-037X.2007.00266.x>>. Acesso em: 5 abr. 2018.

CABASSI, G.; CAVALLI, D.; FUCCELLA, R. Evaluation of four NIR spectrometers in the analysis of cattle slurry. **Biosystems Engineering**, v. 133, p. 1-13, May 2015. Disponível em: <<http://dx.doi.org/10.1016/j.biosystemseng.2015.02.011>>. Acesso em: 9 abr. 2018.

CORREIA, R. M.; TOSATO, F.; DOMINGOS, E.; RODRIGUES, R. R. T.; AQUINO, L. F. M.; FILGUEIRAS, P. R.; LACERDA, V.; ROMÃO, W. Portable near infrared spectroscopy applied to quality control of Brazilian coffee. **Talanta**, v. 176, p. 59-68, Jan. 2018. Disponível em: <<http://dx.doi.org/10.1016/j.talanta.2017.08.009>>. Acesso em: 5 abr. 2018.

COZZOLINO, D. Use of infrared spectroscopy for in-field measurement and phenotyping of plant properties: instrumentation, data analysis, and examples. **Applied Spectroscopy Reviews**, v. 49, n. 7, p. 564-584, 2014. Disponível em: <<http://dx.doi.org/10.1080/05704928.2013.878720>>. Acesso em: 9 abr. 2018.

FERNÁNDEZ PIERNA, J. A.; ABBAS, O.; LECLER, B.; HOGREL, P.; DARDENNE, P.; BAETEN, V. NIR fingerprint screening for early control of non-conformity at feed mills. **Food Chemistry**, v. 189, p. 2-12, Dec. 2015. Disponível em: <<https://doi.org/10.1016/j.foodchem.2014.09.105>>. Acesso em: 18 mar. 2016.

GUILLEMAIN, A.; DÉGARDIN, K.; ROGGO, Y. Performance of NIR handheld spectrometers for the detection of counterfeit tablets. **Talanta**, v. 165, p. 632-640, Apr. 2017. Disponível em: <<http://dx.doi.org/10.1016/j.talanta.2016.12.063>>. Acesso em: 5 abr. 2018.

HERBERHOLZ, L.; KOLOMIETS, O.; SIESLER, H. W. Quantitative analysis by a portable near infrared spectrometer: can it replace laboratory instrumentation for in situ analysis? **NIR News**, v. 21, n. 4, p. 6-8, 2010. Disponível em: <<http://journals.sagepub.com/doi/pdf/10.1255/nirn.1185>>. Acesso em: 16 abr. 2018.

HONORATO, F. A.; BARROS NETO, B.; MARTINS, M. N.; GALVÃO, R. K. H.; PIMENTEL, M. F. Transferência de calibração em métodos multivariados. **Química Nova**, v. 30, n. 5, p. 1301-1312, 2007.

KOLOMIETS, O.; SIESLER, H. W. The influence of spectral resolution on the quantitative near infrared spectroscopic determination of an active ingredient in a solid drug formulation. **Journal of Near Infrared Spectroscopy**, v. 12, p. 271-277, 2004. Disponível em: <<http://journals.sagepub.com/doi/pdf/10.1255/jnirs.435>>. Acesso em: 9 abr. 2018.

KUMAGAI, M.; OHISA, N.; AMANO, T.; OGAWA, N. Canonical discriminant analysis of cadmium content levels in unpolished rice using a portable near-infrared spectrometer. **Analytical sciences : the international journal of the Japan Society for Analytical Chemistry**, v. 19, n. 11, p. 1553-1555, 2003. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/14640458>>. Acesso em: 5 abr. 2018.

LUO, W.; WU, J.; WANG, X.; LIN, X.; LI, H. Near infrared spectroscopy combination with PLS to monitor the parameters of naproxen tablet preparation process. **Analytical Methods**, v. 5, n. 5, p. 1337, 2013. Disponível em: <<https://doi.org/10.1039/C2AY26260F>>. Acesso em: 9 abr. 2018.

LIMA, F. S. G.; BORGES L. E. P. Evaluation of Standardization Methods of near Infrared Calibration Models. **Journal of Near Infrared Spectroscopy**, v. 10, n. 4, p. 269-278, 2002. Disponível em: <<http://journals.sagepub.com/doi/pdf/10.1255/jnirs.343>>. Acesso em: 9 abr. 2018.

LYSAGHT, M. J.; VAN ZEE, J. A.; CALLIS, J. B. Laptop chemistry: A fiberoptic, field portable, near infrared spectrometer. **Review of Scientific Instruments**, v. 62, n. 2, p. 507-515, 1991. Disponível em: <<https://doi.org/10.1063/1.1142095>>. Acesso em: 9 abr. 2018.

- MALEGORI, C.; MARQUES, E. J. N.; FREITAS, S. T.; PIMENTEL, M. F.; PASQUINI, C.; CASIRAGHI, E. Comparing the analytical performances of Micro-NIR and FT-NIR spectrometers in the evaluation of acerola fruit quality, using PLS and SVM regression algorithms. **Talanta**, v. 165, p. 112-116, Apr. 2017. Disponível em: <<http://dx.doi.org/10.1016/j.talanta.2016.12.035>>. Acesso em: 9 abr. 2018.
- MARCHESINI, G.; SERVA, L.; GARBIN, E.; MIRISOLA, M.; ANDRIGHETTO, I. Near-infrared calibration transfer for undried whole maize plant between laboratory and on-site spectrometers. **Italian Journal of Animal Science**, v. 17, n. 1, p. 66-72, 2018. Disponível em: <<http://dx.doi.org/10.1080/1828051X.2017.1345660>>. Acesso em: 5 abr. 2018.
- MITCHELL, A. R.; PINTER JR., P. J.; GUERRERO, J. N.; HERNANDEZ, C. B.; MARBLE, V. L. Spectral reflectance measurements of alfalfa under sheep grazing. **Agronomy Journal**, v. 82, n. 6, p. 1098-1103, 1990. Disponível em: <<http://dx.doi.org/10.2134/agronj1990.00021962008200060014x>>. Acesso em: 9 abr. 2018.
- MOSCETTI, R.; HAFF, R. P.; SARANWONG, S.; MONARCA, D.; CECCHINI, M.; MASSANTINI, R. Nondestructive detection of insect infested chestnuts based on NIR spectroscopy. **Postharvest Biology and Technology**, v. 87, p. 88-94, Jan. 2014a. Disponível em: <<http://dx.doi.org/10.1016/j.postharvbio.2013.08.010>>. Acesso em: 5 abr. 2018.
- MOSCETTI, R.; MONARCA, D.; CECCHINI, M.; HAFF, R. P.; CONTINI, M.; MASSANTINI, R. Detection of mold-damaged chestnuts by near-infrared spectroscopy. **Postharvest Biology and Technology**, v. 93, p. 83-90, July 2014b. Disponível em: <<http://dx.doi.org/10.1016/j.postharvbio.2014.02.009>>. Acesso em: 5 abr. 2018.
- 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. Disponível em: <<http://dx.doi.org/10.1590/S0103-50532003000200006>>. Acesso em: 26 jan. 2017.
- PLANS, M.; SIMÓ, J.; CASAÑAS, F.; SABATÉ, J.; RODRIGUEZ-SAONA, L. Characterization of common beans (*Phaseolus vulgaris* L.) by infrared spectroscopy: comparison of MIR, FT-NIR and dispersive NIR using portable and benchtop instruments. **Food Research International**, v. 54, n. 2, p. 1643-1651, 2013. Disponível em: <<http://dx.doi.org/10.1016/j.foodres.2013.09.003>>. Acesso em: 5 abr. 2018.
- ROGEL-CASTILLO, C.; BOULTON, R.; OPASTPONGKARN, A.; HUANG, G.; MITCHELL, A. E. Use of near-infrared spectroscopy and chemometrics for the nondestructive identification of concealed damage in raw almonds (*Prunus dulcis*). **Journal of Agricultural and Food Chemistry**, v. 64, n. 29, p. 5958-5962, 2016. Disponível em: <<http://dx.doi.org/10.1021/acs.jafc.6b01828>>. Acesso em: 5 abr. 2018.



SANTOS, C. A. T.; PÁSCOA, R. N. M. J.; LOPES, J. A. A review on the applications of portable near-infrared spectrometers in the agro-food industry. **Applied Spectroscopy**, v. 67, n. 11, p. 1215-1233, 2013. Disponível em: <<https://doi.org/10.1366/13-07228>>. Acesso em: 9 abr. 2018.

SILVA, N. C.; CAVALCANTI, C. J.; HONORATO, F. A.; AMIGO, J. M.; PIMENTEL, M. F.; Standardization from a benchtop to a handheld NIR spectrometer using mathematically mixed NIR spectra to determine fuel quality parameters. **Analytica Chimica Acta**, v. 954, p. 32-42, Feb. 2017. Disponível em: <<http://dx.doi.org/10.1016/j.aca.2016.12.018>>. Acesso em: 9 abr. 2018.

SORAK, D.; HERBERHOLZ, L.; IWASCEK, S.; ALTINPINAR, S.; PFEIFER, F.; SIESLER, H. W. New developments and applications of handheld raman, mid-infrared, and near-infrared spectrometers. **Applied Spectroscopy Reviews**, v. 47, n. 2, p. 83-115, 2012. Disponível em: <<https://doi.org/10.1080/05704928.2011.625748>>. Acesso em: 5 abr. 2018.

WANG, Y.; VELTKAMP, D. J.; KOWALSKI, B. R. Multivariate instrument standardization. **Analytical Chemistry**, v. 63, n. 23, p. 2750-2756, 1991. Disponível em: <<http://dx.doi.org/10.1021/ac00023a016>>. Acesso em: 9 abr. 2018.

WANG, Y.; LYSAGHT, M. J.; KOWALSKI, B. R. Improvement of multivariate calibration through instrument standardization. **Analytical Chemistry**, v. 64, n. 5, p. 562-564, 1992. Disponível em: <<http://dx.doi.org/10.1021/ac00029a021>>. Acesso em: 9 abr. 2018.

WILLIAMS, P.; NORRIS, K. **Near-infrared technology in the agricultural and food industries**. 2nd ed. St. Paul: American Association of Cereal Chemists, 2001. 296 p.

# Aplicações da Espectroscopia no Infravermelho Próximo na Cadeia Produtiva de Grãos

Maicon Andre de Paula  
Oneida Vasconcelos Vieira  
Sidnei Menossi  
José Elias Carvalho  
Cristina Rocha Vieira Abucáter

## Cocamar

### Características da Empresa

A Cocamar Cooperativa Agroindustrial foi fundada em 27 de março de 1963, em Maringá, PR. Reuniu, inicialmente, um grupo de 46 fundadores, todos produtores de café. O objetivo era organizar a produção regional, receber e beneficiar o produto. Com o tempo, a cooperativa diversificou os negócios e cresceu. Hoje, a Cocamar está presente em vários municípios em mais de 60 unidades operacionais espalhadas pelo Norte e Noroeste do Paraná, Oeste paulista e Sudoeste do Mato Grosso do Sul. Conta com 13 mil associados que atuam na produção de soja, milho, trigo, café e laranja. Teve um planejamento de esmagamento de 770.000 t/ano em 2017 e comercializa farelo de soja, óleo degomado e óleo refinado.

A primeira indústria construída pela cooperativa foi de farelos e óleos vegetais, inaugurada em 1979, sendo também a primeira, até então, do cooperativismo paranaense. Tem capacidade de moagem de 3.250 mil toneladas de soja/dia. Além da indústria, possui estruturas de refino e envase.

A Cocamar conta com duas plantas de farelo de soja integral, denominadas Fábrica 1 e Fábrica 2, de esmagamento de soja, com capacidade de produção instalada de 3.100 t/dia, sendo 2.500 t/dia (80,65%) processadas pela Fábrica 1 e 600 t/dia (19,35%) processadas na Fábrica 2.

## Preparo das amostras

Na indústria estão instaladas três *probes* (sondas do NIR *on line* da Bruker), na saída da produção e na saída dos armazéns. A implantação de controle nestes pontos auxilia a área de produção com informações adicionais aos dados coletados pelos demais equipamentos instalados no processo, auxiliando no controle da qualidade do farelo de entrada e saída dos armazéns. Dois itens têm ação imediata na qualidade do farelo: a retirada de casca após o quebrador, que retorna ao farelo antes de passar pelo tostador, e a umidade de saída do secador, para garantir que esta fique abaixo de 12,5%. Assim, alteram-se os parâmetros da amostra e aumenta-se a distribuição na elaboração da curva. Para determinar em que momento ocorreria a coleta do farelo de soja após a alteração dos parâmetros, realizou-se um estudo coletando amostras para análise na via tradicional a cada 5 minutos, até determinar o momento em que ocorresse uma mudança significativa nos parâmetros e, confrontando com o desvio padrão da metodologia tradicional. Foi definido um período mínimo de 30 minutos após a alteração no processo para iniciar a coleta da amostra e foram adquiridas alíquotas do farelo de soja durante todo o período de aquisição do espectro (400 leituras em 30 segundos por ciclo). O farelo de soja coletado é homogeneizado e quarteado para uma amostra, de aproximadamente, 500 gramas.

A amostra é quarteada em dois volumes de, aproximadamente, 250 gramas. Uma amostra é direcionada para moagem utilizando moinho de martelo com peneira de 0,5 mm por, no máximo, 3 minutos, e a segunda amostra é encaminhada para demais análises (fibras, proteína solúvel, atividade ureática, cinzas, etc.), que não possuem modelos de calibração multivariada utilizando a espectroscopia no infravermelho próximo.

## Equipamentos NIRS disponíveis

No complexo de Maringá, PR, na fábrica de extração de óleos, está instalado o espectrômetro MATRIX-F FT-NIR On-Line da Bruker, com capacidade de seis *probes*. Atualmente, estão implantadas na fábrica uma *probe* na fita transportadora de farelo de soja para os armazéns e duas *probes* estão instaladas nas fitas que direcionam o produto estocado para a expedição de farelo de soja, sendo as leituras alternadas entre os três pontos de controle. O equipamento NIR possui comunicação por fibra ótica entre o espectrômetro e as *probes*, não ultrapassando uma distância máxima de 50 metros de fibra ótica, evitando-se que ocorra uma possível perda de informação coletada.

## Faixa de medida espectral e resolução

Na elaboração dos modelos de calibração multivariada utilizando a espectroscopia NIR, utilizamos a faixa espectral entre  $9.450\text{ cm}^{-1}$  e  $4.450\text{ cm}^{-1}$ , com resolução de  $2\text{ cm}^{-1}$ .

## Metodologia

### Resultados de referência

A utilização da tecnologia NIR na Cocamar teve origem em 2012, com a instalação de um NIR de Bancada da marca NIR XDS da Foss como ação do processo de melhoria contínua desenvolvido pelo programa Lean Six Sigma, com a ferramenta DMAIC (sigla para os termos *Define, Measure, Analyse, Improve e Control*) (De Feo; Barnard, 2005). Com o uso dessa ferramenta, ocorreu progressivamente a melhoria dos resultados obtidos pelos métodos de referência, diminuindo os desvios padrão pelo controle sistemático da repetibilidade e reprodutibilidade e pela atualização dos métodos analíticos. Após a melhoria dos métodos de referência, desenvolvemos os primeiros modelos de calibração utilizando a espectroscopia NIR no laboratório da Cocamar. Assim, para a implantação do NIR *on line* da Bruker, a Cocamar já possuía uma estrutura consolidada no laboratório físico-químico e adequada para desenvolver os modelos multivariados utilizando os espectros obtidos com o novo equipamento NIR. Os resultados das análises realizadas pela tecnologia NIR são acompanhados em reuniões semanais de S&OP (*Sales and Operations Planning*) (Thomé et al., 2012). Quando necessário, são apresentados o planejamento de coleta de amostras para o desenvolvimento de modelo de calibração e a avaliação do monitoramento das análises de verificação do controle de qualidade.

### Metodologia de referência utilizada

Teor de Umidade e Voláteis – Método Cocamar, baseado no método oficial AOCS Ca 2c-25 (American Oil Chemists' Society, 1997c).

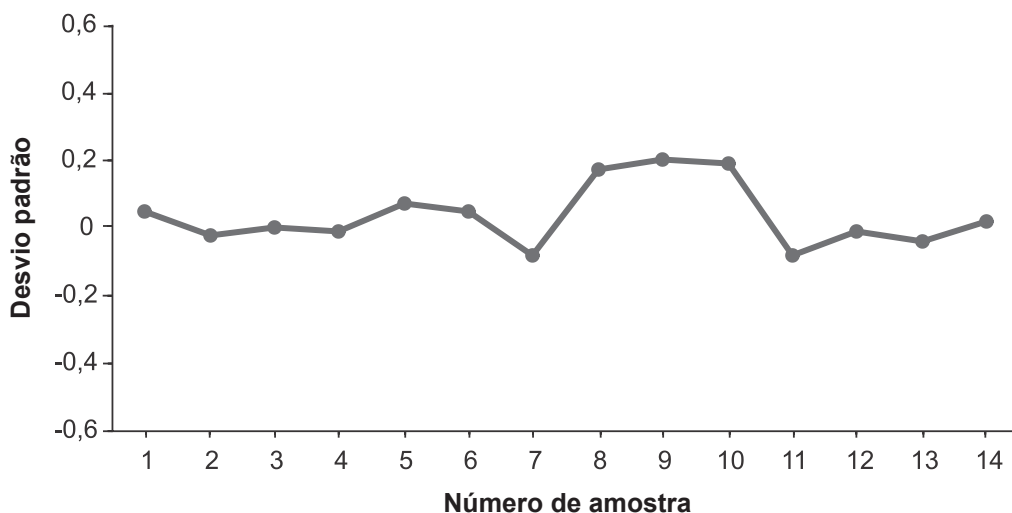
Teor de Proteína Bruta – Método Cocamar, baseado no método oficial AOCS BA 4e-93 (American Oil Chemists' Society, 1997b).

Teor de Óleo – Método Cocamar, baseado no método oficial AOCS BA3-38 (American Oil Chemists' Society, 1997a).

## Resultados

O desenvolvimento de modelos é realizado pela associação dos espectros obtidos no NIR *on line* aos resultados obtidos pelos métodos de referência utilizando métodos de calibração multivariada com validação cruzada. No início de uma nova safra é realizada a atualização dos modelos, com a inclusão de novas amostras e a verificação do RMSECV (Raiz QAdrada do Erro Médio Quadrático de Validação Cruzada, RPD (Desvio de Predição Residual), RSQ (Coeficiente de Determinação da Calibração) e a Distância Limite de Mahalanobis. As inclusões de novas amostras são avaliadas pelo método de PCA (Análise de Componentes Principais) para determinar a ocorrência de *outliers* de espectro que possam distorcer uma calibração. Em determinados casos acontece uma avaliação do modelo com atualização de amostras, por exemplo, quando ocorre uma alteração no processo industrial, nas condições de armazenamento da soja, e o aumento do resíduo entre o valor real e o predito pelo modelo supera  $\pm 3$  desvios padrão do método tradicional na carta CEP (Controle Estatístico de Processo) (Woodall; Montgomery, 1999), ou mesmo quando se verifica a ocorrência de viés na carta CEP.

Durante o período de monitoramento dos modelos de calibração, as coletas de controle são realizadas de segunda-feira a sexta-feira, três vezes ao dia, nas três *probes* instaladas, com espaçamento de uma a duas horas e alternando entre os períodos da manhã e tarde para alimentar a carta CEP (Figura 1).



**Figura 1.** Carta de controle de processo para determinação de umidade em farelo de soja.

Fonte: Adaptado da norma ISO 12099: 2017.

Para a construção dos modelos, utilizou-se o software Opus Versão 7 – Copyright© Bruker Optik GmbH 2011. Os dados espectrais foram centrados na média e pré-processados utilizando primeira derivada com janela de 17 pontos. Utilizou-se a validação cruzada com as configurações do software de calibração. Os resultados obtidos para os indicadores estatísticos (ASTM, 2012) utilizados no desenvolvimento dos modelos estão descritos na Tabela 1.

**Tabela 1.** Resultados de análise de óleo, proteína e umidade realizadas em farelo de soja.

Parâmetro	Óleo	Proteína	Umidade
Número de amostras Calibração	191	568	417
Média <sup>1</sup>	2,21%	45,80%	13,09%
Faixa do componente	0,85 - 6,7	40,88 - 49,38	10,57 - 15,25
Tratamento dos espectros	SNV <sup>2</sup>	SNV <sup>2</sup>	MSC <sup>3</sup>
Máximo de componentes	10	9	10
Faixa espectral (cm <sup>-1</sup> )	9.427 – 5.446,4	9.411,2 – 4.921,6	8.871,2 – 4.474,2
Distância limite de Mahalanobis	0,26	0,12	0,16
RMSECV <sup>4</sup>	0,19	0,39	0,23
RSQ <sup>5</sup>	97,11	87,07	85,01
Bias	-0,00133	-0,00141	-0,00224
RPD <sup>6</sup>	5,88	2,78	2,58
Desvio padrão método de referência	0,09%	0,15%	0,05%

<sup>1</sup> Média – média dos resultados, <sup>2</sup> SNV – Variação padrão normal, <sup>3</sup> MSC – Correção multiplicativa de signal, <sup>4</sup> RMSECV – Raiz quadrada do erro médio quadrático de validação cruzada, <sup>5</sup> RSQ – Coeficiente de determinação da calibração, <sup>6</sup> RPD – Relação de desempenho do desvio.

## Considerações finais

A utilização do NIR *on line* possibilita a inspeção de quase 100% da produção conforme suas configurações, podendo minimizar falhas e reprocesso. A partir deste melhor acompanhamento, é possível estabelecer limites de controle mais apropriados a oscilações de processo. Antes o número de amostras coletadas no processo utilizando o NIR de bancada era limitado pelo tempo de coleta e preparo da amostra para realizar a leitura nas dependências do laboratório. Assim, com a introdução do NIR *on line*, conseguimos agilizar a resposta da análise para a produção a quase um tempo real

e direcionar os colaboradores para realizarem uma produção mais precisa em termos de processo, evitando perdas e retrabalho.

A desvantagem é o aumento do desvio padrão do resultado, que é minimizado pela capacidade de análises a cada 40 segundos. Assim, pode-se avaliar as oscilações do processo em comparação com as análises tradicionais, que demoram horas e têm caráter destrutivo, impossibilitando a reanálise da mesma amostra. Com isso, estudos estatísticos do processo como o DOE (*Design of Experiments*) podem ter uma resposta mais rápida e refinada para cada conjunto de parâmetros estudado.

## Referências

AMERICAN OIL CHEMISTS' SOCIETY. **Method BA 4e-93**. 4th ed. Champaign, 1997. Official Methods and Recommended Practices.

AMERICAN OIL CHEMISTS' SOCIETY. **Method BA3-38**. 4th ed. Champaign, 1997. Official Methods and Recommended Practices.

AMERICAN OIL CHEMISTS' SOCIETY. **Method Ca 2c-25**. 4th ed. Champaign, 1997. Official Methods and Recommended Practices.

DE FEO, J. A.; BARNARD, W. W. **Juran Institute's Six Sigma Breakthrough and Beyond: quality performance breakthrough methods**. New York: McGraw-Hill, 2004. 374 p.

THOMÉ, A. M. T.; SCAVARDA, L. F.; FERNANDEZ, N. S.; SCAVARDA, A. J. Sales and operations planning: A research synthesis. **International Journal of Production Economics**, v. 138, n. 1, p. 1-13, 2012.

WOODALL, W. H.; MONTGOMERY, D. C. Research issues and ideas in statistical process control. **Journal of Quality Technology**, v. 31, n. 4, p. 376-386, 1999.

## CBO Análises Laboratoriais

### Características da Empresa

Fundado em 2001, o laboratório CBO iniciou suas atividades na cidade de Campinas, SP, e, através da qualidade e confiança do seu trabalho, foi reconhecido em todo o Brasil.

Enfrentando diversos desafios do segmento, a empresa manteve sua busca pela excelência e continuou seus investimentos, o que possibilitou capacitar colaboradores, aumentar o portfólio de análises e potencializar ainda mais a confiança dos clientes. Atualmente, a empresa conta com uma infraestrutura de ponta localizada na cidade de Valinhos, SP.

Inaugurada em setembro de 2016, a nova sede tem mais de 2.000 m<sup>2</sup> de estrutura que potencializam sua qualidade através de três laboratórios (Físico-Químico, Cromatografia, Microbiologia) totalmente integrados, com grande capacidade produtiva e que atendem às áreas de nutrição animal, nutrição humana, cosmética, farmacêutica, veterinária e universidades no apoio às suas pesquisas.

A CBO Análises Laboratoriais está sempre preocupada em investir em tecnologia, garantindo cada vez mais precisão, confiabilidade e imparcialidade nos seus resultados.

O laboratório atualmente conta com softwares de controle e aplicativos totalmente integrados que agilizam todo o processo operacional e garantem a segurança da informação.

Nossa qualidade e nosso crescimento estão associados diretamente ao alto conhecimento e capacidade dos nossos colaboradores e empenho constante deles na busca pela excelência.

E é com essa energia que a CBO mantém seus treinamentos atualizados por profissionais certificados e dentro de todos os padrões internacionais de qualidade.

Preparo das amostras: moagem de aproximadamente 100 g de amostra. A amostra moída fica com tamanho de partícula entre 0,5 mm e 1,0 mm.

Equipamento NIR disponível: Foss DS2500. Faixa de comprimento de onda: 400 nm a 2.500 nm; resolução: 0,5 nm; número de pontos: 4.200.

Faixa de medida espectral: 400 nm a 2.500 nm; resolução: 0,5 nm.

Produtos analisados através do NIR tradicional: milho, farelo de soja, sorgo, farelo de trigo, soja extrusada, soja desativada, soja semi-integral, farinha de carne, farinha de vísceras, farinha de pena.

Análises realizadas no NIR tradicional: umidade, proteína, extrato etéreo, fibra bruta, matéria mineral. Em farinha de carne, realizamos também as determinações de cálcio e fósforo.



## Metodologia

Resultados de referência: foram desenvolvidos na empresa, a partir de resultados de referência próprios.

Metodologia de referência utilizada: foram utilizados os métodos do Compêndio Brasileiro de Alimentação Animal (Compêndio..., 2017).

Informações quanto à manutenção do modelo de calibração: a atualização do modelo de calibração é realizada a cada 2 anos.

## Resultados

Utilização das análises via NIR: as análises via NIR são solicitadas pelos clientes do laboratório que precisam de resultados rápidos para a tomada de decisão em seus processos produtivos.

As principais análises realizadas em milho, sorgo, soja (farelo, extrusada e semi-integral), farelo de trigo, farinha de carne, farinha de vísceras e farinha de pena pelo Laboratório CBO utilizando a tecnologia NIR estão apresentadas na Tabela 2.

**Tabela 2.** Resultados obtidos para os modelos de calibração multivariada utilizando a espectroscopia NIR para a realização de análises químicas em milho, sorgo, soja, trigo, farinha de carne, farinha de vísceras e farinha de pena.

Produto e parâmetro analisado	Nº	Média	SD	SEC	RSQ	SECV/ SEP	1-VR	Variável latentes
Milho - Umidade	1.206	12,35	0,82	0,28	0,83	0,34	0,83	8
Milho - Proteína	1.489	7,60	0,72	0,18	0,91	0,21	0,91	8
Milho - Extrato etéreo	997	3,92	0,38	0,16	0,78	0,17	0,78	8
Milho - Fibra bruta	284	1,34	0,25	0,05	0,91	0,07	0,91	8
Milho - Matéria mineral	1.074	1,13	0,13	0,06	0,79	0,06	0,79	8
Sorgo - Umidade	849	11,50	1,05	0,14	0,98	0,16	0,98	8
Sorgo - Proteína	931	8,17	0,72	0,13	0,96	0,14	0,96	8
Sorgo - Extrato etéreo	587	2,32	0,47	0,18	0,78	0,22	0,78	8
Sorgo - Fibra bruta	494	1,15	0,28	0,08	0,88	0,10	0,88	8
Sorgo - Matéria mineral	380	1,26	0,13	0,04	0,86	0,05	0,86	8

continua...

Tabela 2. Continuação.

Produto e parâmetro analisado	Nº	Média	SD	SEC	RSQ	SECV/ SEP	1-VR	Variável latentes
Farelo de soja - Umidade	1.046	12,19	0,66	0,35	0,66	0,38	0,66	8
Farelo de soja - Proteína	1.652	47,09	1,50	0,31	0,95	0,34	0,95	8
Farelo de soja - Extrato etéreo	857	1,64	0,46	0,23	0,72	0,24	0,72	8
Farelo de soja - Fibra bruta	1.107	5,02	1,06	0,33	0,88	0,37	0,88	8
Farelo de soja - Matéria mineral	350	6,47	0,36	0,13	0,66	0,21	0,66	8
Soja desativada - Umidade	588	9,82	2,35	0,24	0,98	0,28	0,98	8
Soja desativada - Proteína	622	36,65	2,02	0,23	0,98	0,26	0,98	8
Soja desativada - Extrato etéreo	635	21,82	1,39	0,50	0,84	0,55	0,84	8
Soja desativada - Fibra bruta	232	4,58	0,99	0,24	0,92	0,28	0,92	8
Soja desativada - Matéria mineral	167	5,04	0,34	0,10	0,85	0,13	0,85	8
Soja extrusada - Umidade	686	6,80	1,10	0,20	0,95	0,23	0,95	8
Soja extrusada - Proteína	636	37,61	1,35	0,26	0,95	0,29	0,95	8
Soja extrusada - Extrato etéreo	591	21,49	2,07	0,34	0,96	0,41	0,96	8
Soja semi-integral - Umidade	265	7,11	0,91	0,11	0,98	0,14	0,98	8
Soja semi-integral - Proteína	763	42,98	1,99	0,31	0,97	0,33	0,97	8
Soja semi-integral - Extrato etéreo	642	12,49	2,14	0,22	0,98	0,24	0,98	8
Farelo de trigo - Umidade	859	12,13	0,70	0,26	0,83	0,28	0,83	8
Farelo de trigo - Proteína	1.008	15,77	1,16	0,21	0,96	0,22	0,96	8
Farelo de trigo - Extrato etéreo	321	3,04	0,33	0,10	0,84	0,13	0,84	8
Farelo de trigo - Fibra bruta	903	8,46	2,05	0,51	0,93	0,53	0,93	8
Farelo de trigo - Matéria mineral	207	4,52	0,36	0,06	0,95	0,08	0,95	8
Farinha de carne - Umidade	1.338	5,73	0,94	0,27	0,91	0,27	0,91	8
Farinha de carne - Proteína	1.076	44,83	3,28	0,35	0,99	0,37	0,99	8
Farinha de carne - Extrato etéreo	1.027	11,49	1,85	0,35	0,96	0,37	0,96	8
Farinha de carne - Matéria mineral	746	37,61	3,22	0,47	0,97	0,51	0,97	8
Farinha de carne - Cálcio	914	13,07	1,32	0,32	0,93	0,34	0,93	8
Farinha de carne - Fósforo	1.517	6,13	0,59	0,18	0,90	0,18	0,90	8
Farinha de vísceras - Umidade	1.832	4,59	1,60	0,49	0,89	0,52	0,89	8
Farinha de vísceras - Proteína	1.392	63,20	3,57	0,49	0,98	0,54	0,98	8
Farinha de vísceras - Extrato etéreo	1.657	12,94	2,53	0,47	0,96	0,50	0,96	8
Farinha de vísceras - Matéria mineral	1.487	14,48	3,02	0,50	0,96	0,56	0,96	8
Farinha de pena - Umidade	846	7,28	1,89	0,25	0,98	0,28	0,98	8
Farinha de pena - Proteína	1.111	83,86	3,00	0,59	0,95	0,63	0,95	8

continua...

**Tabela 2.** Continuação.

Produto e parâmetro analisado	Nº	Média	SD	SEC	RSQ	SECV/ SEP	1-VR	Variável latentes
Farinha de pena - Extrato etéreo	862	6,08	2,04	0,36	0,96	0,39	0,96	8
Farinha de pena - Matéria mineral	910	2,38	0,79	0,22	0,91	0,23	0,91	8

\*Todos os espectros foram pré-processados utilizando o algoritmo SNV (Variação padrão normal); N (nº amostras); Média (Média dos resultados); SD (Desvio padrão); SEC (Erro padrão de calibração); RSQ (Coeficiente de determinação da calibração); SECV/SEP (Erro padrão da validação cruzada/Erro padrão de predição); 1-VR (1-Variance Ratio - Coeficiente de determinação da validação cruzada); e Variável latente (nº de fatores).

## Considerações finais

Na nossa atividade (laboratório comercial), a utilização do NIR é vantajosa para atender aos clientes que solicitam um grande número de análises de ingredientes para ração. Para estes clientes, o NIR possibilita custos mais baixos, com resultados confiáveis e que atendem às necessidades de controle de qualidade na formulação.

## Perspectivas futuras

Acreditamos que, com o avanço desta tecnologia, poderemos ter novas aplicações no nosso negócio.

## Referências

COMPÊNDIO brasileiro de alimentação animal. 5. ed. São Paulo: Sindirações, 2017.

## Cooperativa Agroindustrial Alfa - Cooperalfa

### Características da Empresa

Fundada em 1967, em Chapecó, SC, por 39 agricultores, com o objetivo de promover o desenvolvimento econômico através da cooperação, a Cooperalfa construiu uma trajetória que demonstra a força da união em prol de um objetivo comum. Mais de 80% da base associativa da cooperativa é de agricultores familiares.

Acreditando que cooperar é evoluir, a Cooperalfa desenvolve um primoroso trabalho de desenvolvimento técnico, social e econômico com seus associados, colaboradores e comunidades de seu entorno.

Raízes fortes em uma região com vocação para a produção de alimentos proporcionaram à Cooperalfa a base necessária para solidificar-se e buscar a expansão. Através de fusões e incorporações, a Alfa levou sua experiência para outras regiões.

Atualmente, atua em Santa Catarina, Mato Grosso do Sul, Paraná e Rio Grande do Sul, nos segmentos de recepção, comercialização e industrialização de milho, soja e trigo, entre outras atividades. Tem capacidade de processamento de 1.200 toneladas de soja por dia e uma indústria de trigo com capacidade de moagem de 650 toneladas por dia.

Atuando em um setor estratégico para a economia nacional, a Cooperalfa promove suas atividades visando à sustentabilidade futura.

Preparo das amostras:

- É realizada a coleta em determinados horários, de acordo com a necessidade do processo (por exemplo, de hora em hora);
- A moagem é feita em moinho faca com peneira 3 mm;
- As amostras com granulometria menor que 3 mm são analisadas de forma natural;
- A quantidade de amostra coletada é em torno de 500 gramas e, para análise, são utilizadas 50 gramas.

Equipamento NIR disponível: NIR marca Buchi modelo Nirflex N-500 Solids, reflectância com 32 scans por leitura.

Faixa de medida espectral e resolução: faixa de medida 1.000 nm a 2.500 nm; resolução de 4 cm<sup>-1</sup>.

Produtos analisados através do NIR tradicional: soja, farelo de soja, farinha de trigo, farelo de trigo, milho extrusado, soja desativada e nutrisoja.

Análises realizadas no NIR tradicional: umidade, proteína bruta, residual de óleo, resíduo mineral, fibra bruta.

## Metodologia

Resultados de referência: os resultados de referência são obtidos em laboratório próprio.

Metodologia de referência utilizada: AACC (*American Association of Cereal Chemists*) (AACC International, 1995a, 1995b, 1995c) e, para os métodos modificados são realizadas validação e participação em ensaios de proficiência.

Informações quanto à manutenção do modelo de calibração:

- Comparar correlação dos resultados obtidos no NIR com métodos tradicionais de referência, pois mesmo em modelos de calibração utilizados na rotina, são analisados em torno de 10% das amostras em via tradicional.
- Os resultados por métodos tradicionais encontrados nos extremos de amplitude de calibração são adicionados ao banco de dados, a fim de melhorar os modelos de calibração.

## Resultados

As análises utilizando a espectroscopia no infravermelho próximo associada a métodos de calibração multivariada são utilizadas na rotina das indústrias de trigo, soja e milho para controle de processo.

As principais análises realizadas são: umidade, cinzas, proteína, extrato etéreo e fibra bruta.

## Considerações finais

Vantagens:

- Possibilidade de obter vários resultados (umidade, proteína bruta, extrato etéreo), com o mesmo equipamento e a mesma matriz;
- Ótima correlação com métodos de referência para ensaios quantitativos;
- Agilidade nos resultados;
- Não necessita diluições e/ou utilização de reagentes;
- Menor influência do analista.

Limitações:

- Impossibilidade de analisar substâncias com baixa concentração, como as micotoxinas.

## Perspectivas futuras

Instalação de NIR *on line* no processo industrial.

## Referências

AACC International. **Method 08-12**: ash in farina and semolina. 9th ed. St. Paul, 1995a. v. 1. Approved methods of analysis.

AACC International. **Method 44-15A**: moisture. 9th ed. St. Paul, 1995b. Approved methods of analysis.

AACC International. **Method 46-12**: crude protein. 9th ed. St. Paul, 1995c. v. 1. Approved methods of analysis.

# Vilma Alimentos

## Características da Empresa

Em outubro de 2015, a Vilma completou 90 anos. Em sua trajetória, a empresa fortaleceu-se, fez aquisições e reinventou-se de várias maneiras. Isso tudo sem perder a essência do trabalho em família. Atualmente, a Vilma é uma das maiores indústrias alimentícias do País.

A Vilma Alimentos atua nos seguintes setores: moinhos de trigo, fabricação de massas, fabricação de misturas de bolos, refrescos, fabricação de biscoitos e fabricação de molhos, temperos e especiarias. A capacidade de produção é de aproximadamente 198.000 t/ano de farinha de trigo branca, 60.000 t/ano de farelo de trigo e de, aproximadamente, 90.000 t/ano de massas.

Preparo das amostras: moagem para os produtos granulados.

Equipamento NIR disponível: NIR MPA – Bruker

O software utilizado para realizar o processamento dos dados é o OPUS com o sistema OVP (*OPUS Validation Program*).

## Faixa de medida espectral e resolução

Faixa de medida espectral: 12.000  $\text{cm}^{-1}$  a 3.600  $\text{cm}^{-1}$

Resolução: 2  $\text{cm}^{-1}$

Produtos analisados através do NIR tradicional: trigo em grãos, farinha de trigo, farelo, semolina, especiarias em geral, creme de milho, fubá, ovo em pó, clara em pó.

Análises realizadas no NIR tradicional: umidade, cinzas, proteínas

## **Metodologia**

Resultados de referência: foram aproveitadas algumas curvas do próprio equipamento. E na Vilma Alimentos foram realizadas mais de 1.000 análises nos métodos de referência.

### **Metodologia de referência utilizada**

- 1) Umidade: American Association of Cereal Chemists (AACC International, 1995b).
- 2) Cinzas: American Association of Cereal Chemists (AACC International, 1995a).
- 3) Proteínas: American Association of Cereal Chemists (AACC International, 1995c).

## **Resultados**

### **Utilização das análises via NIRS**

As análises via NIRS são utilizadas na rotina da empresa. Quando ocorre algum resultado fora da especificação, realizamos a análise no método de referência.

Os principais parâmetros analisados via NIRS são: umidade, cinzas, proteínas (trigo em grãos, farinhas de trigo e semolina), lipídios e umidade (macarrão instantâneo), umidade e proteínas (creme de milho e fubá), umidade (especiarias em geral), umidade e proteínas (massas secas), umidade e proteínas (achocolatados em pó), umidade e cinzas (farelos de trigo) e umidade (biscoitos doces e salgados).

### **Considerações finais**

**Vantagens:** agilidade nas análises, o que é um fator positivo para a indústria para as tomadas de decisão de compra, processo e investimentos.

**Limitações:** é necessária a atualização dos modelos a cada ano, o que demanda um número considerável de amostras para análise nos métodos de referência. Nem todos os usuários possuem equipamentos para realizar todos os métodos de referência, e para realizá-los externamente o custo é muito alto.



Outra questão são os parâmetros que não apresentaram resultados satisfatórios utilizando a tecnologia NIR como Falling number, cor e reologia.

## Perspectivas futuras

A empresa tem muitos segmentos diversificados, a perspectiva é desenvolver mais curvas que atendam às nossas necessidades de análises.

## Referências

AACC International. **Method 08-12**: ash in farina and semolina. 9th ed. St. Paul, 1995a. v. 1. Approved methods of analysis.

AACC International. **Method 44-15A**: moisture. 9th ed. St. Paul, 1995b. v. 2. Approved methods of analysis.

AACC International. **Method 46-12**: crude protein. 9th ed. St. Paul, 1995c. v. 1. Approved methods of analysis.

ASTM. **E1655-05**: standards practices for infrared multivariate quantitative analysis. West Conshohocken, 2012. Annual Book of ASTM Standards.

## Anexo

### Lista de Siglas

#### A

---

ABNT – Associação Brasileira de Normas Técnicas

AL – Limite de Ação (*Action Limit*)

ASTM – *American Society for Testing and Materials*

#### B

---

BIAS – Erro sistemático, viés

#### C

---

CAS – *Chemical Abstracts Service*

CEP – Controle Estatístico de Processo

CL – Linha Central (*Central Line*)

CLS – *Classical Least Squares*

CPs – Componentes Principais

CRM – Material de Referência Certificado (*Certified Reference Material*)

CV – Coeficiente de Variação

#### D

---

D – *Detrend*

DD – *Data-driven*

DM – *Density Modeling*

DON – Deoxinivalenol

DPR – Desvio Padrão Relativo

## E

---

EE – Extrato Etéreo

ELISA – *Enzyme-Linked Immunosorbent Assay*

EPTIS – Sistema de Informação Europeu de Testes de Proficiência (*European Proficiency Testing Information System*)

FB – Fibra Bruta

## H

---

HCA – *Hierarchical Cluster Analysis*

HPLC – *High Performance Liquid Chromatography*

HSI – *Hiperespectral Image*

HSI-NIR – *Hiperespectral Image – Near Infrared*

## I

---

ICH – Conferência Internacional de Harmonização de Requisitos Técnicos (*International Conference on Harmonization of Technical Requirements*).

INMETRO – Instituto Nacional de Metrologia

IPD – Índice Preliminar de DON

IQC – Controle de Qualidade Intralaboratorial (*Internal Quality Control*)

IR – Imagem de Referência

ISO – *International Organization for Standardization*

IUPAC – *International Union of Pure and Applied Chemistry*

## L

---

LAL – Limite Inferior de Ação (*Lower Action Limit*)

LD – Limite de Detecção

LDA – *Linear Discriminant Analysis*

LOD – *Limit of Detection*

LOO – *Leave-one-out*

LOQ – *Limit of Quantification*

LQ – Limite de quantificação

LWL – Limite Inferior de Alerta (*Lower Warning Limit*)

## **M**

---

Mapa – Ministério da Agricultura, Pecuária e Abastecimento

MCR-ALS – *Multivariate Curve Resolution – Alternating Least Square*

MIR – *Medium Infrared*

MLR – *Multiple Linear Regression*

MM – Matéria Mineral

MPLS – *Modified Partial Least Square*

MR – Método Referência

MSC – *Multiplicative Scatter Correction*

MSPC – Controle de Processo Estatístico Multivariado (*Multivariate Statistical Process Control*)

## **N**

---

NAS – Sinal Analítico Líquido (*Net Analytical Signal*)

NIR – Infravermelho Próximo (*Near Infrared*)

NIR – *Near Infrared OC – One-class*

NIR-CI – *Near-Infrared Chemical Imaging*

NIRS – Espectroscopia no Infravermelho Próximo (*Near Infrared Spectroscopy*)

**P**

---

PAT – Técnicas Analíticas de Processo

PB – Proteína Bruta

PCA – Análise de Componentes Principais (*Principal Components Analysis*)

PCR – *Principal Component Regression*

PLS – Regressão Parcial por Mínimos Quadrados (*Partial Least Square*)

PLS-DA – *Partial Least Square – Discriminant Analysis*

PLSR – *Partial Least Square Regression*

PRESS – Soma Quadrática dos Resíduos de Previsão (*Predicted Residual Error Sum of Square*)

PT – Teste de Proficiência (*Proficiency Test*)

**Q**

---

QCM – Material para Controle de Qualidade (*Quality Control Material*)

QDA – *Quadratic Discriminant Analysis*

**R**

---

R<sup>2</sup> – Coeficiente de Determinação

RER – *Range Error Ratio*

RGB – *Red-Green-Blue*

RLM – *Regression Linear Multiple*

RM – Material de Referência (*Reference Material*)

RMSEC – Raiz Quadrada do Erro Médio Quadrático de Calibração (*Root Mean Square Error of Calibration*)

RMSECV – Raiz Quadrada do Erro Médio Quadrático de Validação Cruzada (*Root Mean Square Error of Cross Validation*)

RMSEP – Raiz Quadrada do Erro Médio Quadrático de Previsão (*Root Mean Square Error of Prediction*)

ROI – *Region of Interest* RPD – *Residual Prediction Deviation*

RSDPB – *Relative Spectral Discriminatory Probability*

RSEP – *Relative Standard Error of Prediction*

RSQ – *R squared*

## S

---

SAM – *Spectral Angle Mapper*

SEC – *Erro Padrão de Calibração (Standard Error of Calibration).*

SECV – *Standard Error of Cross Validation*

SEP – *Erro padrão de previsão (Standard Error of Prediction)*

SI – *Sistema Internacional de Unidades (Système International d'Unités)*

SID – *Spectral Information Divergence*

SIMCA – *Soft Independent Modeling of Class Analogy*

SNV – *Standard Normal Variation*

SPA – *Successive Projections Algorithm*

SVM – *Support Vector Machines*

## U

---

UAL – *Limite Superior de Ação (Upper Action Limit)*

UM – *Umidade*

UWL – *Limite Superior de Alerta (Upper Warning Limit)*

## V

---

VANT – *Veículo Autônomo Não Tripulado*

VIP – *Variable Importance in Projection*

VLs – *Variáveis Latentes*

---

## **W**

---

WL – Limite de Alerta (*Warning Limit*)

## **Z**

---

ZEA – Zearalenona





*Impressão e acabamento*

**Nome da gráfica**

**Embrapa**

---

**Trigo**

MINISTÉRIO DA  
**AGRICULTURA, PECUÁRIA  
E ABASTECIMENTO**



978-85-7035-839-4

**CGPE 14674**