



UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA DE ALIMENTOS

MARIA LUCIMAR DA SILVA MEDEIROS

**APLICAÇÃO DE ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO E
IMAGENS HIPERESPECTRAIS PARA QUANTIFICAR O TEOR DE ÓLEO E
CLASSIFICAR SEMENTES DO GÊNERO *BRASSICA***

**APPLICATION OF NEAR INFRARED SPECTROSCOPY AND HYPERSPECTRAL
IMAGING TO QUANTIFY THE OIL CONTENT AND CLASSIFY *BRASSICA* SEEDS**

CAMPINAS – SP

2021

MARIA LUCIMAR DA SILVA MEDEIROS

**APLICAÇÃO DE ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO E
IMAGENS HIPERESPECTRAIS PARA QUANTIFICAR O TEOR DE ÓLEO E
CLASSIFICAR SEMENTES DO GÊNERO *BRASSICA***

**APPLICATION OF NEAR INFRARED SPECTROSCOPY AND HYPERSPECTRAL
IMAGING TO QUANTIFY THE OIL CONTENT AND CLASSIFY *BRASSICA* SEEDS**

Dissertação apresentada à Faculdade de Engenharia de Alimentos da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestra em Engenharia de Alimentos.

Dissertation presented to the School of Food Engineering of the University of Campinas in partial fulfillment of the requirements for the degree of Master, in Food Engineering.

Supervisor/Orientador: Prof. Dr. Douglas Fernandes Barbin

Este trabalho corresponde à versão final da dissertação defendida pela aluna Maria Lucimar da Silva Medeiros, e orientada pelo Prof. Dr. Douglas Fernandes Barbin.

CAMPINAS - SP

2021

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Faculdade de Engenharia de Alimentos
Claudia Aparecida Romano - CRB 8/5816

M467a Medeiros, Maria Lucimar da Silva, 1995-
Aplicação de espectroscopia no infravermelho próximo e imagens
hiperespectrais para quantificar o teor de óleo e classificar sementes do gênero
Brassica / Maria Lucimar da Silva Medeiros. – Campinas, SP : [s.n.], 2021.

Orientador: Douglas Fernandes Barbin.
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade
de Engenharia de Alimentos.

1. Canola. 2. Oleaginosas. 3. Quimiometria. 4. Tecnologias emergentes. I.
Barbin, Douglas Fernandes. II. Universidade Estadual de Campinas. Faculdade
de Engenharia de Alimentos. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Application of Near Infrared spectroscopy and hyperspectral
imaging to quantify the oil content and classify *Brassica* seeds

Palavras-chave em inglês:

Canola

Oilseeds

Chemometrics

Emerging technologies

Área de concentração: Engenharia de Alimentos

Titulação: Mestra em Engenharia de Alimentos

Banca examinadora:

Douglas Fernandes Barbin [Orientador]

Renato Grimaldi

Raul Benito Siche Jara

Data de defesa: 20-01-2021

Programa de Pós-Graduação: Engenharia de Alimentos

Identificação e informações acadêmicas do(a) aluno(a)

- ORCID do autor: <https://orcid.org/0000-0002-7293-756X>

- Currículo Lattes do autor: <http://lattes.cnpq.br/0516919681097451>

COMISSÃO EXAMINADORA

Prof. Dr. Douglas Fernandes Barbin

Orientador - Faculdade de Engenharia de Alimentos, UNICAMP

Prof. Dr. Raúl Benito Siche Jara

Membro Titular - Universidad Nacional de Trujillo

Dr. Renato Grimaldi

Membro Titular - Faculdade de Engenharia de Alimentos, UNICAMP

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

DEDICATÓRIA

*Aos pilares da minha vida: mainha, painho e
Ludemario.*

AGRADECIMENTOS

À Deus, por me dar forças e sabedoria para superar as dificuldades encontradas no caminho.

À minha família, em especial aos meu pais, Irimar Pereira e Lourival Lauro, que nunca mediram esforços para que eu atingisse meus objetivos; e ao meu irmão, Ludemario Medeiros, por sempre acreditar em mim e me incentivar a viver meus os sonhos. Como costume dizer: Essa vitória é nossa!

Ao meu orientador, Dr. Douglas Fernandes Barbin, pelo acolhimento no grupo de pesquisa, orientação no desenvolvimento do trabalho, oportunidades de crescimento e conhecimentos transmitidos.

À professora Dr. Ana Paula Badan, por ter doado as amostras e cedido o Laboratório de Óleos e Gorduras para realização da análises.

Ao meu amigo Jonas e à minha prima Amanda, que mesmo de longe continuam me acompanhando e incentivando. Aos amigos que fiz durante o mestrado, Paulinho, Raquel, Marcos, Bia e Renann, pelo companheirismo, palavras de incentivo e bons momentos compartilhados.

Aos colegas do grupo de pesquisa em Tecnologias Analíticas Emergentes do Laboratório de Inovação em Alimentos – LINA, em especial, aos que estiveram diretamente envolvidos na realização desta pesquisa, Yasmin, Marciano e Pier.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código Financeiro 001; à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (Projeto número 2019/06846-5 e 2015/24351-2) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo financiamento da pesquisa.

E a todos que torceram por mim e/ou contribuíram para a concretização de mais uma etapa importante em minha vida, os meus sinceros agradecimentos!

RESUMO

As sementes do gênero *Brassica* constituem uma das fontes mais importantes de óleos vegetais do mundo e possui inúmeras aplicações nas indústrias alimentícia, farmacêutica e química. O teor de óleo, principal parâmetro quantitativo para avaliar a sua adequação para o mercado, é geralmente mensurado por técnicas analíticas tradicionais, que demandam tempo, são onerosas e destroem a amostra. Técnicas vibracionais, como a espectroscopia no infravermelho próximo (NIR) e a imagem hiperespectral (NIR-HSI), possibilitam a quantificação simples, rápida e precisa de componentes químicos em alimentos, sem destruir a amostra e sem geração de resíduos químicos e podem ser aplicadas como uma alternativa aos métodos convencionais. Diante disso, o presente estudo teve como objetivo investigar o potencial de espectros adquiridos em um equipamento portátil (NIRS) (900 – 1700 nm) e em uma câmera hiperespectral (NIR-HSI) (928 – 2524 nm), em conjunto com ferramentas de quimiometria, na autenticação de espécies e predição do teor de óleo em sementes *Brassicas* intactas. A Análise de Componentes Principais (PCA) foi usada como uma análise exploratória dos dados. A análise Discriminante por mínimos quadrados parciais (PLS-DA) foi aplicada para discriminar entre as espécies *Brassica napus*, *Brassica rapa* e *Brassica juncea* e demonstrou ser eficiente, com acurácias entre 75,0 – 93,6% e 91,0 – 100% para os dados NIRS e NIR-HSI. A Regressão por mínimos quadrados parciais (PLS) foi utilizada para prever o teor de óleo nas sementes intactas e demonstrou boa capacidade de predição, especialmente para os espectros NIR pré-processados com primeira derivada e para os espectros NIR-HSI brutos, onde foram obtidos coeficientes de determinação de 0,682 e 0,764, erros quadráticos médios na predição menores que 1,0 e RPD de 2,0 e 2,2, respectivamente. A seleção de comprimentos de onda relevantes a partir do algoritmo do intervalo PLS (iPLS) aumentou a capacidade discriminativa do modelo NIRS baseado nos espectros suavizados com Savitzky-Golay, onde foi atingida acurácia de 94,9%, e proporcionou uma melhoria nos resultados da calibração PLS para os espectros NIR-HSI, especialmente quando pré-processados com segunda derivada, atingindo RPD de 2,31. Os resultados obtidos neste estudo mostraram que ambos os dispositivos de espectroscopia NIR, combinados com ferramentas de quimiometria, podem ser aplicados para discriminar sementes das espécies *Brassicas* analisadas e para fornecer uma estimativa rápida do teor de óleo nas sementes.

Palavras-chave: Canola; oleaginosas; espectroscopia; quimiometria; tecnologias emergentes; métodos não invasivos

ABSTRACT

Brassica seeds are one of the most important sources of vegetable oils in the world and have numerous applications in the food, pharmaceutical and chemical industries. The oil content, the main quantitative parameter to assess its suitability for the market, is usually measured by traditional analytical techniques, which are time consuming, costly and destroy the sample. Vibrational techniques, such as near-infrared spectroscopy (NIR) and hyperspectral imaging (NIR-HSI), allow simple, fast, and accurate quantification of chemical components in food, without destroying the sample and without generating chemical residues and can be applied as an alternative to conventional methods. Therefore, the present study aimed to investigate the potential of spectra acquired in a portable equipment (NIRS) (900 - 1700 nm) and in a hyperspectral camera (NIR-HSI) (928 - 2524 nm), together with chemometrics, species authentication and prediction of oil content in intact *Brassic*as seeds. Principal Component Analysis (PCA) was used as an exploratory analysis of the data. Discriminant Analysis by Partial Least Squares (PLS-DA) was applied to discriminate between *Brassica napus*, *Brassica rapa* and *Brassica juncea* species and proved to be efficient, with accuracy between 75.0 - 93.6% and 91.0 - 100% for NIRS and NIR-HSI data. Partial Least Squares regression (PLS) was used to predict the oil content in intact seeds and demonstrated good predictive ability, especially for the pre-processed NIR spectra with first derivative and for the raw NIR-HSI spectra, where they were obtained determination coefficients of 0.682 and 0.764, mean square errors in the prediction less than 1.0 and RPD of 2.0 and 2.2, respectively. The selection of relevant wavelengths from the PLS interval algorithm (iPLS) increased the discriminative capacity of the NIRS model based on the spectra smoothed with Savitzky-Golay, where an accuracy of 94.9% was achieved, and provided an improvement in the results of the PLS calibration for the NIR-HSI spectra, especially when pre-processed with second derivative, reaching RPD of 2.31. The results obtained in this study showed that both NIR spectroscopy devices, combined with chemometric tools, can be applied to discriminate seeds of the analyzed *Brassic*as species and to provide a quick estimate of the oil content in the seeds.

Keywords: Canola; oilseeds; spectroscopy; chemometric; emerging technologies; non-invasive methods

LISTA DE FIGURAS

Capítulo 2 – Revisão de Literatura

Figura 2.1. Representação de um hipercubo de imagem hiperespectral (x, y, λ), que pode ser entendido como uma imagem espacial separada (x, y) em cada comprimento de onda individual (λ). Fonte: O autor (2020).25

Capítulo 3 – NIR spectroscopy and hyperspectral imaging for investigation of *Brassicas* seeds quality and species

Figure 3.1. Mean spectra of Brassicas seeds after Savitzky-Golay smoothing (a and b), Savitzky-Golay smoothing and MSC (c and d) e 2nd derivative (e and f) to NIR and NIR-HSI data, respectively.43

Figure 3.2. Scores plots of PCA performed on spectra with Savitzky-Golay smoothing for NIR (a and b) and HSI data (c) and their correspondent Loadings plots (d and e).45

Figure 3.3. iPLS-DA model: (a and b) Important wavelengths selected and (c and d) iPLS-DA classification plots to discriminate varieties of Brassicas seeds: *B. juncea* (red diamonds), *B. rapa* (green squares) and *B. napus* (blue triangles) for NIRS and NIR-HIS data. The test set samples of each class are represented by the same symbols on the right side.49

Figure 3.4. Reduced PLSR model: (a and b) Important wavelengths selected by iPLS algorithm and (c and d) Predicted versus measured oil content values in Brassicas seeds.52

LISTA DE TABELAS

Capítulo 2 – Revisão de Literatura

Tabela 2.1. Principais ácidos graxos presentes em óleos vegetais.	19
Tabela 2.2. Aplicações recentes (2015 – 2020) da espectroscopia NIR na quantificação de parâmetros composicionais de sementes Brassicas.	23

Capítulo 3 – NIR spectroscopy and hyperspectral imaging for investigation of *Brassicas* seeds quality and species

Table 3.1. Oil content (g/100g) for different species of <i>Brassicas</i> seeds.	42
Table 3.2. Performance of PLS-DA model for classification the species of Brassicas seeds using NIR spectra.	46
Table 3.3. Performance of PLSR model for the prediction of oil content in Brassicas seeds using NIR and NIR-HSI spectra.	50

SUMÁRIO

Capítulo 1 – Introdução Geral	13
1.1. INTRODUÇÃO GERAL	14
Capítulo 2 – Revisão de Literatura.....	16
2.1. REVISÃO DE LITERATURA	17
2.1.1. BRASSICAS	17
2.1.2. ÓLEO DE BRASSICAS	18
2.1.3. QUANTIFICAÇÃO DO TEOR DE ÓLEO	21
2.1.4. MÉTODOS NÃO DESTRUTIVOS.....	22
2.1.4.1. Espectroscopia no Infravermelho Próximo (NIR).....	22
2.1.4.2. Imagem Hiperespectral no Infravermelho Próximo (HSI - NIR).....	24
2.2.5. QUIMIOMETRIA.....	25
2.2.5.1. Pré-processamento dos dados.....	26
2.2.5.2. Análise exploratória	26
2.2.5.3. Análises de multivariada de predição/regressão.....	27
2.2.5.4. Análise multivariada de classificação/discriminação.....	29
Capítulo 3 - NIR spectroscopy and hyperspectral imaging for investigation of <i>Brassicas</i> seeds quality and species	34
3.1. INTRODUCTION	36
3.2. MATERIAL AND METHODS	38
3.2.1. SAMPLES COLLECTION.....	38
3.2.2. SPECTRAL DATA COLLECTION.....	38
3.2.2.1. Near infrared spectroscopy (NIR)	38
3.2.2.2. Near infrared hyperspectral imaging (NIR-HSI).....	38
3.2.3. MEASUREMENT OF OIL CONTENT	39
3.2.4. STATISTICAL DATA ANALYSIS.....	39
3.2.5. SPECTRAL PRE-PROCESSING	39
3.2.6. SELECTION OF CLASSIFICATION AND CALIBRATION SETS.....	40
3.2.7. MULTIVARIATE ANALYSES.....	40
3.2.7.1. Principal component analysis (PCA)	40
3.2.7.2. Classification analysis	40
3.2.7.3. PLSR for prediction of oil content	41
3.3. RESULTS AND DISCUSSION	42
3.3.1. OIL CONTENT OF SAMPLES	42
3.3.2. SPECTRA PROFILE	42

3.3.4. CLASSIFICATION MODELS	46
3.3.5. PREDICTION MODELS.....	49
3.4. CONCLUSION	52
ACKNOWLEDGEMENTS	53
CREDIT AUTHORSHIP CONTRIBUTION STATEMENT	53
DECLARATION OF COMPETING INTEREST.....	53
REFERENCES.....	54
SUPPLEMENTARY MATERIAL	60
Capítulo 4 – Conclusão Geral.....	72
Capítulo 5 - Referências	74

Capítulo 1 – Introdução Geral

1.1. INTRODUÇÃO GERAL

As sementes oleaginosas do gênero *Brassica*, especialmente *Brassica napus*, *Brassica rapa* e *Brassica juncea*, fornecem cerca de 12% do óleo vegetal total do mundo e constituem a principal safra de oleaginosas da União Europeia, Canadá, Austrália, China e Índia (Iftikhar et al., 2018). As aplicações industriais dos óleos dessas espécies são determinadas pela sua composição de ácidos graxos, especialmente pela presença de ácido erúico. Óleos com alta concentração de ácido erúico são destinados para a indústria oleoquímica, produção de fármacos, lubrificantes, plastificantes, cosméticos, surfactantes e outros (Bassegio et al., 2016), enquanto os óleos com baixa proporção ou isentos são utilizados para aplicações alimentares.

A avaliação da qualidade das sementes *Brassicas* e a estimativa do seu rendimento em termos de teor óleo, que é parâmetro químico mais importante para a indústria processadora, é realizada por métodos analíticos tradicionais, os quais são onerosos e demandam tempo, inviabilizando a análise rápida no controle de qualidade das sementes e em grandes lotes de amostras (Olivos-Trujillo et al., 2015). Neste sentido, têm-se aumentado a busca por métodos analíticos rápidos e precisos para aplicações online na determinação da qualidade dos produtos e de parâmetros de processamento (Kaufmann et al., 2019).

A espectroscopia no infravermelho próximo (NIR) tem sido amplamente investigada como um método não destrutivo e demonstrou resultados promissores para a avaliação composicional (Badaró et al., 2020; Barbin et al., 2018; Caporaso et al., 2018a, 2018b; Kaur et al., 2016) e da autenticidade de alimentos (Cruz-Tirado et al., 2020; Oliveira et al., 2020; Orrillo et al., 2019). Nessa técnica, a radiação na região do infravermelho (780 a 2500 nm) que incide na amostra interage com as moléculas que contém H ligado a C, N, O e S resultando em um espectro característico do alimento, a partir do qual podem ser extraídas informações relacionadas a sua composição (Font et al., 2006). A espectroscopia NIR garante a integridade da amostra, permite a obtenção de informações químicas de forma rápida e sem o uso de solvente orgânico, apresenta baixo custo e pode ser aplicado em todas as etapas do processamento, desde a colheita a inspeção do produto.

Diante disso, o objetivo deste estudo foi investigar as potencialidades da espectroscopia NIR para autenticar espécies e quantificar o teor de óleo em sementes *Brassicas* intactas. Foram construídos modelos de classificação e modelos de calibração, com o auxílio de ferramentas de quimiometria, usando toda a faixa espectral e comprimentos de onda

importantes, para comparar o desempenho de espectros adquiridos em um equipamento NIR portátil e em uma câmera hiperspectral.

Capítulo 2 – Revisão de Literatura

2.1. REVISÃO DE LITERATURA

2.1.1. BRASSICAS

Brassica é um gênero de plantas pertencente à família Brassicaceae e compreende um grande número de plantas com importância econômica, incluindo vegetais (repolho, couve-flor, nabo, couve de Bruxelas e rabanetes), forragens, plantas ornamentais, condimentos e commodities agrícolas (Rani et al., 2013). Do ponto de vista econômico e agrônômico, as sementes oleaginosas são as commodities mais importantes desse gênero, sobretudo pelo elevado teor de óleo e farelo rico em proteína que sobra após a extração do óleo, e porque são capazes de germinar e crescer a baixas temperaturas, com formas de hábitos na primavera e inverno.

O desenvolvimento de sementes com qualidade alimentar na década de 1960 (baixo teor de ácido erúico e baixo teor de glucosinolatos) (Downey et al., 1967) proporcionou um aumento considerável na produção de sementes *Brassicas*, de forma que se tornaram a terceira cultura oleaginosa mais importante do mundo. Nos últimos 15 anos a quantidade de sementes colhidas dobrou, e atingiu uma produção superior a 70 milhões de toneladas em uma área de 34 milhões de hectares em 2019 (FAO, 2019; Li et al., 2013). Além disso, a existência de diversas cultivares adaptadas a zonas semitropicais e a zonas temperadas, e com maior tolerância à seca, umidade, calor, geadas e pragas (Oram et al., 2005; Singh et al., 2014), permitiu o cultivo em diversos países. O Canadá é o principal produtor, responsável por 26,4% da produção mundial, seguido pela China (19,1%), Índia (13,1%), França (5,0%) e Ucrânia (4,7%).

As principais espécies utilizadas para a produção de óleo são *Brassica napus*, *Brassica rapa* e *Brassica juncea*. Essas sementes são prensadas para produzir o óleo, cerca de 45,7% da semente, e geram como subproduto o farelo de canola/colza. O óleo é constituído por 3,5 - 5,6% de ácidos graxos saturados, 62,8 - 70,5% de ácidos graxos monoinsaturados e 21,0 - 29,7% de poliinsaturados (Cartea et al., 2019), e possui diferentes aplicações, incluindo desde o uso como alimento até a produção de biocombustíveis, as quais são definidas pela proporção dos seus principais ácidos graxos.

O farelo resultante da extração do óleo contém elevado teor de proteína, 38,1 – 44,0%, cerca de 9,79% de açúcares solúveis, 7,03 - 12,2% de fibra bruta, além de componentes minoritários, como glucosinolatos (5,49 – 139,09 $\mu\text{mol/g}$), fitatos e compostos fenólicos (1,35 – 2,26%) (Chen et al., 2015; Mawlong et al., 2017; Sen et al., 2018; Woyengo et al., 2016). O

farelo de canola/colza é considerado a segunda fonte de proteína alimentar, depois do farelo de soja, e tem grande potencial de utilização como fonte de proteína de qualidade alimentar. Entretanto, a presença de componentes antinutricionais, impede a sua utilização para consumo humano. Os produtos da hidrólise dos glucosinolatos são tóxicos e podem causar problemas na tireoide, além de contribuir para o odor pungente e sabor ardente da farinha (Diederichsen and McVetty, 2011; Zhou et al., 2018). Os fitatos atuam como agentes quelantes e se ligam ao Zn e ao Fe, tornando esses micronutrientes essenciais indisponíveis para o metabolismo. E os produtos da degradação dos compostos fenólicos, que incluem ácidos fenólicos e taninos, conferem a farinha sabor amargo e adstringente. Atualmente, o seu uso é destinado para a produção de alimentação animal e, quanto apresenta alto teor de glucosinolatos, de fertilizantes (Thiyam-Holländer et al., 2012).

2.1.2. ÓLEO DE BRASSICAS

As sementes *Brassicas* constituem uma das fontes mais importantes de óleo vegetal em todo o mundo e são responsáveis por cerca de 12% da produção mundial (24,6 milhões de toneladas de óleo), sendo superada apenas pelo dendê (71,5 milhões de toneladas) (*Elaeis guineensis* Jacq) e pela soja (*Glycine max* L.) (56,9 milhões de toneladas) (FAO, 2018). Essas sementes podem apresentar teor de óleo entre 17,6 e 54,9% (Oblath et al., 2016; Velasco et al., 1999) e, assim como a sua composição, difere entre as variedades, híbridos e linhas avançadas de *Brassicas* e é influenciado por condições ambientais, interações genótipo x ambiente e práticas culturais (Long et al., 2012).

O óleo é acumulado nas sementes, principalmente, durante a fase de maturação e fornece energia para a subsequente germinação das sementes e crescimento das mudas. O óleo de *Brassicas* é constituído majoritariamente por triacilgliceróis (TAGs), cerca de 95%, que consistem em três ácidos graxos ligados a uma estrutura de glicerol. Suas propriedades nutricionais, tecnológicas e a estabilidade, são definidas pelo comprimento da cadeia de carbono, nível de saturação e posição dos ácidos graxos na estrutura do glicerol (Akoh, 2017; Kinney, 2017; Thiyam-Holländer et al., 2012). Outros componentes, como fosfolipídios, ácidos graxos livres, tocoferóis, compostos fenólicos e pigmentos, estão presentes no óleo em menores proporções (Barthet and Daun, 2011).

Os óleos das variedades tradicionais, comumente denominados óleo de colza, são constituídos majoritariamente por ácidos graxos monoinsaturados de cadeia longa, onde 42,3 a 54,1% do total corresponde ao ácido erúico (HEAR). Este ácido, possui em sua

composição 22 átomos de carbonos e apenas uma ligação dupla, e tem ocorrência natural nas sementes da família Brassicaceae (No et al., 2013), principal característica que o diferencia dos óleos de soja, milho, palma e girassol (Tabela 2.1).

Tabela 2.1. Principais ácidos graxos presentes em óleos vegetais.

Óleo vegetal	Composição de ácidos graxos (%)						
	C12:0	C16:0	C18:0	C18:1	C18:2	C18:3	C22:1
Canola							
HEAR	-	2,3 - 3,5	0,4 - 0,8	7,9 - 18,7	7,9 - 15,7	6,5 - 10,9	42,3 - 54,1
LEAR	-	4,3 - 4,8	1,1 - 1,3	59,9 - 63,7	20,2 - 20,7	6,9 - 10,7	0,004 - 0,1
HOLLCO	-	3,7 - 3,9	1,7 - 1,8	70,4 - 74,4	14,4 - 17,9	1,8 - 3,0	0,1
HLCO	31,3 - 46,7	3,3 - 3,6	1,1 - 1,3	35,1 - 30,9	3,3 - 14,6	0,7 - 8,8	-
HSCO	-	3,6 - 5,9	27,5 - 44,0	28,8 - 37,2	4,0 - 18,4	3,4 - 17,4	-
Outras oleaginosas							
Oliva	-	11,0 - 15,8	2,2 - 4,3	62,4 - 75,7	4,8 - 16,2	0,3 - 0,7	-
Girassol	-	5,5 - 7,1	3,6 - 4,2	18,9 - 26,2	64,1 - 67,5	0,8	0,02
Milho	-	10,8 - 12,9	1,9 - 2,2	22,3 - 25,4	57,6 - 59,5	1,2 - 5,3	-
Palma	-	39,0 - 39,6	4,5 - 5,4	42,9 - 44,5	8,9 - 11,0	0,2	-
Soja	-	11,0 - 12,2	4,7 - 6,5	21,8 - 29,1	44,1 - 53,9	2,7 - 7,8	-

HEAR: Colza com alto Teor de Ácido Erúico; LEAR: Canola com baixo Teor de Ácido Erúico; HOLLCO: Canola com alto teor de ácido oléico e baixo teor de ácido linolênico; HLCO: Canola com alto teor de ácido láurico; HSCO: Canola com alto teor de ácido esteárico. Fontes: Alves et al. (2019); Byrdwell and Neff (2001); Cartea et al. (2019); Jabeur et al. (2014); Neff et al. (1997); Sarkaya et al. (2016); Xu et al. (2020).

Os óleos com alto teor de ácido erúico são úteis para aplicações na indústria oleoquímica, como matérias-primas de filmes plásticos, como componente em óleos lubrificantes, surfactantes, revestimentos de superfície e nylon (Reginato et al., 2013), bem como nas indústrias farmacêutica e cosmética. Na indústria de alimentos, o seu uso é limitado, pois o consumo de ácido erúico causa lesões cardíacas e interrompe outros processos fisiológicos em animais (No et al., 2013). Entretanto, a hidrogenação completa desse óleo resulta na conversão do ácido erúico (C22:1) em ácido behênico (C22:0), que por sua vez, pode ser consumido sem restrições nutricionais (Ribeiro et al., 2017). Estudos demonstraram que o ácido behênico apresenta baixa capacidade de absorção intestinal (Kojima et al., 2010), que triacilgliceróis estruturados contendo ácido behênico e ácidos graxos de cadeia média contém aproximadamente metade das calorias do triacilglicerol natural (5kcal/g) (Webb and Sanders, 1991) e que a sua ingestão suprime a absorção de gordura e não induz a obesidade (Moreira et al., 2017; Silva et al., 2019). Além disso, Kojima et al. (2010) e Kojima et al. (2013) verificaram que a ingestão desses lipídios não influencia na absorção de minerais e

vitaminas lipossolúveis, demonstrando que o ácido behênico pode ser usado como um componente funcional de gorduras de baixa caloria.

A presença de elevadas concentrações de ácido erúcico limitou durante anos a utilização do óleo de colza para fins alimentares, devido as suas características como antinutriente. Na década de 1960, diversos programas de melhoramento genético buscaram produzir cultivares de *Brassicas* com baixo teor de ácido erúcico (LEAR), as quais foram disponibilizadas pela primeira vez na cultivar *B. napus* 'Oro' em 1968 e na cultivar *B. rapa* 'Span' em 1971 (Gupta, 2016; Downey, 1990). Posteriormente, essa qualidade também foi alcançada em *B. juncea*. A partir disso, o termo canola (*Canadian oil low acid*) passou a ser usado para descrever o perfil de óleo de qualquer uma das espécies *B. napus*, *B. rapa* e *B. juncea* que contém baixo teor de ácido erúcico (<2,0%) e baixo teor de glucosinolatos (< 30,0 µmol/g de sólido livre de óleo) (Cartea et al., 2019).

Atualmente, o óleo de canola é amplamente utilizado na culinária (Rani et al., 2013) e por conter menores níveis de ácidos graxos saturados, resultou em um aumento nutricionalmente desejável no conteúdo de ácido oleico e nas frações de ácido linoleico (Cartea et al., 2019). A presença de ácido oleico proporciona benefícios a saúde, sabor agradável e maior termoestabilidade ao óleo, tornando-o mais adequado para aplicações onde são utilizadas altas temperaturas em comparação com os ácidos graxos polinsaturados, que aumentam a suscetibilidade do óleo à oxidação, dando origem a odor desagradável e redução da sua vida útil (Bocianowski et al., 2012).

Após essa modificação, os programas de melhoramento de sementes *Brassicas* disponibilizaram diversos genótipos de qualidade alimentar com perfil lipídico adaptado a finalidade de aplicação. Óleos com maior teor de ácido oleico (HOCO) e com teor de ácido linolênico reduzido (<3,0%) (LLCO) foram desenvolvidos para aplicações em frituras como substituto as gorduras que contêm isômeros trans. A combinação de ácido oleico alto e ácido linoléico baixo (LLHOCO) com o teor de ácido graxo saturado convencionalmente baixo, proporcionou um óleo com composição próxima a do azeite de oliva (Tabela 2.1) e melhorou a estabilidade oxidativa, resultando em vida útil e de fritura mais longa, perfil de sabor limpo e suave e melhor qualidade nutricional dos produtos fritos, menor teor de gordura saturada e de gorduras trans (Thiyam-Holländer et al., 2012). Além desses, foi desenvolvido óleo de canola com um nível elevado de ácido esteárico (HSCO) para substituir as gorduras hidrogenadas nos mercados de pão e padaria e óleo com alto teor de ácido láurico (HLCO) para uso em revestimentos de confeitaria, branqueadores de café e coberturas batidas (Vecchio, 1996).

2.1.3. QUANTIFICAÇÃO DO TEOR DE ÓLEO

As sementes *Brassicas* são comercializadas, sobretudo, para a produção de óleo. Por esse motivo, a informação sobre o seu rendimento é o parâmetro mais relevante para a indústria processadora. A seleção de sementes com alto rendimento, tanto para os programas de melhoramento quanto para as indústrias de processamento, é realizada através de uma triagem dos genótipos existentes no germoplasma e avaliação por métodos analíticos convencionais (Kaur et al., 2016).

O teor de óleo (também chamado de extrato etéreo, gordura neutra ou gordura bruta) pode ser definido como a quantidade máxima de lipídio que pode ser extraída das sementes usando solventes específicos e deve fornecer uma estimativa da quantidade de óleo que poderia ser obtida de maneira ideal em um processo de extração industrial (unidade de prensagem com extração por solvente) (Thiyam-Holländer et al., 2012). O método analítico empregado como referência para mensurar o teor de óleo nas sementes é o método Am 2-93 especificado pela American Oil Chemists' Society (AOCS, 2009).

O objetivo desse método é produzir uma extração exaustiva de lipídios neutros e baseia-se na extração do óleo com solventes orgânicos. Antes de iniciar o processo de extração, a umidade das sementes deve ser ajustada para a faixa de 7 - 9%. A umidade alta implica na redução da eficiência da extração, recuperação de óleo e do solvente do resíduo, além de reduzir a qualidade do óleo e a eficiência de degomagem, pois alguns solventes apolares, como éter dietílico e hexano, não penetram facilmente em tecidos úmidos. Em seguida, as sementes são trituradas para reduzir o tamanho de partícula. Apesar do pequeno tamanho da semente de canola, o tegumento e a estrutura celular são robustos e devem ser rompidos para permitir que as partículas de óleo migrem da estrutura celular durante o processo de prensagem e que tenha maior contato com o solvente e, consequentemente, maior eficiência extração (Akoh, 2017; Unger, 2011).

A separação do óleo dos demais constituintes da semente, como proteínas e carboidratos, é possível devido a insolubilidade dos lipídios na água e afinidade por solventes orgânicos. Os componentes majoritários do óleo, como TAGs e ésteres de colesterol, são altamente solúveis em solventes de hidrocarbonetos, como hexano, e solventes mais hidrofóbicos, como éter de petróleo, e por isso esses solventes são os mais empregados para extrair e determinar o conteúdo lipídico em sementes. Entretanto, frações de compostos solúveis em óleo, como fosfolipídios, ceras, esteróis, glicolipídios, ácidos graxos livres, vitaminas e pigmentos, também podem ser extraídos e determinados como lipídios, enquanto

os lipídios que estão covalentemente ligados a grupos polipeptídicos e polissacarídeos permanecerão no farelo de canola (Akoh, 2017; Barthet and Daun, 2011).

Apesar de fornecer uma estimativa confiável da quantidade de óleo presente nas sementes (97 - 99%) (Barthet and Daun, 2011), esse método é inviável para aplicação em análises de recepção da matéria-prima e em muitos programas de melhoramento de plantas, onde as sementes de linhagens avançadas selecionadas são escassas e existe a necessidade analisar muitas amostras em um tempo muito curto, pois é um método demorado e destrói as amostras testadas. Além disso, exige o uso de grandes quantidades de solventes químicos, que podem ser perigosos e prejudiciais aos analistas e ao ambiente (Cheng et al., 2017; Jin et al., 2016).

2.1.4. MÉTODOS NÃO DESTRUTIVOS

O desenvolvimento de métodos inovadores, não invasivos e fáceis de manusear para possibilitar a determinação simples, rápida e precisa de compostos químicos em alimentos tem aumentado nos últimos anos e está atraindo cada vez mais a atenção de indústrias alimentícias. Entre as tecnologias emergentes, a espectroscopia NIR e imagem hiperespectral têm sido amplamente estudadas e implementadas como alternativa aos métodos analíticos convencionais e tem demonstrado resultados promissores na autenticação e garantia da qualidade de uma ampla variedade de produtos (Badaró et al., 2019).

2.1.4.1. Espectroscopia no Infravermelho Próximo (NIR)

A espectroscopia no infravermelho próximo (do inglês, *Near Infrared Spectroscopy*, NIRS) é uma técnica analítica rápida e não destrutiva, frequentemente empregada para quantificar e avaliar informações químicas (Ravazzi, 2019) em alimentos e produtos agrícolas, bem como em indústrias farmacêuticas, petroquímicas e químicas (Blanco and Villarroya, 2002).

Essa técnica é baseada na interação entre a radiação eletromagnética e a amostra, que é influenciada por sobretons, combinações e ressonâncias de vibrações fundamentais das ligações envolvendo carbono, nitrogênio, oxigênio e enxofre ligado a um átomo de hidrogênio, as quais são detectáveis na região do infravermelho de 780 a 2500 nm (Su et al., 2017).

Cada alimento apresenta um espectro característico (Barbin et al., 2015), uma vez que os níveis de absorção de radiação em cada comprimento de onda estão associados à sua composição química (Rady and Adedeji, 2018). Desta forma, as informações espectrais fornecem uma “impressão digital” do alimento e podem ser utilizadas para quantificar compostos específicos, como gordura, açúcares, proteínas e umidade (Badaró et al., 2019) e para diferenciar e classificar os alimentos de acordo com fatores relacionados a condições da cultura/processo, origem geográfica, variedades ou genótipos, desde que estes influenciem na sua composição química (Cozzolino, 2016).

Diversos estudos têm avaliado o potencial da espectroscopia NIR como alternativa aos métodos convencionais para análises composicionais em sementes *Brassicas* intactas, conforme apresentado na Tabela 2.1.

Tabela 2.2. Aplicações recentes (2015 – 2020) da espectroscopia NIR na quantificação de parâmetros composicionais de sementes *Brassicas*.

Parâmetro	Equipamento	Quimiometria	Resultados	Referência
Teor de óleo, proteína, ácidos graxos, glucosinolatos, clorofila e valor de iodo	NIR Portátil (908 - 1676 nm)	PLSR	$R_p^2 = 0,985$, RPD= 5,41*	(Barthet et al., 2020)
α -Tocoferol, ($\gamma+\beta$)-Tocoferol e Tocoferol total	NIR (1100 - 2498 nm)	PCA, MPLS	R_p^2 : 0,428 – 0,635, SEP: 20,02 – 29,451	(Xu et al., 2019)
Teor de óleo	NIR (400 - 2500 nm)	MPLS	<i>B. juncea</i> : $R_p^2 = 0,92$, SEP=0,88; <i>B. napus</i> : $R_p^2 = 0,55$, SEP=1,64	(Kaur et al., 2017)
Teor de óleo, ácidos graxos, fenóis, glucosinolatos	NIR (400 - 2500 nm)	MPLS	<i>B. juncea</i> : $R_p^2 = 0,766$, SEP= 0,849; <i>B. napus</i> : $R_p^2 = 0,865$, SEP=0,379	(Sen et al., 2018)
Teor de óleo	FT-NIR (2000 - 4000cm ⁻¹)	-	$R_p^2 = 0,96$, Erro relativo médio: 2,03%	(Wan et al., 2018)
Ácidos graxos	NIR (400 - 2500 nm)	MPLS	R_p^2 : 0,20 - 0,90, SEP: 0,51 - 2,34	(Kaur et al., 2016)
Teor de óleo, umidade, ácidos graxos, proteína, glucosinolatos e clorofila	NIR (400–2500 nm)	PLS	R_C^2 : 0,69 - 1,00, SEC: 0,09 - 9,68	(Oblath et al., 2016)
Teor de óleo	NIR (680 a 2500 nm)	MLR, SVR, ANN	Erro de predição: 1,3%	(Olivos-Trujillo et al., 2015)

*Figuras de mérito para o teor de óleo; MPLS: Regressão de mínimos quadrados parciais modificada; MLR: Regressão múltipla linear; ANN: Rede Neural Artificial; SVR: Regressão por Máquina de Vetor de Suporte.

Foi demonstrado que a técnica pode fornecer uma estimativa confiável do conteúdo de óleo, umidade, proteína e fibras e de componentes minoritários, como ácidos graxos, glucosinolatos e fenóis. Além disso, os equipamentos NIR portáteis podem ser úteis para análises de controle de qualidade, online ou in situ, e apresentam vantagens como baixo custo, tamanho reduzido, robustez e simplicidade de análise, e foram utilizados sem perda de desempenho para identificar fraude em especiarias (Oliveira et al., 2020) e para quantificar acidez em óleo de palma (Kaufmann et al., 2019) e componentes importantes em sementes de canola (Barthet et al., 2020).

Embora a espectroscopia NIR tenha sido reconhecida como uma ferramenta eficiente e confiável para a detecção simultânea de constituintes do alimento, o seu uso em matrizes heterogêneas é complexo, uma vez que oferece informações espectrais de uma pequena porção da amostra e não fornece informações sobre a distribuição espacial dos componentes químicos (Badaró et al., 2019). Para superar esse desafio e obter informações espaciais da amostra, são necessárias técnicas que combinem a espectroscopia e visão computacional, como a imagem hiperespectral (Feng et al., 2019).

2.1.4.2. Imagem Hiperespectral no Infravermelho Próximo (HSI - NIR)

A imagem hiperespectral (HSI - NIR) é uma técnica inovadora, rápida, não-invasiva e confiável para inspeção da qualidade e autenticidade de alimentos. Essa técnica integra a espectroscopia NIR clássica e um sistema de captura de imagem, possibilitando obter simultaneamente as informações espectrais e espaciais de um objeto (Cheng et al., 2017).

As imagens são constituídas por regiões menores denominadas *pixels*, sobre as quais é obtido um espectro na região do infravermelho próximo. O conjunto de dados gerados (“hipercubo”), possui um arranjo tridimensional (x , y e λ), em que x e y definem as coordenadas espaciais e o λ as informações espectrais (Amigo, 2010) (Figura 1). Consequentemente, é possível caracterizar as propriedades químicas inerentes de uma amostra e observar a sua distribuição no espaço (Tibola et al., 2018).

Essa técnica mostrou bom desempenho para prever o conteúdo de acidez, índice de peróxido e umidade em azeite de oliva (Martínez Gila et al., 2015) e quantificar o teor de óleo, proteínas (Cheng et al., 2017) e umidade em grãos de amendoim (Jin et al., 2015). A técnica demonstrou aptidão para identificar a origem de grãos de arroz de diferentes regiões da China (Sun et al., 2017) e discriminar variedades de óleos de gergelim (Xie et al., 2014) e de sementes de uva (Zhao et al., 2018). Orrillo *et al.*, (2019) demonstraram o potencial da

técnica na detecção e visualização de adulteração de pimenta preta por sementes de mamão, inteiras e em pó.

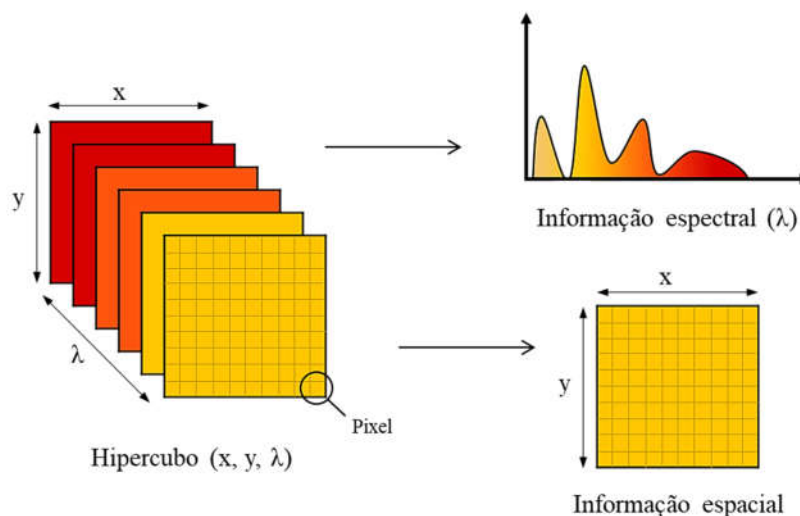


Figura 2.1. Representação de um hipercubo de imagem hiperespectral (x, y, λ), que pode ser entendido como uma imagem espacial separada (x, y) em cada comprimento de onda individual (λ). Fonte: O autor (2020).

Os estudos mencionados anteriormente, mostraram resultados pertinentes de previsão ou classificação usando diferentes métodos quimiométricos. Assim, a tecnologia HSI - NIR pode fornecer mais informações e pode ser usada para estabelecer modelos mais confiáveis para inspeção rápida e não invasiva de composições de alimentos e produtos agrícolas (Cheng et al., 2017). Até agora, não foi relatado o uso da técnica de imagem hiperespectral para a determinação não destrutiva dos parâmetros de qualidade e composicionais de espécies de sementes do gênero *Brassica*.

2.2.5. QUIMIOMETRIA

A expansão da espectroscopia NIR como ferramenta analítica foi possibilitada pelo desenvolvimento de equipamentos de baixo custo, com componentes eletrônicos e ópticos aprimorados e, sobretudo, pelo surgimento de ferramentas capazes de processar e extrair informações contidas nos espectros (Pasquini, 2018; Singh et al., 2014). Os espectros NIR consistem em uma elevada quantidade de variáveis com dados complexos, correlacionados e que geralmente não estão prontamente disponíveis para fins analíticos (Ravazzi, 2019). A extração de informações é possível através do uso da quimiometria, que envolve o uso de

métodos matemáticos e estatísticos para tratar, interpretar e prever dados químicos, através da análise multivariada (Ferreira, 2015).

A análise multivariada permite tratar conjuntos de dados que contenham qualquer número de variáveis e extrair informações relevantes na forma de modelos empíricos, que podem ser utilizados para medições quantitativas e qualitativas (Kemsley et al., 2019).

2.2.5.1. Pré-processamento dos dados

Os espectros brutos nem sempre podem ser interpretados diretamente ou usados para criar modelos robustos, porque podem conter fontes de variabilidade espectral que não estão relacionadas com a composição química do alimento, como ruído aleatório, dispersão de luz, variação da inclinação e deslocamento da linha de base. Esses efeitos podem ser minimizados ou corrigidos empregando métodos de pré-processamento espectral e são atribuídos a mudanças no caminho ótico, na sensibilidade do detector e do amplificador e a diferenças no tamanho das partículas ou propriedades físicas (morfologia e porosidade) do alimento (Ferreira, 2015; Pasquini, 2018).

Os algoritmos de pré-processamento espectral podem melhorar a capacidade preditiva ou discriminatória do modelo ao eliminar fontes de variabilidade espectral que mascaram informações relevantes para o estudo. Entretanto, a sua aplicação deve ser realizada com cautela, analisando a característica do espectro e a finalidade de cada método, pois o excesso ou má aplicação do pré-processamento pode causar perda de informações importantes (Pasquini, 2018).

A suavização de Savitzky-Golay é aplicada para reduzir o ruído instrumental; a correção de espalhamento multiplicativa (Multiplicative Scatter Correction - MSC) e a variável normal padrão (Standard Normal Variate - SNV) são usadas para reduzir os efeitos aditivos e multiplicativos causados pelo espalhamento e as derivadas corrigem o deslocamento e tendência da linha de base (Nolasco, 2019). Esses algoritmos de pré-processamento podem ser usados individualmente e em combinação, desde que sejam projetados para superar fontes de variabilidade espectral diferentes (Ferreira, 2015).

2.2.5.2. Análise exploratória

A análise exploratória é um método não supervisionado que utiliza informações espectrais para mostrar relação entre as amostras e entre as amostras e as variáveis,

permitindo visualizar a formação de grupos naturais ou tendência entre elas, além de identificar amostras anômalas, “*outliers*” (Corgozino, 2009). Entre as técnicas de análise exploratória dos dados, a Análise de Componentes Principais (PCA) é a mais difundida e tem como princípio projetar os dados multivariados em comum em um espaço dimensional menor, visando a redução e interpretação dos dados (Tibola et al., 2018). Para isso, as variáveis originais são combinadas linearmente em componentes principais (PC's), que condensam a maior parte da informação (variância) em poucas variáveis, diminuindo o número de dimensões dos dados sem perder significativamente a informação química (Kemsley et al., 2019).

A obtenção das componentes principais é realizada a partir da decomposição da matriz original (X) em uma matriz de *scores* (T) e na transposta de *loadings* (P^T). Cada *score* é composto de informações de todo o espectro e os *loadings*, que são multiplicadores de vetores, são calculados para atribuir pesos às variáveis originais para que as pontuações maximizem sucessivamente a variância. Além dessas, é gerada uma matriz de resíduos, que engloba todas as variâncias não explicadas, conforme descrito matematicamente pela seguinte equação (Saeys et al., 2019).

$$X = TP^T + E$$

Após aplicação da PCA, as variáveis originais são reduzidas ao número de componentes principais relevantes para descrever a variabilidade dos dados, de modo que o primeiro componente (PC1) detém maior porcentagem de informação estatística, seguido do segundo componente (PC2), que contém informação que não foi explicado pelo primeiro componente principal e assim sucessivamente. Desta forma, a PCA permite visualizar e inferir sobre as variáveis semelhantes, além de identificar comportamentos atípicos que podem estar relacionadas a falhas no processo de obtenção de informações (Ferreira, 2015).

2.2.5.3. Análises de multivariada de predição/regressão

A calibração multivariada busca estabelecer um modelo matemático que relacione um conjunto de medidas instrumentais realizadas em amostras (variáveis independentes, X), com determinadas propriedades de interesse, como a concentração de um composto (variáveis dependentes, Y) (Saeys et al., 2019). Os parâmetros desse modelo são estimados a partir de um conjunto de amostras de calibração ou treinamento (cerca de 2/3 do universo amostral)

para os quais os espectros foram adquiridos e os traços de qualidade de interesse foram quantificados por métodos de referência (Duarte, 2015).

Diversos métodos lineares podem ser usados para a construção dos modelos de predição, como a Regressão Linear Múltipla (MLR), Regressão das Componente Principais (PCR) e Regressão por Mínimos Quadrados Parciais (PLS). Entre esses, o PLSR é o mais utilizado para monitorar a composição química de sementes *Brassicac*s (Tabela 2.2).

A Regressão Linear Múltipla (MLR) é um dos métodos mais antigos e é cada vez menos utilizado em aplicações devido ao aprimoramento do poder de computação. Esta regressão considera que as concentrações de interesse são função das absorbâncias, permitindo estabelecer uma ligação entre os comprimentos de onda e uma propriedade das amostras. O método MLR, apesar de simples, apresenta a desvantagem de requerer uma seleção prévia entre as variáveis preditoras para viabilizar o cálculo da matriz inversa (número de variáveis na calibração deve ser menor que o número de amostras) e que a correlação entre elas não seja alta (Xiaobo et al., 2010; Ferreira, 2015).

A regressão de componente principal (PCR) busca aproximar o espaço das medidas originais por um espaço vetorial reduzido. Para isso, é implementado uma decomposição PCA dos dados do espectro X e em seguida as informações das concentrações são projetadas nas pontuações dos componentes principais. Esse método pode ser utilizado para dados com um alto grau de covariância nas variáveis preditoras ou, onde matrizes mal condicionadas estão presentes e apresenta a vantagem de reduzir o ruído presente no sistema ao definir o espaço de dimensão reduzida. Entretanto, como a decomposição dos dados é feita usando apenas as informações espectrais, os fatores podem capturar outras fontes de variação além daquelas causadas por mudanças na concentração do constituinte de interesse (Xiaobo et al., 2010).

O método PLS tem o mesmo princípio do PCR, com adição de que uma proposta de restrição é utilizada para que a matriz de dados seja decomposta direcionando a solução para uma propriedade de interesse. Para isso, as informações das propriedades de interesse são incluídas no cálculo das novas variáveis, chamadas de variáveis latentes (VL), que são combinações lineares ortogonais das variáveis originais e capturam ao máximo a covariância entre as variáveis definidas e indefinidas. O PLS tem a capacidade de analisar um grande número de variáveis com informações ruidosas e correlacionadas, e reúne características positivas da regressão PCR e MLR. O PCR visa encontrar os fatores que capturam a maior parte da variância nos dados e o MLR busca um único fator que correlaciona os dados e suas concentrações, enquanto o PLS tenta maximizar a covariância, capturando assim a variância e correlacionando os dados juntos (Ferreira, 2015).

A qualidade dos modelos obtidos, para ambos os métodos, é avaliada por meio de uma etapa de validação. Inicialmente, o modelo é verificado usando uma validação cruzada *leave-one-out*, em que as amostras do conjunto de calibração são retiradas uma de cada vez e usadas como amostra de validação e, em seguida, o modelo pode ser revalidado utilizando amostras que não participaram do conjunto de calibração (validação externa) e por meio de ferramentas estatísticas de diagnóstico. Por fim, o modelo validado pode ser empregado para estimar as propriedades requeridas em amostras desconhecidas, a partir dos seus espectros (Duarte, 2015). A calibração multivariada apresenta como vantagens a realização de determinações diretas na presença de interferentes desde que presentes no conjunto de calibração e a possibilidade de quantificar simultaneamente diferentes analitos a partir do mesmo conjunto de espectros, além de apresentar uma diminuição do erro estimado no modelo devido a utilização de múltiplas variáveis (Tibola et al., 2018).

2.2.5.4. Análise multivariada de classificação/discriminação

Os métodos de classificação permitem atribuir uma amostra a uma classe ou categoria previamente definida, usando as informações contidas no espectro (Oliveira et al., 2019). Diversos métodos de reconhecimento de padrões são utilizados para classificar as amostras, como Análise Discriminante Linear (LDA), a Análise Discriminante pelo Método de Mínimos Quadrados Parciais (PLS-DA), método dos k-ésimos Vizinhos mais próximos (k-NN) e Modelagem Independente Suave da Analogia de Classes (SIMCA). Esses métodos partem do pressuposto de que quanto mais as amostras se assemelham entre si em relação às variáveis medidas, mais próximas elas estarão no espaço multidimensional gerado por tais variáveis (Ferreira, 2015).

A Análise Discriminante Linear (LDA) é um dos métodos de reconhecimento de padrões supervisionados mais usados e baseia-se na determinação de funções discriminantes lineares, que maximizam a razão da variância entre as classes e minimizam a variância dentro das classes. A classificação de objetos desconhecidos é realizada inserindo os dados de características nas funções discriminantes. É definido um hiperplano, de modo que as amostras de classes diferentes fiquem separadas e o objeto é atribuído à classe para a qual seu centróide tem a menor distância euclidiana. As desvantagens do LDA incluem o fato de que a associação simultânea de um objeto em várias classes não é detectada e que os outliers, que não pertencem a nenhuma das classes, sempre serão categorizados, além disso o LDA está

sujeito à restrição de que o número de variáveis não deve exceder o número de objetos (Berrueta et al., 2007; Otto, 2016).

A Análise Discriminante pelo Método de Mínimos Quadrados Parciais (PLS-DA) visa encontrar as variáveis e direções no espaço multivariado que discriminam as classes estabelecidas no conjunto de calibração. No PLS-DA, uma matriz Y é construída com zeros e uns, em que os elementos da classe é representada pelo número um e os elementos das demais classes é codificado pelo zero. Quanto mais próximo um elemento de uma determinada coluna em Y estiver de 1 e os elementos das outras colunas de 0, mais provável é que um objeto seja membro de uma classe específica. Uma característica importante do PLS é que ele leva em consideração os erros em ambas as matrizes, X e Y , e assume que eles são igualmente distribuídos. Além disso, o PLS é adequado para conjuntos de dados com menos objetos do que variáveis e um alto grau de intercorrelação entre as variáveis independentes (Ferreira, 2015; Berrueta et al., 2007).

O método dos k -enésimos Vizinhos mais próximos (k -NN) é um método simples de classificação não paramétrica, que leva em consideração as distâncias entre um objeto desconhecido e cada um dos objetos do conjunto de treinamento para designá-lo a uma classe. A distância mais baixa é selecionada e o objeto é atribuído à classe correspondente. Em k -NN, os k -objetos mais próximos à amostra desconhecida são selecionados e a incógnita é classificada no grupo ao qual a maioria dos k objetos pertence. O método apresenta vantagem pela sua simplicidade matemática, o que não o impede de obter bons resultados de classificação. Por outro lado, essa técnica tem limitações semelhantes à LDA e não funciona bem se houver grandes diferenças no número de amostras em cada classe e depende da escolha do número de vizinhos mais próximos (Berrueta et al., 2007; Otto, 2016).

A modelagem independente suave da analogia de classes (SIMCA) é um método de modelagem de classe, que constrói modelos independentes para cada classe com base nas pontuações dos Componentes Principais. O PCA é calculado apenas em amostras pertencentes à classe de interesse e o número de componentes principais é determinado por validação cruzada, de forma que um número suficiente de componentes principais é retido para contabilizar a maior parte da variação dentro de cada classe. A distância da classe pode ser calculada como a distância geométrica dos modelos de componentes principais ou considerando que cada classe é limitada por uma região do espaço, que representa uma porcentagem do nível de confiança (normalmente 95%) de que um determinado objeto pertence a uma classe (Otto, 2016; Wold and Sjöström, 1977).

2.1.5.5. Seleção de variáveis

A utilização de espectroscopia de infravermelho próximo implica na obtenção de uma grande quantidade de variáveis e de informações sobrepostas. Para melhorar o desempenho do modelo de previsão são utilizados métodos de seleção de variáveis, os quais eliminam comprimentos de onda que não possuem informações relevantes para as condições avaliadas (referidas como não informativas), bem como interferentes e variáveis responsáveis por produzir não linearidade do modelo, gerando um subconjunto contendo as variáveis que melhor definem a classe da amostra ou melhor se correlacionam com as propriedades de interesse (variáveis informativas) (Pasquini, 2018).

Esses métodos selecionam regiões específicas do espectro (um comprimento de onda ou um conjunto de comprimentos de onda) em que a colinearidade não é tão importante, para minimizar os erros de previsão e gerar modelos mais estáveis, robustos e mais simples de interpretar (focada em um pequeno conjunto de variáveis relevantes) (Xiaobo et al., 2010). A seleção de variáveis informativas é útil, sobretudo, para viabilizar aplicações industriais online com imagens hiperespectrais, pois a elevada quantidade de comprimentos de onda obtido para cada imagem aumenta o tempo para processamento da imagem e, consequentemente, para a obtenção de resultados (Sun et al., 2017).

Diversos métodos de seleção de variáveis são descritos na literatura, como Algoritmo de Projeções Sucessivas (SPA), Algoritmo Genético (GA), iPLS (Interval Partial Least Square), entre outros, conforme mencionado por Pasquini (2018). Esses métodos utilizam diferentes critérios e mecanismos de seleção, produzindo diferentes conjuntos de variáveis selecionadas com diversos tamanhos, comprimentos de onda e relevância para o modelo. Podem ser utilizadas abordagens univariadas, em que variáveis com maior correlação com a resposta são selecionadas e abordagens sequenciais, que selecionam a melhor variável e então o melhor par formado pela primeira e segunda e assim por diante (Leardi et al., 2002).

O iPLS é uma extensão interativa desenvolvida internamente para o PLS, que fragmenta o espectro em pequenos intervalos equidistantes entre si e constrói modelos PLS locais utilizando o mesmo número de variáveis latentes para todas as partes. Inicialmente, as variáveis originais são selecionadas aleatoriamente e calcula-se o PLS local, que é comparado com o modelo completo usando como parâmetro a raiz quadrática média do erro de validação cruzada (RMSECV). A variável a ser adicionada ou retirada do modelo é escolhida aleatoriamente e um novo modelo PLS é construído e avaliado. Se o novo RMSECV for inferior ao original, o novo conjunto de variáveis substituirá o original. O algoritmo é

encerrado quando cada variável foi testada uma vez sem oferecer melhorias e, então, a região que obtém o menor erro para a predição da variável de interesse é selecionada. O iPLS é um método rápido, que fornece uma imagem geral das informações importantes em diferentes subdivisões espectrais, focando nas regiões autônomas mais relevantes e removendo interferências de outras regiões (Leardi and Nørgaard, 2004; Nørgaard et al., 2000; Osborne et al., 1997).

O Algoritmo genético (GA) utiliza uma abordagem probabilística baseada em mecanismos de seleção natural para selecionar comprimentos de onda relevantes. Esse método explora o conhecimento contido em uma população de soluções chamadas cromossomos (que são comprimentos de onda específicos) para gerar novas e melhores soluções. Durante cada geração, os cromossomos são avaliados individualmente usando o RMSEC e os cromossomos com melhores figuras de mérito podem sobreviver, sofrer mutações e se recombinar para formar uma nova população que muitas vezes produz uma solução melhor para o problema. Uma vez que as variáveis selecionadas dependem da população inicial gerada aleatoriamente, o algoritmo é executado diversas vezes e modelo final é obtido seguindo uma abordagem *stepwise*, em que as variáveis são inseridas de acordo com a frequência das seleções (ou seja, no modelo com n variáveis, as n variáveis selecionadas com mais frequência estão presentes) (Forina et al., 2004; Leardi, 2000; Leardi and González, 1998). O algoritmo GA é um bom método de otimização que otimiza o espaço variável globalmente, entretanto é fácil causar sobreajuste quando os dados de espectros NIR contêm mais de 200 variáveis (Yu et al., 2020) e devido à sua natureza estocástica, os resultados são dependentes da realização e as seleções de variáveis podem não ser reproduzíveis (Araújo et al., 2001). Além disso, apresenta um enorme desafio de configuração para o usuário devido aos inúmeros fatores ajustáveis (função de aptidão, os critérios de convergência, a taxa de mutação, o esquema de cruzamento, o número de cromossomos considerados, a população inicial e o número de gerações) que afetam o resultado do GA (Xiaobo et al., 2010).

O SPA emprega operações de projeção simples em um espaço vetorial para obter subconjuntos de variáveis com o mínimo de colinearidade. O objetivo do SPA é selecionar comprimentos de onda cujo conteúdo de informação seja minimamente redundante e baseia-se no princípio de que a nova variável selecionada tem o maior valor de projeção no subespaço ortogonal da variável selecionada anteriormente. É um método de seleção direta, que inicia com um comprimento de onda e incorpora um novo a cada iteração, até que um determinado número N de comprimentos de onda seja atingido. Apesar de proporcionar bons

resultados, as variáveis selecionadas por SPA podem ter baixo sinal de ruído para relação (S/N) ou ser insuficientes para calibração multivariada, o que pode afetar a precisão da previsão do modelo (Araújo et al., 2001; Xiaobo et al., 2010).

Métodos baseados na aparência dos parâmetros dos modelos de predição, tais como Loadings (gráfico de carregamentos), Importância da Variável para a Projeção (VIP) e coeficientes de regressão, também podem ser utilizados para identificar as variáveis que fornecem o melhor modelo. Normalmente, as variáveis que têm carregamentos relativamente baixos em todos os componentes são viáveis de serem removidas, bem como variáveis com coeficientes de regressão baixos. O VIP é uma medida combinada de quanto uma variável contribui para descrever os dois conjuntos de dados (variáveis dependentes e independentes). e reflete quão importante essa informação é para o modelo das variáveis independentes. Um VIP menor que um indica uma variável não importante (Andersen e Bro, 2010).

**Capítulo 3 - NIR spectroscopy and hyperspectral imaging for investigation
of *Brassicas* seeds quality and species**

NIR spectroscopy and hyperspectral imaging for investigation of *Brassicas* seeds quality and species

Maria Lucimar da Silva Medeiros^a; J.P. Cruz-Tirado^a; José Marcelino de Souza Netto^b;
Ana Paula Badan Ribeiro^c; Doglas Bassegio^d; Douglas Fernandes Barbin^{a,*}

^aDepartment of Food Engineering, School of Food Engineering, University of Campinas, Campinas, SP, Brazil. ^bDepartment of Fundamental Chemistry, Federal University of Pernambuco, Recife, PE, Brazil. ^cDepartment of Food Technology, School of Food Engineering, University of Campinas, Campinas, SP, Brazil. ^dDepartment of Energy Engineering in Agriculture, State University of Western Paraná, Cascavel, PR, Brazil.

Abstract

Brassica is a genus of plants with high industrial interest in the production of edible oils, modified lipids, industrial oils, and biofuels. The most important chemical indicator for assessing *Brassicas* seeds quality for the processing industry is oil content, as it varies according to seed varieties. The industry demands fast and efficient analytical methods to quantify oil content and select seeds with high yield. In the present work, Near Infrared Spectroscopy (NIR) devices: portable NIR spectrometer (NIRS) and NIR-hyperspectral image (NIR-HSI) in tandem with chemometrics were investigated to (1) to discriminate seeds species and (2) predict the oil content in intact *Brassicas* seeds. PLS-DA analysis showed satisfactory discrimination between *Brassicas* species, with classification rates of 94.9 and 100% for the NIR spectrometer and NIR-HSI devices, respectively. The best PLSR models presented RMSEP < 1.0, R² de 0,682 and 0.751, and RPD of 2.0 and 2.31 for the NIRS and NIR-HSI, indicating that both devices could be adapted to provide an estimate of oil content in *Brassicas* seeds, according to the industry or consumer needs.

Keywords: portable spectrometer; hyperspectral imaging; canola; mustard; rapeseed; machine learning.

3.1. INTRODUCTION

Brassica is a plant genus belonging to the Brassicaceae Family with high industrial interest in the oil extracted from seeds of *B. napus*, *B. rapa* and *B. juncea*. These seeds constitute one of the most important sources of vegetable oils worldwide, surpassing 75 million tons of production in 2018 (FAO, 2019; Li et al., 2013). *Brassica* seeds have an oil content between 17.6 and 54.9% (Oblath et al., 2016; Velasco et al., 1999), which is the most relevant parameter of commercial seed quality (Wan et al., 2018; Welle et al., 2007).

Oils of traditional varieties of these species are characterized by a high concentration of erucic acid and have numerous applications in the pharmaceutical, cosmetic and oleochemical industries (Reginato et al., 2013), while new cultivars, developed with low erucic acid (canola quality), are used for food purposes. Oils containing high levels of erucic acid can be fully hydrogenated to obtain behenic acid, which has demonstrated an important anti-obesity effect. Oils from *Brassicac*s represent the only potential alternative for obtaining lipid bases containing high levels of behenic acid, of great interest for application in food (Moreira et al., 2017; Silva et al., 2020).

The oil content and composition is influenced by several factors, including environmental conditions, genotypes, genotype x environment interactions and cultural practices (Long et al., 2012). The identification of *Brassicac*s seeds with higher oil content requires an extensive germplasm evaluation for this characteristic (Kaur et al., 2016) and its quantification is commonly performed by mechanical pressing or solvent extraction (Gaber et al., 2018). Although these methods provides an effective and accurate measurement of the oil content in the seed, it is time consuming, costly and uses large amounts of chemicals that can be harmful to the analyst and the environment (Cheng et al., 2017; Jin et al., 2016; Kaur et al., 2016).

Recently, the new industrial era called “Industry 4.0” has required the development of non-destructive and environmental friendly techniques for the simple, rapid, and effective determination of oil content. NIR spectroscopy is a vibrational technique based on the absorption of electromagnetic radiation in the NIR range (780 - 2500 nm) and provides information associated with molecular vibrations of chemical bonds involving C-H, N-H, S-H and O-H, which are the primary structural components of organic molecules (Su et al., 2017). This technique is useful for simultaneous determination of major constituents and has been previously reported to predict oil

content in soybean (Jiang, 2020), sugar beet seed (Martínez-Arias et al., 2018), sesame seed (Xu et al., 2017), *Brassica napus* seed (Olivos-Trujillo et al., 2015) and others. Nowadays, new sensors such as portable NIR spectrometers, are being tested in several agricultural products (Barbin et al., 2018; Mendez et al., 2019). Due to its small size, these sensors have an ergonomic and portable design, making it possible to monitor the industry at various stages of the supply chain, from harvesting to the processing industry. Previous works reported the use of portable equipment for prediction of oil content in intact canola seeds (Barthet et al., 2020) and peanut seeds (Bilal et al., 2020). However, NIR spectroscopy is a spot measurement technique, and in some cases, due to food heterogeneity (Mendez et al., 2019), the spectral information obtained may be insufficient or not representative of the entire sample. A promising alternative to overcome this limitation is the application of techniques that include traditional NIR spectroscopy and computer vision in a single device, such as hyperspectral imaging (NIR-HSI), providing information about the chemical composition and its disposition on the sample surface (Amigo, 2010). Hyperspectral imaging (HSI) has demonstrated good performance to guarantee authenticity and predict quality attributes in grape seed (Zhao et al., 2018), cotton seed (Soares et al., 2016), cocoa beans (Caporaso et al., 2018a) and others (Feng et al., 2019), however studies involving prediction of oil content in *Brassica* seeds have not been reported.

Due to the large amount of data generated, multivariate analysis techniques are often used to interpret important information from NIR spectra (Kaur et al., 2016). Principal Component Analysis (PCA) are used to obtain an initial overview of the spectra, while multivariate calibration methods, such as Discriminant Analysis (PLS-DA) and Partial Least Squares Regression (PLSR), allow the classification and prediction of desired parameters in samples, respectively (Ferreira, 2015). In addition, variable selection methods can be applied to eliminate redundant information present in the spectrum and to identify important spectral regions.

The objective of this study was to compare the performance of portable NIR spectrometer and NIR-HSI to discriminate *Brassicac*s species and to quantify the oil content in intact seeds. We built calibration models with the entire spectral range and simplified models using important wavelengths, related to the chemical information of interest.

3.2. MATERIAL AND METHODS

3.2.1. SAMPLES COLLECTION

Nineteen varieties of *Brassica* seeds were harvested at the experimental farm Lageado, belonging to Universidade do Estado de São Paulo (UNESP), Botucatu, SP, Brazil, comprising nine varieties of the species *B. juncea* (27 samples), nine varieties of *B. rapa* (27 samples) and one of *B. napus* (03 samples). Six samples of *B. rapa* had yellow seeds, while the others had dark seeds. The particle size of samples was between 1 and 2 mm, and moisture content was less than 6%. Accessions were obtained from the North Central Regional Plant Introduction Station (NCRPIS-USDA-USA).

3.2.2. SPECTRAL DATA COLLECTION

3.2.2.1. Near infrared spectroscopy (NIR)

NIR spectra were recorded using a portable spectrometer (DLPR NIRscan™ Nano, Texas Instruments, USA), in the spectral range from 900 to 1700 nm, with average intervals of 3.51 nm. *Brassicas* seeds were placed in Petri dishes ($\varnothing = 60$ mm) and five spectra were recorded at different points, in order to obtain a better representation. Finally, 54 mean spectra were recorded for obtained for *B. juncea*, 54 for *B. rapa* and 24 for *B. napus*.

3.2.2.2. Near infrared hyperspectral imaging (NIR-HSI)

Hyperspectral images of *Brassicas* seeds were acquired using a hyperspectral camera SisuCHEMA SWIR (Specim Spectral Imaging Ltd, FIN-90571 Oulu, Finland) in pushbroom mode, in the wavelength range of 928 - 2524 nm, at 6.23 nm intervals and 256 wavelength channels, with spectral resolution of 10 nm and spatial resolution of 320 pixels by line scanning. The intact seeds were spread on Petri dishes and then placed on the translation base and moved to reach the equipment's field of view. The images were acquired using a 50 mm lens, with an exposure time of 2.1 ms and controlled by the Evinco software (UmBio AB, Sweden). The program automatically subtracts the white (~99% reflectance) and dark (0% reflectance) references from the

acquired images. The hyperspectral images were segmented using the Evince software and the medium spectra were extracted using the Matlab R2019b software (Mathworks, Natick, USA). As previously described for the portable NIR equipment, 132 mean spectra were obtained for all species.

3.2.3. MEASUREMENT OF OIL CONTENT

After image acquisition, *Brassicas* seeds were ground (model IKA A11 basic), and the oil was extracted with petroleum ether in a Butt extractor, according to AM 2-93 method (AOCS, 2009). In this method, 2 g of ground samples were weighed in a paper cartridge and then subjected to direct extraction at a temperature of 90° C for 4 hours. After extraction, the solvent was removed by evaporation in a rotary evaporator and the samples were kept in an oven to eliminate residual solvent, until constant weight. All analysis was performed in duplicate.

3.2.4. STATISTICAL DATA ANALYSIS

Differences in oil content in *Brassicas* species were compared using analysis of variance (ANOVA), followed by the Tukey multiple-comparison test ($p < 0.05$). Statistical analysis was performed using Sisvar software, Version 5.7 (Lavras, Minas Gerais, Brazil).

3.2.5. SPECTRAL PRE-PROCESSING

Different pre-processing methods were investigated to correct the effects of light scattering caused by seed morphology and other undesirable effects caused by external factors, such as random noise and changes in the baseline. Multiplicative Scatter Correction (MSC), Savitzky-Golay smoothing, First Savitzky-Golay Derivative and Second Savitzky-Golay Derivative were tested individually and in combinations. Both derivatives were used with a gap of 13 points and second order polynomial filtering. All pre-processing methods and data analysis were performed using the PLS Toolbox 8.9.1 from Eigenvector Research, Inc. (Manson, WA, USA) for Matlab R2019a (Mathworks, Natick, USA).

3.2.6. SELECTION OF CLASSIFICATION AND CALIBRATION SETS

For both the classification and calibration set, the samples were divided into calibration (70%) and prediction (30%) using the KS algorithm (Kennard-Stone, 1969). In this algorithm, the Euclidean distances between the vectors of the instrumental responses (X) of the selected samples were maximized. For the calibration, the KSXY algorithm was used, which considers the information of both X (matrix of instrumental responses) and Y (matrix of classes in the classification and concentrations in the calibration) in the calculation of the inter-sample distances. Both KS and KSXY algorithms are part of a routine developed in a Matlab R2010a environment (Mathworks, Natick, USA).

3.2.7. MULTIVARIATE ANALYSES

3.2.7.1. Principal component analysis (PCA)

Principal Component Analysis (PCA) is an unsupervised multivariate analysis technique used to reduce spectral dimensions in principal components (PC), which capture maximum data variability and maintain relevant spectral information (Xu et al., 2018). In this work, PCA was applied to the NIR and NIR-HSI raw and pre-processed data as an exploratory analysis to visualize the clustering of *Brassicas* species according to their spectral variations. PCA models were developed using singular value decomposition algorithm (SVD) (95% confidence level) and outliers (anomalous samples) were detected and eliminated using the residual Q and Hotelling T² values.

3.2.7.2. Classification analysis

Partial Least Squares discriminant analysis (PLS-DA) was applied to build classification models for NIR and NIR-HSI spectra to discriminate between the varieties of *Brassicas* seeds. The optimum number of Latent variables (LV) was chosen using the lowest root mean cross-validation square error (RMSECV). The interval PLS (iPLS) variable selection algorithm was used to identify the important wavelengths for the classification of *Brassicas* seeds and to build simplified models. Model performance was evaluated by sensitivity (a measure of how often the model correctly identified a

positive sample as positive), specificity (a measure of how accurate the model was against false-positives), accuracy (ability of the model to identify actually true and false samples out from all samples including the false positive and negative samples) and classification error.

$$\begin{aligned}
 \text{Sensitivity (\%)} &= \frac{TP}{TP + FN} \\
 \text{Specificity (\%)} &= \frac{TN}{FP + TN} \\
 \text{Accuracy (\%)} &= \frac{TP + TN}{TP + TN + FP + FN} \\
 \text{Error (\%)} &= \frac{FP + FN}{TP + FP + TN + FN}
 \end{aligned}$$

where TP is true positive, TN is true negative, FN is false negative, and FP is false positive.

3.2.7.3. PLSR for prediction of oil content

Partial Least Squares Regression (PLSR) was applied to assess the ability of the NIR and NIR-HSI data to predict the oil content of *Brassicas* Seeds. This technique allows establishing a linear mathematical relationship between two sets of data, X (spectra) and Y (oil content), using orthogonal linear combinations of the original variables, called latent variables (LV), which capture the maximum covariance between the variables X and Y (Ferreira, 2015). The ideal number of LV was chosen using the lowest root mean square error of cross-validation (RMSECV), where adding a new LV did not substantially improve the model's performance. Calibration models were developed using full spectra, and the iPLS algorithm was used to select the relevant wavelengths to build reduced models. The performance of the calibration models was assessed by the coefficient of determination and Root of the Mean Square Error for Calibration (R_c^2 , RMSEC) and cross validation (R_{cv}^2 , RMSECV). Predictive capacity of the models was assessed using the coefficient of determination (R_p^2), Root of the Mean Square Error for prediction (RMSEP) and the Performance Ratio for Deviation (RPD = SD/RMSEP, where SD is the standard deviation of the oil content).

3.3. RESULTS AND DISCUSSION

3.3.1. OIL CONTENT OF SAMPLES

Brassicas seeds showed oil content between 28.06 and 36.90 % and the average values did not differ statistically (Tukey's test, $p < 0.05$) between species (Table 3.1). Similar values for *Brassicas* species were reported by Oblath et al. (2016) (17.6 – 51.0 %) e Velasco, Möllers and Becker (1999) (28.5 - 54.9 %). The concentration range, as well as the number of samples and their distribution in this range, are important parameters for building calibration models, as they ensure that the model contemplates a wide variation among the samples and that they are evenly distributed (Fearn, 1992). The oil contents obtained in this paper included different *Brassicas* species used for oil production and are close to the data reported by authors who obtained good prediction models for compositional attributes in seeds (Montes et al., 2007; Petisco et al., 2010).

Table 3.1. Oil content (g/100g) for different species of *Brassicas* seeds.

<i>Brassicas</i> Species	Range	Mean	SD
<i>B. juncea</i>	28.06 – 36.90	32.66 ^a	2.08
<i>B. rapa</i>	29.35 – 36.71	32.90 ^a	1.81
<i>B. napus</i>	32.60 – 34.89	33.64 ^a	0.91
<i>Brassicas</i> (all species)	28.06 – 36.90	32.82	1.91

Same letters indicate that there is no significant difference ($p > 0.05$) among *Brassicas* species, by Tukey test.

3.3.2. SPECTRA PROFILE

The mean absorbance spectra obtained for intact *Brassicas* seeds (Fig. 3.1) from both NIR spectrometer and NIR-HSI showed a similar shape and differed only in the absorbance intensity along the spectral region. The differences observed at the baseline, mainly in the spectra obtained by NIR spectrometer, correspond to changes in the light scattering properties between the samples, probably associated with the seeds morphology and granulometry (Manley, 2014).

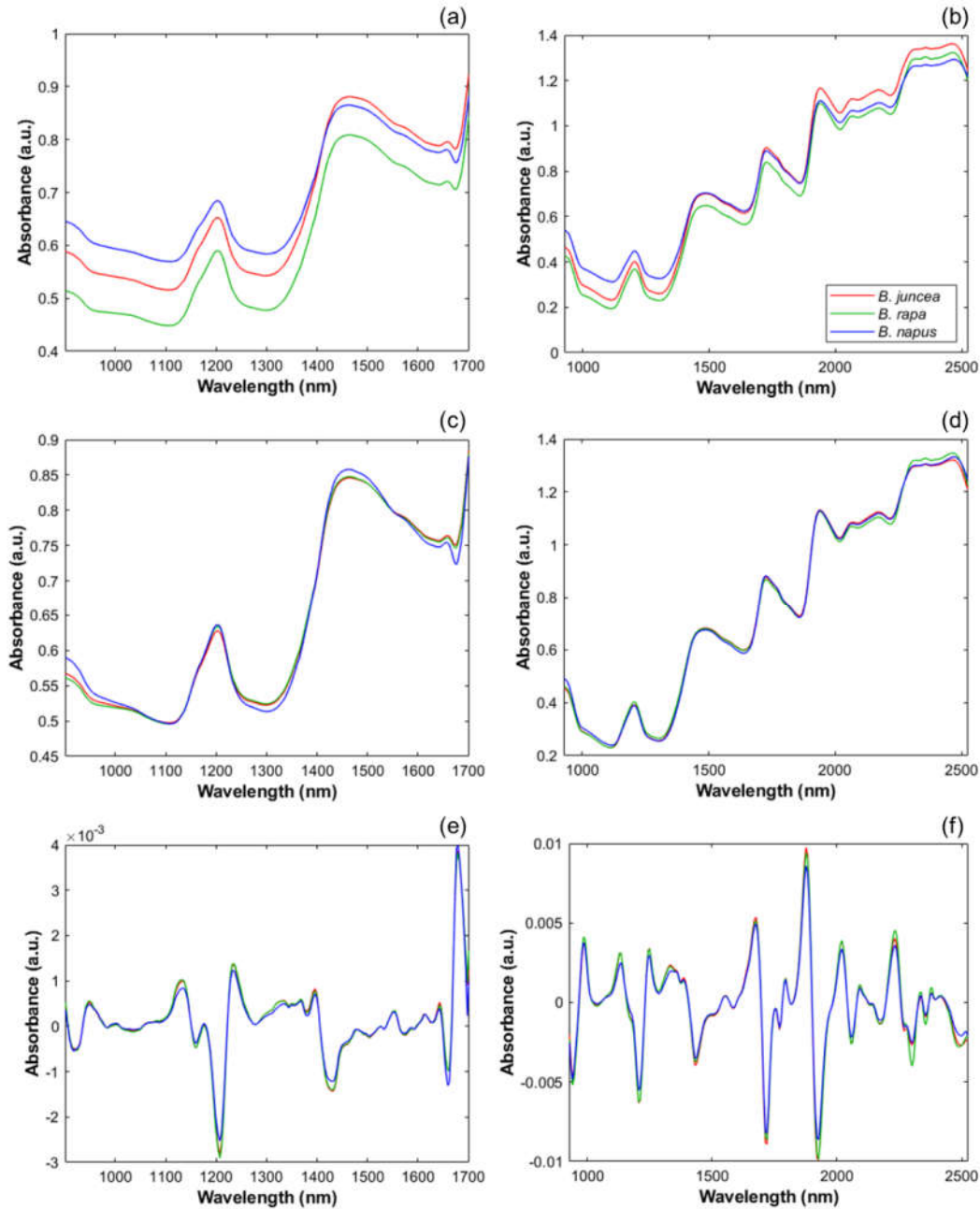


Figure 3.1. Mean spectra of *Brassic*as seeds after Savitzky-Golay smoothing (a and b), Savitzky-Golay smoothing and MSC (c and d) e 2nd derivative (e and f) to NIR and NIR-HSI data, respectively.

The main absorption bands in the NIR spectra after smoothing (Fig. 3.1a) and smoothing with MSC (Fig.3.1c) were observed at 1215 nm and 1450 nm, which correspond to the C-H stretching vibration in the second overtone (-CH₂) and O-H vibrations in the first overtone, attributed to oil, water and fiber, respectively. In addition to the peaks at 1215 and 1450 nm, some local absorption maximums are

observed at 1725, 1940 and 2310 nm. The peak at 1725 nm is related to C-H stretching vibrations in the first harmonic ($-\text{CH}_2$), in 1940 nm corresponds to the stretching vibration or O-H combination of deformation and at 2310 nm, the stretching vibration or C-H combination of deformation, attributable to the main macronutrients present in seeds, such as water, oil and fiber (Xu et al., 2019; Osborne et al., 1993). The presence of these peaks in the region above 1700 nm can significantly contribute to the performance of the prediction and classification models and the lack of data in this range can lead to a lower performance of the spectrum in the portable NIR compared to the NIR-HSI spectrum.

After combining the second derivative and MSC preprocessing (Fig.3.1 e and f), other peaks and valleys were highlighted throughout the spectra. The peaks and valleys around 930, 1037, 1360, 1765 nm are associated with stretching vibration C-H in the first and third overtones or C-H combination of deformation, characteristic of the oil and fiber. The peak at 2230 and the valley at 2294 nm correspond to the N-H stretching vibration or NH_3^+ combination of deformation and stretching vibrations involving N-H and C=O, respectively, which may be associated with amino acids, and the peaks at 2050 and 2180 nm are related to N-H vibrations present in peptide bonds and amide groups of proteins (Xu et al., 2019; Osborne et al., 1993).

3.3.3. PRINCIPAL COMPONENTS ANALYSIS (PCA)

The principal component analysis was applied to the pre-processed NIRS and NIR-HSI spectra, in order to visualize the variation between the samples and to evaluate the effect of spectra pre-processing for grouping of *Brassicacae* species. The best separation of groups was obtained by applying Savitzky - Golay smoothing (Fig. 3.2).

The bidimensional representation of the PC1 and PC2 scores (Fig. 3.2a and 3.2c) for the NIRS and NIR-HSI spectra showed satisfactory discrimination between *Brassicacae* species. In general, the seeds of *B. juncea* and *B. rapa* were characterized by positive and negative scores on PC1, respectively, and the seeds of *B. napus*, were located on the positive part of PC2. Scores plots of the PC1 and PC3 (Fig. 2b) for NIRS data, showed a clear discrimination between accessions of *B. rapa* and the other samples. The yellow colored *B. rapa* seeds were spread on the negative side of PC1 and PC3, while dark colored samples were characterized by positive scores. This formation of clusters, also observed by other authors (Petisco et al., 2010), may be related to

differences in absorbance values due to differences in composition of the samples, since yellow color of *Brassicas* seeds is associated with a lower fiber content (Badani et al., 2006). These results indicate that the spectral information in the NIR region can provide enough information to differentiate *Brassicas* species (Petisco et al., 2010), as well as accessions that present differences in fiber contents.

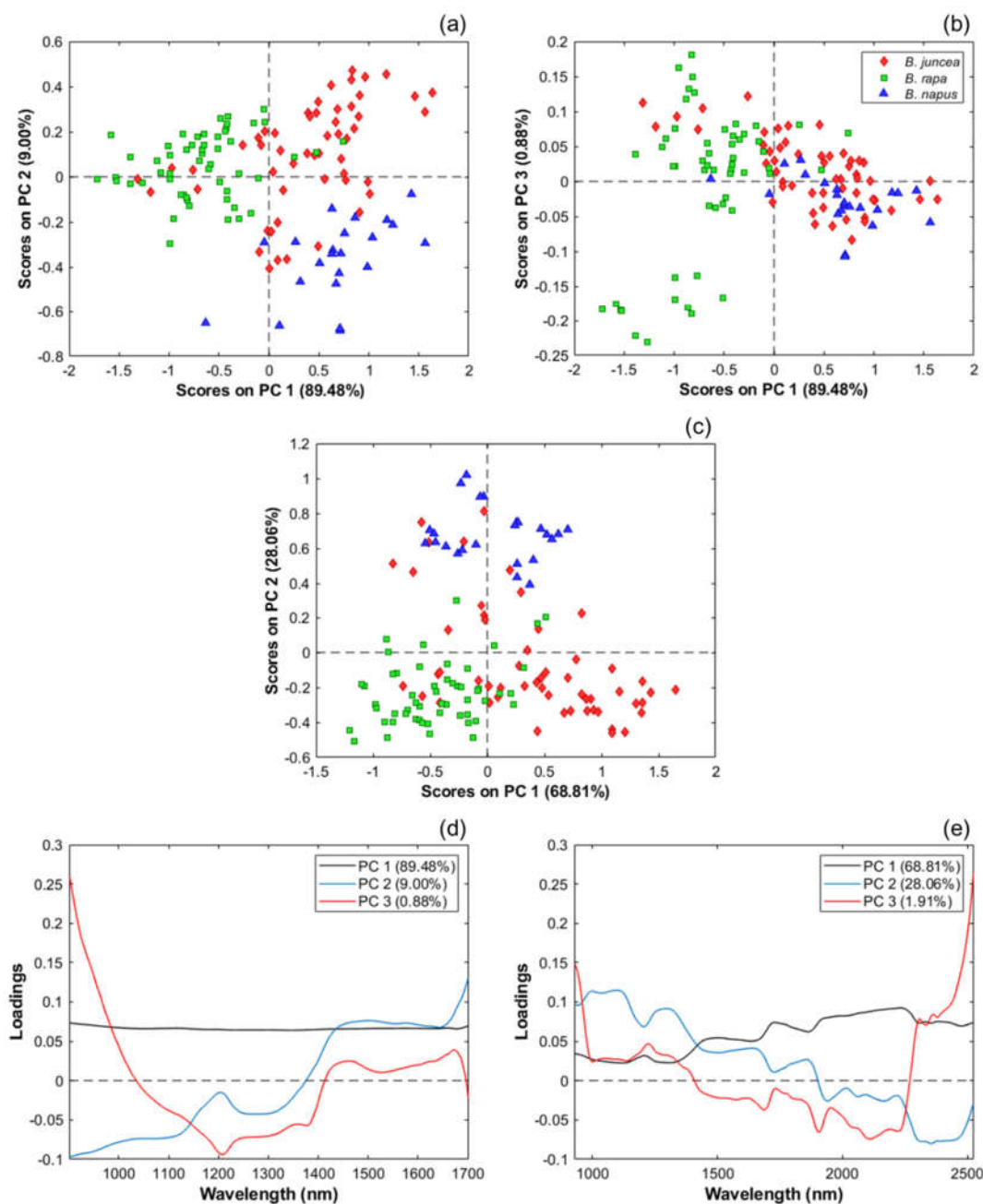


Figure 3.2. Scores plots of PCA performed on spectra with Savitzky–Golay smoothing for NIR (a and b) and HSI data (c) and their correspondent Loadings plots (d and e).

The first three main components, in both techniques, explained more than 98% of the total sample variance (Fig. 3.2d and 3.2e). Loadings of PC1 for NIR data (Fig. 3.2d) were similar to spectrum of *Brassic*s seeds and, consequently, the same load was observed for all wavelengths. This behavior was also observed by Golebiowski, Leong e Panozzo (2005), who reported that although the seed consists of high levels of oil, the main information contained in the spectrum is related to the physical properties of the seed. In general, the absorption maximums observed in the PCA loadings plots are similar to those highlighted in the spectrum and are attributed mainly to water (970, 1450 and 1940 nm), aliphatic chains of fat components (937, 1200 - 1360 nm, 1720, 2101 and 2300 nm), aromatic compounds or CH₃ (1670 nm) and amino acids (2242 nm) (Osborne et al., 1993).

3.3.4. CLASSIFICATION MODELS

PLS-DA models (Table 3.2) obtained for discrimination of species of *Brassic*s seeds with different pre-processing methods were built using between 3 and 11 latent variables, and were evaluated according to sensitivity, specificity, error rate and accuracy (Table 1 in supplementary material).

Table 3.2. Performance of PLS-DA model for classification the species of *Brassic*s seeds using NIRS spectra.

Spectral pre-treatment	LV	Species	Calibration			Validation		
			Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
NIRS								
Full wavelength								
Raw data	4	B. juncea	0.763	0.741	0.894	0.750	0.917	0.873
		B. rapa	0.919	0.891		1.000	0.957	
		B. napus	1.000	0.920		1.000	0.909	
Smoothing	4	B. juncea	0.763	0.722	0.894	0.750	0.917	0.936
		B. rapa	0.919	0.891		1.000	0.957	
		B. napus	1.000	0.920		1.000	0.879	
MSC	5	B. juncea	0.816	0.685	0.839	0.938	0.583	0.784
		B. rapa	0.811	0.764		1.000	0.565	
		B. napus	1.000	0.920		0.857	1.000	
Smoothing + MSC	5	B. juncea	0.842	0.648	0.804	0.875	0.667	0.832
		B. rapa	0.811	0.727		1.000	0.609	
		B. napus	1.000	0.920		0.857	1.000	
1st derivative	9	B. juncea	0.842	0.833	0.878	0.875	0.792	0.871
		B. rapa	0.892	0.873		0.882	0.826	
		B. napus	0.941	0.947		0.714	1.000	
2nd derivative	7	B. juncea	0.868	0.833	0.922	0.750	0.792	0.800
		B. rapa	0.865	0.909		0.824	0.783	
		B. napus	1.000	0.933		1.000	0.909	
MSC + 1st derivative	9	B. juncea	0.868	0.778	0.922	1.000	0.750	0.822
		B. rapa	0.865	0.909		0.824	0.783	

		<i>B. napus</i>	1.000	0.933		0.857	1.000	
MSC +		<i>B. juncea</i>	0.868	0.704		0.938	0.375	
2nd	4	<i>B. rapa</i>	0.757	0.600	0.828	0.588	0.739	0.753
derivative		<i>B. napus</i>	0.941	0.907		1.000	0.970	
Selected wavelength (iPLS)								
		<i>B. juncea</i>	0.868	0.870		0.813	0.917	
Smoothing	8	<i>B. rapa</i>	0.865	0.964	0.960	1.000	0.913	0.949
		<i>B. napus</i>	1.000	0.947		1.000	0.970	
NIR-HIS								
Full wavelength								
		<i>B. juncea</i>	0.947	0.870		0.938	0.958	
Raw data	9	<i>B. rapa</i>	0.973	0.982	0.959	1.000	1.000	0.954
		<i>B. napus</i>	1.000	0.973		1.000	0.970	
		<i>B. juncea</i>	0.939	0.814		1.000	0.958	
Smoothing	8	<i>B. rapa</i>	0.929	0.940	0.951	1.000	0.957	1.000
		<i>B. napus</i>	1.000	0.960		1.000	1.000	
		<i>B. juncea</i>	0.789	0.796		0.875	0.792	
MSC	3	<i>B. rapa</i>	0.892	0.927	0.912	1.000	0.913	0.919
		<i>B. napus</i>	1.000	0.973		0.857	0.970	
		<i>B. juncea</i>	0.789	0.796		0.875	0.792	
Smoothing + MSC	3	<i>B. rapa</i>	0.892	0.927	0.905	1.000	0.913	0.918
		<i>B. napus</i>	1.000	0.947		0.857	0.970	
		<i>B. juncea</i>	0.947	0.889		0.875	0.917	
1st derivative	9	<i>B. rapa</i>	0.973	0.945	0.986	1.000	0.957	0.967
		<i>B. napus</i>	1.000	0.987		1.000	1.000	
		<i>B. juncea</i>	0.895	0.889		1.000	1.000	
2nd derivative	9	<i>B. rapa</i>	0.973	0.964	0.978	1.000	0.957	0.982
		<i>B. napus</i>	1.000	0.987		1.000	1.000	
		<i>B. juncea</i>	0.895	0.907		1.000	0.958	
MSC + 1st derivative	11	<i>B. rapa</i>	0.946	0.982	0.972	1.000	1.000	1.000
		<i>B. napus</i>	1.000	0.987		1.000	1.000	
		<i>B. juncea</i>	0.816	0.852		0.750	0.833	
MSC +		<i>B. rapa</i>	0.865	0.927	0.907	1.000	0.913	0.918
2nd derivative	4	<i>B. napus</i>	0.941	0.973		1.000	0.970	
Selected wavelength (iPLS)								
		<i>B. juncea</i>	0.921	0.926		0.875	0.958	
MSC	9	<i>B. rapa</i>	1.000	0.964	0.979	1.000	0.957	0.982
		<i>B. napus</i>	1.000	0.973		1.000	1.000	

In general, the models built with the NIR-HSI spectra showed better classification performances, with higher sensitivity and specificity values, lower error rates and greater accuracy compared to the NIRS spectra. The pre-processing applied to the NIR-HSI data did not significantly improve the discriminatory capacity of the models. The best classification indexes were obtained for the models build with raw data, pre-treated with first derivative, second derivative and a combination of MSC with first derivative, where the sensitivity and specificity values were greater than 0.870 and the accuracy greater than 95.0%. For the NIRS data, pre-processing with Savitzky - Golay smoothing provided the best classification model, with accuracy of 0.963, and high values of sensitivity (0.722 and 1.0) and specificity (0.750 and 1.0), demonstrating good ability to discriminate between the *Brassicas* species. The high efficiency values in the discrimination and classification of *B. napus*, in both spectrum acquisition methods,

samples may be related to the small variation between samples, since only a variety of seeds were included in this species.

The results presented for PLS-DA classification models obtained with the complete spectra were consistent with other studies involving the classification of seeds using spectroscopy techniques, such as grape seeds (Zhao et al., 2018), tomato seeds (Shrestia, Deleuran and Gislum, 2016), and cotton seeds (Soares et al., 2016).

Simplified models using important wavelengths selected by the iPLS algorithm, are presented in full in Table 2 (supplementary material). These models were obtained using between 5 and 14 wavelengths and presented performance close to or higher to the complete models. The best iPLS-DA model for the NIRS spectra was obtained with the Savitzky-Golay smoothing, using 10 wavelengths (1117, 1131, 1174, 1243, 1312-1315, 1371, 1391, 1637, 1673) (Fig. 3.3.a), and presented an accuracy of 0.949 and high values of sensitivity (0.813 and 1.0) and specificity (0.870 and 0.970). The distribution of samples in classes, seen in Fig. 3.3.c, shows that 2/16 samples of *B. juncea* were incorrectly classified as *B. rapa* and 1/16 samples of *B. juncea* were classified as *B. napus*.

Spectra processed with MSC provided the best iPLS-DA model for the NIR-HSI data using 10 wavelengths (1167, 1180, 1343-1350, 1394, 1644, 1713, 1895, 2163, 2269) (Fig. 3.3.b). The model performed better than using the full spectra, with higher sensitivity, higher specificity, minor classification errors and accuracy of 98.2% (Fig. 3.3.d). Only 1/16 sample of *B. juncea* was incorrectly classified as *B. rapa*, which demonstrates the potential of simplified models based on NIR-HSI data to discriminate between different species of seeds *Brassicacae*. The wavelengths selected by iPLS that presented the greatest contribution to class separation may be related to major components present in seeds, such as aliphatic fat chain (1167, 1180, 1243, 1312 - 1394, 1637, 1713 nm), protein (2160 nm), structural polysaccharides (1895 and 2269 nm) and aromatic compounds (1131 and 1673 nm) (Osborne et al., 1993).

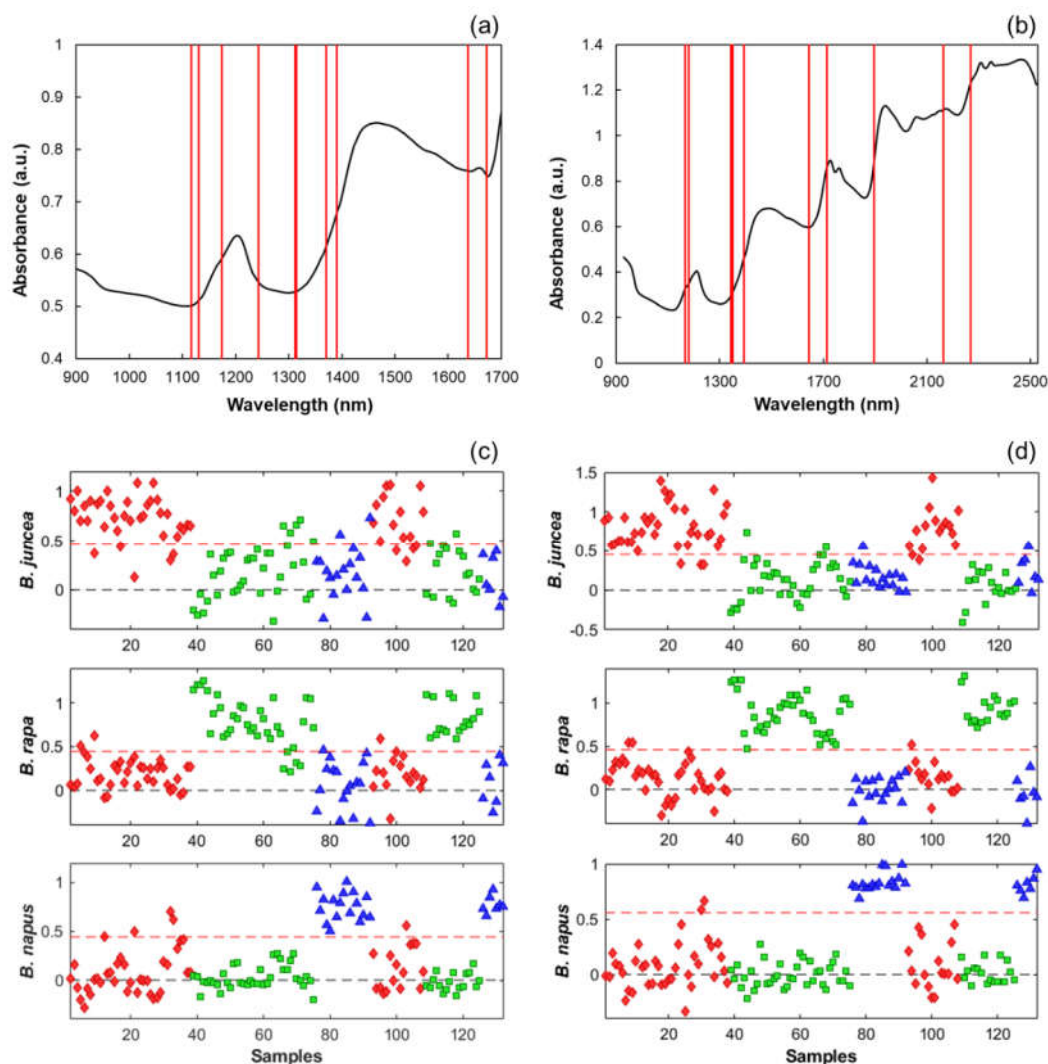


Figure 3.3. iPLS-DA model: (a and b) Important wavelengths selected and (c and d) iPLS-DA classification plots to discriminate varieties of Brassicas seeds: *B. juncea* (red diamonds), *B. rapa* (green squares) and *B. napus* (blue triangles) for NIRS and NIR-HIS data. The test set samples of each class are represented by the same symbols on the right side.

3.3.5. PREDICTION MODELS

PLSR models for oil content prediction were obtained using a maximum of 7 latent variables, which could be an indication of good models (Table 3.3), without the occurrence of overfitting or noise interference (Faber and Rajkó, 2007). Adding too many latent variables can make the calibration look unrealistically good, but unstable and inaccurate when used to predict on the external set (Cazes, 2005).

Table 3.3. Performance of PLSR model for the prediction of oil content in Brassicas seeds using NIRS and NIR-HSI spectra.

Spectral pre-treatment	LV	Calibration		Cross-validation		Prediction		
		R_C^2	RMSEC	R_{CV}^2	RMSECV	R_P^2	RMSEP	RPD
NIRS								
Full wavelength								
Raw data	7	0.809	0.813	0.711	1.006	0.602	1.064	1.79
Smoothing	7	0.792	0.848	0.709	1.007	0.594	1.061	1.79
MSC	5	0.761	0.876	0.712	0.964	0.627	1.104	1.72
Smoothing + MSC	5	0.738	0.917	0.666	1.041	0.655	1.070	1.78
1 st derivative	5	0.748	0.927	0.660	1.083	0.682	0.953	2.00
2 nd derivative	7	0.796	0.836	0.657	1.091	0.579	1.072	1.77
MSC + 1st derivative	4	0.717	0.950	0.635	1.080	0.651	1.087	1.75
MSC + 2nd derivative	7	0.775	0.868	0.571	1.216	0.573	1.109	1.71
Selected wavelength								
1 st derivative	7	0.766	0.894	0.722	0.976	0.673	0.962	1.98
NIR-HIS								
Full wavelength								
Raw data	6	0.792	0.809	0.734	0.918	0.764	0.864	2.20
Smoothing	6	0.791	0.810	0.734	0.917	0.764	0.863	2.20
MSC	4	0.819	0.804	0.779	0.890	0.591	0.949	2.00
Smoothing + MSC	4	0.818	0.806	0.778	0.891	0.590	0.951	2.00
1 st derivative	4	0.766	0.872	0.707	0.979	0.746	0.870	2.18
2 nd derivative	4	0.777	0.875	0.722	0.978	0.706	0.873	2.18
MSC + 1st derivative	3	0.791	0.852	0.734	0.962	0.689	0.888	2.14
MSC + 2nd derivative	3	0.793	0.868	0.739	0.977	0.581	0.955	1.99
Selected wavelength								
2 nd derivative	4	0.820	0.788	0.814	0.801	0.751	0.823	2.31

In general, the complete models based on the NIR-HSI spectra showed more efficient calibration and internal validation, and greater predictive capabilities in the external set compared to the NIRS. The pre-processing methods applied to the NIR-HSI spectra did not provide significant improvements in the predictive capacity of the models compared to that obtained with the original data, which presented R^2 of 0.764, RMSEP of 0.864 and RPD of 2.20, suggesting that both multiplicative and additives effects were minimal during image acquisition (Orrillo et al., 2019). Regarding the portable NIR spectrometer, the best model was obtained with the pre-processing first derivative, which resulted in the lowest RMSEP (0.953), highest determination coefficient (0.682) and highest RPD (2.00), probably due to the reduction of baseline offset effects caused by the particle size distribution of the sample.

According to RPD values, NIR spectrometer and NIR-HSI models obtained are not recommended for application in quality control analyzes (> 2.5), however, the values obtained for the NIR model preprocessed with the first derivative and for the NIR-HSI models are indicated for screening purposes (between 2.0 - 2.5) (Nicolai et al., 2007). These results are similar with those reported by other authors for NIR

spectrometers applied for prediction of oil content in rapeseed (Montes, Paul and Melchinger, 2007), *Brassica* seeds (Kaur, Sangha e Kaur, (2017); Sen et al., (2018), Oblath et al., (2016), Font, Del Río-Celestino and Haro-Bailón (2006), Petisco et al. (2010)) and corn seeds (Weinstock et al., 2006).

The inclusion of different *Brassicas* species in the construction of a model can increase its robustness, because the interspecific variability is increased (Font et al., 2003). On the other hand, it is possible that the inclusion of species with high variability increases SEP values and, consequently, reduces RPD values (Font et al., 2006). The narrow range of oil in the samples can also increase the error and reduce RPD value. In addition, these models were built using the entire spectrum and may contain non-essential information (noise, collinearity, and overfitting) that reduces predictive capacity (Kamruzzaman et al., 2015). In contrast, the selection of important wavelengths, related to the chemical information of interest, can reduce the data processing time, and provide more representative and viable models for industrial applications (Orrillo et al., 2019).

The iPLS models based on the NIR-HSI data showed good predictive capacity, with determination coefficients between 0.572 and 0.751 (Table 4 in supplementary material) and prediction errors lower than the complete models, especially in internal validation, regardless of pre-processing applied. For the NIRS models, determination coefficients between 0.537 and 0.681 were obtained and only the pre-processed models with the MSC, first derivative, or the combination of these pre-processing, had a higher predictive capability than the complete models. In general, NIRS models required a greater number of wavelengths (9 to 37) compared to NIR-HSI models (5 to 14).

The best iPLS models for the NIRS and NIR-HSI data (Table 3.2) were obtained with pre-processed spectra with first and second derivative, respectively, using 9 (933, 1040, 1261, 1290, 1433, 1512, 1603, 1659, 1692 nm) and 11 wavelengths (928, 1356, 1444, 1488, 1582, 1888, 2032, 2318-2325, 2399-2407 nm) (Fig. 3.3a and 3.3b) and presented similar or superior performance to the respective complete models, indicating that simples models are able to predict the seed oil content without significant loss of performance.

Most of the wavelengths that made the greatest contribution to the prediction of seed oil content are the same as those observed in the seed spectra and highlighted in the PCA loadings and are attributed mainly to water (1444 nm) such as aliphatic fat chain (930, 1040, 1350, 1430, 1659, 1692, 2318-2325 nm), protein (1488, 1512 and 2032

nm), cellulose or carbohydrates (1888 nm) and aromatic compounds (1131 and 1673 nm) (Osborne et al., 1993). Similar results have been reported by Weinstock et al., (2006), to predict the oil content in corn kernels and Olivos-Trujillo et al. (2015), for prediction of oil content in *B. napus*.

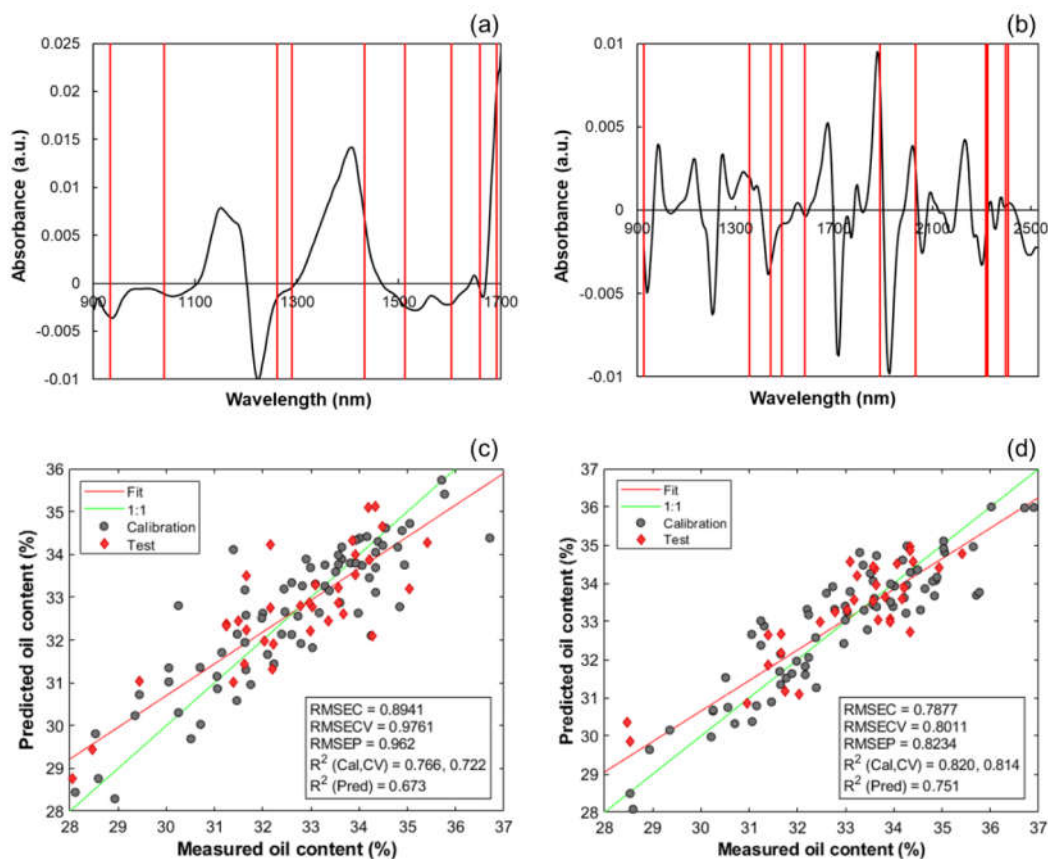


Figure 3.4. Reduced PLSR model: (a and b) Important wavelengths selected by iPLS algorithm and (c and d) Predicted versus measured oil content values in Brassicas seeds.

3.4. CONCLUSION

This work demonstrated that both a portable NIR spectrometer and NIR-HSI devices in tandem with chemometrics can be used to classify *Brassicac*s species and to provide a quick estimate of the oil content of the seeds. PLS-DA models showed a high performance to identify and discriminate species, reaching 94.9% and 100% of correct classification rate for NIRS and NIR-HSI, respectively. On the other hand, the performance of PLSR models to predict oil content in *Brassicac*s seeds was better for NIR-HSI. Moreover, the selection of important wavelengths made it possible to reduce

the size of the spectral data without impairing the performance of the models. This is important since allowing the development of cheaper multispectral NIR devices to classify the species and for the online assessment of oil content in *Brassicas* seeds during the harvest or reception of the raw material by processing industries. The results of this study may be useful for analytical purposes, screening seeds in breeding programs, quality analysis during harvest and checking the raw material before oil extraction, through the implementation of cheaper multispectral NIR equipment.

ACKNOWLEDGEMENTS

This study was financed in part by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001; and São Paulo Research Foundation (FAPESP) (project number 2019/06846-5, 2015/24351-2). Maria Lucimar Medeiros acknowledges scholarship funding from FAPESP, grant n° 2019/06846-5. Luis Jam Pier Cruz-Tirado acknowledges scholarship funding from FAPESP, grant n° 2018/02500-4, 2020/09198-1.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Maria Lucimar da Silva Medeiros: Software, Validation, Formal analysis, Investigation, Writing - original draft, Visualization. **J.P. Cruz-Tirado:** Software, Validation, Formal analysis, Investigation, Writing - original draft, Visualization. **Marcelino de Souza Netto:** Software, Validation, Formal analysis, Investigation, Writing - original draft, Visualization. **Ana Paula Badan Ribeiro:** Software, Methodology, Validation, Writing - original draft, Writing - review & editing. **Douglas Bassegio:** Investigation and Resources. **Douglas Fernandes Barbin:** Writing - review & editing, Resources, Supervision, Project administration, Funding acquisition.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

- Amigo, J.M., 2010. Practical issues of hyperspectral imaging analysis of solid dosage forms. *Anal. Bioanal. Chem.* 398, 93–109. <https://doi.org/10.1007/s00216-010-3828-z>
- AOCS. 2009. Official methods and recommended practices of the American Oil Chemists' Society. Champaign: American Oil Society.
- Badani, A.G., Snowdon, R.J., Wittkop, B., Lipsa, F.D., Baetzel, R., Horn, R., De Haro, A., Font, R., Lühs, W., Friedt, W., 2006. Colocalization of a partially dominant gene for yellow seed colour with a major QTL influencing acid detergent fibre (ADF) content in different crosses of oilseed rape (*Brassica napus*). *Genome* 49, 1499–1509. <https://doi.org/10.1139/G06-091>
- Barbin, D.F., Maciel, L.F., Bazoni, C.H.V., Ribeiro, M.S., Carvalho, R.D.S., Bispo, E.S., Miranda, M.P.S., Hirooka, E.Y., 2018. Classification and compositional characterization of different varieties of cocoa beans by near infrared spectroscopy and multivariate statistical analyses. *J. Food Sci. Technol.* 55, 2457–2466. <https://doi.org/10.1007/s13197-018-3163-5>
- Barthet, V.J., Petryk, M.W.P., Siemens, B., 2020. Rapid Nondestructive Analysis of Intact Canola Seeds Using a Handheld Near-Infrared Spectrometer. *JAOCs, J. Am. Oil Chem. Soc.* 97, 577–589. <https://doi.org/10.1002/aocs.12335>
- Bilal, M., Xiaobo, Z., Arslan, M., Tahir, H.E., Azam, M., Junjun, Z., Basheer, S., Abdullah, 2020. Rapid determination of the chemical compositions of peanut seed (*Arachis hypogaea*.) Using portable near-infrared spectroscopy. *Vib. Spectrosc.* 110, 103138. <https://doi.org/10.1016/j.vibspec.2020.103138>
- Caporaso, N., Whitworth, M.B., Fowler, M.S., Fisk, I.D., 2018. Hyperspectral imaging for non-destructive prediction of fermentation index, polyphenol content and antioxidant activity in single cocoa beans. *Food Chem.* 258, 343–351. <https://doi.org/10.1016/j.foodchem.2018.03.039>
- Carreiro Soares, S.F., Medeiros, E.P., Pasquini, C., De Lelis Morello, C., Harrop Galvão, R.K., Ugulino Araújo, M.C., 2016. Classification of individual cotton seeds with respect to variety using near-infrared hyperspectral imaging. *Anal. Methods* 8, 8498–8505. <https://doi.org/10.1039/c6ay02896a>

Cazes, J, Ewing, G. W. 2005. Ewing's analytical instrumentation handbook. Marcel Dekker, New York.

Cheng, J.H., Jin, H., Xu, Z., Zheng, F., 2017. NIR hyperspectral imaging with multivariate analysis for measurement of oil and protein contents in peanut varieties. *Anal. Methods* 9, 6148–6154. <https://doi.org/10.1039/c7ay02115a>

Faber, N.M., Rajkó, R., 2007. How to avoid over-fitting in multivariate calibration-The conventional validation approach and an alternative. *Anal. Chim. Acta* 595, 98–106. <https://doi.org/10.1016/j.aca.2007.05.030>

FAOstat. Food and agriculture organization of the United statistics division. Disponível em: <http://www.fao.org/faostat/es/#home>, acesso em 03 de novembro de 2019

Fearn, T., 1992. Near infrared spectroscopy—bridging the gap between data analysis and NIR applications. Ellis Horwood Limited, Nova Iorque, pp 61 – 66.

Ferreira, M. M. C. 2015. Quimiometria: Conceitos, Métodos e Aplicações. Campinas, SP: Editora da Unicamp.

Feng, L., Zhu, S., Liu, F., He, Y., Bao, Y., Zhang, C., 2019. Hyperspectral imaging for seed quality and safety inspection: a review. *Plant Methods* 1–25. <https://doi.org/10.1186/s13007-019-0476-y>

Font, R., del Río-Celestino, M., de Haro-Bailón, A., 2006. The use of near-infrared spectroscopy (NIRS) in the study of seed quality components in plant breeding programs. *Ind. Crops Prod.* 24, 307–313. <https://doi.org/https://doi.org/10.1016/j.indcrop.2006.06.012>

Font, R., del Río, M., Fernández, J.M., Haro, A. de, 2003. Acid Detergent Fiber Analysis in Oilseed Brassicas by Near-Infrared Spectroscopy. *J. Agric. Food Chem.* 51, 2917–2922.

Gaber, M.A.F.M., Tujillo, F.J., Mansour, M.P., Juliano, P., 2018. Improving Oil Extraction from Canola Seeds by Conventional and Advanced Methods. *Food Eng. Rev.* 10, 198–210. <https://doi.org/10.1007/s12393-018-9182-1>

Golebiowski, T., Leong, A.S., Panozzo, J.F., 2005. Near infrared reflectance spectroscopy of oil in intact canola seed (*Brassica napus* L.). II. Association between

principal components and oil content. *J. Near Infrared Spectrosc.* 13, 255–264. <https://doi.org/10.1255/jnirs.556>

Jiang, G.L., 2020. Comparison and application of non-destructive NIR evaluations of seed protein and oil content in soybean breeding. *Agronomy* 10. <https://doi.org/10.3390/agronomy10010077>

Jin, H., Ma, Y., Li, L., Cheng, J.H., 2016. Rapid and Non-destructive Determination of Oil Content of Peanut (*Arachis hypogaea* L.) Using Hyperspectral Imaging Analysis. *Food Anal. Methods* 9, 2060–2067. <https://doi.org/10.1007/s12161-015-0384-3>

Kamruzzaman, M., Makino, Y., Oshita, S., 2015. Hyperspectral imaging in tandem with multivariate analysis and image processing for non-invasive detection and visualization of pork adulteration in minced beef. *Anal. Methods* 7, 7496–7502. <https://doi.org/10.1039/c5ay01617g>

Kaur, B., Sangha, M.K., Kaur, G., 2017. Development of Near-Infrared Reflectance Spectroscopy (NIRS) Calibration Model for Estimation of Oil Content in Brassica juncea and Brassica napus. *Food Anal. Methods* 227–233. <https://doi.org/10.1007/s12161-016-0572-9>

Kaur, B., Sangha, M.K., Kaur, G., 2016. Calibration of NIRS for the Estimation of Fatty Acids in Brassica Juncea. *J. Am. Oil Chem. Soc.* 93, 673–680. <https://doi.org/10.1007/s11746-016-2802-0>

Kennard, R. W. and Stone, L. A., 1969. Computer aided design of experiments. *Technometrics* 11(1), 137-148.

Li, Y., Hussain, N., Zhang, L., Chen, X., Ali, E., Jiang, L., 2013. Correlations between Tocopherol and Fatty Acid Components in Germplasm Collections of Brassica Oilseeds. *J. Agric. Food Chem.* 61, 34–40. <https://doi.org/10.1021/jf3042837>

Long, D.S., McCallum, J.D., Young, L., Lenssen, A.W., 2012. In-stream measurement of canola (*Brassica napus* L .) seed oil concentration using in-line near infrared reflectance spectroscopy 395, 387–395. <https://doi.org/10.1255/jnirs.993>

Manley, M., 2014. Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. *Chem. Soc. Rev.* 43, 8200–8214. <https://doi.org/10.1039/c4cs00062e>

- Martínez-Arias, R., Ronquillo-López, M.G., Schechert, A., 2018. Quantification of oil content in intact sugar beet seed by near-infrared spectroscopy. *Agronomy* 8, 7–9. <https://doi.org/10.3390/agronomy8110254>
- Mendez, Jaffrey, Mendonza, L., Cruz-Tirado, J.P., Quevedo, R., Ciche, R., 2019. Trends in application of NIR and hyperspectral imaging for food authentication. *Sci. Agropecu.* 9, 153–156. <https://doi.org/10.17268/sci.agropecu.2018.01.16>
- Montes, J.M., Paul, C., Melchinger, A.E., 2007. Quality assessment of rapeseed accessions by means of near-infrared spectroscopy on combine harvesters. *Plant Breed.* <https://doi.org/10.1080/0144039X.2013.791174>
- Moreira, D.K.T., Souza, P., Gambero, A., Alves, G., 2017. Evaluation of structured lipids with behenic acid in the prevention of obesity. *Food Res. Int.* 95, 52–58. <https://doi.org/10.1016/j.foodres.2017.03.005>
- Nicolai, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I., Lammertyn, J., 2007. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biol. Technol.* 46, 99–118. <https://doi.org/10.1016/j.postharvbio.2007.06.024>
- Oblath, E.A., Isbell, T.A., Berhow, M.A., Allen, B., Archer, D., Brown, J., Gesch, R.W., Hatfield, J.L., Jabro, J.D., Kiniry, J.R., Long, D.S., 2016. Development of near-infrared spectroscopy calibrations to measure quality characteristics in intact Brassicaceae germplasm. *Ind. Crops Prod.* 89, 52–58. <https://doi.org/https://doi.org/10.1016/j.indcrop.2016.03.022>
- Olivos-Trujillo, M., Gajardo, H.A., Salvo, S., Gonzalez, A., Muñoz, C., 2015. Assessing the stability of parameters estimation and prediction accuracy in regression methods for estimating seed oil content in *Brassica napus* L. using NIR spectroscopy, in: 2015 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON). <https://doi.org/10.1109/Chilecon.2015.7400347>
- Orrillo, I., Cruz-Tirado, J.P., Cardenas, A., Oruna, M., Carnero, A., Barbin, D.F., Siche, R., 2019. Hyperspectral imaging as a powerful tool for identification of papaya seeds in black pepper. *Food Control* 101, 45–52. <https://doi.org/10.1016/j.foodcont.2019.02.036>

- Petisco, C., García-Criado, B., Vázquez-de-Aldana, B.R., de Haro, A., García-Ciudad, A., 2010. Measurement of quality parameters in intact seeds of Brassica species using visible and near-infrared spectroscopy. *Ind. Crops Prod.* 32, 139–146. <https://doi.org/https://doi.org/10.1016/j.indcrop.2010.04.003>
- Reginato, P., de Souza, C.M.A., da Silva, C.J., Rafull, L.Z.L., 2013. Desempenho agrônômico e qualidade de sementes de crambe em diferentes épocas e profundidades de semeadura. *Pesqui. Agropecu. Bras.* 48, 1410–1413. <https://doi.org/10.1590/S0100-204X2013001000013>
- Sen, R., Sharma, S., Kaur, G., Banga, S.S., 2018. Near-infrared reflectance spectroscopy calibrations for assessment of oil, phenols, glucosinolates and fatty acid content in the intact seeds of oilseed Brassica species. *J. Sci. Food Agric.* 98, 4050–4057.
- Shrestha, S., Deleuran, L.C., Gislum, R., 2016. Classification of different tomato seed cultivars by multispectral visible-near infrared spectroscopy and chemometrics. *J. Spectr. Imaging* 5, 1–9. <https://doi.org/10.1255/jsi.2016.a1>
- Silva, R.M., Zarricueta, M.L., Moreira, D.K.T., de Moraes, T.R., Rizzardi, K.F., Parisotto, T.M., de Pádua Gandra, R.L., Zuin, J.C., e Paiva Caria, C.R., Macedo, J.A., Macedo, G.A., Gambero, A., 2020. Structured Lipid Containing Behenic Acid versus Orlistat for weight loss: An experimental study in mice. *PharmaNutrition* 14, 100213. <https://doi.org/https://doi.org/10.1016/j.phanu.2020.100213>
- Su, W.H., He, H.J., Sun, D.W., 2017. Non-Destructive and rapid evaluation of staple foods quality by using spectroscopic techniques: A review. *Crit. Rev. Food Sci. Nutr.* 57, 1039–1051. <https://doi.org/10.1080/10408398.2015.1082966>
- Velasco, L., Möllers, C., Becker, H.C., 1999. Estimation of seed weight, oil content and fatty acid composition in intact single seeds of rapeseed (*Brassica napus* L.) by near-infrared reflectance spectroscopy. *Euphytica* 106, 79–85. <https://doi.org/10.1023/A:1003592115110>
- Wan, L., Sun, H., Ni, Z., Yan, G., 2018. Rapid determination of oil quantity in intact rapeseeds using near-infrared spectroscopy. *J. Food Process Eng.* 41. <https://doi.org/10.1111/jfpe.12594>

Welle, R., Zähle, K., Hildebrand, C., Kräling, K., Greten, W., 2007. Application of near infrared spectroscopy on-combine for canola breeding. *J. Near Infrared Spectrosc.* 15, 317–325.

Weinstock, B.A., Janni, J., Hagen, L., Wright, S., 2006. Prediction of oil and oleic acid concentrations in individual corn (*Zea mays* L.) kernels using near-infrared reflectance hyperspectral imaging and multivariate analysis. *Appl. Spectrosc.* 60, 9–16. <https://doi.org/10.1366/000370206775382631>

Xu, J., Nwafor, C.C., Shah, N., Zhou, Y., Zhang, C., 2019. Identification of genetic variation in *Brassica napus* seeds for tocopherol content and composition using near-infrared spectroscopy technique. *Plant Breed.* 138, 624–634.

Xu, J.L., Gowen, A.A., Sun, D.W., 2018. Time series hyperspectral chemical imaging (HCI) for investigation of spectral variations associated with water and plasticizers in casein based biopolymers. *J. Food Eng.* 218, 88–105. <https://doi.org/10.1016/j.jfoodeng.2017.09.006>

Xu, Y.D., Zhou, Y.P., Chen, J., 2017. Near-infrared spectroscopy combined with multivariate calibration to predict the yield of sesame oil produced by traditional aqueous extraction process. *J. Food Qual.* 2017. <https://doi.org/10.1155/2017/2515476>

Zhao, Y., Zhang, C., Zhu, S., Gao, P., Feng, L., He, Y., 2018. Non-destructive and rapid variety discrimination and visualization of single grape seed using near-infrared hyperspectral imaging technique and multivariate analysis. *Molecules* 23. <https://doi.org/10.3390/molecules23061352>

SUPPLEMENTARY MATERIAL

Table 1. PLS-DA model for classification the species of *Brassicas* seeds using NIRS and NIR-HSI spectra.

Pre-processing	LV	Model	Sensitivity			Specificity			Classification error			Accuracy
			<i>B. juncea</i>	<i>B. rapa</i>	<i>B. napus</i>	<i>B. juncea</i>	<i>B. rapa</i>	<i>B. napus</i>	<i>B. juncea</i>	<i>B. rapa</i>	<i>B. napus</i>	
NIRS												
Raw data	4	C	0.763	0.919	1.000	0.741	0.891	0.920	0.248	0.095	0.040	0.894
		CV	0.763	0.892	1.000	0.741	0.873	0.920	0.248	0.118	0.040	0.873
		Pred	0.750	1.000	1.000	0.917	0.957	0.909	0.167	0.022	0.045	0.873
Smoothing	4	C	0.763	0.919	1.000	0.722	0.891	0.920	0.257	0.095	0.040	0.894
		CV	0.737	0.919	1.000	0.741	0.891	0.907	0.261	0.095	0.047	0.879
		Pred	0.750	1.000	1.000	0.917	0.957	0.879	0.167	0.022	0.061	0.936
MSC	5	C	0.816	0.811	1.000	0.685	0.764	0.920	0.250	0.213	0.040	0.839
		CV	0.789	0.784	0.941	0.593	0.745	0.920	0.309	0.235	0.069	0.784
		Pred	0.938	1.000	0.857	0.583	0.565	1.000	0.240	0.217	0.071	0.784
SNV	5	C	0.816	0.811	1.000	0.704	0.727	0.920	0.240	0.231	0.040	0.847
		CV	0.789	0.784	0.941	0.611	0.709	0.920	0.300	0.254	0.069	0.784
		Pred	0.938	1.000	0.857	0.583	0.609	1.000	0.240	0.196	0.071	0.784
Smoothing + MSC	5	C	0.842	0.811	1.000	0.648	0.727	0.920	0.255	0.231	0.040	0.804
		CV	0.816	0.757	0.941	0.574	0.727	0.907	0.305	0.258	0.076	0.791
		Pred	0.875	1.000	0.857	0.667	0.609	1.000	0.229	0.196	0.071	0.832
Smoothing + SNV	5	C	0.842	0.784	1.000	0.648	0.745	0.920	0.255	0.235	0.040	0.817
		CV	0.842	0.757	0.941	0.574	0.727	0.907	0.292	0.258	0.076	0.791
		Pred	0.875	1.000	0.857	0.667	0.609	1.000	0.229	0.196	0.071	0.832
1st derivative	9	C	0.842	0.892	0.941	0.833	0.873	0.947	0.162	0.118	0.056	0.878
		CV	0.789	0.838	0.824	0.759	0.836	0.893	0.226	0.169	0.142	0.814
		Pred	0.875	0.882	0.714	0.792	0.826	1.000	0.167	0.146	0.143	0.871

(Continua)

Table 1. PLS-DA model for classification the species of *Brassicas* seeds using NIRS and NIR-HIS spectra.

(Continuação)

Pre-processing	LV	Model	Sensitivity			Specificity			Classification error			Accuracy
			<i>B. juncea</i>	<i>B. rapa</i>	<i>B. napus</i>	<i>B. juncea</i>	<i>B. rapa</i>	<i>B. napus</i>	<i>B. juncea</i>	<i>B. rapa</i>	<i>B. napus</i>	
2nd derivative	7	C	0.868	0.865	1.000	0.833	0.909	0.933	0.149	0.113	0.033	0.922
		CV	0.789	0.757	0.647	0.704	0.727	0.933	0.253	0.258	0.210	0.724
		Pred	0.750	0.824	1.000	0.792	0.783	0.909	0.229	0.197	0.045	0.800
MSC + 1st derivative	9	C	0.868	0.865	1.000	0.778	0.909	0.933	0.177	0.113	0.033	0.922
		CV	0.842	0.811	0.824	0.704	0.855	0.907	0.227	0.167	0.135	0.787
		Pred	1.000	0.824	0.857	0.750	0.783	1.000	0.125	0.197	0.071	0.822
SNV + 1st derivative	9	C	0.895	0.865	1.000	0.778	0.909	0.920	0.164	0.113	0.040	0.914
		CV	0.842	0.811	0.765	0.722	0.873	0.907	0.218	0.158	0.164	0.813
		Pred	1.000	0.824	0.857	0.708	0.739	1.000	0.146	0.219	0.071	0.822
MSC + 2nd derivative	4	C	0.868	0.757	0.941	0.704	0.600	0.907	0.214	0.322	0.076	0.828
		CV	0.842	0.703	0.765	0.648	0.545	0.893	0.255	0.376	0.171	0.785
		Pred	0.938	0.588	1.000	0.375	0.739	0.970	0.344	0.336	0.015	0.753
SNV + 2nd derivative	6	C	0.789	0.838	0.941	0.722	0.764	0.920	0.244	0.199	0.069	0.746
		CV	0.789	0.757	0.647	0.667	0.618	0.893	0.272	0.313	0.230	0.700
		Pred	0.938	0.882	1.000	0.542	0.652	0.970	0.260	0.233	0.015	0.770
NIR-HSI												
Raw data	9	C	0.947	0.973	1.000	0.870	0.982	0.973	0.091	0.023	0.013	0.959
		CV	0.842	0.892	1.000	0.815	0.927	0.947	0.172	0.090	0.027	0.884
		Pred	0.938	1.000	1.000	0.958	1.000	0.970	0.052	0.000	0.015	0.954
Smoothing	8	C	0.939	0.929	1.000	0.814	0.940	0.960	0.124	0.066	0.020	0.951
		CV	0.879	0.857	1.000	0.780	0.920	0.960	0.171	0.111	0.020	0.906
		Pred	1.000	1.000	1.000	0.958	0.957	1.000	0.021	0.022	0.000	1.000
MSC	3	C	0.789	0.892	1.000	0.796	0.927	0.973	0.207	0.090	0.013	0.912
		CV	0.763	0.865	1.000	0.796	0.891	0.947	0.220	0.122	0.027	0.879
		Pred	0.875	1.000	0.857	0.792	0.913	0.970	0.167	0.043	0.087	0.919

(Continua)

Table 1. PLS-DA model for classification the species of *Brassicas* seeds using NIRS and NIR-HIS spectra.

(Conclusão)

Pre-processing	LV	Model	Sensitivity			Specificity			Classification error			Accuracy
			<i>B. juncea</i>	<i>B. rapa</i>	<i>B. napus</i>	<i>B. juncea</i>	<i>B. rapa</i>	<i>B. napus</i>	<i>B. juncea</i>	<i>B. rapa</i>	<i>B. napus</i>	
SNV	3	C	0.789	0.892	1.000	0.796	0.927	0.947	0.207	0.090	0.027	0.905
		CV	0.763	0.865	1.000	0.778	0.891	0.947	0.230	0.122	0.027	0.885
		Pred	0.875	1.000	0.857	0.792	0.913	0.970	0.167	0.043	0.087	0.918
Smoothing + MSC	3	C	0.789	0.892	1.000	0.796	0.927	0.947	0.207	0.090	0.027	0.905
		CV	0.763	0.865	1.000	0.778	0.891	0.947	0.230	0.122	0.027	0.885
		Pred	0.875	1.000	0.857	0.792	0.913	0.970	0.167	0.043	0.087	0.918
Smoothing + SNV	3	C	0.789	0.892	1.000	0.796	0.927	0.947	0.207	0.090	0.027	0.905
		CV	0.763	0.865	1.000	0.778	0.891	0.947	0.230	0.122	0.027	0.885
		Pred	0.875	1.000	0.857	0.792	0.913	0.970	0.167	0.043	0.087	0.918
1st derivative	9	C	0.947	0.973	1.000	0.889	0.945	0.987	0.082	0.041	0.007	0.986
		CV	0.816	0.946	0.941	0.833	0.982	0.987	0.175	0.036	0.036	0.914
		Pred	0.875	1.000	1.000	0.917	0.957	1.000	0.104	0.022	0.000	0.967
2nd derivative	9	C	0.895	0.973	1.000	0.889	0.964	0.987	0.108	0.032	0.007	0.978
		CV	0.816	0.946	1.000	0.815	0.909	0.987	0.185	0.072	0.007	0.919
		Pred	1.000	1.000	1.000	1.000	0.957	1.000	0.000	0.022	0.000	0.982
MSC + 1st derivative	11	C	0.895	0.946	1.000	0.907	0.982	0.987	0.099	0.036	0.007	0.972
		CV	0.842	0.946	1.000	0.870	0.909	0.960	0.144	0.072	0.020	0.916
		Pred	1.000	1.000	1.000	0.958	1.000	1.000	0.021	0.000	0.000	1.000
SNV + 1st derivative	10	C	0.895	0.946	1.000	0.889	0.964	0.987	0.108	0.045	0.007	0.972
		CV	0.842	0.919	1.000	0.796	0.945	0.960	0.181	0.068	0.020	0.915
		Pred	1.000	1.000	1.000	0.875	1.000	1.000	0.063	0.000	0.000	1.000
MSC + 2nd derivative	4	C	0.816	0.865	0.941	0.852	0.927	0.973	0.166	0.104	0.043	0.907
		CV	0.816	0.865	0.941	0.833	0.927	0.973	0.175	0.104	0.043	0.894
		Pred	0.750	1.000	1.000	0.833	0.913	0.970	0.208	0.043	0.015	0.918
SNV + 2nd derivative	3	C	0.816	0.865	0.941	0.833	0.909	0.987	0.175	0.113	0.036	0.905
		CV	0.789	0.838	0.941	0.815	0.890	0.987	0.198	0.136	0.036	0.873
		Pred	0.813	1.000	0.857	0.708	0.913	0.970	0.240	0.043	0.087	0.918

Table 2. iPLS-DA model for classification the species of *Brassicac*s seeds using NIRS and NIR-HSI spectra.

Pre-processing	Spectral range	LV	Model	Sensitivity			Specificity			Classification error			Accuracy
				B. juncea	B. rapa	B. napus	B. juncea	B. rapa	B. napus	B. juncea	B. rapa	B. napus	
				NIRS									
Raw data	1250.93-1257.91, 1286.84, 1664.94-1667.87	3	C	0.763	0.919	0.941	0.815	0.855	0.907	0.211	0.113	0.076	0.866
			CV	0.763	0.919	0.941	0.815	0.836	0.920	0.211	0.122	0.069	0.873
			Pred	0.750	1.000	1.000	0.917	0.826	0.879	0.167	0.087	0.061	0.868
Smoothing	1117.29, 1131.9, 1174.11, 1243.92, 1312.05-1315.47, 1371.81, 1391.79, 1637.33, 1673.73	8	C	0.868	0.865	1.000	0.870	0.964	0.947	0.131	0.086	0.027	0.960
			CV	0.868	0.892	1.000	0.852	0.945	0.933	0.135	0.081	0.033	0.928
			Pred	0.813	1.000	1.000	0.917	0.913	0.970	0.135	0.043	0.015	0.949
MSC	949.411, 1102.61, 1109.96, 1147.64, 1240.41, 1343.79, 1446.49, 1700.83	7	C	0.868	0.838	1.000	0.759	0.800	0.920	0.186	0.181	0.040	0.891
			CV	0.895	0.838	0.941	0.741	0.800	0.920	0.182	0.181	0.069	0.832
			Pred	0.938	0.941	0.571	0.583	0.522	1.000	0.240	0.269	0.214	0.790
SNV	1082.93, 1128.25, 1188.44, 1232.21, 1329.1	4	C	0.921	0.838	0.941	0.574	0.673	0.880	0.252	0.245	0.089	0.801
			CV	0.921	0.838	0.941	0.593	0.673	0.893	0.243	0.245	0.083	0.793
			Pred	1.000	1.000	0.714	0.458	0.478	1.000	0.271	0.261	0.143	0.820
Smoothing + MSC	953.265, 1090.32, 1117.29, 1192.02-1195.59, 1236.9, 1337.02	4	C	0.921	0.865	1.000	0.630	0.709	0.907	0.225	0.213	0.047	0.842
			CV	0.947	0.838	1.000	0.630	0.727	0.907	0.212	0.217	0.047	0.815
			Pred	0.938	0.941	0.857	0.583	0.565	1.000	0.240	0.247	0.071	0.898
Smoothing + SNV	1106.28, 1113.62, 1165.71, 1254.42, 1272.99, 1312.05-1315.47, 1329.1, 1343.79, 1371.81-1375.15, 1381.82, 1682.48, 1692.16	9	C	0.921	0.919	1.000	0.870	0.909	0.933	0.104	0.086	0.033	0.943
			CV	0.921	0.919	0.941	0.852	0.891	0.933	0.114	0.095	0.063	0.893
			Pred	0.750	0.941	1.000	0.750	0.739	1.000	0.250	0.160	0.000	0.865
1st derivative	1177.7-1181.28, 1269.52, 1337.02, 1353.91, 1360.64, 1371.81, 1443.24	6	C	0.816	0.865	1.000	0.870	0.891	0.947	0.157	0.122	0.027	0.913
			CV	0.816	0.865	1.000	0.833	0.873	0.920	0.175	0.131	0.040	0.907
			Pred	0.813	0.941	0.857	0.750	0.870	0.970	0.219	0.095	0.087	0.902

(Continua)

Table 2. iPLS-DA model for classification the species of *Brassicas* seeds using NIRS and NIR-HSI spectra.

(Continuação)

Pre-processing	Spectral range	LV	Model	Sensitivity			Specificity			Classification error			Accuracy
				<i>B. juncea</i>	<i>B. rapa</i>	<i>B. napus</i>	<i>B. juncea</i>	<i>B. rapa</i>	<i>B. napus</i>	<i>B. juncea</i>	<i>B. rapa</i>	<i>B. napus</i>	
2nd derivative	985.208, 1120.95, 1128.25, 1329.1, 1337.02, 1347.17, 1388.47, 1567.69, 1643.28	5	C	0.842	0.919	0.941	0.870	0.891	0.893	0.144	0.095	0.083	0.896
			CV	0.842	0.919	0.941	0.852	0.891	0.893	0.153	0.095	0.083	0.874
			Pred	0.625	0.882	1.000	0.750	0.652	0.879	0.313	0.233	0.061	0.839
MSC + 1st derivative	1158.5-1162.11, 1221.62, 1315.47, 1463.74-1470.17, 1503.14, 1516.83	3	C	0.921	0.838	0.941	0.704	0.782	0.880	0.188	0.190	0.089	0.831
			CV	0.921	0.838	0.941	0.704	0.800	0.893	0.188	0.181	0.083	0.819
			Pred	1.000	1.000	1.000	0.500	0.696	1.000	0.250	0.152	0.000	0.850
SNV + 1st derivative	930.077, 1147.64, 1214.54, 1232.21, 1254.42, 1269.52, 1375.15, 1466.96	4	C	0.921	0.784	0.882	0.685	0.836	0.907	0.197	0.190	0.105	0.851
			CV	0.921	0.811	0.882	0.685	0.855	0.893	0.197	0.167	0.112	0.805
			Pred	0.938	0.941	0.857	0.417	0.565	1.000	0.323	0.247	0.071	0.781
MSC + 2nd derivative	930.077, 1016.85, 1128.25, 1214.54, 1312.05, 1347.17, 1659.05	5	C	0.868	0.892	1.000	0.722	0.800	0.920	0.205	0.154	0.040	0.863
			CV	0.895	0.892	1.000	0.759	0.800	0.907	0.173	0.154	0.047	0.855
			Pred	1.000	0.941	1.000	0.792	0.522	0.970	0.104	0.269	0.015	0.870
SNV + 2nd derivative	909.335, 1040.69, 1124.6, 1243.92, 1272.99, 1340.41, 1350.54, 1637.33, 1659.05, 1670.8	8	C	0.947	0.919	1.000	0.815	0.891	0.920	0.119	0.095	0.040	0.921
			CV	0.974	0.892	0.941	0.796	0.891	0.920	0.115	0.109	0.069	0.888
			Pred	0.813	0.941	1.000	0.792	0.826	0.970	0.198	0.116	0.015	0.867
HSI-NIR													
Raw data	991.43, 1155.19, 1381.49, 1394.05, 1406.61-1412.89, 1913.81	6	C	0.816	0.919	1.000	0.870	0.964	0.947	0.157	0.059	0.027	0.936
			CV	0.816	0.919	1.000	0.907	0.964	0.947	0.138	0.059	0.027	0.915
			Pred	1.000	1.000	1.000	0.958	1.000	1.000	0.021	0.000	0.000	1.000
Smoothing	940.99, 1111.12-1117.42, 1155.19, 1387.77, 1419.16, 1513.27, 2088.53, 2461.91, 2486.75	6	C	0.909	0.905	1.000	0.780	0.900	0.973	0.156	0.098	0.013	0.930
			CV	0.879	0.905	1.000	0.780	0.900	0.973	0.171	0.098	0.013	0.906
			Pred	1.000	1.000	1.000	0.833	0.957	1.000	0.083	0.022	0.000	1.000

(Continua)

Table 2. iPLS-DA model for classification the species of *Brassicas* seeds using NIRS and NIR-HSI spectra.

(Continuação)

Pre-processing	Spectral range	LV	Model	Sensitivity			Specificity			Classification error			Accuracy
				<i>B. juncea</i>	<i>B. rapa</i>	<i>B. napus</i>	<i>B. juncea</i>	<i>B. rapa</i>	<i>B. napus</i>	<i>B. juncea</i>	<i>B. rapa</i>	<i>B. napus</i>	
MSC	1167.77, 1180.36, 1343.81-1350.09, 1394.05, 1644.87, 1713.74, 1895.07, 2163.32, 2269.18	9	C	0.921	1.000	1.000	0.926	0.964	0.973	0.077	0.018	0.013	0.979
			CV	0.921	0.973	1.000	0.926	0.964	0.973	0.077	0.032	0.013	0.964
			Pred	0.875	1.000	1.000	0.958	0.957	1.000	0.083	0.022	0.000	0.982
SNV	959.9-966.21, 1167.77, 1375.21, 1657.4, 2076.06-2082.3, 2188.24	7	C	0.921	0.892	1.000	0.889	0.945	0.987	0.095	0.081	0.007	0.935
			CV	0.895	0.892	0.941	0.907	0.964	0.973	0.099	0.072	0.043	0.929
			Pred	0.938	1.000	0.857	0.750	0.957	1.000	0.156	0.022	0.071	0.935
Smoothing + MSC	1035.54, 1167.77, 1199.23-1205.52, 1218.1, 1375.21, 1657.4, 1726.26, 1745.03, 1920.05, 2076.06, 2262.95	10	C	0.868	0.190	1.000	0.907	0.964	0.987	0.112	0.059	0.007	0.971
			CV	0.816	0.892	1.000	0.870	0.964	0.987	0.157	0.072	0.007	0.943
			Pred	0.938	1.000	1.000	0.833	0.957	1.000	0.115	0.022	0.000	0.982
Smoothing + SNV	959.9-966.21, 1167.77, 1375.21, 1657.4, 2076.06-2082.3, 2188.24	7	C	0.921	0.892	1.000	0.889	0.945	0.987	0.095	0.081	0.007	0.935
			CV	0.895	0.892	0.941	0.907	0.964	0.973	0.099	0.072	0.043	0.929
			Pred	0.938	1.000	0.857	0.750	0.957	1.000	0.156	0.022	0.071	0.935
1st derivative	1331.25, 1488.18, 1644.87, 1895.07, 2306.51, 2337.61	4	C	0.842	0.946	1.000	0.815	0.927	0.960	0.172	0.063	0.020	0.920
			CV	0.868	0.919	1.000	0.815	0.945	0.947	0.158	0.068	0.027	0.908
			Pred	0.875	1.000	1.000	1.000	0.957	0.970	0.063	0.022	0.015	0.967
2nd derivative	928.37, 1079.64, 1111.12, 1324.97, 1431.72, 1688.7, 1763.79, 1826.33, 2219.38, 2318.95, 2350.05	6	C	0.842	0.946	1.000	0.926	0.927	0.973	0.116	0.063	0.013	0.942
			CV	0.816	0.946	1.000	0.889	0.927	0.973	0.148	0.063	0.013	0.935
			Pred	0.875	1.000	1.000	0.958	0.870	1.000	0.083	0.065	0.000	0.947
MSC + 1st derivative	959.9-966.21, 1318.68, 1331.25, 1832.58-1838.83, 1851.33, 2206.92, 2262.95, 2300.29	8	C	0.921	0.973	1.000	0.944	0.964	0.960	0.068	0.032	0.020	0.957
			CV	0.921	0.946	1.000	0.944	0.945	0.947	0.067	0.054	0.027	0.957
			Pred	0.938	1.000	1.000	0.917	0.913	1.000	0.073	0.043	0.000	0.982

(Continua)

Table 2. iPLS-DA model for classification the species of *Brassicac*s seeds using NIRS and NIR-HSI spectra.

(Conclusão)

Pre-processing	Spectral range	LV	Model	Sensitivity			Specificity			Classification error			Accuracy
				<i>B. juncea</i>	<i>B. rapa</i>	<i>B. napus</i>	<i>B. juncea</i>	<i>B. rapa</i>	<i>B. napus</i>	<i>B. juncea</i>	<i>B. rapa</i>	<i>B. napus</i>	
SNV + 1st derivative	953.6, 1218.1, 1287.26, 1318.68-1324.97, 1676.18, 2138.4, 2244.28, 2281.62, 2300.29, 2343.83	11	C	0.947	1.000	1.000	0.963	0.945	0.987	0.045	0.027	0.007	0.963
			CV	0.947	1.000	1.000	0.963	0.927	0.987	0.045	0.036	0.007	0.956
			Pred	1.000	0.941	1.000	0.917	0.957	1.000	0.042	0.051	0.000	0.966
MSC + 2nd derivative	934.68, 1293.55, 1318.68, 2169.55, 2269.18	5	C	0.921	0.919	1.000	0.907	0.982	0.973	0.086	0.050	0.013	0.957
			CV	0.921	0.946	1.000	0.907	0.982	0.973	0.086	0.036	0.013	0.949
			Pred	0.875	1.000	0.857	0.958	0.957	1.000	0.083	0.022	0.071	0.936
SNV + 2nd derivative	934.68, 1293.55, 1318.68, 2169.55, 2269.18	5	C	1.000	0.919	1.000	0.907	0.964	1.000	0.046	0.059	0.000	0.962
			CV	1.000	0.919	1.000	0.889	0.964	1.000	0.056	0.059	0.000	0.962
			Pred	0.938	1.000	1.000	0.917	1.000	0.970	0.073	0.000	0.015	0.936

Table 3. PLSR model for the prediction of oil content in *Brassicac*s seeds using NIRS and NIR-HSI spectra.

Pre-processing	LV	R_C^2	R_{CV}^2	R_P^2	RMSEC	RMSECV	RMSEP	% Relative error	RPD
NIRS									
Raw spectra	7	0.809	0.711	0.602	0.813	1.006	1.064	3.263	1.786
Smoothing	7	0.792	0.709	0.594	0.848	1.007	1.061	3.253	1.792
MSC	5	0.761	0.712	0.627	0.876	0.964	1.104	3.385	1.722
SNV	5	0.780	0.677	0.672	0.895	1.023	1.046	3.206	1.818
Smoothing + MSC	5	0.738	0.666	0.655	0.917	1.041	1.070	3.280	1.777
Smoothing + SNV	6	0.753	0.677	0.650	0.890	1.024	1.067	3.273	1.781
1st derivative	5	0.748	0.660	0.682	0.927	1.083	0.953	2.921	1.995
2nd derivative	7	0.796	0.657	0.579	0.836	1.091	1.072	3.286	1.774

(Continua)

Table 3. PLSR model for the prediction of oil content in *Brassicas seeds* using NIRS and NIR-HSI spectra.

(Conclusão)

Pre-processing	LV	R_C^2	R_{CV}^2	R_p^2	RMSEC	RMSECV	RMSEP	% Relative error	RPD
MSC + 1st derivative	4	0.717	0.635	0.651	0.950	1.080	1.087	3.334	1.748
SNV + 1st derivative	4	0.721	0.659	0.611	0.944	1.046	1.131	3.469	1.680
MSC + 2nd derivative	7	0.775	0.571	0.573	0.868	1.216	1.109	3.401	1.713
SNV + 2nd derivative	7	0.772	0.541	0.586	0.888	1.292	1.054	3.231	1.804
NIR-HSI									
Raw spectra	6	0.792	0.734	0.764	0.809	0.918	0.864	2.631	2.200
Smoothing	6	0.791	0.734	0.764	0.810	0.917	0.863	2.629	2.202
MSC	4	0.819	0.779	0.591	0.804	0.890	0.949	2.891	2.002
SNV	4	0.819	0.779	0.591	0.804	0.890	0.949	2.891	2.002
Smoothing + MSC	4	0.818	0.778	0.590	0.806	0.891	0.951	2.896	1.998
Smoothing + SNV	4	0.818	0.778	0.590	0.806	0.891	0.951	2.896	1.998
1st derivative	4	0.766	0.707	0.746	0.872	0.979	0.870	2.650	2.184
2nd derivative	4	0.777	0.722	0.706	0.875	0.978	0.873	2.661	2.175
MSC + 1st derivative	3	0.791	0.734	0.689	0.852	0.962	0.888	2.704	2.141
SNV + 1st derivative	3	0.791	0.734	0.690	0.852	0.962	0.884	2.693	2.149
MSC + 2nd derivative	3	0.793	0.739	0.581	0.868	0.977	0.955	2.908	1.990
SNV + 2nd derivative	3	0.794	0.744	0.608	0.862	0.962	0.899	2.738	2.114

Table 4. PLSR model for the prediction of oil content in Brassicas seeds using NIRS and NIR-HSI spectra.

Pre-processing	Spectral range	LV	R_C^2	R_{CV}^2	R_P^2	RMSEC	RMSECV	RMSEP	Relative error (%)	RPD
NIRS										
Raw spectra	981.391-985.208, 1207.45, 1214.54, 1228.68, 1236.9, 1247.43, 1279.93, 1286.84, 1301.76, 1318.88, 1343.79-1350.54, 1360.64, 1396.21, 1409.43-1412.72, 1457.28-1463.74, 1470.17-1482.99, 1499.97, 1512.63, 1523.12, 1538.77, 1549.14, 1631.37-1634.35, 1643.28, 1664.94, 1692.16, 1697.95-1700.84	9	0.815	0.696	0.537	0.799	1.032	1.205	3.696	1.577
Smoothing	1203.9-1211, 1360.64, 1409.43, 1446.49, 1460.51, 1649.21, 1664.94, 1697.95	6	0.799	0.740	0.549	0.834	0.951	1.141	3.500	1.665
MSC	901.525, 1005.49, 1016.85, 1028.17, 1036.94, 1044.44, 1055.66, 1063.12, 1075.51, 1102.61-1109.96, 1131.9-1136.75, 1144.02, 1207.45, 1378.49, 1612.37, 1618.39, 1631.37, 1652.17-1655.12	6	0.706	0.634	0.617	0.973	1.089	1.116	3.423	1.703
SNV	901.525, 945.553, 957.114, 981.391, 1024.4, 1033.18-1040.69, 1055.66, 1063.12, 1207.45, 1225.15, 1232.21, 1283.39, 1290.29-1293.74, 1301.76, 1308.62-1312.05, 1333.63-1337.02, 1350.54, 1357.28, 1381.82, 1391.79, 1555.34, 1618.39, 1652.17, 1673.73, 1679.57-1686.36, 1697.95-1700.83	6	0.783	0.725	0.681	0.833	0.943	1.028	3.153	1.848
Smoothing + MSC	901.525, 1120.95-1124.6, 1203.9-1207.45, 1221.62, 1532.52, 1538.77, 1596.23, 1618.39, 1637.33, 1655.12, 1700.83	7	0.738	0.650	0.634	0.919	1.068	1.089	3.340	1.745

(Continua)

Table 4. PLSR model for the prediction of oil content in Brassicas seeds using NIRS and NIR-HSI spectra.

(Continuação)

Pre-processing	Spectral range	LV	R_C^2	R_{CV}^2	R_P^2	RMSEC	RMSECV	RMSEP	Relative error (%)	RPD
Smoothing + SNV	901.525, 1199.15-1207.45, 1236.9-1240.41, 1247.43, 1574.86, 1655.12-1659.05, 1664.94-1667.87, 1682.48-1686.36, 1700.83	7	0.766	0.698	0.613	0.867	0.989	1.120	3.435	1.697
1st derivative	933.953, 1040.69, 1261.4, 1290.29, 1433.47, 1512.63, 1603.3, 1659.05, 1692.16	7	0.766	0.722	0.673	0.894	0.976	0.962	2.950	1.976
2nd derivative	941.691, 977.569-985.208, 997.901, 1033.18, 1040.69-1048.19, 1059.4-1063.12, 1071.8, 1086.63-1094.01, 1165.71, 1174.11, 1269.52-1272.99, 1290.29-1293.74, 1360.64, 1416.01, 1463.74, 1538.77-1541.89, 1552.24, 1612.37, 1621.39, 1673.73, 1682.48, 1689.26	4	0.841	0.783	0.630	0.738	0.864	1.002	3.071	1.898
MSC + 1st derivative	1109.96, 1211, 1283.39, 1371.81, 1436.73, 1446.49-1449.73, 1558.44, 1634.35, 1646.25-1649.21, 1682.48, 1692.16-1695.06	5	0.740	0.689	0.614	0.911	0.997	1.124	3.447	1.691
SNV + 1st derivative	1109.96, 1211, 1371.81, 1446.49, 1558.44, 1634.35-1637.33, 1692.16, 1700.83	7	0.735	0.694	0.606	0.921	0.990	1.119	3.431	1.699
MSC + 2nd derivative	901.525, 933.953, 949.411-953.265, 977.569-981.391, 989.021, 1001.7, 1016.85, 1033.18, 1071.8, 1090.32, 1151.27, 1162.11-1170.51, 1269.52, 1381.82, 1460.51, 1493.62-1496.8, 1512.63, 1519.98-1523.12, 1535.65, 1541.89, 1549.14, 1564.61, 1580.99, 1612.37, 1618.39, 1664.94, 1682.48	5	0.776	0.675	0.558	0.866	1.044	1.136	3.483	1.673

(Continua)

Table 4. PLSR model for the prediction of oil content in Brassicas seeds using NIRS and NIR-HSI spectra.

(Continuação)

Pre-processing	Spectral range	LV	R_C^2	R_{CV}^2	R_P^2	RMSEC	RMSECV	RMSEP	Relative error (%)	RPD
SNV + 2nd derivative	921.018, 930.077, 957.114, 966.078, 977.569, 1001.7, 1016.85, 1028.17, 1040.69, 1144.02, 1151.27, 1158.5, 1177.7, 1203.9-1207.45, 1218.09, 1236.9, 1243.92, 1254.42, 1261.4, 1269.52, 1312.05, 1318.88, 1371.81, 1381.82-1385.15, 1402.83, 1446.49, 1476.59, 1493.62, 1519.98, 1538.77, 1558.44, 1593.19, 1618.39, 1643.28, 1682.48	6	0.790	0.648	0.583	0.853	1.122	1.068	3.274	1.780
NIR-HSI										
Raw spectra	947.29, 959.9, 1142.6, 1293.55, 1318.68, 1657.4, 1726.26, 1745.03-1751.28, 1870.08, 1901.31, 2094.77, 2113.47	6	0.818	0.762	0.704	0.756	0.868	0.974	2.968	1.950
Smoothing	928.37, 940.99, 1682.44, 1726.26, 1738.77, 1888.82, 2094.77, 2312.73-2318.95, 2424.64-2430.85, 2443.28-2449.49	5	0.803	0.760	0.751	0.786	0.871	0.886	2.698	2.145
MSC	1180.36, 1726.26, 1770.05-1776.3, 1807.57, 1882.57, 2088.53, 2188.24, 2393.57, 2406	5	0.825	0.779	0.606	0.790	0.893	0.936	2.851	2.030
SNV	1619.82, 1726.26, 1738.77, 1770.05, 2069.83, 2107.24	5	0.820	0.799	0.572	0.802	0.846	0.973	2.964	1.953
Smoothing + MSC	928.37, 1657.4, 1732.51, 1782.56, 2076.06, 2468.12	5	0.827	0.795	0.643	0.785	0.856	0.888	2.706	2.139
Smoothing + SNV	928.37, 1657.4, 1732.51-1738.77, 1782.56, 2076.06	5	0.827	0.800	0.581	0.790	0.846	0.979	2.983	1.940
1st derivative	928.37, 1098.53, 1613.56, 1701.22, 1763.79, 1845.08-1857.58, 1945.03, 2057.36-2063.59, 2101, 2343.83, 2461.91, 2524.01	7	0.785	0.731	0.717	0.837	0.937	0.936	2.851	2.030

(Continua)

Table 4. PLSR model for the prediction of oil content in Brassicas seeds using NIRS and NIR-HSI spectra.*(Conclusão)*

Pre-processing	Spectral range	LV	R_C^2	R_{CV}^2	R_P^2	RMSEC	RMSECV	RMSEP	Relative error (%)	RPD
2nd derivative	928.37, 1356.37, 1444.27, 1488.18, 1582.23, 1888.82, 2032.41, 2318.95-2325.17, 2399.78-2407	4	0.820	0.814	0.751	0.788	0.801	0.823	2.508	2.307
MSC + 1st derivative	1136.31, 1701.22, 1838.83, 2063.59, 2219.39	4	0.811	0.791	0.696	0.809	0.851	0.896	2.728	2.122
SNV + 1st derivative	1136.31, 1701.22, 1838.83, 2063.59, 2219.38	4	0.811	0.792	0.698	0.801	0.850	0.890	2.710	2.136
MSC + 2nd derivative	1205.52, 1280.98-1287.26, 1494.46-1500.73, 1582.23, 1669.92, 1770.05, 1807.57, 1882.57, 2206.92, 2399.78-2406, 2449.49	5	0.823	0.791	0.633	0.803	0.874	0.895	2.725	2.124
SNV + 2nd derivative	1205.52, 1293.55, 1544.62, 1720, 1820.08, 2406, 2424.64-2430.85, 2449.49	4	0.818	0.788	0.615	0.810	0.877	0.915	2.787	2.076

Capítulo 4 – Conclusão Geral

CONCLUSÃO GERAL

Os modelos PLS-DA obtidos para o espectrômetro NIR portátil e para o sistema NIR-HSI apresentaram bons resultados e demonstraram ser adequados para selecionar espécies *Brassicac*s. Os modelos PLSR, especialmente os baseados nos espectros NIR-HSI, podem ser aplicados para auxiliar em análises de controle de qualidade, fornecendo uma estimativa do teor de óleo presente nas sementes, independente da espécie. A seleção de comprimentos de onda importantes aumentou a capacidade preditiva dos modelos, viabilizando a aplicação de sensores NIR multiespectrais com menor custo em aplicações industriais online.

Capítulo 5 - Referências

REFERÊNCIAS

- Akoh, C.C., 2017. Food Lipids: Chemistry, Nutrition and Biotechnology.
- Alves, A.Q., da Silva, V.A.J., Góes, A.J.S., Silva, M.S., de Oliveira, G.G., Bastos, I.V.G.A., de Castro Neto, A.G., Alves, A.J., 2019. The Fatty Acid Composition of Vegetable Oils and Their Potential Use in Wound Care. *Adv. Skin Wound Care* 32.
- Amigo, J.M., 2010. Practical issues of hyperspectral imaging analysis of solid dosage forms. *Anal. Bioanal. Chem.* 398, 93–109. <https://doi.org/10.1007/s00216-010-3828-z>
- Andersen, C.M., Bro, R., 2010. Variable selection in regression—a tutorial. *J. Chemom.* 24, 728–737. <https://doi.org/https://doi.org/10.1002/cem.1360>
- André Weinstock, B., Janni, J., Hagen, L., Wright, S., 2006. Prediction of oil and oleic acid concentrations in individual corn (*Zea mays* L.) kernels using near-infrared reflectance hyperspectral imaging and multivariate analysis. *Appl. Spectrosc.* 60, 9–16. <https://doi.org/10.1366/000370206775382631>
- AOCS. 2009. Official methods and recommended practices of the American Oil Chemists' Society. Champaign: American Oil Society.
- Araújo, M.C.U., Saldanha, T.C.B., Galvão, R.K.H., Yoneyama, T., Chame, H.C., Visani, V., 2001. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemom. Intell. Lab. Syst.* 57, 65–73. [https://doi.org/https://doi.org/10.1016/S0169-7439\(01\)00119-8](https://doi.org/https://doi.org/10.1016/S0169-7439(01)00119-8)
- Badani, A.G., Snowdon, R.J., Wittkop, B., Lipsa, F.D., Baetzel, R., Horn, R., De Haro, A., Font, R., Lühs, W., Friedt, W., 2006. Colocalization of a partially dominant gene for yellow seed colour with a major QTL influencing acid detergent fibre (ADF) content in different crosses of oilseed rape (*Brassica napus*). *Genome* 49, 1499–1509. <https://doi.org/10.1139/G06-091>
- Badaró, A.T., Garcia-Martin, J.F., López-Barrera, M. del C., Barbin, D.F., Alvarez-Mateos, P., 2020. Determination of pectin content in orange peels by near infrared hyperspectral imaging. *Food Chem.* 323, 126861. <https://doi.org/https://doi.org/10.1016/j.foodchem.2020.126861>

- Badaró, A.T., Morimitsu, F.L., Ferreira, A.R., Clerici, M.T.P.S., Fernandes Barbin, D., 2019. Identification of fiber added to semolina by near infrared (NIR) spectral techniques. *Food Chem.* 289, 195–203. <https://doi.org/10.1016/j.foodchem.2019.03.057>
- Barbin, D.F., Kaminishikawahara, C.M., Soares, A.L., Mizubuti, I.Y., Grespan, M., Shimokomaki, M., Hirooka, E.Y., 2015. Prediction of chicken quality attributes by near infrared spectroscopy. *Food Chem.* 168, 554–560. <https://doi.org/10.1016/j.foodchem.2014.07.101>
- Barbin, D.F., Maciel, L.F., Bazoni, C.H.V., Ribeiro, M.S., Carvalho, R.D.S., Bispo, E.S., Miranda, M.P.S., Hirooka, E.Y., 2018. Classification and compositional characterization of different varieties of cocoa beans by near infrared spectroscopy and multivariate statistical analyses. *J. Food Sci. Technol.* 55, 2457–2466. <https://doi.org/10.1007/s13197-018-3163-5>
- Barthet, V.J., Daun, J.K., 2011. Seed Morphology, Composition, and Quality, Canola: Chemistry, Production, Processing, and Utilization. AOCS Press. <https://doi.org/10.1016/B978-0-9818936-5-5.50009-7>
- Barthet, V.J., Petryk, M.W.P., Siemens, B., 2020. Rapid Nondestructive Analysis of Intact Canola Seeds Using a Handheld Near-Infrared Spectrometer. *JAOCS, J. Am. Oil Chem. Soc.* 97, 577–589. <https://doi.org/10.1002/aocs.12335>
- Bassegio, D., Zanotto, M.D., Santos, R.F., Werncke, I., Dias, P.P., Olivo, M., 2016. Oilseed crop crambe as a source of renewable energy in Brazil. *Renew. Sustain. Energy Rev.* 66, 311–321. <https://doi.org/10.1016/j.rser.2016.08.010>
- Berrueta, L.A., Alonso-Salces, R.M., Héberger, K., 2007. Supervised pattern recognition in food analysis. *J. Chromatogr. A* 1158, 196–214. <https://doi.org/https://doi.org/10.1016/j.chroma.2007.05.024>
- Bilal, M., Xiaobo, Z., Arslan, M., Tahir, H.E., Azam, M., Junjun, Z., Basheer, S., Abdullah, 2020. Rapid determination of the chemical compositions of peanut seed (*Arachis hypogaea*.) Using portable near-infrared spectroscopy. *Vib. Spectrosc.* 110, 103138. <https://doi.org/10.1016/j.vibspec.2020.103138>
- Blanco, M., Villarroya, I., 2002. NIR spectroscopy: A rapid-response analytical tool. *TrAC - Trends Anal. Chem.* 21, 240–250. [https://doi.org/10.1016/S0165-9936\(02\)00404-1](https://doi.org/10.1016/S0165-9936(02)00404-1)

- Bocianowski, J., Mikołajczyk, K., Bartkowiak-Broda, I., 2012. Determination of fatty acid composition in seed oil of rapeseed (*Brassica napus* L.) by mutated alleles of the FAD3 desaturase genes. *J. Appl. Genet.* 53, 27–30. <https://doi.org/10.1007/s13353-011-0062-0>
- Byrdwell, W.C., Neff, W.E., 2001. Autoxidation products of normal and genetically modified canola oil varieties determined using liquid chromatography with mass spectrometric detection. *J. Chromatogr. A* 905, 85–102. [https://doi.org/10.1016/S0021-9673\(00\)00958-4](https://doi.org/10.1016/S0021-9673(00)00958-4)
- Caporaso, N., Whitworth, M.B., Fowler, M.S., Fisk, I.D., 2018a. Hyperspectral imaging for non-destructive prediction of fermentation index, polyphenol content and antioxidant activity in single cocoa beans. *Food Chem.* 258, 343–351. <https://doi.org/10.1016/j.foodchem.2018.03.039>
- Caporaso, N., Whitworth, M.B., Grebby, S., Fisk, I.D., 2018b. Rapid prediction of single green coffee bean moisture and lipid content by hyperspectral imaging. *J. Food Eng.* 227, 18–29. <https://doi.org/10.1016/j.jfoodeng.2018.01.009>
- Carreiro Soares, S.F., Medeiros, E.P., Pasquini, C., De Lelis Morello, C., Harrop Galvão, R.K., Ugulino Araújo, M.C., 2016. Classification of individual cotton seeds with respect to variety using near-infrared hyperspectral imaging. *Anal. Methods* 8, 8498–8505. <https://doi.org/10.1039/c6ay02896a>
- Cartea, E., Haro-bail, A. De, Padilla, G., Obreg, S., Rio-celestino, M., Ord, A., 2019. Seed Oil Quality of *Brassica napus* and *Brassica rapa* Germplasm from Northwestern Spain. *Foods* 8.
- Cazes, J, Ewing, G. W. 2005. Ewing's analytical instrumentation handbook. Marcel Dekker, New York.
- Chen, X., Parr, C., Utterback, P., Parsons, C.M., 2015. Nutritional evaluation of canola meals produced from new varieties of canola seeds for poultry. *Poult. Sci.* 94, 984–991. <https://doi.org/10.3382/ps/pev043>
- Cheng, J.H., Jin, H., Xu, Z., Zheng, F., 2017. NIR hyperspectral imaging with multivariate analysis for measurement of oil and protein contents in peanut varieties. *Anal. Methods* 9, 6148–6154. <https://doi.org/10.1039/c7ay02115a>

Corgozino, C.N.C., 2009. Utilização de técnicas espectrofluorimétricas e quimiométricas na análise de biodiesel e de suas misturas com óleo diesel. Tese doutorado em Ciências Exatas. Univ. Fed. Minas Gerais 164.

Cozzolino, D., 2016. Authentication of Cereals and Cereal Products, *Advances in Food Authenticity Testing*. Elsevier Ltd. <https://doi.org/10.1016/b978-0-08-100220-9.00016-3>

Cruz-Tirado, J.P., Fernández Pierna, J.A., Rogez, H., Barbin, D.F., Baeten, V., 2020. Authentication of cocoa (*Theobroma cacao*) bean hybrids by NIR-hyperspectral imaging and chemometrics. *Food Control* 118, 107445. <https://doi.org/https://doi.org/10.1016/j.foodcont.2020.107445>

Diederichsen, A., McVetty, P.B.E., 2011. Botany and Plant Breeding, in: *Canola: Chemistry, Production, Processing, and Utilization*. AOCS Press, pp. 29–56. <https://doi.org/10.1016/B978-0-9818936-5-5.50006-1>

Downey, R.K., Craig, N.M., Youngs, C.G., 1967. Breeding Rapeseed for Oil and Meal Quality. *J. Am. Oil Chem. Soc.* 46, 121–123.

Downey, R.K. Brassica oilseed breeding-achievements and opportunities. *Plant Breed. Abstr.* 60, 1165–1170, 1990.

Duarte, L.M., 2015. Modelagem multivariada para determinação de propriedades físico-químicas de petróleo e para quantificação de edulcorantes artificiais em adoçantes de mesa. Diss. Mestr. em Química. Univ. Fed. Juiz Fora. 120.

Faber, N.M., Rajkó, R., 2007. How to avoid over-fitting in multivariate calibration-The conventional validation approach and an alternative. *Anal. Chim. Acta* 595, 98–106. <https://doi.org/10.1016/j.aca.2007.05.030>

FAOstat. Food and agriculture organization of the United statistics division. Disponível em: <http://www.fao.org/faostat/es/#home>, acesso em 03 de novembro de 2019

Fearn, T., 1992. Near infrared spectroscopy-bridging the gap between data analysis and NIR applications. Ellis Horwood Limited, Nova Iorque, pp 61 – 66.

Ferreira, M. M. C. 2015. Quimiometria: Conceitos, Métodos e Aplicações. Campinas, SP: Editora da Unicamp

- Feng, L., Zhu, S., Liu, F., He, Y., Bao, Y., Zhang, C., 2019. Hyperspectral imaging for seed quality and safety inspection: a review. *Plant Methods* 1–25. <https://doi.org/10.1186/s13007-019-0476-y>
- Font, R., del Río-Celestino, M., de Haro-Bailón, A., 2006. The use of near-infrared spectroscopy (NIRS) in the study of seed quality components in plant breeding programs. *Ind. Crops Prod.* 24, 307–313. <https://doi.org/https://doi.org/10.1016/j.indcrop.2006.06.012>
- Font, R., del Río, M., Fernández, J.M., Haro, A. de, 2003. Acid Detergent Fiber Analysis in Oilseed Brassicas by Near-Infrared Spectroscopy. *J. Agric. Food Chem.* 51, 2917–2922.
- Forina, M., Lanteri, S., Oliveros, M.C.C., Millan, C.P., 2004. Selection of useful predictors in multivariate calibration. *Anal. Bioanal. Chem.* 380, 397–418. <https://doi.org/10.1007/s00216-004-2768-x>
- Gaber, M.A.F.M., Tujillo, F.J., Mansour, M.P., Juliano, P., 2018. Improving Oil Extraction from Canola Seeds by Conventional and Advanced Methods. *Food Eng. Rev.* 10, 198–210. <https://doi.org/10.1007/s12393-018-9182-1>
- Golebiowski, T., Leong, A.S., Panozzo, J.F., 2005. Near infrared reflectance spectroscopy of oil in intact canola seed (*Brassica napus* L.). II. Association between principal components and oil content. *J. Near Infrared Spectrosc.* 13, 255–264. <https://doi.org/10.1255/jnirs.556>
- Gupta, S.K. Brassicas. In *Breeding Oilseed Crops for Sustainable Production: Opportunities and Constraints*; Academic press: New York, NY, USA, 2016.
- Iftikhar, R., Wang, X., Rahman, H., 2018. Broadening the genetic base of *Brassica napus* canola by interspecific crosses with different variants of *B. oleracea*. *Euphytica* 214, 133. <https://doi.org/10.1007/s10681-018-2213-4>
- Jabeur, H., Zribi, A., Makni, J., Rebai, A., Abdelhedi, R., Bouaziz, M., 2014. Detection of Chemlali Extra-Virgin Olive Oil Adulteration Mixed with Soybean Oil, Corn Oil, and Sunflower Oil by Using GC and HPLC. *J. Agric. Food Chem.* 62, 4893–4904. <https://doi.org/10.1021/jf500571n>

- Jiang, G.L., 2020. Comparison and application of non-destructive NIR evaluations of seed protein and oil content in soybean breeding. *Agronomy* 10. <https://doi.org/10.3390/agronomy10010077>
- Jin, H., Li, L., Cheng, J., 2015. Rapid and Non-destructive Determination of Moisture Content of Peanut Kernels Using Hyperspectral Imaging Technique. *Food Anal. Methods* 8, 2524–2532. <https://doi.org/10.1007/s12161-015-0147-1>
- Jin, H., Ma, Y., Li, L., Cheng, J.H., 2016. Rapid and Non-destructive Determination of Oil Content of Peanut (*Arachis hypogaea* L.) Using Hyperspectral Imaging Analysis. *Food Anal. Methods* 9, 2060–2067. <https://doi.org/10.1007/s12161-015-0384-3>
- Kamruzzaman, M., Makino, Y., Oshita, S., 2015. Hyperspectral imaging in tandem with multivariate analysis and image processing for non-invasive detection and visualization of pork adulteration in minced beef. *Anal. Methods* 7, 7496–7502. <https://doi.org/10.1039/c5ay01617g>
- Kaufmann, K.C., Favero, F. de F., de Vasconcelos, M.A.M., Godoy, H.T., Sampaio, K.A., Barbin, D.F., 2019. Portable NIR Spectrometer for Prediction of Palm Oil Acidity. *J. Food Sci.* 84, 406–411. <https://doi.org/10.1111/1750-3841.14467>
- Kaur, B., Sangha, M.K., Kaur, G., 2017. Development of Near-Infrared Reflectance Spectroscopy (NIRS) Calibration Model for Estimation of Oil Content in Brassica juncea and Brassica napus. *Food Anal. Methods* 227–233. <https://doi.org/10.1007/s12161-016-0572-9>
- Kaur, B., Sangha, M.K., Kaur, G., 2016. Calibration of NIRS for the Estimation of Fatty Acids in Brassica Juncea. *J. Am. Oil Chem. Soc.* 93, 673–680. <https://doi.org/10.1007/s11746-016-2802-0>
- Kemsley, E.K., Defernez, M., Marini, F., 2019. Multivariate statistics: Considerations and confidences in food authenticity problems. *Food Control* 105, 102–112. <https://doi.org/10.1016/j.foodcont.2019.05.021>
- Kinney, A.J., 2017. Oils. *Encycl. Appl. Plant Sci.* 2, 87–92. <https://doi.org/10.1016/B978-0-12-394807-6.00224-0>

Kojima, M., Arishima, T., Shimizu, R., Kohno, M., Kida, H., Hirotsuka, M., Ikeda, I., 2013. The Consumption of a Structured Triacylglycerol Containing Behenic and Oleic Acids Increases Fecal Fat Excretion in Humans 1001, 997–1001.

Kojima, M., Tachibana, N., Yamahira, T., Seino, S., Izumisawa, A., Sagi, N., Arishima, T., Kohno, M., Takamatsu, K., Hirotsuka, M., Ikeda, I., 2010. Structured triacylglycerol containing behenic and oleic acids suppresses triacylglycerol absorption and prevents obesity in rats 2–7.

Leardi, R., 2000. Application of genetic algorithm-PLS for feature selection in spectral data sets. *J. Chemom.* 14, 643–655. [https://doi.org/10.1002/1099-128X\(200009/12\)14:5/6<643::AID-CEM621>3.0.CO;2-E](https://doi.org/10.1002/1099-128X(200009/12)14:5/6<643::AID-CEM621>3.0.CO;2-E)

Leardi, R., González, A.L., 1998. Genetic algorithms applied to feature selection in PLS regression: how and when to use them. *Chemom. Intell. Lab. Syst.* 41, 195–207. [https://doi.org/https://doi.org/10.1016/S0169-7439\(98\)00051-3](https://doi.org/https://doi.org/10.1016/S0169-7439(98)00051-3)

Leardi, R., Nørgaard, L., 2004. Sequential application of backward interval partial least squares and genetic algorithms for the selection of relevant spectral regions. *J. Chemom.* 18, 486–497. <https://doi.org/https://doi.org/10.1002/cem.893>

Leardi, R., Seasholtz, M.B., Pell, R.J., 2002. Variable selection for multivariate calibration using a genetic algorithm: prediction of additive concentrations in polymer films from Fourier transform-infrared spectral data. *Anal. Chim. Acta* 461, 189–200. [https://doi.org/https://doi.org/10.1016/S0003-2670\(02\)00272-6](https://doi.org/https://doi.org/10.1016/S0003-2670(02)00272-6)

Li, Y., Hussain, N., Zhang, L., Chen, X., Ali, E., Jiang, L., 2013. Correlations between Tocopherol and Fatty Acid Components in Germplasm Collections of Brassica Oilseeds. *J. Agric. Food Chem.* 61, 34–40. <https://doi.org/10.1021/jf3042837>

Long, D.S., McCallum, J.D., Young, L., Lenssen, A.W., 2012. In-stream measurement of canola (*Brassica napus* L .) seed oil - concentration using - in-line near infrared reflectance s - pectroscopy 395, 387–395. <https://doi.org/10.1255/jnirs.993>

Manley, M., 2014. Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. *Chem. Soc. Rev.* 43, 8200–8214. <https://doi.org/10.1039/c4cs00062e>

- Martínez-Arias, R., Ronquillo-López, M.G., Schechert, A., 2018. Quantification of oil content in intact sugar beet seed by near-infrared spectroscopy. *Agronomy* 8, 7–9. <https://doi.org/10.3390/agronomy8110254>
- Martínez Gila, D.M., Cano Marchal, P., Gámez García, J., Gómez Ortega, J., 2015. On-line system based on hyperspectral information to estimate acidity, moisture and peroxides in olive oil samples. *Comput. Electron. Agric.* 116, 1–7. <https://doi.org/10.1016/j.compag.2015.06.002>
- Mawlong, I., Sujith Kumar, M.S., Kandpal, B.K., Premi, O.P., Gurung, B., Singh, D., 2017. Meal and oil quality among genotypes of indian mustard (*Brassica juncea*) varies under recommended dose of nitrogen fertilizer. *Appl. Ecol. Environ. Res.* 15, 1427–1445. https://doi.org/10.15666/aeer/1504_14271445
- Mendez, Jaffrey, Mendonza, L., Cruz-Tirado, J.P., Quevedo, R., Ciche, R., 2019. Trends in application of NIR and hyperspectral imaging for food authentication. *Sci. Agropecu.* 9, 153–156. <https://doi.org/10.17268/sci.agropecu.2018.01.16>
- Mendez, Jeffrey, Mendoza, L., Cruz-Tirado, J.P., Quevedo, R., Siche, R., 2019. Trends in application of NIR and hyperspectral imaging for food authentication. *Sci. Agropecu.* 10, 143–161. <https://doi.org/10.17268/sci.agropecu.2019.01.16>
- Montes, J.M., Paul, C., Melchinger, A.E., 2007. Quality assessment of rapeseed accessions by means of near-infrared spectroscopy on combine harvesters. *Plant Breed.* <https://doi.org/10.1080/0144039X.2013.791174>
- Moreira, D.K.T., Souza, P., Gambero, A., Alves, G., 2017. Evaluation of structured lipids with behenic acid in the prevention of obesity. *Food Res. Int.* 95, 52–58. <https://doi.org/10.1016/j.foodres.2017.03.005>
- Neff, W.E., Mounts, T.L., Rinsch, W.M., 1997. Oxidative Stability as Affected by Triacylglycerol Composition and Structure of Purified Canola Oil Triacylglycerols from Genetically Modified Normal and High Stearic and Lauric Acid Canola Varieties. *LWT - Food Sci. Technol.* 30, 793–799. <https://doi.org/https://doi.org/10.1006/fstl.1997.0274>
- Nicolaï, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I., Lammertyn, J., 2007. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biol. Technol.* 46, 99–118. <https://doi.org/10.1016/j.postharvbio.2007.06.024>

- No, D.S., Zhao, T., Kim, B.H., Choi, H.D., Kim, I.H., 2013. Enrichment of erucic acid from crambe oil in a recirculated packed bed reactor via lipase-catalyzed ethanolysis. *J. Mol. Catal. B Enzym.* 87, 6–10. <https://doi.org/10.1016/j.molcatb.2012.10.001>
- Nolasco, I.M.P., 2019. Caracterização e Identificação de carne de aves por métodos não-destrutivos. Universidade Estadual de Campinas.
- Nørgaard, L., Saudland, A., Wagner, J., Nielsen, J.P., Munck, L., Engelsen, S.B., 2000. Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. *Appl. Spectrosc.* 54, 413–419. <https://doi.org/10.1366/0003702001949500>
- Oblath, E.A., Isbell, T.A., Berhow, M.A., Allen, B., Archer, D., Brown, J., Gesch, R.W., Hatfield, J.L., Jabro, J.D., Kiniry, J.R., Long, D.S., 2016. Development of near-infrared spectroscopy calibrations to measure quality characteristics in intact Brassicaceae germplasm. *Ind. Crops Prod.* 89, 52–58. <https://doi.org/https://doi.org/10.1016/j.indcrop.2016.03.022>
- Oliveira, M.M., Cruz-Tirado, J.P., Barbin, D.F., 2019. Nontargeted Analytical Methods as a Powerful Tool for the Authentication of Spices and Herbs: A Review. *Compr. Rev. Food Sci. Food Saf.* 18, 670–689. <https://doi.org/https://doi.org/10.1111/1541-4337.12436>
- Oliveira, M.M., Cruz-Tirado, J.P., Roque, J. V, Teófilo, R.F., Barbin, D.F., 2020. Portable near-infrared spectroscopy for rapid authentication of adulterated paprika powder. *J. Food Compos. Anal.* 87, 103403. <https://doi.org/https://doi.org/10.1016/j.jfca.2019.103403>
- Olivos-Trujillo, M., Gajardo, H.A., Salvo, S., Gonzalez, A., Muñoz, C., 2015. Assessing the stability of parameters estimation and prediction accuracy in regression methods for estimating seed oil content in Brassica napus L. using NIR spectroscopy, in: 2015 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON). <https://doi.org/10.1109/Chilecon.2015.7400347>
- Oram, R.N., Kirk, J.T.O., Veness, P.E., Hurlstone, C.J., Edlington, J.P., Halsall, D.M., 2005. Breeding Indian mustard [Brassica juncea (L.) Czern.] for cold-pressed, edible oil productiona review. *Aust. J. Agric. Res.* 56, 581–596.

- Orrillo, I., Cruz-Tirado, J.P., Cardenas, A., Oruna, M., Carnero, A., Barbin, D.F., Siche, R., 2019. Hyperspectral imaging as a powerful tool for identification of papaya seeds in black pepper. *Food Control* 101, 45–52. <https://doi.org/10.1016/j.foodcont.2019.02.036>
- Osborne, S.D., Jordan, R.B., Künnemeyera, R., 1997. Method of Wavelength Selection for Partial Least Squares. *Analyst* 122, 1531–1537. <https://doi.org/10.1039/A703235H>
- Otto, M., 2016. Pattern Recognition and Classification, in: *Chemometrics (Statistics and Computer Application in Analytical Chemistry)*. pp. 135–211. <https://doi.org/10.1002/9783527699377.ch5>
- Pasquini, C., 2018. Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. *Anal. Chim. Acta* 1026, 8–36. <https://doi.org/https://doi.org/10.1016/j.aca.2018.04.004>
- Petisco, C., García-Criado, B., Vázquez-de-Aldana, B.R., de Haro, A., García-Ciudad, A., 2010. Measurement of quality parameters in intact seeds of Brassica species using visible and near-infrared spectroscopy. *Ind. Crops Prod.* 32, 139–146. <https://doi.org/https://doi.org/10.1016/j.indcrop.2010.04.003>
- Rady, A., Adedeji, A., 2018. Assessing different processed meats for adulterants using visible-near-infrared spectroscopy. *Meat Sci.* 136, 59–67. <https://doi.org/10.1016/j.meatsci.2017.10.014>
- Rani, T., Yadav, R.C., Yadav, N.R., Rani, A., Singh, D., 2013. Genetic transformation in oilseed brassicas -A review. *Indian J. Agric. Sci.* 83 (4), 367–73.
- RAVAZZI, C.G., 2019. Identificação e quantificação de adulterantes de Whey protein concentrado empregando espectroscopia no infravermelho próximo e resolução multivariada de curvas, in: *Dissertação de Mestrado (Química Analítica)*. Universidade Estadual de Campinas. p. 75.
- Reginato, P., de Souza, C.M.A., da Silva, C.J., Rafull, L.Z.L., 2013. Desempenho agrônômico e qualidade de sementes de crambe em diferentes épocas e profundidades de semeadura. *Pesqui. Agropecu. Bras.* 48, 1410–1413. <https://doi.org/10.1590/S0100-204X2013001000013>
- Ribeiro, M.D.M.M., Ming, C.C., Lopes, T.I.B., Grimaldi, R., Marsaioli, A.J., Gonçalves, L.A.G., 2017. Synthesis of structured lipids containing behenic acid from

fully hydrogenated *Crambe abyssinica* oil by enzymatic interesterification. *J. Food Sci. Technol.* 54, 1146–1157. <https://doi.org/10.1007/s13197-017-2540-9>

Saeys, W., Nguyen Do Trong, N., Van Beers, R., Nicolaï, B.M., 2019. Multivariate calibration of spectroscopic sensors for postharvest quality evaluation: A review. *Postharvest Biol. Technol.* 158, 110981. <https://doi.org/10.1016/j.postharvbio.2019.110981>

Sarıkaya, İ., Bilgen, S., Akbaş, H., 2016. The production of fuel from, and the thermal stability of, sunflower oil, corn oil, and canola oil. *Energy Sources, Part A Recover. Util. Environ. Eff.* 38, 3514–3520. <https://doi.org/10.1080/15567036.2016.1180330>

Sen, R., Sharma, S., Kaur, G., Banga, S.S., 2018. Near-infrared reflectance spectroscopy calibrations for assessment of oil, phenols, glucosinolates and fatty acid content in the intact seeds of oilseed Brassica species. *J. Sci. Food Agric.* 98, 4050–4057.

Shrestha, S., Deleuran, L.C., Gislum, R., 2016. Classification of different tomato seed cultivars by multispectral visible-near infrared spectroscopy and chemometrics. *J. Spectr. Imaging* 5, 1–9. <https://doi.org/10.1255/jsi.2016.a1>

Silva, R.M., Kono, D., Moreira, T., Luzzi, M., Rabelo, C., Alves, J., Alves, G., Gambero, A., 2019. The postprandial inflammatory response is attenuated by a dietary structured lipid containing behenic acid. *J. Funct. Foods* 58, 350–354. <https://doi.org/10.1016/j.jff.2019.05.013>

Silva, R.M., Zarricueta, M.L., Moreira, D.K.T., de Morais, T.R., Rizzardi, K.F., Parisotto, T.M., de Pádua Gandra, R.L., Zuin, J.C., e Paiva Caria, C.R., Macedo, J.A., Macedo, G.A., Gambero, A., 2020. Structured Lipid Containing Behenic Acid versus Orlistat for weight loss: An experimental study in mice. *PharmaNutrition* 14, 100213. <https://doi.org/https://doi.org/10.1016/j.phanu.2020.100213>

Singh, J., Sharma, P.C., Sharma, S.K., Rai, M., 2014. Assessing the effect of salinity on the oil quality parameters of Indian mustard (*Brassica juncea* L. Czern & Coss) using Fourier Transform Near-Infrared Reflectance (FT-NIR) spectroscopy. *Grasas y Aceites* 65, 1–9. <https://doi.org/10.3989/gya.063413>

- Su, W.H., He, H.J., Sun, D.W., 2017. Non-Destructive and rapid evaluation of staple foods quality by using spectroscopic techniques: A review. *Crit. Rev. Food Sci. Nutr.* 57, 1039–1051. <https://doi.org/10.1080/10408398.2015.1082966>
- Sun, J., Lu, X., Mao, H., Jin, X., Wu, X., 2017. A Method for Rapid Identification of Rice Origin by Hyperspectral Imaging Technology. *J. Food Process Eng.* 40. <https://doi.org/10.1111/jfpe.12297>
- Thiyam-Holländer, U., Eskin, N.A.M., Matthäus, B., 2012. *Canola and Rapeseed: Production, Processing, Food Quality, and Nutrition*, CRC Press: Taylor & Francis Group. <https://doi.org/10.1201/b13023>
- Tibola, C.S., de MEdeiros, E.P., Simeone, M.L.F., de Oliveira, M.A., 2018. *Espectroscopia no Infravermelho Próximo para Avalia Indicadores de Qualidade Tecnológica e Contaminantes em Grãos*, Brasília, DF: Embrapa Trigo.
- Unger, E.H., 2011. Processing, in: *Canola: Chemistry, Production, Processing, and Utilization*. pp. 163–188. <https://doi.org/10.1016/B978-0-9818936-5-5.50010-3>
- Vecchio, A.J. High-laurate canola; *INFORM* 1996, 7, 230–231, 235–236, 239–240, 242–243.
- Velasco, L., Möllers, C., Becker, H.C., 1999. Estimation of seed weight, oil content and fatty acid composition in intact single seeds of rapeseed (*Brassica napus* L.) by near-infrared reflectance spectroscopy. *Euphytica* 106, 79–85. <https://doi.org/10.1023/A:1003592115110>
- Wan, L., Sun, H., Ni, Z., Yan, G., 2018. Rapid determination of oil quantity in intact rapeseeds using near-infrared spectroscopy. *J. Food Process Eng.* 41. <https://doi.org/10.1111/jfpe.12594>
- Webb, D.R., Sanders, R.A., 1991. Caprenin 1. Digestion, Absorption, and Rearrangement in Thoracic Duct-Cannulated Rats 10.
- Welle, R., Zähle, K., Hildebrand, C., Kräling, K., Greten, W., 2007. Application of near infrared spectroscopy on-combine for canola breeding. *J. Near Infrared Spectrosc.* 15, 317–325.
- Wold, S., Sjöström, M., 1977. SIMCA: A Method for Analyzing Chemical Data in Terms of Similarity and Analogy, in: *Chemometrics: Theory and Application*, ACS

Symposium Series. AMERICAN CHEMICAL SOCIETY, pp. 12–243.
<https://doi.org/doi:10.1021/bk-1977-0052.ch012>

Woyengo, T.A., Jha, R., Beltranena, E., Zijlstra, R.T., 2016. In vitro digestion and fermentation characteristics of canola co-products simulate their digestion in the pig intestine. *Animal* 10, 911–918. <https://doi.org/10.1017/S1751731115002566>

Xiaobo, Z., Jiewen, Z., Povey, M.J.W., Holmes, M., Hanpin, M., 2010. Variables selection methods in near-infrared spectroscopy. *Anal. Chim. Acta* 667, 14–32. <https://doi.org/https://doi.org/10.1016/j.aca.2010.03.048>

Xie, C., Wang, Q., He, Y., 2014. Identification of different varieties of sesame oil using near-infrared hyperspectral imaging and chemometrics algorithms. *PLoS One* 9. <https://doi.org/10.1371/journal.pone.0098522>

Xu, J., Nwafor, C.C., Shah, N., Zhou, Y., Zhang, C., 2019. Identification of genetic variation in *Brassica napus* seeds for tocopherol content and composition using near-infrared spectroscopy technique. *Plant Breed.* 138, 624–634.

Xu, J.L., Gowen, A.A., Sun, D.W., 2018. Time series hyperspectral chemical imaging (HCI) for investigation of spectral variations associated with water and plasticizers in casein based biopolymers. *J. Food Eng.* 218, 88–105. <https://doi.org/10.1016/j.jfoodeng.2017.09.006>

Xu, Y.D., Zhou, Y.P., Chen, J., 2017. Near-infrared spectroscopy combined with multivariate calibration to predict the yield of sesame oil produced by traditional aqueous extraction process. *J. Food Qual.* 2017. <https://doi.org/10.1155/2017/2515476>

Xu, Z., Ye, Z., Li, Y., Li, J., Liu, Y., 2020. Comparative study of the oxidation stability of high oleic oils and palm oil during thermal treatment. *J. Oleo Sci.* 69, 573–584. <https://doi.org/10.5650/jos.ess19307>

Yu, H.-D., Yun, Y.-H., Zhang, W., Chen, H., Liu, D., Zhong, Q., Chen, Wenxue, Chen, Weijun, 2020. Three-step hybrid strategy towards efficiently selecting variables in multivariate calibration of near-infrared spectra. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* 224, 117376. <https://doi.org/https://doi.org/10.1016/j.saa.2019.117376>

Zhao, Y., Zhang, C., Zhu, S., Gao, P., Feng, L., He, Y., 2018. Non-destructive and rapid variety discrimination and visualization of single grape seed using near-infrared

hyperspectral imaging technique and multivariate analysis. *Molecules* 23. <https://doi.org/10.3390/molecules23061352>

Zhou, Q., Tang, H., Jia, X., Zheng, C., Huang, F., Zhang, M., 2018. Distribution of glucosinolate and pungent odors in rapeseed oils from raw and microwaved seeds. *Int. J. Food Prop.* 21, 2296–2308. <https://doi.org/10.1080/10942912.2018.1514632>