

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

PATRICIA CASARIN DE LIMA

**DETECÇÃO DE ADULTERAÇÃO EM FARINHAS DE TEFF USANDO  
ESPECTROSCOPIAS NIR E EDXRF EM CONJUNTO COM QUIMIOMETRIA**

DISSERTAÇÃO DE MESTRADO

CAMPO MOURÃO

2022

PATRICIA CASARIN DE LIMA

**DETECÇÃO DE ADULTERAÇÃO EM FARINHAS DE TEFF USANDO  
ESPECTROSCOPIAS NIR E EDXRF EM CONJUNTO COM QUIMIOMETRIA**

**Detection of adulteration in teff flours using NIR e EDXRF spectroscopies with  
chemometrics**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Tecnologia de Alimentos da Universidade Tecnológica Federal do Paraná como requisito parcial para a obtenção do título de Mestre em Tecnologia de Alimentos.

Orientador: Prof. Dr. Evandro Bona  
Coorientador: Prof. Dr. Fabio Luiz Melquiades

CAMPO MOURÃO

2022



[4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Esta licença permite que outros remixem, adaptem e criem a partir do seu trabalho para fins não comerciais, desde que atribuam o devido crédito e que licenciem as novas criações sob termos idênticos. Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.



**Ministério da Educação  
Universidade Tecnológica Federal do Paraná  
Campus Medianeira**



PATRICIA CASARIN DE LIMA

**DETECÇÃO DE ADULTERAÇÃO EM FARINHAS DE TEFF USANDO ESPECTROSCOPIAS NIR E EDXRF EM CONJUNTO COM QUIMIOMETRIA**

Trabalho de pesquisa de mestrado apresentado como requisito para obtenção do título de Mestre Em Tecnologia De Alimentos da Universidade Tecnológica Federal do Paraná (UTFPR). Área de concentração: Tecnologia De Alimentos.

Data de aprovação: 30 de Maio de 2022

Evandro Bona, - Universidade Tecnológica Federal do Paraná

Dra. Angela Claudia Rodrigues, Doutorado - Universidade Tecnológica Federal do Paraná

Dr. Frederico Luis Felipe Soares, Doutorado - Universidade Federal do Paraná (Ufpr)

Documento gerado pelo Sistema Acadêmico da UTFPR a partir dos dados da Ata de Defesa em 30/05/2022.

## AGRADECIMENTOS

Inicio meus agradecimentos a Deus e N. Sra. Aparecida, pois acredito nessa força superior que, com certeza, levantou-me nos piores momentos e me lembrou do meu amor pela docência em todas as vezes que pensei em desistir.

Minha gratidão à minha família: minha mãe, pai e irmão, que fizeram o possível e o impossível para me animar e me erguer e, principalmente, continuaram me dando todo o carinho e amor do mundo. Amo vocês mais do que tudo.

Sou grata aos profissionais incríveis que são os professores do PPGTA-UTFPR. Obrigada pela dedicação de vocês à ciência e aos alunos, por todo o conhecimento compartilhado com tanta qualidade e entusiasmo. Vocês são meus modelos.

Agradeço ao meu orientador Prof. Dr. Evandro Bona que me acompanha desde o IC, sempre me incentivou e fez com que eu me apaixonasse pela Quimiometria. Obrigada por toda a paciência e compreensão durante esse caminho. Obrigada por todo o ensinamento nesses 7 anos de parceria. Obrigada por não me deixar desistir e acreditar no meu potencial. A você, toda a minha admiração.

Agradeço também ao meu coorientador Prof. Dr. Fábio Luiz Melquiades por toda ajuda, contribuição e conhecimento valioso repassado. Muito obrigada pela recepção em Londrina, foi uma experiência gratificante.

Claro que não posso deixar de agradecer à família que a gente ganha ao longo da vida: os amigos. É difícil mencionar todos, mas mesmo se seu nome não está aqui, tenha a certeza de que está no meu coração, você fez a diferença na minha vida. Meu obrigada aos meus amigos de graduação Sid, Gi e Flávia, que mesmo longe nunca me deixaram sozinha. Obrigada às minhas amigas de PPGTA Fran, Ju e Amarílis por todo o apoio e ajuda. Obrigada Talitinha pelas melhores risadas, pelo melhor SLACA, pelos melhores broches e, claro, por todo o incentivo, o nosso encontro é pra vida toda. E eu não poderia deixar de expressar minha gratidão pela minha gêmea de alma Luana. Você me acolheu com tanto carinho nos piores dias, jamais poderei agradecer o suficiente tudo o que você fez por mim. Te guardo no coração para sempre.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Por fim, agradeço a UTFPR, Laboratório Camulti-CM, Departamento de Física da UEL pela contribuição instrumental e ao CNPq pela contribuição financeira.

## RESUMO

O cereal *Eragrostis tef* (Zucc.), conhecido por teff, ganhou atenção mundial por ser livre de glúten e rico em minerais, principalmente ferro. Devido a isso, a farinha de teff é atrativa às fraudes destacando-se a adição de farinhas de cereais mais baratos. Nesse contexto, o estudo de técnicas rápidas e baratas como a espectroscopia NIR (infravermelho próximo) e EDXRF (fluorescência de raios X por dispersão de energia) é importante para a identificação desse tipo de adulteração. Nesse trabalho, foi avaliada a capacidade dos métodos NIR e EDXRF na identificação de farinhas de teff adulteradas com farinhas de arroz, trigo integral, aveia e centeio. Foram utilizadas duas safras de teff (A e B) adulteradas seguindo um planejamento experimental {5,4} simplex-lattice, variando a proporção de teff de 65% a 100% e adulterantes de 0 a 35%. Os espectros NIR foram pré-tratados com correção do espalhamento multiplicativo (MSC), 2ª derivada ou combinação (MSC+2ªd) e analisados por componentes principais (PCA). Os espectros EDXRF de 15 kV e 50 kV foram alisados e pré-tratados com MSC, 1ª derivada ou combinação. Então, os dados das condições de 15 kV e 50 kV foram fundidos em único conjunto pela análise de dimensões comuns (ComDim), a qual também possibilita a análise exploratória multiblocos. Construíram-se modelos de regressão por mínimos quadrados parciais (PLSR) para relacionar os espectros NIR ao teor de teff nas amostras, e modelos por regressão linear múltipla (MLR) para relacionar os *scores* ComDim-EDXRF. Foram utilizadas as metodologias Kennard-Stone e SPXY para separar os conjuntos de calibração e teste. A definição da quantidade de variáveis latentes (LV) e dimensões comuns (CD) foi realizada usando validação cruzada 10-fold e 8-fold, respectivamente. A PCA dos espectros NIR com MSC+2ªd mostrou clara separação entre as farinhas adulterantes e as safras de teff, devido aos teores de cinzas e carboidratos. Com base nas figuras de mérito e número de LV, o melhor modelo PLSR foi obtido com os espectros MSC+2ªd e amostragem SPXY, que usou 4 LV e apresentou  $R^2_{\text{previsão}} = 92\%$ ,  $RMSEP < RMSEC$  (0,0209 e 0,0299) que sugere ausência de sobreajuste dos dados, e  $RPD_{\text{previsão}} > 3$  indicando capacidade preditiva quantitativa excelente do modelo. Para os espectros EDXRF, o melhor modelo MLR utilizou 6 CD e os dados MSC-SPXY, apresentando  $R^2_{\text{previsão}}$  equivalente ao modelo NIR, menor RMSEP (0,0181), ausência de sobreajuste e aplicação quantitativa. Destaca-se a elevada robustez dos modelos NIR e EDXRF dada a previsão do percentual de adulteração na farinha de teff mesmo com as diferenças entre teff A e B indicadas pelas análises exploratórias. Portanto, as espectroscopias NIR e EDXRF em conjunto com quimiometria se mostraram métodos eficientes para o controle de adulteração de farinhas de teff.

Palavras-chave: controle de fraude; métodos espectroscópicos rápidos; fusão de dados; calibração multivariada.

## ABSTRACT

The cereal *Eragrostis tef* (Zucc.), known as teff, gained worldwide attention because it is gluten-free and rich in minerals, mainly iron. Because of this, teff flour is attractive to frauds, highlighting the addition of cheaper cereal flours. In this context, the study of fast and inexpensive techniques such as NIR (near-infrared) and EDXRF (energy scattering X-ray fluorescence) spectroscopy are important to identify this type of adulteration. This work evaluated the potential of the NIR and EDXRF methods to identify teff flours adulterated with rice, whole wheat, oat, and rye flours. Two teff crops were adulterated following a {5,4} simplex-lattice experimental design, varying the proportion of teff from 65% to 100% and adulterants from 0 to 35%. The NIR spectra were pre-treated with multiplicative scattering correction (MSC), 2nd derivative or combination (MSC+2nd), and analyzed by principal components (PCA). The 15 kV and 50 kV EDXRF spectra were smoothed and pretreated with MSC, 1st derivative, or combination. Then, the data from the 15 kV and 50 kV conditions were fused in a single set by the common dimension analysis (ComDim), which allows the multiblock exploratory analysis. Partial least squares regression (PLSR) models were constructed to relate the NIR spectra to the teff content in the samples, and multiple linear regression (MLR) models were built to link the ComDim-EDXRF scores. Kennard-Stone and SPXY methodologies were used to separate the calibration and test sets. The number of latent variables (LV) and common dimensions (CD) were defined using 10-fold and 8-fold cross-validation, respectively. Due to the ash and carbohydrate contents, the PCA of the MSC+2nd spectra showed a clear separation between the adulterant flours and the teff crops. Based on the figures of merit and LV number, the best PLSR model was obtained with the MSC+2nd spectra and SPXY sampling, which used 4 LV and presented  $R^2_{\text{prediction}} = 92\%$ ,  $\text{RMSEP} < \text{RMSEC}$  (0.0209 and 0.0299) which suggests no overfitting of the data, and  $\text{RPD}_{\text{prediction}} > 3$  indicating an excellent quantitative predictive capacity of the model. For EDXRF spectra, the best MLR model used 6 CD and MSC-SPXY data, presenting  $R^2_{\text{prediction}}$  equivalent to the NIR model, lower RMSEP (0.0181), absence of overfitting, and quantitative application. The high robustness of the NIR and EDXRF models are highlighted, given the prediction of the percentage of adulteration in the teff flour, even with the differences between teff A and B indicated by the exploratory data analysis. Therefore, NIR and EDXRF spectroscopies with chemometrics proved to be efficient methods for controlling the adulteration of teff flours.

Keywords: fraud control; fast spectroscopic methods; exploratory analysis; data fusion; multivariate calibration.



## LISTA DE FIGURAS

<b>Figura 1 – Cultivo de teff.....</b>	<b>16</b>
<b>Figura 2 – Teff marrom (a) Grãos (b) Farinha.....</b>	<b>17</b>
<b>Figura 3 – Fluxograma do algoritmo ComDim.....</b>	<b>24</b>
<b>Figura 4 – Espectros NIR das amostras de farinhas e adulterantes pré-tratados com MSC .....</b>	<b>40</b>
<b>Figura 5 – Gráficos dos scores da PCA dos espectros NIR-MSc com separação das amostras de acordo com o percentual de farinha de (a) Teff (b) Arroz (c) Aveia (d) Trigo Integral (e) Centeio .....</b>	<b>43</b>
<b>Figura 6 – Gráfico de loadings da PCA para os espectros NIR-MSc.....</b>	<b>44</b>
<b>Figura 7 – Gráfico APARP do modelo NIR-MSc-2d-SPXY.....</b>	<b>48</b>
<b>Figura 8 – Gráfico (a) VIP para modelo com 4 LV (b) Coeficientes de regressão do modelo PLS com 4 LV .....</b>	<b>50</b>
<b>Figura 9 - Espectros EDXRF das amostras puras de teff e adulterantes (a) 15 kV (b) 50 kV .....</b>	<b>51</b>
<b>Figura 10 – Gráficos dos scores do ComDim-EDXRF com separação das amostras de acordo com o percentual de farinha de (a) Teff (b) Arroz (c) Aveia (d) Trigo Integral (e) Centeio.....</b>	<b>54</b>
<b>Figura 11 - Gráfico de loadings do ComDim para os espectros EDXRF originais .....</b>	<b>55</b>
<b>Figura 12 - Gráfico APARP do modelo MSc-MLR.....</b>	<b>58</b>
<b>Figura 13 – Gráficos dos loading das 6 CDs utilizadas pelo melhor modelo ComDim-MLR. Contribuição dos espectros (a) 15 kV na CD1 (b) 50 kV na CD1 (c) 15 kV nas CDs 2-6 (d) 50 kV nas CDs 2-6 .....</b>	<b>60</b>

## LISTA DE TABELAS

<b>Tabela 1 - Fatores de conversão de nitrogênio em proteínas para cereais .....</b>	<b>31</b>
<b>Tabela 2 – Composição proximal das farinhas puras .....</b>	<b>39</b>
<b>Tabela 3 – Modelos de calibração NIR-PLS validados utilizando figuras de mérito.....</b>	<b>47</b>
<b>Tabela 4 – Saliências das condições das análises EDXRF nas dimensões comuns.....</b>	<b>52</b>
<b>Tabela 5 – Modelos de calibração ComDim-MLR validados utilizando figuras de mérito .....</b>	<b>57</b>
<b>Tabela 6 – Figuras de mérito analíticas do modelo ComDim-MLR para os dados MSC.....</b>	<b>59</b>

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	12
<b>2 OBJETIVOS</b> .....	15
2.1 Objetivo Geral .....	15
2.2 Objetivos Específicos .....	15
<b>3 REVISÃO BIBLIOGRÁFICA</b> .....	16
3.1 Teff.....	16
3.2 Planejamento de Misturas .....	17
3.3 Espectroscopia no Infravermelho Próximo.....	19
3.4 Espectroscopia de Fluorescência de Raios X por Dispersão de Energia.....	19
3.5 Quimiometria.....	20
3.5.1 Pré-processamento dos Espectros .....	21
3.5.2 Análise por Componentes Principais – PCA .....	23
3.5.3 Análise por Dimensões Comuns – ComDim.....	23
3.5.4 Regressão Linear Múltipla – MLR .....	24
3.5.5 Regressão por Mínimos Quadrados Parciais – PLSR .....	25
3.5.6 Avaliação da Não-Linearidade dos Dados .....	26
3.5.7 Importância da Variável na Projeção .....	27
3.5.8 Figuras de Mérito .....	27
<b>4 MATERIAL E MÉTODOS</b> .....	30
4.1 Preparo e Adulteração das Farinhas .....	30
4.2 Composição Proximal.....	31
4.3 Espectroscopia no Infravermelho Próximo.....	32
4.4 Espectroscopia de Fluorescência de Raios X por Dispersão de Energia.....	32
4.5 Análises Quimiométricas .....	32
4.5.1 Pré-processamento dos Espectros .....	33
4.5.2 Análise por Componentes Principais – PCA .....	33
4.5.3 Análise por Dimensões Comuns – ComDim.....	33
4.5.4 Regressão por Mínimos Quadrados Parciais – PLSR .....	34
4.5.5 Regressão Linear Múltipla – MLR .....	34
4.5.6 Avaliação da Não-Linearidade dos Dados .....	35
4.5.7 Importância da variável na projeção .....	36
<b>5 RESULTADOS E DISCUSSÃO</b> .....	37

5.1 Composição Proximal.....	37
5.2 Análise exploratória dos espectros NIR.....	40
5.3 Calibração multivariada usando os espectros NIR.....	45
5.4 Análise exploratória dos espectros EDXRF.....	50
5.5 Calibração multivariada usando os espectros EDXRF.....	56
<b>6 CONCLUSÃO.....</b>	<b>62</b>
<b>REFERÊNCIAS.....</b>	<b>63</b>
<b>APÊNDICE A – Planejamento de misturas para a adulteração.....</b>	<b>71</b>
<b>APÊNDICE B – Bandas identificadas na espectroscopia NIR.....</b>	<b>76</b>
<b>APÊNDICE C – Energias dos elementos químicos identificados na espectroscopia EDXRF.....</b>	<b>79</b>

## 1 INTRODUÇÃO

Os *millet*s são cereais com tamanho de grão muito reduzido, dentre os quais está o teff (MUTHAMILARASAN; PRASAD, 2020). *Eragrostis tef* (Zucc.) Trotter é um cereal nativo da Etiópia onde sua produção já é consolidada. Por ser um grão pequeno, seu consumo ocorre na forma integral, mesmo quando em farinhas, o que mantém a composição de macro e micronutrientes no produto (ABEBE; RONDA, 2014; GEBREMARIAM; ZARNKOW; BECKER, 2014; VIELL et al., 2020b).

Nos últimos anos, o cultivo desse grão vem se disseminando para países como Austrália, Estados Unidos e China (BARRETTO et al., 2020). O atual interesse no teff se deve às suas características nutricionais que, de modo geral, são superiores aos cereais comuns na alimentação mundial, como trigo e centeio. Este alimento, possui proteínas de melhor qualidade nutricional dentre os cereais devido a maior concentração de aminoácidos essenciais, perfil lipídico com predomínio de ácidos graxos insaturados oléico e linoléico, e ainda alto percentual de minerais cujo conteúdo de cálcio, ferro, magnésio e zinco supera a maioria dos cereais conhecidos, como cevada e trigo (ABEBE; RONDA, 2014; BARRETTO et al., 2020; GEBREMARIAM; ZARNKOW; BECKER, 2014; ZHU, 2018). Além disso, o teff é livre de glúten, apresentando-se como alternativa interessante para o segmento crescente de produtos *gluten-free* na conferência de melhores valores nutricionais ao produto (ABEBE; RONDA, 2014; VIELL et al., 2020a; ZHU, 2018).

A popularização do teff e a produção em escala ainda reduzida tornam sua farinha um alimento atrativo à prática de adulteração por motivação econômica, principalmente através da adição de produtos mais baratos, como cereais de baixo custo. Essas fraudes resultam em prejuízos ao consumidor, que além de ser induzido ao erro pode sofrer danos à saúde dado o desconhecimento da substância adicionada (GALVIN-KING; HAUGHEY; ELLIOTT, 2018). No caso do teff, a adição de cereais com glúten pode levar os consumidores portadores da doença celíaca à sérias reações alérgicas.

Por isso, o desenvolvimento de métodos rápidos e confiáveis para a detecção de fraude em alimentos se faz necessário para garantir a segurança de alimentos. Assim, um parâmetro precisa ser medido de modo que sua variação se relacione com a presença de adulterante, por exemplo o conteúdo mineral ou orgânico da amostra (GALVIN-KING; HAUGHEY; ELLIOTT, 2018; MEDINA et al., 2019).

Para a determinação de conteúdo mineral, as metodologias comumente empregadas são as espectrometrias de absorção e emissão atômicas, como espectrometria de absorção

atômica com chama (FAAS), espectrometria de emissão óptica com plasma de indutivamente acoplado (ICP-OES) e espectrometria de massa com plasma de acoplamento induzido (ICP-MS) (HE et al., 2017; MELQUIADES et al., 2011). Apesar de serem técnicas de aplicação bem estabelecidas, demandam significativo tempo operacional devido à necessidade de preparo da amostra por digestão em ácido e impossibilidade de análise multielementar. As técnicas por ICP contornam este último problema, porém aumentam o seu custo (MASOTTI et al., 2020; MELQUIADES et al., 2011; MIR-MARQUÉS; CERVERA; DE LA GUARDIA, 2016).

Desse modo, os métodos espectroscópicos por fluorescência de raios X vêm se mostrando uma alternativa mais rápida na quantificação de minerais, pois a preparação da amostra é mais simples e vários elementos são identificados simultaneamente. Além disso, são facilmente adaptáveis para uso em linha para determinações em tempo real, sendo vantajoso para a indústria (MASOTTI et al., 2020; MELQUIADES et al., 2011; MIR-MARQUÉS; CERVERA; DE LA GUARDIA, 2016).

A espectroscopia de fluorescência de raios X por dispersão de energia (EDXRF) se baseia na emissão de fótons pelos átomos constituintes da amostra. A energia dos raios X incidentes na amostra é suficiente para retirar elétrons das camadas eletrônicas mais internas (menor energia) do átomo, que tendem a preencher essa vacância pela transição de elétrons das camadas mais externas (maior energia). A energia excedente é emitida na transição como fóton característico de raios X, permitindo identificar os elementos químicos, já que essa quantidade de energia é específica para cada transição em cada elemento, e quantificá-los pela relação de intensidade do sinal (DOS SANTOS et al., 2020; RODRIGUES, 2005).

Para a determinação de compostos orgânicos, a espectroscopia no infravermelho próximo (NIR) ganhou ampla aceitação em diferentes campos em virtude de suas vantagens. Destacam-se a capacidade de registrar espectros para amostras sólidas e líquidas sem manipulação prévia e equipamentos portáteis capazes de registrar espectros até mesmo em linhas de produção (BLANCO; VILLARROYA, 2002). A espectroscopia NIR é baseada na absorção de radiação eletromagnética na região de 400-2500 nm resultando em bandas largas associadas às combinações de modos de vibração (O-H, N-H e C-H) e sobretons de vibrações moleculares (WANG et al., 2017). A intensidade das bandas NIR depende da mudança no momento dipolar e da anarmonicidade da ligação, portanto as ligações do hidrogênio com átomos mais eletronegativos apresentam um maior desvio do comportamento harmônico e bandas mais intensas na região do infravermelho próximo. As interações entre átomos em diferentes moléculas (ligações de hidrogênio e/ou interações dipolares) alteram os estados de energia vibracional, deslocando as bandas de absorção existentes e dando origem a novas. Isso

permite que formas cristalinas sejam distinguidas e propriedades físicas correlacionadas com a intensidade espectral, tais como densidade, viscosidade e tamanho de partícula em sólidos, sejam determinadas. Ou seja, o espectro NIR contém não apenas informações químicas de composição, mas também informações físicas que podem ser empregadas para determinar as propriedades físicas das amostras (BLANCO; VILLARROYA, 2002). A análise de alimentos foi uma das primeiras aplicações da espectroscopia NIR (HUANG et al., 2008) que continua sendo amplamente utilizada para a análise, por exemplo, de mel (HUANG et al., 2020), trigo (FIRMANI et al., 2020), gordura hidrogenada (PEREIRA et al., 2018), café (MARQUETTI et al., 2016) e vinho (YU; ZHAN; HUANG, 2017).

A natureza multivariada dos dados de EDXRF e NIR torna indispensável o uso de ferramentas quimiométricas adequadas para extração das informações contidas nos espectros (CALLAO; RUISÁNCHEZ, 2018; DOS SANTOS et al., 2020; ZAREEF et al., 2020). O tratamento quimiométrico dos dados inclui a preparação dos dados (pré-tratamentos), análise exploratória, seleção de amostras (conjuntos de calibração e previsão), calibração e validação dos modelos (BONA; MARÇO; VALDERRAMA, 2018). O emprego da quimiometria para o tratamento de dados espectroscópicos é uma área bem consolidada e várias publicações tratam o tema com variados exemplos de aplicação na área de análise de alimentos (COZZOLINO, 2014; DOS SANTOS et al., 2020, 2021; MARQUETTI et al., 2016; VIEIRA et al., 2020).

A complexidade da matriz alimentar pode, por vezes, tornar as características obtidas por um único método de análise insuficientes para o monitoramento da qualidade do alimento, como a identificação de fraudes. Desse modo, a avaliação simultânea de informações de diferentes técnicas analíticas, compartilhadas entre as técnicas ou únicas, tende a descrever mais fielmente as particularidades e semelhanças do alimento necessárias para reconhecer o produto fraudado (BIANCOLILLO et al., 2019). Para isso, são utilizadas as ferramentas de fusão de dados que unem os blocos de dados para a análise única do conjunto, as quais são classificadas como nível baixo (os dados sem processamentos prévios são fundidos e então modelados para qualificar ou quantificar a variável resposta), nível médio (a fusão ocorre após a extração de características sob a forma de subconjunto de variáveis, por exemplo, as quais são modeladas) e nível alto (os dados são modelados e os resultados previstos obtidos são fundidos para melhorar o desempenho). Um exemplo de técnica de fusão de dados é o ComDim (Análise por Dimensões Comuns) que pertence aos métodos de nível médio (SMOLINSKA et al., 2019).

Assim, o objetivo desse trabalho foi avaliar se as técnicas de espectroscopia NIR e EDXRF em conjunto com quimiometria são capazes de identificar e quantificar adulterações em farinha de teff por adição de farinhas de cereais de menor valor.

## **2 OBJETIVOS**

### **2.1 Objetivo Geral**

Identificar e quantificar a adulteração em farinhas de teff a partir de espectroscopia NIR e EDXRF em conjunto aos métodos quimiométricos.

### **2.2 Objetivos Específicos**

- Determinar a composição proximal das farinhas de teff e farinhas adulterantes (trigo integral, arroz, aveia e centeio);
- Preparar misturas da farinha de teff com os adulterantes de acordo com um planejamento de misturas;
- Identificar e quantificar adulterações em farinha de teff utilizando a informação de espectros NIR e EDXRF, aplicando métodos exploratórios e de calibração multivariada;
- Comparar os dois métodos espectroscópicos na avaliação de adulteração em farinhas de teff;
- Verificar se a fusão de dados é vantajosa para os dados EDXRF obtidos por diferentes condições de análise.

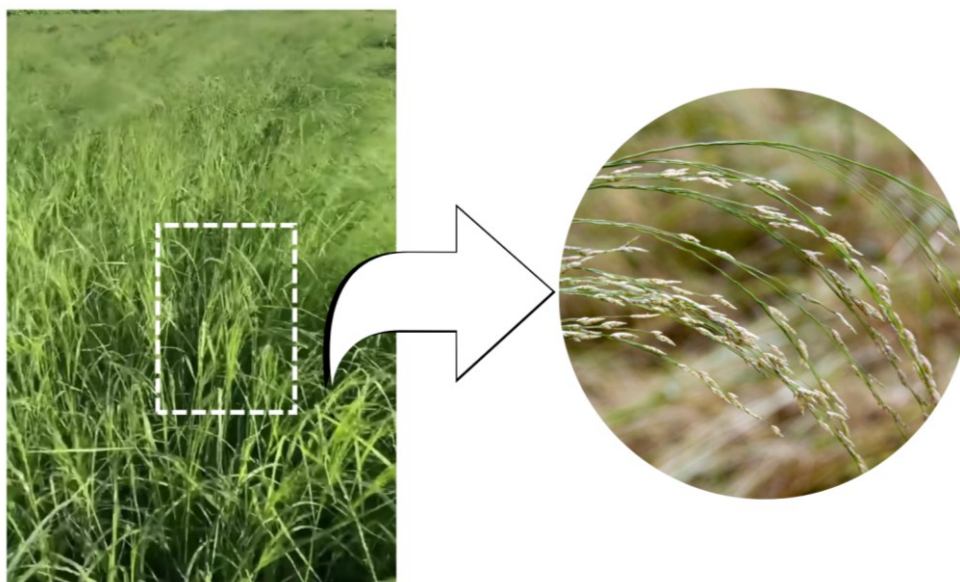


### 3 REVISÃO BIBLIOGRÁFICA

#### 3.1 Teff

Teff ou *Eragrostis tef* (Zucc.) Trotter é um cereal originário da Etiópia, onde se concentra a maior produção mundial com relatos de seu cultivo desde períodos antes de Cristo (BARRETTO et al., 2020; BULTOSA, 2016). A produção em larga escala do teff enfrenta dificuldades devido às diminutas dimensões do grão (Figura 1) que dificulta o uso de maquinários, demandando o uso de colheita manual que culmina em grandes perdas de produção (BARRETTO et al., 2020).

**Figura 1 – Cultivo de teff**



**Fonte: Adaptado de Giroil Alimentos (2022); Ocean Robbins (2021).**

Os cultivares possuem grãos ovais cuja cor varia de branco a marrom escuro, com apenas 0,9 – 1,7 mm de comprimento e 0,7 – 1 mm de diâmetro, sendo esse tamanho reduzido característico de cereais *millets* e responsável pela obtenção de farinhas integrais (Figura 2) (BARRETTO et al., 2020; GEBREMARIAM; ZARNKOW; BECKER, 2014; MUTHAMILARASAN; PRASAD, 2020). As variedades escuras apresentam maiores concentrações de polifenóis que atribuem ao teff propriedades antioxidantes interessantes (BARRETTO et al., 2020).

**Figura 2 – Teff marrom (a) Grãos (b) Farinha**



**Fonte: Autoria própria (2022); Healthy Supplies (2022).**

O teff pode ser usado na alimentação humana e animal. Na Etiópia é comum o seu uso para a produção do *injera*, uma massa fermentada típica na alimentação regional (BARRETTO et al., 2020). Este grão vem ganhando notoriedade por ser um cereal livre de glúten com propriedades nutricionais superiores aos cereais de consumo mundial estabelecido, as quais são mantidas na farinha integral, sendo uma alternativa importante no desenvolvimento de produtos alimentícios desse segmento (GEBREMARIAM; ZARNKOW; BECKER, 2014). Quanto ao teor proteico, o teff é superior ao centeio e arroz integral e similar ao milho e trigo, porém possui melhor composição de aminoácidos essenciais do que a cevada e o trigo, sendo comparável ao ovo. O teff supera o trigo, centeio e arroz integral no conteúdo de lipídeos, cuja composição majoritária é de ácidos graxos insaturados oléico e linoléico (GEBREMARIAM; ZARNKOW; BECKER, 2014). Mesmo em quantidade relativamente baixa, com apenas 6,9 – 9,9 %, a fração de ácido graxo linolênico (n-3) ainda é superior do que o presente na maioria dos cereais (BULTOSA, 2016; CHUNG et al., 2009; ROSENTRATER; EVERS, 2018). O destaque do teff é seu conteúdo mineral, com concentrações elevadas de cálcio, magnésio e principalmente ferro em relação aos demais cereais, contribuindo dessa maneira para a ingestão diária desses micronutrientes (BULTOSA, 2016; GEBREMARIAM; ZARNKOW; BECKER, 2014).

### **3.2 Planejamento de Misturas**

O planejamento de experimentos consiste no conjunto de ferramentas estatísticas utilizadas no entendimento de como as variáveis de processo influenciam em uma determinada variável de resposta. No planejamento de misturas se observa como a proporção dos

componentes que constituem o produto influencia em determinada característica do mesmo, independentemente da quantidade absoluta do componente (CORNELL, 2011).

Em qualquer planejamento de misturas duas condições devem ser obedecidas. A fração dos componentes variam de 0 a 1, ou seja, a mistura pode estar ausente desse componente ou ser formada apenas por ele, logo,  $0 \leq x_i \leq 1$ , e a soma das frações de todos os componentes deve ser 1, ou seja, 100%, logo  $x_1 + x_2 + \dots + x_i = 1$  (CORNELL, 2011). Nota-se que a mistura possui  $i - 1$  graus de liberdade, uma vez que alterando-se a fração de  $i - 1$  constituintes, o  $i$ -ésimo componente tem seu valor definido para a soma das frações ser 1 (BARROS; SPACINO; BRUNS, 2010).

Uma figura geométrica com  $n$  dimensões e  $n + 1$  vértices que correspondem às combinações de frações possíveis para  $q$  componentes respeitando as restrições de mistura é chamada *simplex*. O planejamento de misturas *simplex-lattice* é aquele cujas frações assumem valores igualmente espaçados no *simplex* por um fator  $m$ , logo, é possível ajustar um polinômio de grau  $m$  às proporções do planejamento  $\{q, m\}$ . Cada componente pode apresentar na mistura os  $m + 1$  valores para as frações, conforme a Equação 1 (CORNELL, 2011).

$$x_i = 0, \frac{1}{m}, \frac{2}{m}, \dots, 1 \quad (1)$$

É comum que os componentes devam assumir uma fração mínima na mistura para que essa apresente as características procuradas, por exemplo, a quantidade mínima de ferro para que a farinha seja considerada fortificada. Nesses casos, os valores  $x_i$  do planejamento *simplex-lattice* original são chamados pseudocomponentes, pois não consideram os limites inferior ou superior das proporções. Uma decodificação se faz necessária para saber a real fração do componente  $i$  na mistura, para isso é utilizada a Equação 2 (BARROS; SPACINO; BRUNS, 2010).

$$x_i = \frac{c_i - a_i}{1 - \sum_{i=1}^q a_i} \quad (2)$$

O termo  $c_i$  corresponde à proporção real do componente na mistura;  $a_i$  é a fração do limite inferior do componente  $i$ , enquanto no denominador há o somatório dos limites inferiores de todos os  $q$  componentes presentes na mistura (BARROS; SPACINO; BRUNS, 2010).

### 3.3 Espectroscopia no Infravermelho Próximo

A espectroscopia no infravermelho próximo (NIR) é baseada na relação entre a matéria e a radiação no comprimento de onda de 780 – 2500 nm, sendo essa energia suficiente para promover a vibração de ligações moleculares. Como várias ligações moleculares absorvem na mesma frequência de energia, os espectros são compostos por bandas alargadas devido às combinações de bandas e sobretons. Essa técnica espectroscópica é aplicada a compostos orgânicos, sendo detectadas vibrações de estiramento e angular do átomo de hidrogênio com o carbono, nitrogênio ou oxigênio (BURNS; CURCZAK, 2021; MANLEY, 2014). Desse modo, outras metodologias são necessárias na análise de compostos inorgânicos, como na avaliação de minerais em alimentos.

O NIR já vem sendo aplicado na indústria alimentícia para o controle de qualidade, tendo inúmeros relatos na literatura, sendo vantajoso por ser rápido, multicomponente e não destrutivo (GRASSI; ALAMPRESE, 2018; MCNEIL et al., 2013; SANTOS et al., 2021). Na detecção de fraudes em alimentos, a técnica vem sendo estudada para diferentes matrizes alimentares, incluindo farinhas de cereais (AYVAZ et al., 2021; NDLOVU et al., 2021; XUE et al., 2021).

### 3.4 Espectroscopia de Fluorescência de Raios X por Dispersão de Energia

A técnica EDXRF se baseia na excitação eletrônica pela incidência de radiação no comprimento de onda de raios X. Os raios X incidentes com energia apropriada causam a emissão de elétrons de camadas eletrônicas internas dos elementos químicos que compõem a amostra, resultando na desestabilização do átomo, que para retornar ao seu estado fundamental promove a migração de elétrons mais energéticos para a camada menos energética vaga (vacância) com subsequente emissão de fótons de raios X por fluorescência. A energia desses fótons depende dos níveis eletrônicos e é característica de cada elemento químico, por isso, são chamados de raios X característicos, sendo possível a identificação dos elementos químicos de uma amostra através da avaliação do espectro gerado. Pela intensidade do pico, a análise quantitativa também é possível de ser realizada (DOS SANTOS et al., 2020; RODRIGUES, 2005). A espectroscopia EDXRF apresenta as vantagens de ser não destrutiva, fornecer análise multielementar e demandar pouco ou nenhum preparo de amostra (PERUCHI et al., 2014; SANTANA et al., 2021).

Além da fluorescência dos raios X emitidos pelos átomos, fundamentada no efeito fotoelétrico, podem ocorrer os efeitos de espalhamento da radiação. Dentre os tipos de espalhamento estão o efeito Compton ou espalhamento incoerente e o efeito Rayleigh ou espalhamento coerente, os quais são causados principalmente por elementos leves ( $Z < 8$ ). Por isso, além das informações inorgânicas, a técnica pode fornecer de maneira indireta informações sobre a fração orgânica da amostra devido aos efeitos de interação com a matriz causados pelo espalhamento dos fótons incidentes. Para isso se faz necessário o uso de métodos quimiométricos. (DOS SANTOS et al., 2020, 2021; MARETTI SILVEIRA BUENO et al., 2005).

A identificação da transição eletrônica correspondente ao pico do espectro EDXRF utiliza a letra da camada para a qual o elétron migrou seguido de uma letra grega que indica a camada eletrônica de saída do elétron, onde  $\alpha$  refere-se à camada L e  $\beta$  à camada externa M (DOS SANTOS et al., 2020). Por exemplo, o ferro possui um  $K\alpha$  de 6,4 keV sendo essa a energia contida no fóton emitido durante a transição do elétron na camada L para a K; já a transição M para K produz raios X  $K\beta$  com fótons de 7,06 keV.

Na área de alimentos, o EDXRF vem sendo aplicado para a identificação de traços de Mo e Pb em alimentos (ALI et al., 2014), análise de minerais em frutas e vegetais (GALLARDO et al., 2016; HAZRA et al., 2020; HERREROS-CHAVEZ et al., 2019) e autenticação de chás de ervas (MALYJUREK et al., 2021; ZAWISZA et al., 2021).

### 3.5 Quimiometria

A quimiometria é a junção de ferramentas matemáticas e estatísticas aplicadas na extração de informações químicas provenientes de análises laboratoriais. Esse campo pode ser subdividido em duas grandes áreas, sendo elas o reconhecimento de padrões e a calibração multivariada (FERREIRA, 2015).

O reconhecimento de padrões consiste em métodos que encontram semelhanças entre as amostras a fim de agrupá-las. Esses métodos podem ser supervisionados ou não supervisionados. Os métodos supervisionados são aqueles que utilizam classes conhecidas para a separação, ou seja, já se conhecem os agrupamentos aos quais as amostras serão atribuídas, por exemplo, amostras com adulterante e sem adulterante. Já nos métodos não supervisionados, as classes não são conhecidas e serão construídas a partir da natureza contida nos dados (FERREIRA, 2015).

As técnicas de calibração objetivam a construção de modelos matemáticos para a quantificação de alguma propriedade química, como a concentração de determinado analito ou o percentual de adulteração, por exemplo. A calibração multivariada resulta em modelos de melhor qualidade preditiva para amostras complexas, uma vez que se utiliza um conjunto de medidas para a construção do modelo, por exemplo, a interação da amostra com uma faixa de radiação ao invés de apenas um comprimento de onda como na calibração univariada (FERREIRA, 2015).

A quimiometria é uma importante ferramenta para a extração e compreensão de informações contidas nos dados instrumentais. Assim, são encontrados diversos trabalhos na literatura que utilizam de métodos instrumentais, como as espectroscopias NIR e EDXRF. Há utilização de EDXRF e calibração PLS para monitoramento de minerais em caldo-de-cana durante etapas da fabricação de açúcar e álcool (SPERANÇA; NASCIMENTO; PEREIRA, 2021), na avaliação de parâmetros agrícolas de qualidade do solo (DOS SANTOS et al., 2020) e na verificação do conteúdo mineral de bananas biofortificadas (SPERANÇA et al., 2021). Quanto ao NIR, observou-se o seu uso em conjunto com o método supervisionado PLS-DA na identificação de amostras de erva-mate adulteradas com açúcar e sua classificação regional (VIEIRA et al., 2020). PLS-DA também foi aplicado na classificação de farinhas em trigo durum e comum (UNUVAR; BOYACI; KOKSEL, 2021), identificação de genótipos de trigo (SPOLADORE; BRÍGIDA DOS SANTOS SCHOLZ; BONA, 2021), e há uso da espectroscopia NIR e calibração PLS na quantificação de contaminação por farinha de trigo em farinha de banana verde (NDLOVU et al., 2021).

### **3.5.1 Pré-processamento dos Espectros**

Uma etapa importante que antecede a análise quimiométrica, seja ela exploratória, classificatória ou de regressão, é o pré-processamento dos dados coletados, que consiste na aplicação de métodos matemáticos capazes de reduzir informações indesejáveis contidas no sinal instrumental, as quais não estão relacionadas com as propriedades das amostras, podendo ainda levar a conclusões equivocadas sobre o comportamento do conjunto amostral. A escolha do método de pré-processamento deve ser cuidadosa, a fim de reduzir ao máximo o ruído, mas sem perdas significativas no sinal (FERREIRA, 2015).

A correção do espalhamento multiplicativo (MSC) é uma técnica de correção comumente aplicada em espectros NIR de amostras sólidas, que ajusta os espectros originais ao espectro médio, a partir da regressão por mínimos quadrados (FERREIRA, 2015). Este

método corrige efeitos aditivos e multiplicativos, ambos causadores de deslocamentos da linha de base em relação à absorvância nula (*offset*) referentes ao espalhamento da radiação do infravermelho, o qual se deve à heterogeneidade no tamanho ou forma de amostras particuladas que resulta na reflexão difusa da radiação, e conseqüentemente, em redução da intensidade da radiação que atinge o detector (FERREIRA, 2015; GELADI; MACDOUGALL; MARTENS, 1985; MARTENS; STARK, 1991).

Problemas de *offset* em dados espectrais também podem ser resolvidos pelo uso da primeira derivada, sendo esse método diferente do MSC por não eliminar os efeitos multiplicativos (FERREIRA, 2015).

A utilização de algoritmos de alisamento é uma prática comum em espectroscopias de fluorescência e no infravermelho, a fim de reduzir o ruído e, por consequência, evidenciar a informação contida no sinal pelo aumento da razão sinal/ruído. Dessa forma, torna-se possível revelar picos menos intensos inicialmente interpretados como ruído devido à maior frequência dessa componente aleatória (DOS SANTOS et al., 2021; FERREIRA; BRAGA; SENA, 2013; FERREIRA, 2015; LI et al., 2022; NDLOVU et al., 2021).

Espectroscopias baseadas na reflectância da radiação, como é o caso do NIR, costumam apresentar além do *offset* uma inclinação da linha de base dos espectros. Ambas as informações podem ser corrigidas aplicando-se a segunda derivada, entretanto, assim como para a primeira derivada, o uso de segunda derivada não corrige efeitos multiplicativos como o MSC (FERREIRA, 2015).

Os valores estimados das primeira e segunda derivadas assim como da intensidade do sinal no alisamento costumam ser obtidos pelo método de Savitzky-Golay, o qual ajusta um polinômio de grau  $n$  aos pontos de uma janela móvel de tamanho  $m$ . Essas estimativas correspondem ao resultado no ponto central da janela equivalentes a uma média ponderada das intensidades reais obtidas na análise instrumental. A janela se desloca até o último ponto do espectro para que todos os pontos tenham esses parâmetros previstos. O polinômio e a janela devem ser adequados aos dados a fim de minimizar as perdas de informações do sinal e maximizar a redução de ruídos. Quanto maior o grau do polinômio, maior é a modelagem do sinal original com reduzidas perdas, contudo, o ruído também é modelado. Em contrapartida, o ajuste de polinômios de menor grau eliminam o ruído, mas há maiores deformações do sinal instrumental, podendo-se perder características importantes das amostras. De forma análoga, janelas com mais pontos também reduzem o ruído, mas diminuem a razão S/R (sinal/ruído) (FERREIRA, 2015).

### 3.5.2 Análise por Componentes Principais – PCA

A análise exploratória por componentes principais (PCA) é um método não supervisionado utilizado para reduzir a dimensionalidade do espaço multivariado, a partir de combinações lineares entre as variáveis com alta correlação que constituem o espaço  $n$ -dimensional original, tornando mais evidente as semelhanças entre as amostras. As novas variáveis que descrevem o espaço  $n$ -dimensional reduzido são chamadas componentes principais (PC), as quais descrevem a máxima variância dos dados em uma direção, e cuja ortogonalidade garante que a variância explicada por cada PC não seja redundante (KUCHERYAVSKIY; LOMBORG, 2015; LUYPAERT; MASSART; VANDER HEYDEN, 2007).

### 3.5.3 Análise por Dimensões Comuns – ComDim

O ComDim é um algoritmo não supervisionado de análise exploratória que pode ser utilizado para fusão de dados de nível médio quando combinado a um método de regressão. Esse método se diferencia da PCA por permitir a análise multibloco, ou seja, dados oriundos de diferentes técnicas analíticas aplicadas às mesmas amostras são utilizados, simultaneamente, na investigação de padrões com o intuito de representar maior variância dos dados a partir das informações compartilhadas (CARIOU et al., 2019; DE AGUIAR et al., 2022; MAKIMORI; BONA, 2019). Assim, a relação das diferentes técnicas aplicadas às amostras nas dimensões do novo espaço amostral é dada pela Equação 3.

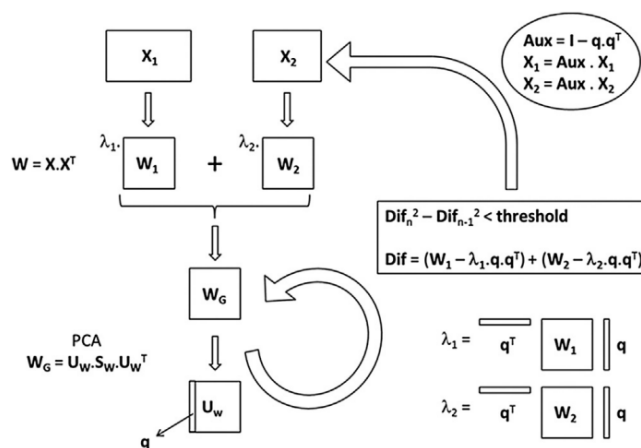
$$\mathbf{W} = \sum_{j=1}^n \lambda_j \mathbf{q}_j \mathbf{q}_j^T \quad (3)$$

O termo  $\mathbf{W}$  é a matriz do produto cruzado da matriz  $\mathbf{X}$  que contém os resultados de cada técnica, logo, todo o método analítico tem seu  $\mathbf{W}$ . Os elementos  $\lambda_j$  são as saliências, isto é, o quanto cada técnica contribui com informações para a construção da dimensão comum  $j$ . Assim, a saliência informa quanto da variância das amostras identificada pela técnica foi possível representar na  $j$ -ésima dimensão comum. Por fim,  $\mathbf{q}_j$  é o vetor com os *scores* das amostras na nova dimensão  $j$  (MAKIMORI; BONA, 2019).



A Figura 3 tem uma representação esquemática do processo, que se inicia com o cálculo da matriz de produto cruzado  $\mathbf{W}$  ( $\mathbf{W} = \mathbf{X}\mathbf{X}^T$ ) para as matrizes  $\mathbf{X}$  com os resultados de cada método analítico, a fim de eliminar a contribuição do diferente número de variáveis preditoras (colunas) utilizadas pelos diferentes métodos analíticos. Então, é realizada uma soma ponderada das matrizes  $\mathbf{W}$  considerando seus pesos ( $\lambda$ ) e a nova matriz é decomposta em valor singular, como na PCA, para obter a componente principal e os *scores* ( $\mathbf{q}$ ) das amostras na nova dimensão. Com os *scores* e as matrizes  $\mathbf{W}$ , são calculados seus novos pesos ( $\lambda$ ) de modo a saber suas contribuições nos *scores*. Esse cálculo é realizado até os pesos convergirem, sendo a dimensão comum definida. Por fim, as matrizes  $\mathbf{X}$  são deflacionadas para que a próxima dimensão comum não contenha informação redundante (DE AGUIAR et al., 2022; MAKIMORI; BONA, 2019).

**Figura 3 – Fluxograma do algoritmo ComDim**



Fonte: Cariou et al. (2019).

Na área de alimentos, há relatos da aplicação do ComDim na descrição sensorial de pães formulados com teff (VIELL et al., 2020a) e bolos formulados com maca peruana e yacon (TORMENA et al., 2017), classificação de blends de café solúvel (MAKIMORI; BONA, 2019) e erva-mate (VIEIRA et al., 2020).

### 3.5.4 Regressão Linear Múltipla – MLR

A regressão linear múltipla (MLR) é um método quimiométrico de calibração multivariada por mínimos quadrados, que consiste em ajustar uma reta aos dados que minimize a soma do quadrado dos resíduos. Matematicamente, o MLR possui baixa complexidade quando comparado a outros métodos da mesma classe, como o PLSR (FERREIRA, 2015).

A projeção da variável independente (vetor  $\mathbf{y}$ ) em um espaço n-dimensional mostra que ela pode ser obtida a partir de contribuições das variáveis dependentes (matriz  $\mathbf{X}$ ) somadas à informação que essas variáveis não abordam, isto é, o resíduo (vetor  $\mathbf{e}$ ), conforme demonstrado na Equação 4. Assim, o objetivo é encontrar essas contribuições, ou seja, os coeficientes (vetor  $\mathbf{b}$ ) da equação que define a reta da regressão. Aplicando-se o método dos mínimos quadrados à Equação 4,  $\mathbf{b}$  é obtido pela Equação 5 (FERREIRA, 2015).

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e} \quad (4)$$

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (5)$$

O uso do MLR na modelagem se restringe à problemas para os quais a matriz  $\mathbf{X}$  possui número de linhas maior do que o número de colunas, logo, é preciso ter mais amostras do que variáveis dependentes. Isso porque a existência da pseudoinversa, necessária para estimar os coeficientes de regressão, está condicionada a essa restrição (FERREIRA, 2015). Por isso, seu uso em dados instrumentais, como NIR e EDXRF, demanda a aplicação prévia de seleção de variáveis. O método de seleção deve considerar a colinearidade das variáveis e eleger, preferencialmente, variáveis não-colineares, a fim de reduzir informações de mesma natureza e melhorar a performance do modelo. O ComDim é uma alternativa que além de reduzir o espaço n-dimensional ao agrupar as variáveis colineares, é capaz de identificar colinearidade entre variáveis de diferentes métodos analíticos e agrupá-las, iniciando o processo de fusão de dados. Assim, é possível correlacionar a matriz de *score* do ComDim com as propriedades de interesse através de modelos MLR (DE AGUIAR et al., 2022).

### 3.5.5 Regressão por Mínimos Quadrados Parciais – PLSR

A metodologia de calibração PLSR possibilita a redução do espaço multivariado de estudo pela combinação linear de variáveis que trazem informação de mesma natureza (colineares), tornando mais simples a interpretação do processo e aumentando a robustez dos modelos. Por isso, é usualmente empregado em dados instrumentais, como espectros NIR ou EDXRF, cuja quantidade de variáveis preditoras é elevada (KUCHERYAVSKIY; LOMBORG, 2015; OLIVIERI, 2015).

O PLSR decompõe a matriz dos espectros ( $\mathbf{X}$ ) de modo a maximizar a covariância entre a matriz  $\mathbf{X}$  com o sinal instrumental (variáveis independentes) e o vetor  $\mathbf{y}$  da característica estudada (variável dependente). Assim, o espaço multivariado reduzido é descrito pelas

chamadas variáveis latentes que atendem essa condição, as quais não são ortogonais (BONA; MARÇO; VALDERRAMA, 2018; OLIVIERI, 2015). Os dados de  $\mathbf{X}$  são projetados no vetor  $\mathbf{y}$ , recebendo novos *scores*  $\mathbf{T}$  que são usados no cálculo dos *loadings*  $\mathbf{L}$  referentes às contribuições das variáveis originais combinadas para obtenção das variáveis latentes. Cada variável latente prevê uma variância relacionada com a variável dependente que a anterior não modelou, ou seja, seus resíduos (FERREIRA, 2015).

Para a construção dos modelos, é necessário separar os dados nos conjuntos de calibração e validação externa (previsão). O conjunto de calibração é usado para encontrar as correlações entre os dados e a propriedade estudada, e na validação cruzada (validação interna), enquanto o conjunto de previsão é utilizado para verificar a capacidade preditiva do modelo diante de amostras desconhecidas, isto é, para as quais ele não foi treinado (FERREIRA, 2015).

É importante realizar uma escolha adequada das amostras que irão compor os conjuntos de calibração e validação. Para a obtenção de conjuntos representativos uma das metodologias mais empregadas é o algoritmo de Kennard-Stone (KS) (WESTAD; MARINI, 2015). No algoritmo KS são selecionadas as amostras mais distantes (geralmente norma Euclidiana) entre si no espaço multivariado, ou seja, são utilizadas apenas informações da matriz  $\mathbf{X}$  (sinal instrumental). Uma alternativa é o algoritmo SPXY (*Sample set Partitioning based on joint  $x$ - $y$  distances*), que nada mais é que uma alteração no algoritmo KS original para considerar também o vetor  $\mathbf{y}$  (propriedade de interesse) no momento da seleção das amostras (GALVAO et al., 2005).

### 3.5.6 Avaliação da Não-Linearidade dos Dados

Ao realizar modelagens matemáticas é necessário avaliar a natureza linear dos dados, de modo a verificar o ajuste dos modelos. Se os dados forem não-lineares, abordagens diferentes do PLSR, como máquinas de vetor suporte, devem ser aplicadas. Sendo assim, métodos de avaliação desse comportamento se fazem necessários (OLIVIERI, 2015).

A técnica APARP (gráfico dos resíduos parciais aumentados) é um método de diagnóstico para modelos PLS capaz de mostrar graficamente se há não-linearidade no conjunto amostral. Ela se baseia na regressão da variável dependente nas  $n$  variáveis latentes que descrevem o modelo PLS e na primeira LV elevada ao quadrado, a fim de verificar a contribuição quadrática da LV com maior variância dos dados, estando este parâmetro intimamente relacionado ao comportamento não-linear dos dados. A variância não modelada (resíduo) somada às contribuições linear e quadrática da primeira LV ( $y_{\text{APARP}}$ ) são plotados em

relação à primeira LV. Se a contribuição linear for mais importante, ou seja, tiver maior peso nos dados, estes terão uma tendência linear, logo, os dados estão livres de não-linearidade. Por fim, são calculados os resíduos aumentados que seriam o valor previsto da variável dependente ( $y_{APARP}$ ) retirada a contribuição linear da primeira LV, logo, se não houver uma tendência não-linear no gráfico dos resíduos, a contribuição quadrática é responsável por uma porção muito pequena do comportamento dos dados (OLIVIERI, 2015; SHAN et al., 2019).

A aplicação de um método estatístico também é importante na pesquisa de não-linearidade a fim de obter conclusões mais confiáveis, sendo o teste de Durbin-Watson uma possível ferramenta. Essa técnica é baseada em um teste de hipótese sobre os resíduos aumentados, onde a hipótese nula afirma existir linearidade nos resíduos e a hipótese alternativa existir não-linearidade. Dessa forma, valores de teste superiores a 0,05 demonstram que o modelo é linear a um nível de significância de 5% (OLIVIERI, 2015).

### 3.5.7 Importância da Variável na Projeção

Os métodos de seleção de variáveis são utilizados para reduzir as variáveis de trabalho em modelos de PLSR, visto que um problema dos dados multivariados é que parte das variáveis não trazem informação quanto à propriedade prevista, sendo uma boa prática a eliminação dessas variáveis do modelo para melhorar sua performance e simplificá-lo matematicamente (FARRÉS et al., 2015).

A importância da variável na projeção (VIP) é um método de seleção comumente usado, no qual uma pontuação é calculada para cada variável independente empregada na construção do modelo preditivo. Essa pontuação considera a contribuição (peso  $w$ ) da variável na construção das variáveis latentes e a variância da propriedade prevista  $y$  que é explicada por essa variável. Nesse método, o parâmetro de corte não é estatístico, sendo selecionadas as variáveis com pontuação superior a 1 (FARRÉS et al., 2015). Além da seleção, o método VIP pode ser utilizado como um indicador das variáveis mais importantes para o modelo PLSR.

### 3.5.8 Figuras de Mérito

As figuras de mérito são parâmetros estatísticos calculados a fim de validar modelos matemáticos através da avaliação da capacidade de previsão de características para conjuntos amostrais desconhecidos, sendo possível comparar o desempenho de diferentes modelos em

determinadas condições de trabalho (BONA; MARÇO; VALDERRAMA, 2018; OLIVIERI, 2014).

A sensibilidade das calibrações multivariadas é uma figura de mérito obtida pelo inverso do módulo do vetor contendo os coeficientes de regressão do modelo multivariado (OLIVIERI, 2014). Esse parâmetro corresponde a quanto o sinal instrumental é alterado quando a concentração do analito avaliado tem um aumento unitário, sendo o modelo considerado sensível para pequenos valores de sensibilidade, os quais indicam que pequenas variações na concentração do analito causam grandes variações do sinal instrumental (OLIVIERI et al., 2006). A sensibilidade é dependente da técnica instrumental utilizada, portanto, modelos preditivos obtidos por métodos analíticos distintos ou dados submetidos a diferentes pré-tratamentos não devem ser comparados por esse parâmetro (OLIVIERI, 2014).

Outra figura de mérito analítica é a sensibilidade analítica, importante por permitir a comparação do parâmetro entre modelos e entre técnicas instrumentais distintas, uma vez que depende apenas da concentração do analito, sendo por isso um parâmetro mais útil na avaliação da sensibilidade do modelo (OLIVIERI, 2014; OLIVIERI et al., 2006). Matematicamente, é a razão entre a sensibilidade e o ruído instrumental. Como sua unidade de medida é concentração<sup>-1</sup>, tem-se que o inverso da sensibilidade analítica ( $\gamma^{-1}$ ) corresponde à menor diferença de concentração do analito entre duas amostras que o modelo de calibração consegue distinguir (OLIVIERI, 2014).

O limite de detecção (LOD) indica a concentração mínima de um analito que o modelo consegue identificar, isto é, distinguir da ausência deste analito a um nível de confiança. É possível obter uma estimativa multivariada para esse parâmetro fixando-se probabilidades de 5% de ocorrência de erros Tipo I (detectar o analito quando ausente) e Tipo II (não detectar o analito quando presente), podendo-se assim medir o limite de detecção de modelos PLSR. Para isso, delimita-se uma faixa de detecção pelos limites inferior (LOD<sub>min</sub>) e superior (LOD<sub>máx</sub>) que consideram a variância da concentração do analito nas amostras de calibração PLS. Concentrações abaixo do limite inferior não são detectáveis com confiança estatística, sendo considerada ausência do analito, enquanto concentrações acima desse limite pode indicar a presença do analito. Para essa situação, se a concentração estiver acima do limite superior a presença é confirmada, porém, para valores entre LOD<sub>min</sub> e LOD<sub>máx</sub> é necessário determinar o LOD específico da amostra e então compará-lo à concentração do analito na amostra, de modo a determinar sua presença (concentração acima do LOD específico) ou ausência (concentração abaixo do LOD específico) (ALLEGRI; OLIVIERI, 2014).

Já o limite de quantificação (LOQ) diz respeito à concentração mínima do analito que pode ser quantificada pelo modelo com considerável precisão, ou seja, com desvio inferior a 10% (FERREIRA, 2015), correspondendo a 3LOD para modelos de calibração (OLIVIERI, 2014). De forma análoga ao LOD, determina-se uma faixa de quantificação para a qual não é possível quantificar o analito se sua concentração for menor que  $LOQ_{min}$ , e a quantificação é possível se a concentração for maior que  $LOD_{máx}$  ou LOD específico.

O coeficiente de determinação ( $R^2$ ) é uma medida da adequação do modelo de regressão aos dados. O RMSEC e RMSEP (raiz quadrada do erro quadrado médio de calibração e previsão, respectivamente) avaliam a exatidão dos modelos, pois indicam o desvio da resposta prevista pelo modelo em relação à resposta real para os conjuntos de calibração e validação externa, respectivamente. Buscam-se valores de RMSEP ligeiramente maiores que o RMSEC, pois indica que o modelo não está sobreajustado ao conjunto amostral, sendo assim aplicável a outras amostras (BONA; MARÇO; VALDERRAMA, 2018).

A precisão dos modelos pode ser verificada pelo parâmetro RPD (desvio residual da predição) que demonstra a qualidade da previsão realizada (IGNE; HURBURGH, 2010).

## 4 MATERIAL E MÉTODOS

### 4.1 Preparo e Adulteração das Farinhas

Grãos de teff marrom das safras 2018 e 2016 foram moídos em misturador comercial (Modelo 51BL31-7011, Waring, EUA) para obtenção das farinhas puras A e B, respectivamente. Os grãos cultivados por produtor comum na região de Santo Ângelo – Brasil foram fornecidos pela indústria Giroil Agroindústria Ltda. As farinhas de arroz, aveia, centeio e trigo integral, utilizadas na adulteração das farinhas de teff, foram obtidas no comércio local de Campo Mourão – Brasil. A granulometria das farinhas foi padronizada utilizando peneira de malha 300  $\mu\text{m}$  (48 mesh) antes do preparo das misturas.

As farinhas de teff A e B foram adulteradas com diferentes proporções dos adulterantes. Essas proporções seguiram um planejamento de mistura  $\{5,4\}$  *simplex-lattice*, no qual os componentes de mistura  $x_1, x_2, x_3, x_4$  e  $x_5$  referem-se ao teff (A ou B), arroz, aveia, trigo integral e centeio, respectivamente (Apêndice A). Os pseudocomponentes assumiram proporções entre 0 e 1 com intervalo de 0,25 ( $m = 4$ ) gerando 70 combinações possíveis. Portanto, são 70 misturas com o teff A, 70 com o teff B e mais quatro amostras das farinhas adulterantes, totalizando 144 amostras. Foi fixado um limite inferior de 65% para o teff a fim de que as adulterações não fossem perceptíveis por inspeção visual devido ao clareamento da amostra. Por consequência, os adulterantes tiveram limite superior de 35%. A descodificação dos pseudocomponentes é realizada através da Equação 2. Considerando o limite mínimo de 65% para o teff são obtidas as Equações 6-10 para a descodificação de cada componente da mistura.

$$x_1 = \frac{c_1 - 0,65}{0,35} \quad (6)$$

$$x_2 = \frac{c_2}{0,35} \quad (7)$$

$$x_3 = \frac{c_3}{0,35} \quad (8)$$

$$x_4 = \frac{c_4}{0,35} \quad (9)$$

$$x_5 = \frac{c_5}{0,35} \quad (10)$$

## 4.2 Composição Proximal

As farinhas puras de arroz, aveia, centeio, trigo integral e teff A e B foram caracterizadas seguindo métodos gravimétricos de referência realizados em triplicata, sendo os constituintes avaliados expressos em g 100 g<sup>-1</sup> em base seca.

A umidade foi determinada a partir da secagem de 5 g de amostra em estufa CE-220/216 (Cienlab, Campinas, Brasil) a 105 °C durante 12 h, sendo atingida massa constante, seguindo o estabelecido pela metodologia padrão (AOAC INTERNATIONAL, 2005). As amostras secas foram calcinadas em forno mufla F2-DM (Fornitec, São Paulo, Brasil) a 600 °C por 4 h para determinação do teor de cinzas (AOAC INTERNATIONAL, 2005).

Os lipídios totais foram obtidos pelo método de extração a quente Soxhlet em extrator de óleos e graxas MA 044/5/50 (Marconi, Piracicaba, Brasil). Cartuchos contendo 5 g de amostra permaneceram por 8 h em contato com 90 mL de hexano em ebulição a 110 °C, seguida da evaporação do solvente e pesagem do extrato (AOAC INTERNATIONAL, 2005).

A fração de proteína bruta foi obtida pelo método de Kjeldahl. No bloco digestor TE-007MP (Tecnal, Limeira, Brasil) a 400 °C por 4 h, foram digeridos 0,3 g de amostra em meio acidificado com 5 mL de ácido sulfúrico P.A na presença de 1 g de mistura catalítica (10 g Na<sub>2</sub>SO<sub>4</sub> + 1 g CuSO<sub>4</sub> + 0,1 g Se). A amostra digerida foi destilada em destilador de nitrogênio TE-0364 (Tecnal, Limeira, Brasil) com 20 mL de NaOH (40%). O destilado foi coletado em 10 mL de ácido bórico (4%) e titulado com ácido clorídrico 0,1 mol L<sup>-1</sup> (AOAC INTERNATIONAL, 2005). O conteúdo de nitrogênio foi corrigido para forma proteica pelo fator de conversão de nitrogênio em alimentos, conforme a Tabela 1.

**Tabela 1 - Fatores de conversão de nitrogênio em proteínas para cereais**

Alimento	Fator
Trigo	5,83
Centeio	5,83
Aveia	5,83
Arroz	5,95
Outros	6,25

Fonte: FAO (2003); Instituto Adolfo Lutz (2008).



O conteúdo de carboidratos foi obtido através do cálculo por diferença em relação ao conteúdo total dos demais componentes (umidade, cinzas, lipídios totais e proteína bruta), conforme a Equação 11 (FAO, 2003).

$$\begin{aligned} \text{Carboidratos} = & 100 - (\text{umidade} + \text{cinzas} + \text{lipídios totais} \\ & + \text{proteína bruta}) \end{aligned} \quad (11)$$

#### 4.3 Espectroscopia no Infravermelho Próximo

As farinhas puras (teffs A e B, arroz, trigo integral, aveia e centeio) e suas misturas tiveram a umidade padronizada em dessecador do tipo *drybox* modelo 0880 (Nalgon, Itupeva, Brasil) durante 24 h. Na sequência, foram compactadas suavemente em suporte para amostras sólidas. Os espectros foram adquiridos à temperatura ambiente (25 °C) no espectrômetro NIRS 6500 (Foss Analytical, Hilleroed, Dinamarca) para o intervalo de comprimentos de onda de 1100 nm a 2500 nm com resolução 2 nm. O software ISIScan (Foss Analytical, Hilleroed, Dinamarca) do equipamento realizou 32 varreduras acumuladas para construção dos espectros (MARQUETTI et al., 2016).

#### 4.4 Espectroscopia de Fluorescência de Raios X por Dispersão de Energia

O conteúdo mineral foi analisado pelo método espectroscópico EDXRF utilizando o espectrofotômetro de bancada Shimadzu EDX 720 com detector silício-lítio resfriado em nitrogênio líquido. Alíquotas de 5 g das amostras foram colocadas em cápsulas poliméricas para análise XRF e fechadas com filme prolene (Chemplex Inc.), as quais foram irradiadas utilizando tubo de raios X com alvo de Rh (tubo de ródio) com aplicação de tensões de 15 kV por 100 s com resolução 0,01, para detecção dos elementos leves, e 50 kV por 100 s com resolução 0,02 para os demais elementos com maior número atômico (DOS SANTOS et al., 2020; MELQUIADES et al., 2012).

#### 4.5 Análises Quimiométricas

O software MATLAB R2021b (The MathWorks Inc., Natick, EUA) e o pacote de ferramentas quimiométricas GAMMA (BONA, 2017) foram usados no desenvolvimento das análises multivariadas aplicadas aos espectros NIR e EDXRF.

#### 4.5.1 Pré-processamento dos Espectros

As abordagens de correção do espalhamento multiplicativo (MSC), 2ª derivada (2ªd) por Savitzky-Golay, e a combinação de ambas as técnicas (MSC + 2ªd) foram aplicadas aos espectros NIR e comparadas buscando a mínima perda de informações relevantes na diferenciação das amostras. Após a realização de testes, foram adotados para os dados NIR janela de 25 pontos e polinômio de 4º grau como parâmetros do método de Savitzky-Golay, por apresentarem melhor relação sinal/ruído e deformação aceitável dos picos espectrais.

Já os espectros EDXRF foram, primeiramente, alisados seguindo a técnica de Savitzky-Golay com janela de 15 pontos e polinômio de 3º grau. Na sequência, foram aplicados o MSC, 1ª derivada (1ªd) e os dois pré-tratamentos combinados (MSC + 1ªd). A derivação desses dados também foi realizada por Savitzky-Golay com janela de 15 pontos e polinômio de 3º grau.

#### 4.5.2 Análise por Componentes Principais – PCA

A PCA foi aplicada sobre a matriz de espectros NIR pré-tratada com MSC e centrada na média para a investigação de padrões entre as amostras, permitindo a visualização de agrupamentos e pesquisa das características importantes para essa separação. Desse modo, o sinal NIR foi decomposto em uma matriz *score* (**T**) contendo as coordenadas da amostra no novo espaço *n*-dimensional, e uma matriz *loading* (**L**) contendo a contribuição de cada banda espectral na formação das PCs (FERREIRA, 2015; KUCHERYAVSKIY; LOMBORG, 2015).

#### 4.5.3 Análise por Dimensões Comuns – ComDim

A análise exploratória ComDim foi adotada ao invés da PCA para os dados EDXRF, pois desejou-se avaliar se as condições de operação de 15 kV e 50 kV apresentam informações que se complementam, de modo a tornar o modelo preditivo mais robusto.

Para aplicação do ComDim, os dados EDXRF brutos foram organizados em duas tabelas contendo os espectros de 15 kV (tabela 1, 144 x 2048) e 50 kV (tabela 2, 144 x 2048). Após centrar as colunas na média, as tabelas foram normalizadas pela norma de Frobenius para apresentarem mesma variância total. Assim, é eliminado o efeito causado pela quantidade diferente de variáveis (energias de fluorescência) de cada tabela.

As matrizes quadradas  $\mathbf{W}_1$  e  $\mathbf{W}_2$  (144 x 144) foram construídas e utilizadas na determinação das contribuições dos espectros de 15 kV e 50 kV na construção das CDs (saliências  $\lambda$ ), e das coordenadas das 144 amostras no espaço das CDs (*scores*  $\mathbf{q}$ ). A quantidade de CDs necessárias foi determinada a partir das saliências e variância total acumulada pelas CDs (MAKIMORI; BONA, 2019).

#### 4.5.4 Regressão por Mínimos Quadrados Parciais – PLSR

Modelos de calibração para a previsão do conteúdo de teff (vetor  $\mathbf{y}$ ) presente nas amostras e, conseqüentemente, identificação de adulterações, foram construídos empregando o método de regressão multivariada por mínimos quadrados parciais (PLSR) sobre os espectros NIR pré-tratados por MSC, 2<sup>a</sup>d e sua combinação (matriz  $\mathbf{X}$ ). Para a normalização dos dados foram testados os métodos de centralização na média, autoescalamamento e pareto.

Para a construção e validação dos modelos de regressão, as amostras foram separadas nos conjuntos de calibração (67%) e previsão (33%) seguindo os algoritmos de seleção Kennard-Stone (KS) ou SPXY.

O número ideal de variáveis latentes foi determinado a partir da validação cruzada pelo método *k-fold* para os seis modelos, escolhendo-se a quantidade resultante dos menores RMSECV (raiz quadrada do erro quadrático médio de validação cruzada) (BONA; MARÇO; VALDERRAMA, 2018). Foi utilizado  $k = 10$ , sendo o conjunto de calibração dividido em 10 subconjuntos para treinar os modelos, desses 10 subconjuntos 1 é excluído do treinamento e usado na validação. Todos os subconjuntos são excluídos uma vez da calibração e usados para avaliar o modelo (ROHANI; TAKI; ABDOLLAHPOUR, 2018).

O desempenho dos modelos PLSR propostos foi avaliado pela determinação das figuras de mérito sensibilidade e inverso da sensibilidade analítica, limites de detecção e quantificação,  $R^2$ , RMSEC, RMSEP e RPD.

#### 4.5.5 Regressão Linear Múltipla – MLR

Os espectros EDXRF 15 kV e 50 kV, com diferentes pré-tratamentos (originais, alisados, MSC, 1<sup>a</sup>d e MSC + 1<sup>a</sup>d) foram fundidos empregando o ComDim, obtendo-se os *scores* (matriz  $\mathbf{Q}$ ) que foram utilizados na construção de modelos de calibração a partir do método MLR, os quais correlacionaram o percentual de teff nas amostras e os dados EDXRF para identificar e quantificar a adulteração nas farinhas.

Os desempenhos dos métodos KS e SPXY foram comparados na seleção das amostras dos conjuntos de calibração (67%) e previsão (33%) necessários para a construção, validação cruzada e teste dos modelos.

A quantidade de CDs necessárias no modelo MLR foi determinada pelo menor RMSECV observado na validação cruzada com método 8-fold (BONA; MARÇO; VALDERRAMA, 2018).

A performance dos modelos MLR foi avaliada a partir das figuras de mérito  $R^2$ , RMSEC, RMSEP e RPD, e figuras de mérito analíticas sensibilidade, sensibilidade analítica, limites de detecção e limite de quantificação.

As variáveis dos espectros EDXRF mais importantes ao melhor modelo foram identificadas a partir da avaliação dos coeficientes de regressão desse modelo em conjunto com os *loadings* do ComDim.

#### 4.5.6 Avaliação da Não-Linearidade dos Dados

Após a aplicação do PLSR e MLR, a natureza não-linear entre as concentrações de *teff* e os espectros NIR e EDXRF foi avaliada pela técnica APARP e teste de Durbin-Watson, (OLIVIERI, 2015; SHAN et al., 2019).

Para a técnica APARP, a concentração de *teff* foi decomposta nas  $n$  variáveis latentes que descrevem os modelos de previsão e no quadrado de sua primeira LV, conforme Equação 12 (SHAN et al., 2019):

$$\%_{teff} = b_0 + b_1LV_1 + b_2LV_2 + \dots + b_nLV_n \dots + b_{11}LV_1^2 + e_{APARP} \quad (12)$$

onde  $e_{APARP}$  é o resíduo não modelado pelas variáveis latentes.

Com base no  $e_{APARP}$ , foram calculadas as concentrações de *teff* modeladas apenas pela contribuição linear da primeira LV e quando inserida sua contribuição quadrática. Os gráficos dos resíduos aumentados foram construídos pela diferença entre as concentrações estimadas de modo linear e não-linear em função da primeira LV para verificar a tendência de sua distribuição.

O APARP foi modificado para as regressões MLR com a substituição dos *scores* das LV pelos *scores* das CD dos modelos na Equação 11.

Por fim, aplicou-se o teste de Durbin-Watson (Equação 13) sobre os resíduos aumentados dos modelos NIR e EDXRF, sendo a não-linearidade estatisticamente significativa para as probabilidades menores que 0,05 (OLIVIERI, 2015).

$$DW = \frac{\sum_{i=1}^{I-1} (e_{i+1} - e_i)^2}{\sum_{i=1}^I (e_i)^2} \quad (13)$$

#### 4.5.7 Importância da variável na projeção

Uma análise das variáveis espectrais mais importantes na previsão do teor de teff para os modelos NIR foi realizada utilizando a técnica de seleção de variáveis VIP (Importância da Variável na Projeção).

O método foi aplicado ao modelo de melhor desempenho de acordo com as figuras de mérito, onde uma pontuação VIP foi calculada para cada variável espectral. Essa pontuação relaciona a contribuição da variável espectral na construção da variável latente do modelo (peso  $w$ ) com a variância do conteúdo de teff explicada por esta variável latente. Nesse método, quanto maior o valor do VIP maior será a importância daquela variável para o modelo (FARRÉS et al., 2015).

## 5 RESULTADOS E DISCUSSÃO

### 5.1 Composição Proximal

Os resultados para a composição proximal de ambas as safras de teff (A e B) e farinhas adulterantes (arroz, aveia, centeio e trigo integral) se encontram na Tabela 2, os quais concordam com os relatados na literatura.

Farinhas de arroz cultivado de modo convencional e orgânico apresentaram cinzas entre 0,42 – 0,63 g 100 g<sup>-1</sup>, lipídeos totais 0,51 – 0,53 g 100 g<sup>-1</sup> e proteína bruta 8,31 – 9,39 g 100 g<sup>-1</sup> (KAUR et al., 2015). Para diferentes cultivares de grão de arroz, foram relatados percentuais de proteína bruta e carboidrato de 7,67% e 79,38%, respectivamente (BHAT; RIAR, 2019). Para farinha de centeio e aveia (TORBICA; BELOVIĆ; TOMIĆ, 2019) foi obtido nessa ordem, 1,53 % e 1,41 % de cinzas, 2,52 % e 4,88 % de lipídeos totais e 12,90 % e 8,33 % de proteína bruta. Para a farinha de aveia, também foram obtidas as composições proximais de 1,28 – 1,40 g 100 g<sup>-1</sup> de cinzas, 7,01 – 7,24 g 100 g<sup>-1</sup> de lipídeos totais e 13,7 – 14,7 g 100 g<sup>-1</sup> de proteína bruta (FERREIRA; TEIXEIRA TARLEY, 2021). Para a farinha integral de trigo duro foram relatados um conteúdo de cinzas de 1,62 %, lipídeos totais de 1,82 % e proteína bruta de 14,4 % (HESHE et al., 2016).

As farinhas apresentaram diferença significativa no teor de umidade, exceto para o arroz e aveia. Apesar dessa diferença, todas as amostras tiveram umidade inferior a 15% conforme estabelecido pela legislação brasileira para farinhas de cereais (ANVISA, 2005). Devido a essa diferença de umidade, as amostras foram mantidas no mesmo ambiente (dessecador *drybox*) para uma padronização das umidades antes da aquisição dos espectros NIR e EDXRF.

O conteúdo mineral superior aos demais cereais característico do teff, apontado pela literatura, ficou demonstrado pelo percentual de cinzas significativamente maior nos teff A e B (2,38% e 2,72%, respectivamente) do que nas demais farinhas. Observou-se ainda, diferença com significância de 5 % no teor de cinzas entre as safras de teff, para o qual a safra de 2016 (teff B) apresentou maior concentração mineral que a safra 2018 (teff A). Essa variação pode decorrer da composição do solo e condições climáticas durante o cultivo das safras (GEBREMARIAM; ZARNKOW; BECKER, 2014; ZHU, 2018).

Centeio e trigo integral apresentaram as maiores frações proteicas, as quais se mostraram estatisticamente equivalentes, assim como as farinhas de aveia e teff B. O teff possui percentual de proteína bruta similar ao trigo e superior ao centeio, diferentemente do observado

neste trabalho, no qual ambas as farinhas de teff tiveram concentrações inferiores de proteína bruta. Entretanto, o teff é composto pela maioria dos aminoácidos essenciais e em maiores quantidades do que o trigo e centeio, sendo nutricionalmente mais interessante (BARRETTO et al., 2020; GEBREMARIAM; ZARNKOW; BECKER, 2014).

A farinha de aveia apresentou maior teor de lipídeos totais (6,79%), seguida das farinhas de teff A e B que foram estatisticamente iguais. O centeio e o trigo integral também foram significativamente similares em relação a esse macronutriente. De acordo com trabalhos anteriores (BULTOSA, 2016; GEBREMARIAM; ZARNKOW; BECKER, 2014), a fração lipídica do teff é inferior à aveia, mas superior ao trigo e centeio, resultados que também foram verificados nesta pesquisa. Embora o teff tenha como ácido graxo predominante o ácido linoleico, sua proporção n-6/n-3 (5,59) é reduzido em comparação ao arroz (42,5), aveia (17,75), trigo (15,21) e centeio (8,13) (BULTOSA, 2016; CHUNG et al., 2009; ROSENTRATER; EVERS, 2018), podendo auxiliar no consumo equilibrado desses ácidos graxos essenciais a fim de reduzir processos inflamatórios decorrentes do excesso de n-6 na dieta (SIMOPOULOS, 2016).

Os carboidratos são os macronutrientes em maior quantidade nos cereais, correspondendo a aproximadamente 75 % de sua composição (LASKOWSKI et al., 2019). Com exceção do arroz, todas as farinhas apresentaram carboidratos próximos desse valor, porém já foi relatado para a farinha de arroz um conteúdo de carboidratos de 80 % (ROSEFL; MARCO, 2008).

Tabela 2 – Composição proximal das farinhas puras

<b>Farinha</b>	<b>Umidade (g 100 g<sup>-1</sup>)</b>	<b>Cinzas (g 100 g<sup>-1</sup>)</b>	<b>Proteína Bruta (g 100 g<sup>-1</sup>)</b>	<b>Lipídios Totais (g 100 g<sup>-1</sup>)</b>	<b>Carboidratos Totais (g 100 g<sup>-1</sup>)</b>
Teff A	8,5142 ± 0,0286 <sup>a</sup>	2,3842 ± 0,0839 <sup>e</sup>	12,0726 ± 0,6263 <sup>c</sup>	3,1639 ± 0,3253 <sup>c</sup>	73,8652 ± 0,6527 <sup>c</sup>
Teff B	11,7383 ± 0,0619 <sup>c</sup>	2,7194 ± 0,0070 <sup>f</sup>	10,7415 ± 0,1309 <sup>b</sup>	3,0219 ± 0,1518 <sup>c</sup>	71,7788 ± 0,0821 <sup>a</sup>
Arroz	9,0625 ± 0,0460 <sup>c</sup>	0,5465 ± 0,0088 <sup>a</sup>	7,7198 ± 0,2182 <sup>a</sup>	0,4739 ± 0,1461 <sup>a</sup>	82,1974 ± 0,3557 <sup>d</sup>
Aveia	9,0868 ± 0,0380 <sup>c</sup>	0,8218 ± 0,0026 <sup>b</sup>	10,8440 ± 0,0940 <sup>b</sup>	6,7914 ± 0,2123 <sup>d</sup>	72,4561 ± 0,3057 <sup>ab</sup>
Centeio	9,5821 ± 0,0172 <sup>d</sup>	1,3437 ± 0,0091 <sup>c</sup>	13,1987 ± 0,1849 <sup>d</sup>	2,0818 ± 0,4163 <sup>b</sup>	73,7936 ± 0,2345 <sup>c</sup>
Trigo Integral	8,9238 ± 0,0802 <sup>b</sup>	1,6769 ± 0,0081 <sup>d</sup>	13,9021 ± 0,0621 <sup>d</sup>	2,1163 ± 0,0452 <sup>b</sup>	73,3810 ± 0,0973 <sup>bc</sup>

Média ± desvio padrão de análises em triplicata. Médias na mesma coluna seguidas por letras diferentes indicam amostras com diferenças estatisticamente significativas, a um nível de 5% de significância, pelo Teste de Tukey.

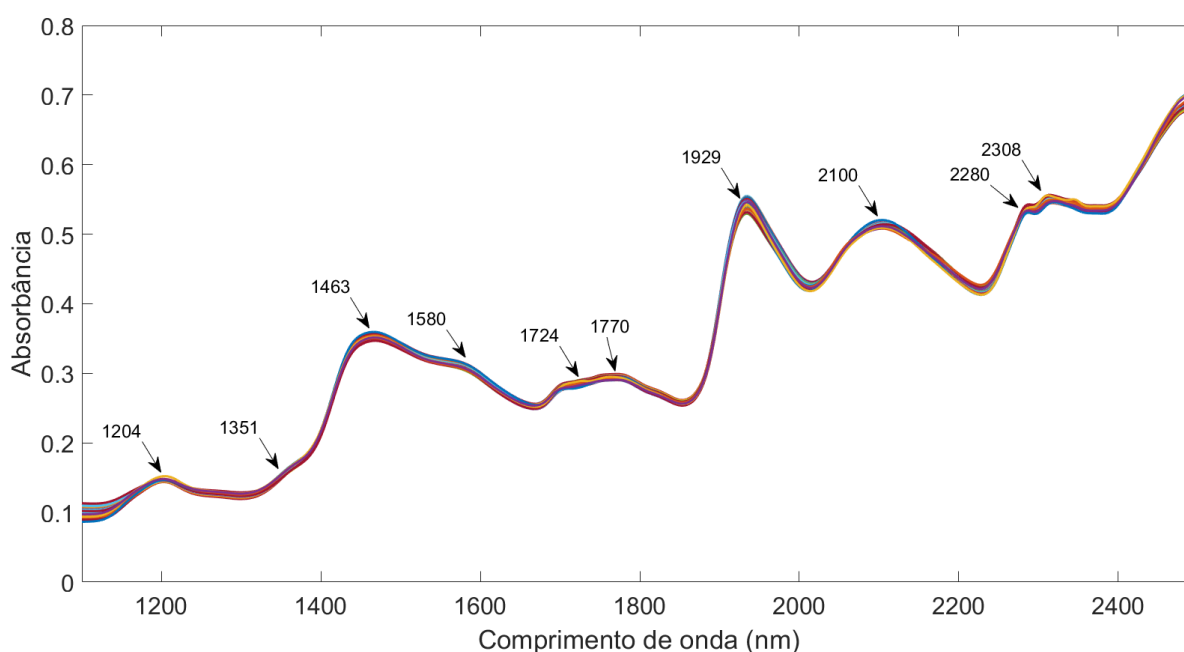
**Fonte: Autoria Própria (2021).**



## 5.2 Análise exploratória dos espectros NIR

Os espectros NIR na região de 1100 – 2500 nm pré-tratados com MSC para as 144 amostras são observados na Figura 4, na qual foram destacadas as principais bandas espectrais verificadas nas farinhas. A relação dos picos identificados com os respectivos grupos funcionais se encontra no Apêndice B.

**Figura 4 – Espectros NIR das amostras de farinhas e adulterantes pré-tratados com MSC**



**Fonte: A autoria Própria (2021).**

Em 1770 nm encontrou-se uma banda relacionada à região do primeiro sobretom da vibração de estiramento dos grupamentos C-H, as bandas 1204 nm e 1351 nm são referentes ao segundo sobretom de estiramento desses grupos, enquanto sua flexão ocorreu em 2308 nm, podendo essas vibrações serem atribuídas à fração lipídica presente nas farinhas (BURNS; CURCZAK, 2021; JIANG; LIU; CHEN, 2020). De acordo com BURNS e CURCZAK (2021), a região de 2310 nm é característica de lipídeos, o que reforça a associação anterior. Já a carbonila (C=O) da estrutura lipídica apresentou vibração em 1724 nm e 2350 nm, sendo este último relacionado ao terceiro sobretom de seu estiramento (BURNS; CURCZAK, 2021; TOMAR et al., 2021).

Nas farinhas de cereais, o constituinte predominante é o amido, logo, este é o principal responsável pelas vibrações em 2100 nm e 1929 nm atribuídas às hidroxilas presentes nos

carboidratos (BURNS; CURCZAK, 2021). O pico em 1929 nm relaciona-se com as vibrações de flexão e estiramento do grupo funcional O-H das moléculas de amilose formadoras do amido, e o pico 1463 nm ao estiramento dessas hidroxilas (TOMAR et al., 2021). A glicose, monômero dos grânulos de amido, gera bandas de combinação em 2280 nm devido às ligações O-H e C-O em sua estrutura (BURNS; CURCZAK, 2021).

As proteínas também contribuem para os sinais no infravermelho próximo, sendo as bandas 1505 nm, 1580 nm e intervalo de 1463 – 1470 nm referentes à ligação N-H. Na região de 2065 nm ocorrem vibrações decorrentes do grupo funcional N-H da amida primária (CONH<sub>2</sub>), a banda em 2150 nm é associado às ligações nitrogenadas N-H e C-N, e à carbonila (BURNS; CURCZAK, 2021).

O gráfico de *scores* da PCA para os dados NIR-MSC se encontra na Figura 5. A variância total explicada pelas três primeiras PCs foi 95,67%, contudo, a terceira PC não trouxe informações relevantes quanto aos agrupamentos, por isso, a distribuição das amostras foi avaliada em PC1 e PC2, que resultaram em 88,72% da variância acumulada.

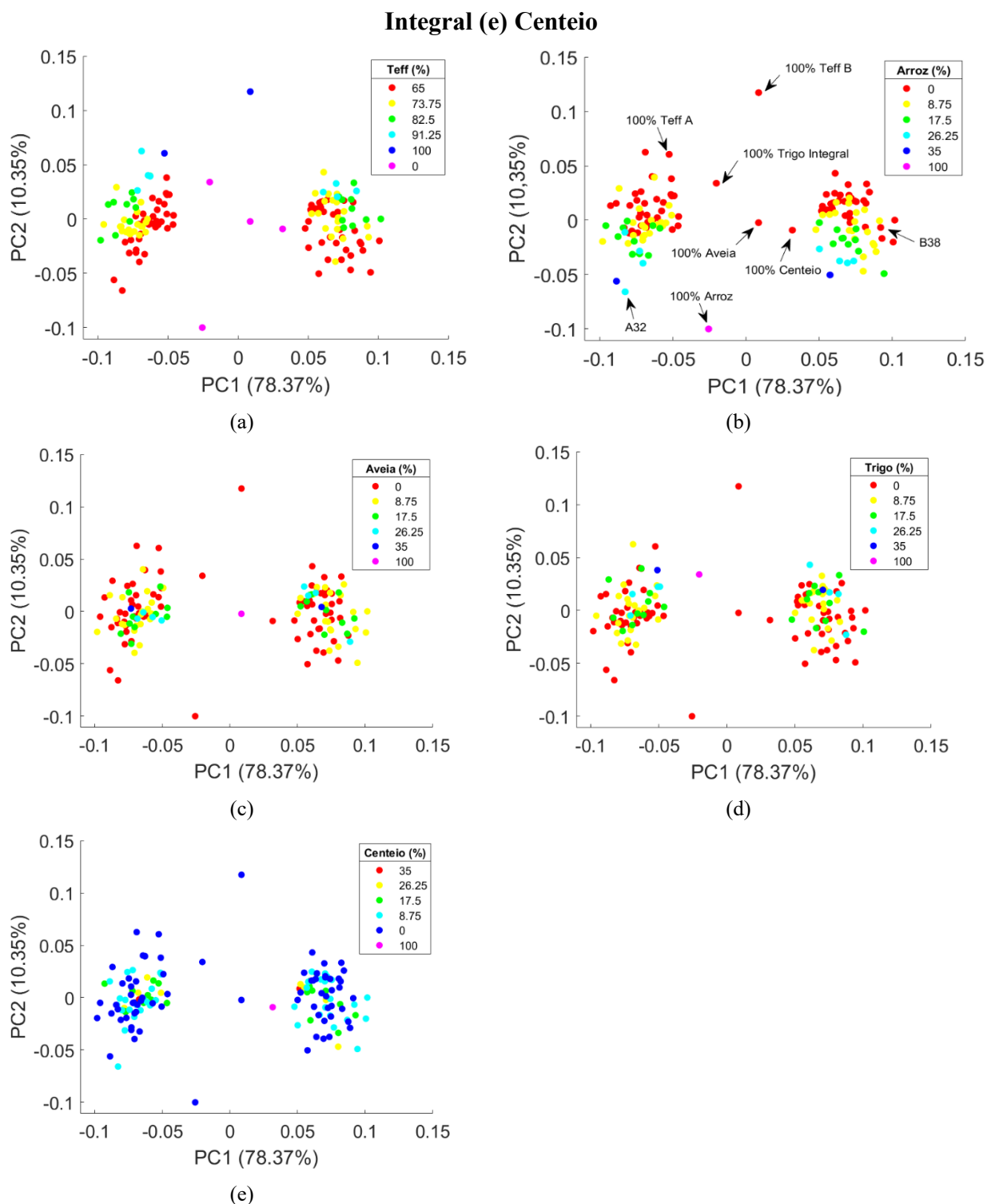
Nas Figuras 5a-e foram destacadas as contribuições de cada farinha pura nos agrupamentos. Observou-se que a PC1 foi responsável por separar as safras de teff, na qual para *scores* negativos encontram-se as amostras da safra 2018, e para *scores* positivos a safra 2016. Os adulterantes foram separados das misturas e distribuídos ao centro da PC1. Conforme a Figura 5a, ambas as PCs contribuem para a separação de acordo com o percentual de teff, dada a distribuição diagonal das amostras. As amostras com maiores teores de teff tendem a apresentar maiores *scores* em PC2.

Com relação ao arroz, as amostras com maior percentual foram deslocadas para *scores* negativos de PC2, o que se deve ao *score* negativo apresentado pela farinha de arroz pura. De forma análoga, as amostras adulteradas com maiores concentrações de aveia e centeio (Figuras 5c e 5e) foram agrupadas centralizadas na PC2, dado o *score* próximo de zero desses adulterantes. Já a adição de trigo integral às farinhas de teff levam as amostras a terem *scores* positivos (Figura 5d).

A posição das amostras no espaço PC1 × PC2 é influenciada pelo teor de teff e pelo tipo de adulterante. Por exemplo, a amostra A32 é composta por 65% de teff A, 26,25% arroz e 8,75% centeio, enquanto a composição da amostra B38 é 73,75% teff B, 8,75% trigo integral e 17,5% aveia. Nota-se na Figura 5b que a amostra A32 se encontra afastada da amostra 100% teff A devido ao menor teor desse componente, assumindo *score* negativo em PC2. Com relação aos adulterantes, essa amostra está mais próxima da amostra 100% arroz com *score* negativo em PC2 e não centralizada como a amostra 100% centeio, isso ocorre devido ao maior teor de

arroz na mistura do que centeio. Por outro lado, a amostra B38 se encontra mais próxima da amostra 100% teff B com *score* em PC2 próximo de valores positivos, afastada da amostra 100% trigo, cujo conteúdo é o menor, e próxima da amostra 100% aveia devido ao maior teor desse adulterante.

**Figura 5 – Gráficos dos scores da PCA dos espectros NIR-MSC com separação das amostras de acordo com o percentual de farinha de (a) Teff (b) Arroz (c) Aveia (d) Trigo**



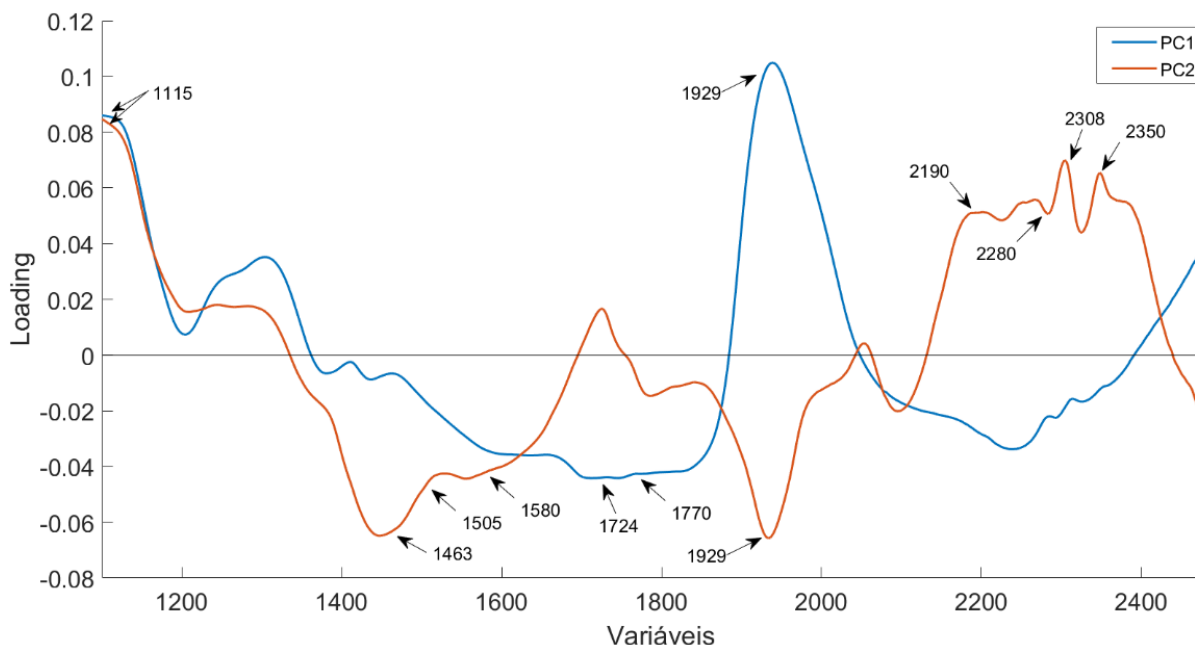
Fonte: Autoria Própria (2021).

A distribuição das farinhas puras em PC2 sugere uma correlação positiva com o teor de cinzas e correlação negativa com carboidratos (Figura 5 e Tabela 2), ou seja, as amostras

com baixa concentração de cinzas e alta de carboidratos apresentaram *scores* mais negativos (arroz, aveia, centeio), enquanto aquelas com alto percentual de cinzas e baixo de carboidratos tiveram *scores* positivos (teff A e B e trigo integral). As correlações do *score* em PC2 com os teores de cinzas e carboidratos foram confirmadas pelos coeficientes de correlação de Pearson de 0,9320 ( $p = 0,0068$ ) e -0,8283 ( $p = 0,0417$ ), respectivamente, comprovando que existe uma relação entre a distribuição das amostras na PC2 com os teores desses componentes (KOTU; DESHPANDE, 2019).

A contribuição dos comprimentos de onda no infravermelho próximo para a formação do espaço das PCs 1 e 2 encontra-se na Figura 6 (FERREIRA, 2015).

**Figura 6 – Gráfico de *loadings* da PCA para os espectros NIR-MSC**



**Fonte: Autoria Própria (2021).**

De acordo com as Figuras 5 e 6, a classificação das amostras em PC1 negativa teve contribuição de bandas de lipídeos (1724 nm e 1770 nm), enquanto para PC1 positiva as regiões de maiores pesos foram de lipídeos (1115 nm) e carboidratos (1929 nm). Já para PC2, as amostras de teff e trigo, *scores* positivos, foram classificadas por bandas de lipídeos (1115 nm, 2308 nm e 2350 nm), carboidratos (2280 nm) e proteínas (2190 nm) que apresentaram pesos positivos, enquanto para os *scores* negativos da aveia, centeio e arroz, foram utilizadas as bandas de carboidratos (1463 nm e 1929 nm) e proteínas (1505 nm e 1580 nm) com *loading* negativos.

O intervalo de 2190 nm – 2350 nm é responsável pela correlação positiva entre o teor de cinzas e os *scores* de PC2. As cinzas provenientes da incineração da matéria orgânica não apresentam banda de absorção características no NIR, contudo, a banda em 2336 nm para celulose pode ser atribuída às cinzas por ser o principal polissacarídeo presente na casca dos cereais, portanto, encontrado em concentrações importantes em farinhas integrais (BURNS; CIURCZAK, 2008). Isso explica os *scores* positivos das farinhas integrais de teff e trigo. Além disso, a utilização de bandas características de proteínas e lipídeos na determinação de cinzas por NIR em farinhas é recorrente devido à maior quantidade desses macronutrientes no farelo, portanto, nas farinhas integrais (BURNS; CIURCZAK, 2008). Isso pode estar relacionado aos minerais quelatos que se encontram nessa forma em cereais por sua ligação com proteínas e carboidratos, principalmente.

### 5.3 Calibração multivariada usando os espectros NIR

Os modelos PLSR obtidos para os dados NIR que apresentaram melhores desempenhos estão dispostos na Tabela 3 com seus respectivos pré-tratamentos, algoritmo de amostragem e figuras de mérito. Para todos os modelos da Tabela 3 foram utilizados dados centrados na média visto que apresentaram melhor desempenho em relação ao autoescalamamento e pareto.

A quantidade mínima de variáveis latentes utilizada para cada um dos modelos (Tabela 3) foi aquela que resultou no menor valor de RMSECV e maior valor de  $R^2_{cv}$  de modo a obter redução no erro de validação externa referente às amostras desconhecidas para o modelo (RMSEP). Os modelos com os espectros NIR pré-tratados com MSC e segunda derivada (2d) apresentaram maior quantidade de LVs, um indicativo de maior complexidade para o modelo e, possivelmente, maior quantidade de ruído modelado o que diminui a robustez do modelo de previsão. Por outro lado, os modelos usando os pré-tratamentos MSC e segunda derivada combinados (MSC-2d) apresentaram um valor de RMSEP ligeiramente superior, porém com uma quantidade menor de LVs utilizadas, sendo, portanto, modelos mais simples e com desempenhos satisfatórios.

Com base nas figuras de mérito (Tabela 3), o modelo NIR-MS-2d com amostragem SPXY mostrou performance com razoável superioridade em relação aos dados amostrados por Kennard-Stone, evidenciada pela redução no erro de previsão  $RMSEP = 0,0209$  e  $RPD_{previsão} = 3,3809$  com  $R^2_{previsão} = 0,9197$ . O valor de RPD superior a 3 para modelos no segmento de

alimentos sugere que todas as regressões são excelentes e podem ser utilizadas para supervisões qualitativas e quantitativas (NICOLAÏ et al., 2007; QUELAL-VÁSCONEZ et al., 2018).

Os valores baixos obtidos para sensibilidade e inverso da sensibilidade analítica (Tabela 3) mostram que este modelo apresentou bons resultados para esses parâmetros. O inverso da sensibilidade analítica sugere que variações 1,05% na concentração de teff nas amostras podem ser distinguidas pelo infravermelho próximo (BRAGA; POPPI, 2004; VALDERRAMA; BRAGA; POPPI, 2007, 2009). A regressão NIR-MS-2d-SPXY também demonstrou resultados satisfatórios de LOD e LOQ, sendo possível detectar concentrações de teff a partir de 3,8% e quantificar conteúdos de 11%.

**Tabela 3 – Modelos de calibração NIR-PLS validados utilizando figuras de mérito**

Modelo	Pré-tratamento	Amostragem	LV	DW		
1	MSC	Kennard-Stone	7	0,6289		
2		SPXY	7	0,3688		
3	2ª derivada	Kennard-Stone	7	0,0047		
4		SPXY	8	0,0265		
5	MSC + 2ª derivada	Kennard-Stone	4	0,2504		
6		SPXY	4	0,5859		
Calibração						
		R <sup>2</sup>	RMSE	RPD		
1	MSC	0,9731	0,0281	6,1321		
2		0,9744	0,0276	6,2771		
3	2ª derivada	0,9695	0,0297	5,7580		
4		0,9796	0,0243	7,0309		
5	MSC + 2ª derivada	0,9706	0,0291	5,8644		
6		0,9694	0,0299	5,7487		
Previsão						
		R <sup>2</sup>	RMSE	RPD		
1	MSC	0,9441	0,0163	4,2242		
2		0,9316	0,0173	3,7675		
3	2ª derivada	0,9156	0,0233	3,1841		
4		0,9239	0,0213	3,4865		
5	MSC + 2ª derivada	0,9236	0,0230	3,2061		
6		0,9197	0,0209	3,3809		
Figuras de Mérito Analíticas						
	Sens.	Sens. Analítica <sup>-1</sup>	LODmin	LODmax	LOQmin	LOQmax
1	0,0600	0,0026	0,0158	0,0749	0,0475	0,2246
2	0,0588	0,0027	0,0157	0,0737	0,0472	0,2210
3	$1,1497 \times 10^{-4}$	0,0057	0,0236	0,0755	0,0709	0,2265
4	$9,1121 \times 10^{-5}$	0,0069	0,0257	0,0659	0,0770	0,1976
5	$1,4230 \times 10^{-4}$	0,0104	0,0372	0,0776	0,1116	0,2328
6	$1,4230 \times 10^{-4}$	0,0105	0,0376	0,0789	0,1129	0,2366

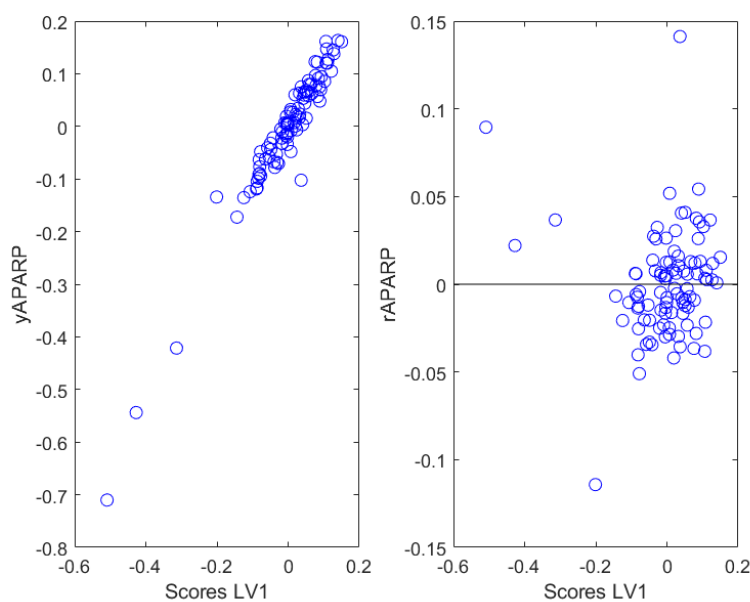
LV: variável latente. DW: teste de Durbin-Watson. R<sup>2</sup>: coeficiente de determinação (previsto vs. Observado). RMSE: raiz quadrada do erro quadrático médio. Sens: sensibilidade analítica expressa em unidade do sinal  $\times$  (m/m)<sup>-1</sup>. Sens. Analítica<sup>-1</sup>: inverso da sensibilidade analítica expresso em (m/m). LOD: limite de detecção expresso em (m/m). LOQ: limite de quantificação expresso em (m/m).

**Fonte: Autoria Própria (2021).**



Pela inspeção visual do gráfico dos resíduos parciais aumentados (APARP) para a regressão NIR-MSC-2d-SPXY (Figura 7), nota-se que as porcentagens previstas ( $y_{APARP}$ ) com 4 LVs e a contribuição quadrática da primeira LV mantém uma relação linear com a primeira LV, o que sugere a linearidade dos dados. De acordo com o resíduo do  $y_{APARP}$  previsto apenas pela contribuição linear da primeira LV ( $r_{APARP}$ ), não há tendência evidente nos resíduos, sendo assim, há indícios da linearidade entre a porcentagem de teff e os dados NIR. Com exceção dos modelos NIR-2d, os modelos não apresentaram não-linearidade significativa sendo os p-valores do teste de Durbin-Watson inferiores a 0,05 (Tabela 3), dessa forma a regressão PLSR é adequada aos dados. Visto que o modelo NIR-MSC-2d-SPXY eleito com melhor performance se mostrou linear, métodos não lineares de regressão, como máquinas de vetor suporte (SVM), não foram aplicados neste estudo (MAGALHÃES DE AGUIAR et al., 2020; OLIVIERI, 2015).

**Figura 7 – Gráfico APARP do modelo NIR-MSC-2d-SPXY**

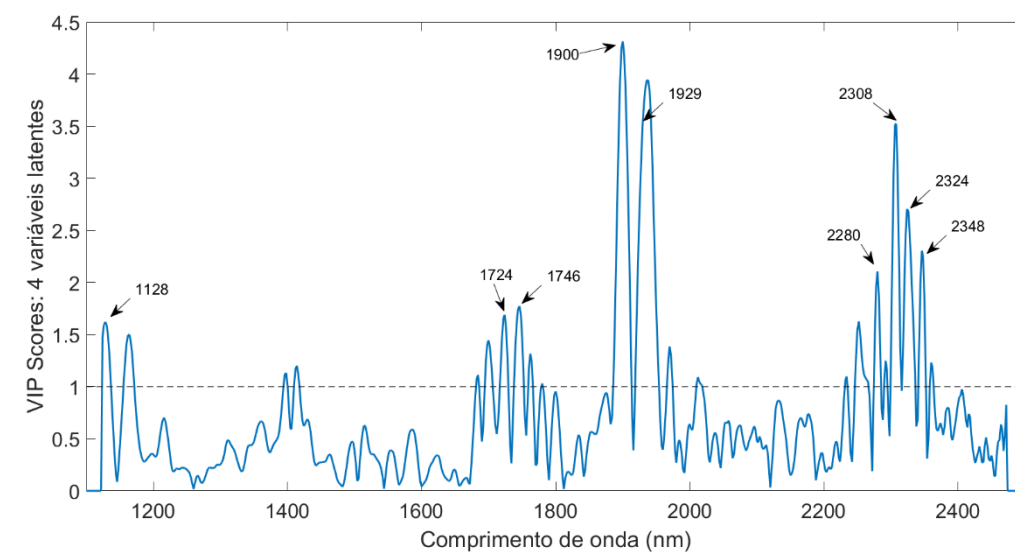


**Fonte: Autoria Própria (2021).**

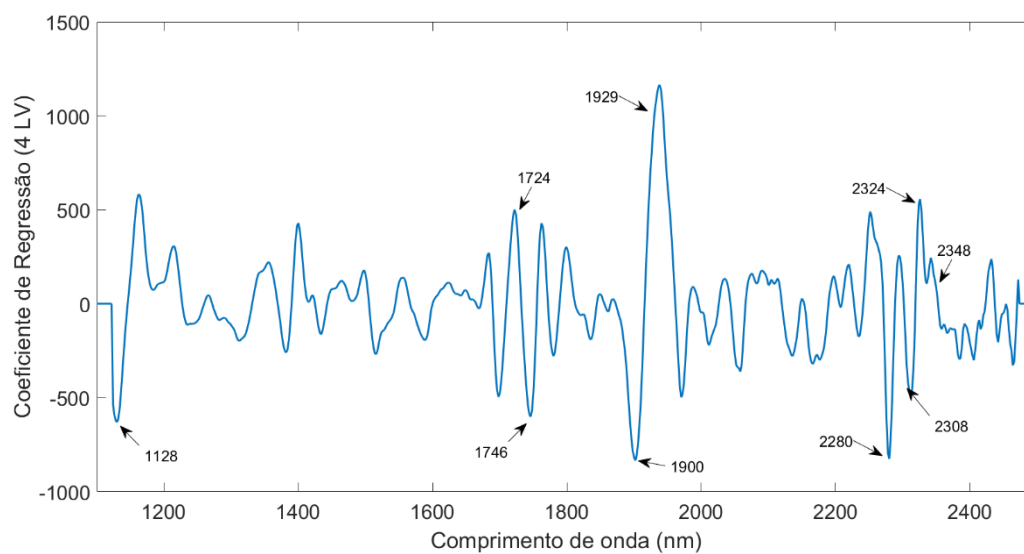
A Figura 8a destaca os comprimentos de onda que apresentaram maiores VIP scores para a regressão NIR-MSC-2d-SPXY, sendo os mais importantes para a previsão de teff pelo modelo. As bandas em 1128 nm, 1724 nm, 2308 nm e 2348 nm estão relacionados às vibrações de moléculas de lipídeos. O pico em 1900 é característico de grupos CO e COOH de ácidos carboxílicos, o qual pode estar relacionado também aos lipídeos (BURNS; CURCZAK, 2021). A região em 1746 nm pode ser atribuída aos grupamentos CH de amins, logo, há relação com proteínas. As demais regiões de 1929 nm, 2280 nm e 2324 nm correspondem aos carboidratos.

De acordo com os coeficientes de regressão do modelo PLS com 4 LVs (Figura 8b), os comprimentos de onda de 1724 nm, 1929 nm, 2324 nm e 2348 nm de vibrações de carboidratos e lipídeos contribuem para o aumento do teor de teff nas amostras, pois apresentaram coeficientes de regressão positivos, enquanto as bandas 1128 nm, 1746 nm, 1900 nm, 2280 nm e 2308 nm estão relacionadas à redução do conteúdo de teff dado os coeficientes negativos.

**Figura 8 – Gráfico (a) VIP para modelo com 4 LV (b) Coeficientes de regressão do modelo PLS com 4 LV**



(a)



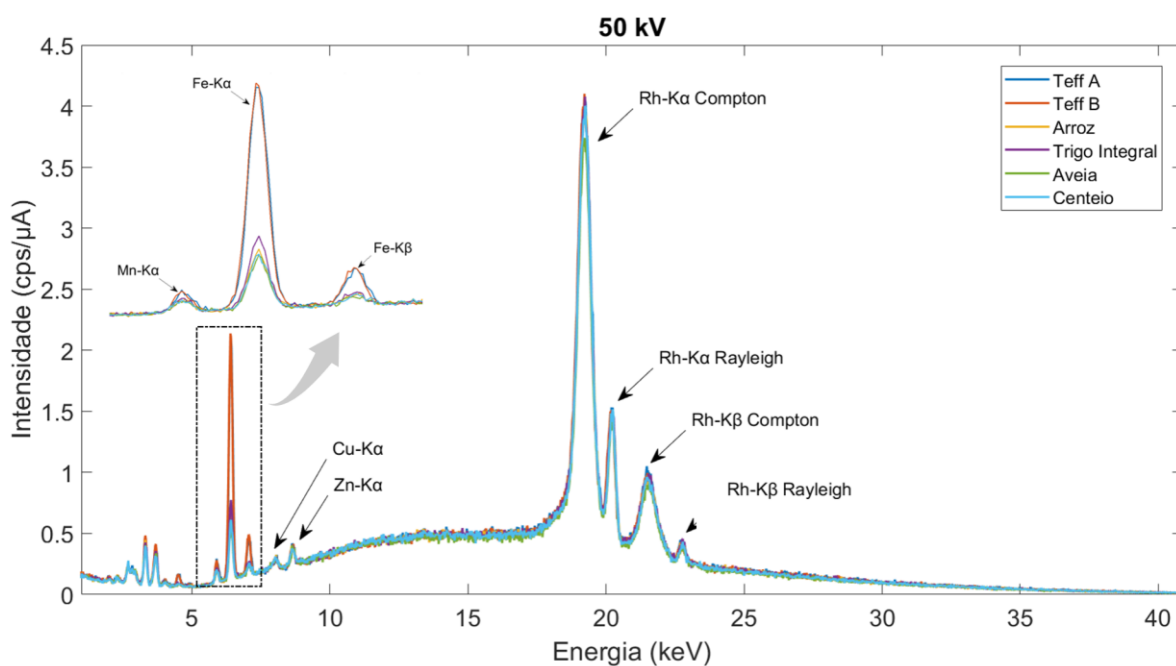
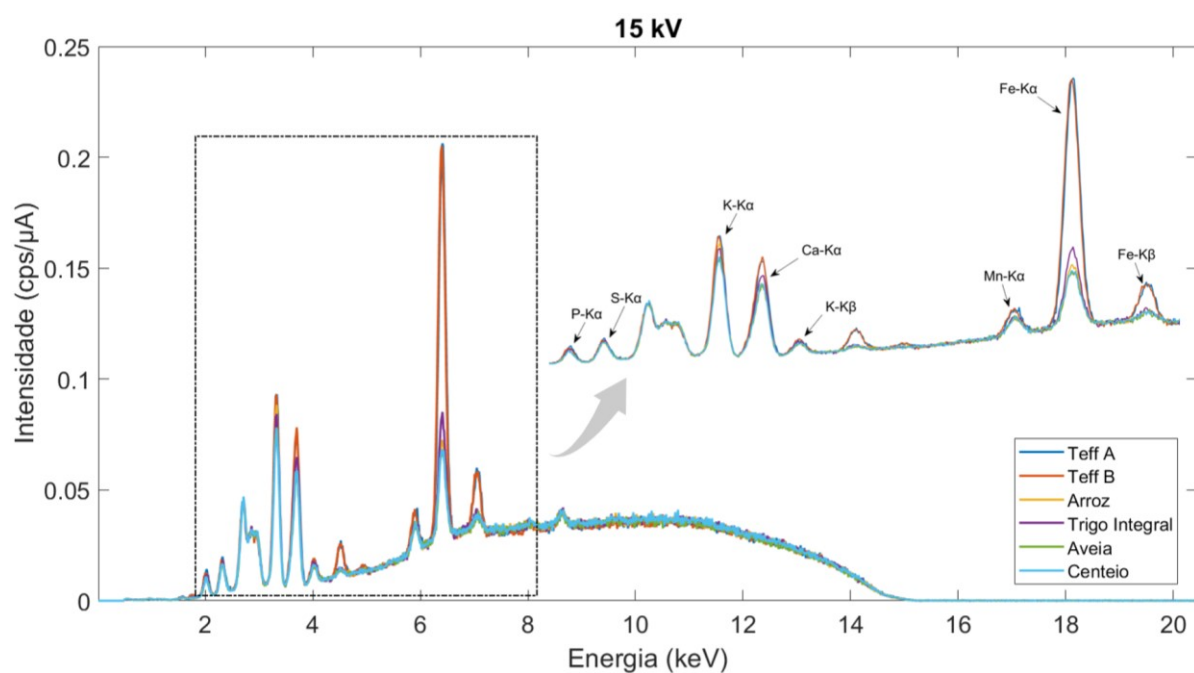
(b)

**Fonte: Autoria Própria (2021).**

#### 5.4 Análise exploratória dos espectros EDXRF

Os espectros, sem pré-processamento, das análises EDXRF para as farinhas puras (adulterantes e teff) se encontram nas Figuras 9a e 9b, nas quais os principais picos foram destacados e identificados. No Apêndice C se encontram os principais elementos químicos presentes nas amostras de farinhas e suas respectivas energias de excitação por raios-X.

**Figura 9 - Espectros EDXRF das amostras puras de teff e adulterantes (a) 15 kV (b) 50 kV**



**Fonte: Autoria Própria (2022).**

A Figura 9a revela a presença de fósforo e enxofre nas farinhas pelos picos em 2,02 keV e 2,31 keV, respectivamente, assim como potássio (3,3 keV), cálcio (3,7 keV e 4 keV), manganês (5,9 keV) e ferro (6,4 keV e 7,07 keV). Além desses minerais, a Figura 9b mostra

picos em 8,05 keV e 8,64 keV mais definidos, os quais são relativos ao cobre e zinco. O ródio emite energia em 20,21 keV e 22,72 keV, logo, esses picos correspondem aos espalhamentos Rayleigh, nos quais toda a energia incidente pelo tubo com alvo Rh é refletida pela matriz. Já os picos próximos à região Rayleigh são característicos do espalhamento Compton, dada a absorção parcial da energia. Nota-se que as transições  $K\beta$  observadas ocorrem com menor frequência do que as transições  $K\alpha$ , isso se deve à menor probabilidade de ocorrência da transição  $K\beta$  dos elementos químicos (RODRIGUES, 2005). As tensões de trabalho produzem perfis espectrais distintos, isso porque a condição de 15 kV tem maior probabilidade de excitar os elementos P, S, K e Ca, enquanto a condição de 50 kV excita de forma mais eficiente os elementos Mn, Fe, Cu e Zn.

De acordo com a literatura, o teff apresenta maior concentração de ferro, cálcio, magnésio, zinco, cobre e fósforo em relação a maioria dos cereais utilizados na alimentação, incluindo o trigo (GEBREMARIAM; ZARNKOW; BECKER, 2014). Entretanto, os resultados do EDXRF (Figura 9) sugerem que a composição mineral do teff quanto ao fósforo, cobre, zinco e enxofre é similar à composição das farinhas de arroz, trigo integral, aveia e centeio, uma vez que esses picos têm intensidades semelhantes para todas as farinhas. Já para o ferro, cálcio e manganês, houve similaridade entre as safras de teff e estas apresentaram concentrações superiores aos adulterantes desses minerais, o que corrobora com os resultados descritos na literatura.

As informações extraídas pelo ComDim referentes aos espectros EDXRF originais (sem pré-tratamento) das 144 amostras estão dispostos na Tabela 4, na qual foram descritas apenas as CDs que extraíram mais de 5% das informações existentes em cada tabela.

**Tabela 4 – Saliências das condições das análises EDXRF nas dimensões comuns**

Tabela	CD1	CD2	Variância acumulada da tabela (%)
15 kV	0,9103	0,0153	92,56
50 kV	0,5607	0,3470	90,77
Variância da CD (%)	92,6911	7,2025	
Variância acumulada (%)	92,6911	99,8936	

Fonte: Autoria Própria (2022).

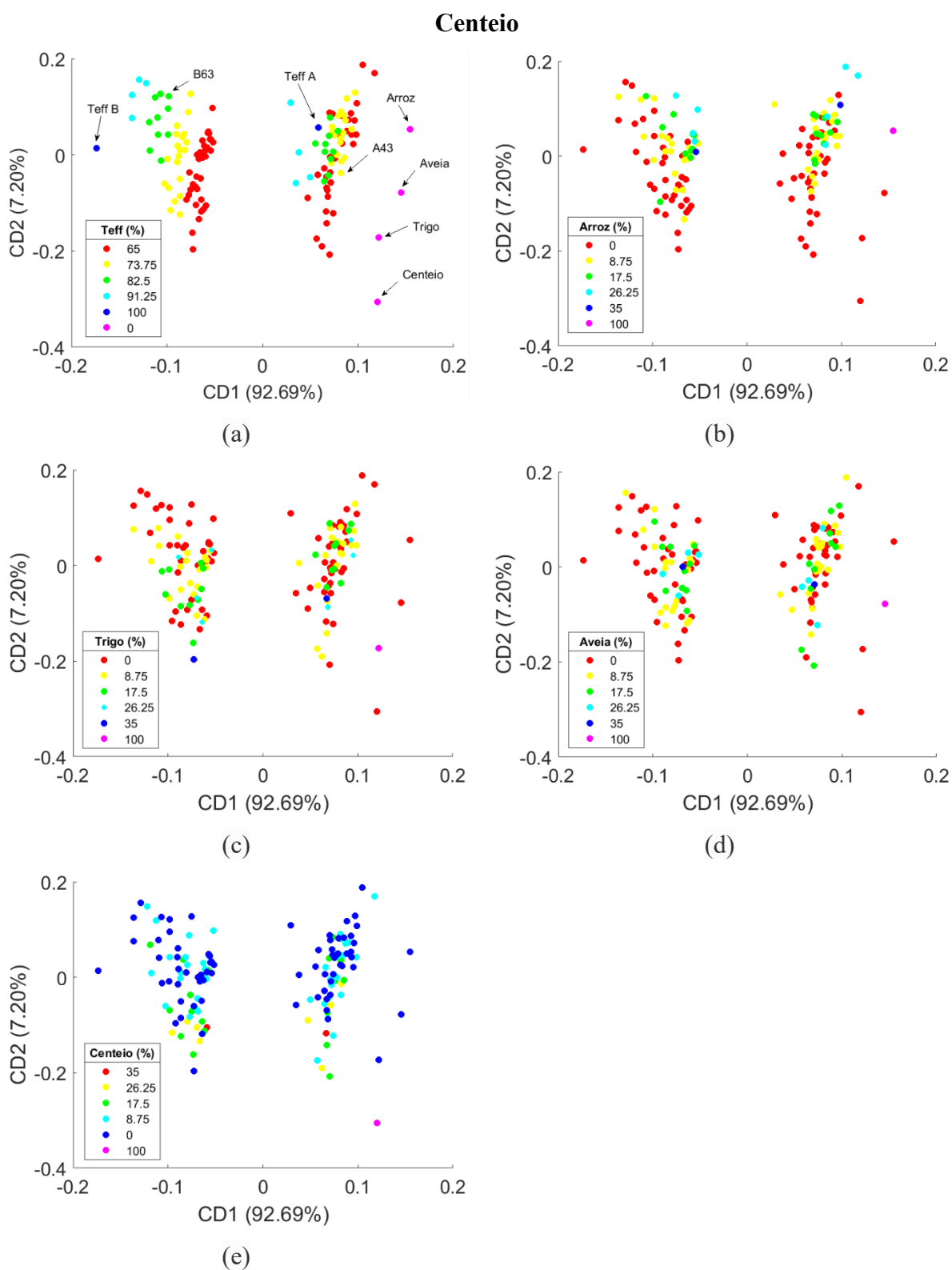
As variâncias acumuladas das tabelas mostram que o ComDim precisou de apenas duas dimensões comuns para reproduzir mais de 90% das informações contidas nas 2048

variáveis dependentes dos espectros de 15 kV e 50 kV. De acordo com as saliências, ambas as condições EDXRF tiveram importantes percentuais de suas variâncias explicadas pela CD1 (91,03% e 56,07%), sendo necessárias na identificação de semelhanças entre as amostras. Enquanto a CD2 agrupou as farinhas a partir dos resultados da técnica 50 kV, dada sua maior influência na construção da CD, logo, esse bloco é importante na investigação de padrões por trazer características das amostras que o espectro 15 kV não identificou ou o fez com baixa eficiência. Portanto, a ferramenta ComDim mostrou que a realização das análises EDXRF 15 kV e 50 kV é vantajosa, uma vez que suas informações se mostraram complementares nas CDs, o que possibilita assim a aplicação da fusão de dados é adequada para os dados do presente trabalho.

A distribuição das farinhas e misturas no novo espaço dimensional formado pela CD1 e CD2 é observada nos gráficos de *score* da Figura 10. Três agrupamentos bem definidos foram visualizados na CD1, os quais revelaram sua capacidade em discriminar misturas, adulterantes e as safras de teff. Os adulterantes apresentaram *scores* positivos na CD1, seguidos pelas misturas com teff da safra 2018, enquanto as misturas da safra 2016 possuem *scores* negativos na CD1. Por outro lado, a CD2 separou os tipos de adulterantes e, por consequência, distribuiu as misturas de acordo com a composição da adulteração. O *score* positivo da farinha de arroz indica que sua adição desloca as amostras para o quadrante positivo da CD2, assim como o centeio tende a deslocar as amostras para o quadrante negativo de CD2.

Para exemplificar o agrupamento das farinhas em  $CD1 \times CD2$ , as amostras B63 (82,5% teff B, 8,75% arroz e 8,75% aveia) e A43 (73,75% teff A, 17,5% trigo integral e 8,75% centeio) foram destacadas na Figura 10a. O alto percentual de teff aliado à presença de arroz atribuiu *score* positivo em CD2 à amostra B63, a qual se deslocou para próximo do teff B puro. A amostra A43 obteve *score* negativo em CD2 devido ao trigo e centeio, sendo encontrada próxima à amostra 100% trigo, o que sugere a maior contribuição do *score* do trigo dado seu maior percentual na mistura.

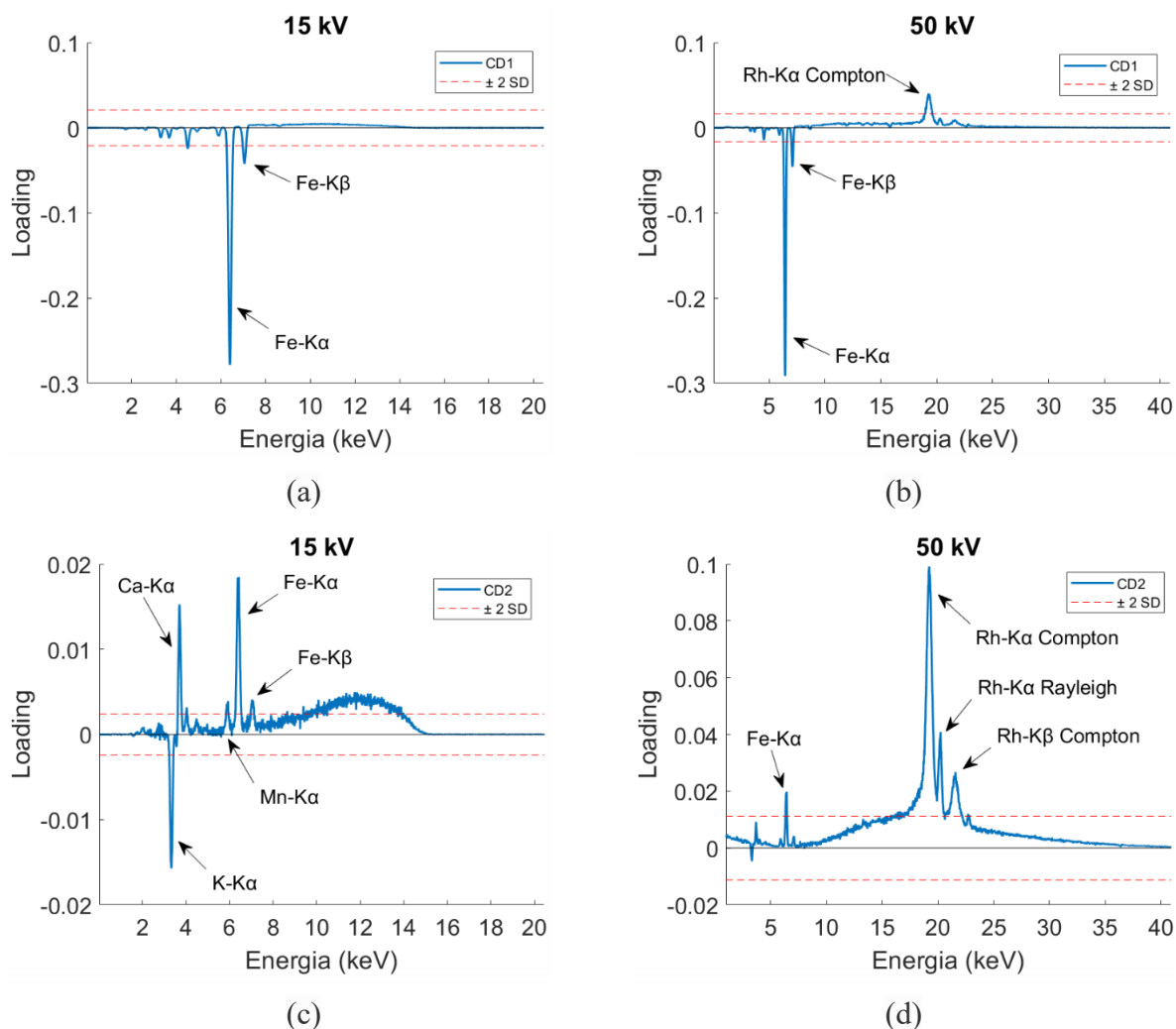
**Figura 10 – Gráficos dos *scores* do ComDim-EDXRF com separação das amostras de acordo com o percentual de farinha de (a) Teff (b) Arroz (c) Aveia (d) Trigo Integral (e)**



Fonte: Autoria Própria (2022).

Os gráficos de *loadings* mostram a natureza das informações dos espectros EDXRF-15kV e EDXRF-50kV utilizadas na construção de CD1 e CD2 (Figura 11).

**Figura 11 - Gráfico de *loadings* do ComDim para os espectros EDXRF originais**



Fonte: Autoria Própria (2022).

Na CD1, a informação compartilhada pelos dados de 15 kV e 50 kV, conforme demonstraram as saliências (Tabela 4), refere-se ao pico do ferro, que apresentou maior peso em ambos os blocos. O *loading* negativo desse pico indica que amostras com alto conteúdo de ferro apresentam menores *scores* em CD1, enquanto o aumento dos *scores* está atrelado à redução desse mineral. Por essa razão, os adulterantes foram separados das misturas (Figura 10) e originaram um agrupamento de farinhas pobres em ferro (*scores* positivos), o que evidencia a singularidade do conteúdo de ferro no teff frente a maioria dos cereais (BULTOSA, 2016; GEBREMARIAM; ZARNKOW; BECKER, 2014). A diferenciação entre os teff A e B



revelou maior teor de ferro na safra 2016, levando à hipótese de que a recuperação do solo para esse mineral não seja realizada entre as colheitas.

A importância dos espectros 15 kV na formação da CD2 estava relacionada aos picos de potássio, cálcio e ferro (Tabela 4 e Figura 11c). Assim, o aumento dos conteúdos de ferro e cálcio em conjunto com a diminuição de potássio contribuíram nos *scores* positivos das amostras. A região com maior peso em CD2 foi atribuída aos espalhamentos Compton e Rayleigh dos raios X identificados pela técnica 50 kV. Uma vez que o espalhamento apresenta relação com a interação da matéria orgânica da matriz com a radiação incidente, a composição de macronutrientes das farinhas é importante para o agrupamento em CD2, portanto, átomos constituintes, dimensão e conformação espacial das moléculas de proteínas, lipídios e carboidratos são fatores que influenciam essa separação (DOS SANTOS et al., 2020, 2021).

Sabendo-se que as farinhas de cereais são compostas majoritariamente por amido, uma possível hipótese é que os *scores* em CD2 são dependentes da conformação do grânulo desse polissacarídeo e como essa estrutura reflete os raios X. Os grânulos dos amidos de arroz e teff são poliédricos, ou seja, possuem várias faces que aumentam a superfície de contato com a radiação e, conseqüentemente, seu espalhamento, o que concorda com os *scores* positivos apresentados por esses cereais (GEBREMARIAM; ZARNKOW; BECKER, 2014; ZHOU et al., 2013). De forma análoga, os grânulos de amido do trigo e centeio são lenticulares o que reduz a superfície de contato, logo, o espalhamento, resultando em *scores* negativos em CD2 (ZHOU et al., 2013). Com relação à aveia, esta possui grânulos poliédricos como o arroz e esféricos como o trigo, obtendo *score* intermediário a esses cereais em CD2 (Figura 10). Além disso, os grânulos podem formar aglomerados que reduzem o espalhamento, o que pode ter tornado a aveia mais parecida ao trigo e centeio, justificando seu *score* negativo (ARENDDT; ZANNINI, 2013).

## 5.5 Calibração multivariada usando os espectros EDXRF

Os modelos MLR com melhor performance na quantificação de teff a partir dos dados ComDim-EDXRF se encontram na Tabela 5. Todas as regressões foram estatisticamente significativas ( $p$ -valor  $< 0,05$ ), portanto, os modelos lineares se adequaram aos dados. Independente do pré-tratamento empregado, o algoritmo de amostragem SPXY gerou os melhores resultados de regressão.

Todos os dados, com exceção do MSC-1d, mostraram ausência estatisticamente significativa de não-linearidade, pois os  $p$ -valores do teste de Durbin-Watson foram inferiores

a 0,05 (Tabela 5). A natureza linear dos dados foi visualizada nos gráficos da técnica APARP de cada modelo, que tiveram comportamento similar à Figura 12. Foi verificado que as porcentagens de teff previstas com as CDs e a contribuição quadrática da CD1 de cada modelo ( $y_{APARP}$ ) apresentaram relação linear com a primeira CD. Além disso, os resíduos dos  $y_{APARP}$  previstos apenas com contribuição linear da CD1 ( $t_{APARP}$ ), mostraram ausência de tendência nos resíduos, logo, há indícios de relação linear entre a porcentagem de teff e os dados EDXRF.

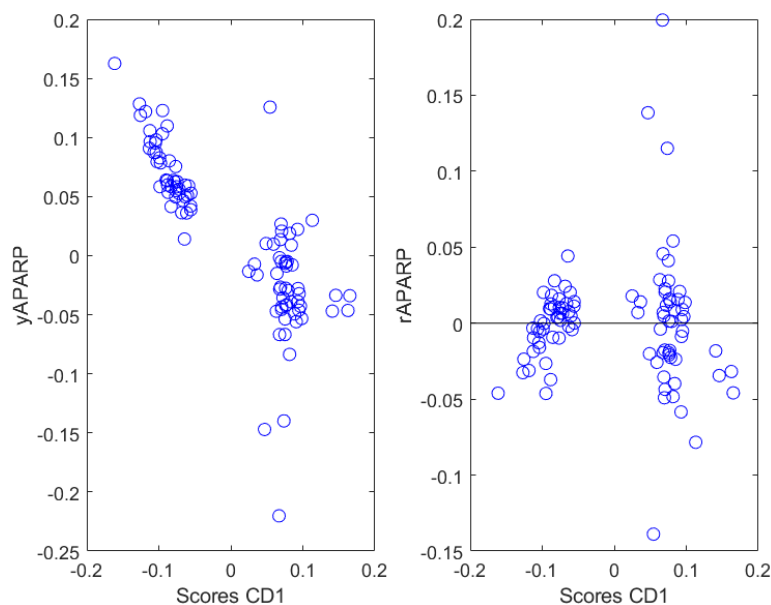
**Tabela 5 – Modelos de calibração ComDim-MLR validados utilizando figuras de mérito**

Modelo	Pré-tratamento	Região Espectral		Amostragem	CD	DW
		15 kV	50 kV			
1	-	0,01 – 20,48	0,02 – 40,96	SPXY	4	0,6453
2	Alisado	0,5 – 15	1 – 25	SPXY	4	0,2837
3	1ª derivada	0,5 – 15	1 – 25	SPXY	4	0,6286
4	MSC	0,5 – 15	1 – 25	SPXY	6	0,0961
5	MSC + 1ª derivada	0,5 – 15	1 – 25	SPXY	4	< 0,05
Calibração						
		R <sup>2</sup>	RMSE	RPD		
1	-	0,9478	0,0391	4,4001		
2	Alisado	0,9538	0,0372	4,6749		
3	1ª derivada	0,9440	0,0411	4,2462		
4	MSC	0,9536	0,0372	4,6675		
5	MSC + 1ª derivada	0,9346	0,0442	3,9298		
Previsão						
		R <sup>2</sup>	RMSE	RPD		
1	-	0,9065	0,0245	2,8444		
2	Alisado	0,7784	0,0298	2,0049		
3	1ª derivada	0,8023	0,0261	2,0475		
4	MSC	0,9167	0,0181	3,4421		
5	MSC + 1ª derivada	0,8331	0,0259	2,4073		

As regiões espectrais usadas para cada modelo e energia de excitação, 15 kV e 50 kV, são expressas em keV. DW: teste de Durbin-Watson. R<sup>2</sup>: coeficiente de determinação (previsto vs. Observado). RMSE: raiz quadrada do erro quadrático médio. RPD: desvio residual da predição.

**Fonte: Autoria Própria (2022).**

**Figura 12 - Gráfico APARP do modelo MSC-MLR**



**Fonte: Autoria Própria (2022).**

A validação cruzada 8-fold mostrou a necessidade de 4 a 6 CDs para minimizar o RMSECV e maximizar o  $R^2_{cv}$  dos modelos (Tabela 5), o que indica previsões do teor de teff com menor erro (RMSEP). Como o MSC corrige o efeito do espalhamento da radiação, é possível que o aumento do número de CDs no modelo ComDim-MLR e dados MSC se deva à redução da contribuição dos picos de espalhamento elástico e inelástico na caracterização da amostra, o que torna necessária a utilização de outras informações que compensem essa perda para manter o bom desempenho. Não foram encontrados modelos sobreajustados, visto que todos tiveram RMSEP ligeiramente menor que RMSEC, logo, podem ser utilizados para quaisquer conjuntos (BONA; MARÇO; VALDERRAMA, 2018).

Os modelos ComDim-MLR com os espectros originais e MSC apresentaram os melhores desempenhos de acordo com as figuras de mérito de previsão (Tabela 5), com explicação da variância dos dados superior a 90% ( $R^2_{previsão}$ ) e as menores diferenças entre os resultados previstos e reais (RMSEP). O uso de 6 CDs pelo modelo ComDim-MLR com os dados MSC aumentou sua complexidade matemática e pode ter modelado mais ruído dos dados, o que interfere na reprodutibilidade dos resultados. Por isso, foi avaliado o *bias* do modelo o qual não teve significância estatística ( $p$ -valor  $> 0,05$ ), ou seja, não foi observado tendência nos dados sendo eliminada a hipótese de inserção de ruído. Diante do exposto, a performance do modelo ComDim-MLR com os dados MSC foi considerada a melhor mesmo com a maior quantidade de CDs devido à redução de 26% no erro de previsão (RMSEP). Assim, dado o

$RPD_{\text{previsão}} > 3$  esse modelo tem confiança estatística para análises qualitativas e quantitativas de adulteração em farinhas de teff (NICOLAÏ et al., 2007; QUELAL-VÁSCONEZ et al., 2018).

O desempenho analítico do modelo ComDim-MLR foi satisfatório conforme demonstram os resultados da Tabela 6. Esse modelo apresentou boa sensibilidade, conforme indicado pelo inverso da sensibilidade analítica, de modo que diferenças de apenas 0,06% de teff nas amostras são identificadas pela regressão, consequentemente, pela análise EDXRF (BRAGA; POPPI, 2004; VALDERRAMA; BRAGA; POPPI, 2007, 2009). As capacidades de detecção e quantificação do modelo também se mostraram satisfatórias, uma vez que permitiram determinar, com segurança estatística, a presença de teff a partir de conteúdos de 2% ( $LOD_{\text{min}}$ ) e quantificá-los a partir de 5% ( $LOQ_{\text{min}}$ ) (ALLEGRINI; OLIVIERI, 2014).

**Tabela 6 – Figuras de mérito analíticas do modelo ComDim-MLR para os dados MSC**

Sens.	Sens. Analítica <sup>-1</sup>	$LOD_{\text{min}}$	$LOD_{\text{máx}}$	$LOQ_{\text{min}}$	$LOQ_{\text{máx}}$
0,5430	$6,0750 \times 10^{-4}$	0,0177	0,0989	0,0530	0,2967

Sens: sensibilidade analítica expressa em frequência  $\times (m/m)^{-1}$ . Sens. Analítica<sup>-1</sup>: inverso da sensibilidade analítica expresso em (m/m). LOD: limite de detecção expresso em (m/m). LOQ: limite de quantificação expresso em (m/m).

**Fonte: Autoria Própria (2022).**

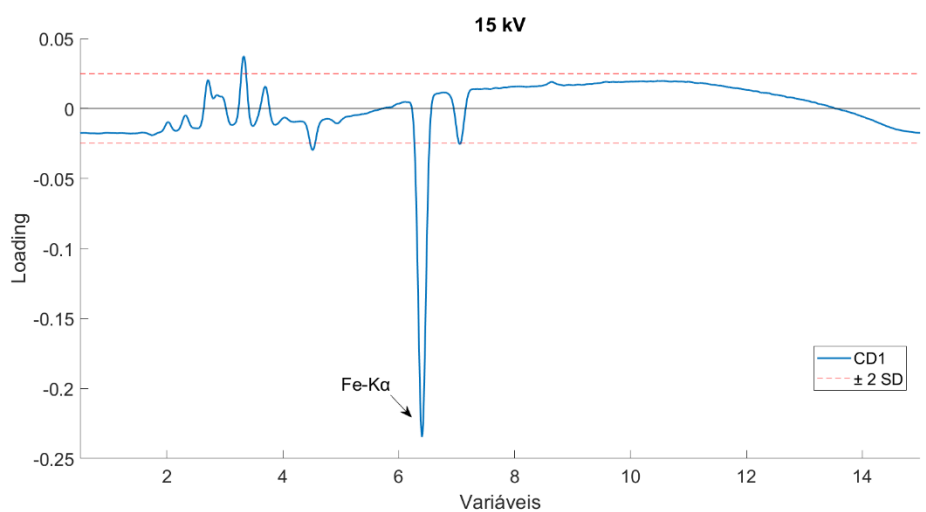
Todos os 7 coeficientes utilizados na construção do modelo linear ComDim-MLR (Equação 13) foram estatisticamente significativos. Os coeficientes negativos indicam relação inversa entre as variáveis independente e dependente, ou seja, a resposta aumenta se a variável dependente diminui. De forma análoga, essa relação é direta quando os coeficientes são positivos, ou seja, o aumento da variável dependente aumenta a resposta. Portanto, baixos *scores* em CD1, CD5 e CD6 e altos em CD2, CD3 e CD4 estão relacionados com amostras com maiores conteúdos de teff.

$$\begin{aligned} \%teff = & 0,70 - 0,60CD1 + 1,12CD2 + 1,10CD3 + 0,11CD4 - 0,19CD5 \\ & - 0,18CD6 \end{aligned} \quad (13)$$

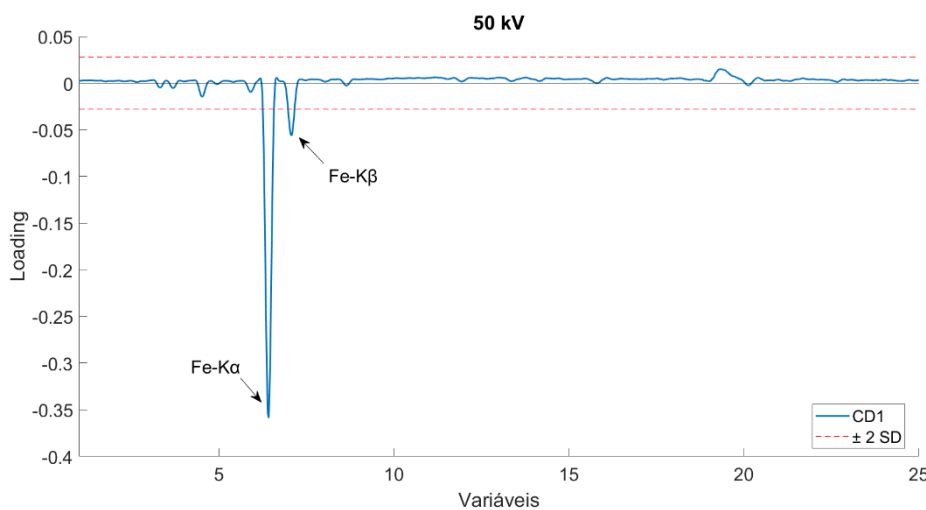
Os *loadings* das seis dimensões comuns (Figura 13) revelaram que a CD1 contribuiu para o aumento da resposta devido à maior intensidade do pico de ferro. Já a CD2 refletiu altos teores de cálcio (mineral também observado no teff com concentrações superiores aos demais cereais) e potássio, enquanto o aumento do *score* de CD3 contribuiu apenas para um maior

conteúdo de cálcio, mas redução do conteúdo de potássio. A CD4 mostrou maior importância do espalhamento remanescente após o MSC, CD5 se relacionou com o aumento do cálcio e espalhamento, mas redução da presença de potássio. Por fim, a CD6 refletiu maiores percentuais de cálcio e menores de potássio.

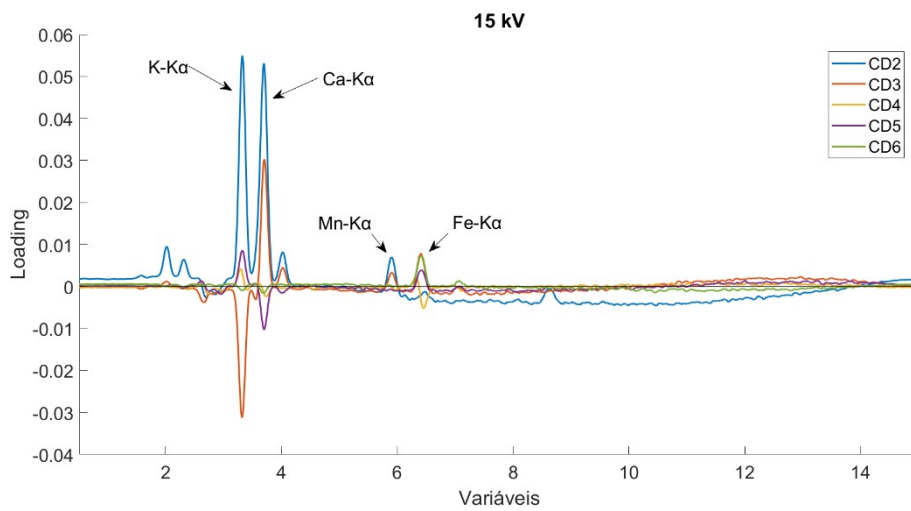
**Figura 13 – Gráficos dos *loading* das 6 CDs utilizadas pelo melhor modelo ComDim-MLR. Contribuição dos espectros (a) 15 kV na CD1 (b) 50 kV na CD1 (c) 15 kV nas CDs 2-6 (d) 50 kV nas CDs 2-6**



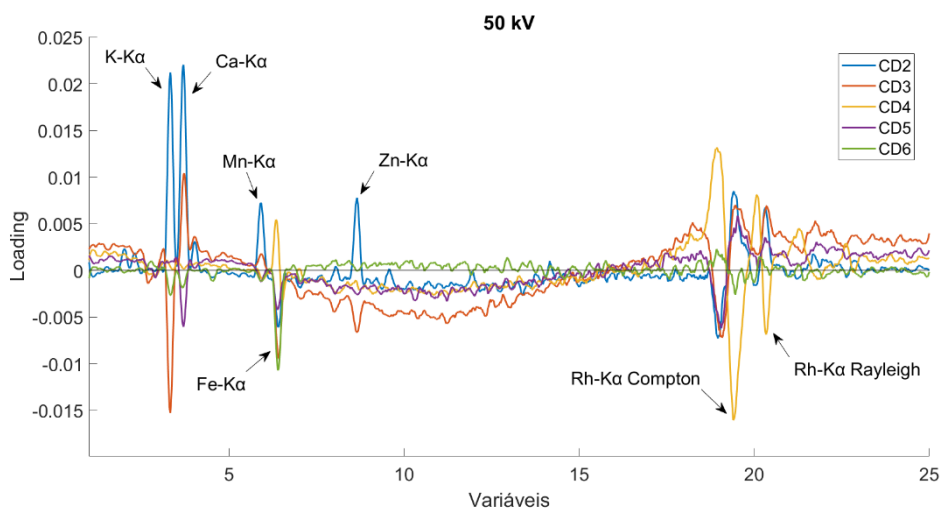
(a)



(b)



(c)



(d)

Os gráficos da CD1 foram separados das demais CDs para eliminar o efeito visual de escala devido às maiores saliências de CD1.

**Fonte: Autoria própria (2022).**

## 6 CONCLUSÃO

As farinhas de teff quando comparadas com as farinhas de arroz, trigo integral, aveia e centeio mostraram diferenças significativas em suas composições proximais, principalmente em relação ao teor de cinzas no qual todas se diferenciaram e o teff se mostrou superior, como era esperado devido aos seus altos valores de minerais.

A aplicação do planejamento de mistura permitiu obter uma amostragem representativa, que possibilitou observar a contribuição de cada adulterante em misturas com dois, três, quatro e cinco componentes.

A PCA revelou a qualidade dos dados NIR em reconhecer as farinhas que foram adulteradas e os próprios adulterantes, além de diferenciar as safras de teff utilizadas.

O ComDim mostrou que alterações na tensão utilizada na análise EDXRF produz informações comuns e complementares entre as condições, as quais contribuiriam positivamente na investigação de fraudes em teff, de modo que o benefício supera o aumento de apenas 2 minutos no tempo total de análise.

Os modelos PLS para os espectros NIR e MLR para os espectros EDXRF demonstraram desempenhos excelentes na quantificação de teff em farinhas, conforme indicou o  $RPD > 3$ , podendo ser utilizados para fins de quantificação e controle de qualidade. Portanto, tanto a espectroscopia NIR quanto EDXRF se mostraram ferramentas eficientes no controle de adulterações em farinhas de teff com desempenho similar. Por isso, não é preciso adquirir ambos os equipamentos, podendo-se optar por aquele que melhor atenda às necessidades da instituição. As técnicas são vantajosas para aplicações na indústria por serem rápidas, não destrutivas e com mínimo preparo de amostras.

## REFERÊNCIAS

- ABEBE, W.; RONDA, F. Rheological and textural properties of tef [*Eragrostis tef* (Zucc.) Trotter] grain flour gels. **Journal of Cereal Science**, v. 60, n. 1, p. 122–130, jul. 2014.
- ALI, M. et al. Determination of traces of molybdenum and lead in foods by x-ray fluorescence spectrometry. **SpringerPlus**, v. 3, n. 1, p. 341, 7 dez. 2014.
- ALLEGRI, F.; OLIVIERI, A. C. IUPAC-Consistent Approach to the Limit of Detection in Partial Least-Squares Calibration. **Analytical Chemistry**, v. 86, n. 15, p. 7858–7866, 5 ago. 2014.
- ANVISA. **RDC Nº 263, DE 22 DE SETEMBRO DE 2005**, 2005.
- AOAC INTERNATIONAL. **Official methods of analysis of AOAC International**. 18th. ed. Gaithersburg: [s.n.].
- ARENDE, E. K.; ZANNINI, E. Oats. In: **Cereal Grains for the Food and Beverage Industries**. [s.l.] Elsevier, 2013. p. 243–283e.
- AYVAZ, H. et al. Detection of einkorn flour adulteration in flour and bread samples using Computer-Based Image Analysis and Near-Infrared Spectroscopy. **Food Control**, v. 127, n. April, 2021.
- BARRETO, R. et al. Teff (*Eragrostis tef*) processing, utilization, and future opportunities: a review. **International Journal of Food Science & Technology**, p. ijfs.14872, nov. 2020.
- BARROS, N. B. DE; SPACINO, S. I.; BRUNS, R. E. **Como fazer experimentos**. 4. ed. Porto Alegre: Bookman, 2010.
- BHAT, F. M.; RIAR, C. S. Effect of composition, granular morphology and crystalline structure on the pasting, textural, thermal and sensory characteristics of traditional rice cultivars. **Food Chemistry**, v. 280, n. October 2018, p. 303–309, maio 2019.
- BIANCOLILLO, A. et al. Data Fusion Strategies in Food Analysis. In: **Data Handling in Science and Technology**. [s.l.: s.n.]. v. 31p. 271–310.
- BLANCO, M.; VILLARROYA, I. NIR spectroscopy: a rapid-response analytical tool. **TrAC Trends in Analytical Chemistry**, v. 21, n. 4, p. 240–250, abr. 2002.
- BONA, E. **GAMMA**. Campo Mourão, 2017.
- BONA, E.; MARÇO, P. H.; VALDERRAMA, P. Chemometrics Applied to Food Control. In: **Food Control and Biosecurity**. [s.l.] Elsevier, 2018. v. 16p. 105–133.
- BRAGA, J. W. B.; POPPI, R. J. Figures of merit for the determination of the polymorphic purity of carbamazepine by infrared spectroscopy and multivariate calibration. **Journal of Pharmaceutical Sciences**, v. 93, n. 8, p. 2124–2134, ago. 2004.



- BULTOSA, G. Teff: Overview. In: WRIGLEY, C. W. et al. (Eds.). **Encyclopedia of Food Grains**. 2. ed. Oxford: Elsevier, 2016. p. 209–220.
- BURNS, D. A.; CIURCZAK, E. W. **Handbook of Near-Infrared Analysis**. 3. ed. Boca Raton: Taylor & Francis, 2008.
- BURNS, D. A.; CURCZAK, E. W. **Handbook of near-infrared analysis**. 4. ed. Boca Raton: Taylor & Francis, 2021.
- CALLAO, M. P.; RUISÁNCHEZ, I. An overview of multivariate qualitative methods for food fraud detection. **Food Control**, v. 86, p. 283–293, abr. 2018.
- CARIOU, V. et al. **ComDim Methods for the Analysis of Multiblock Data in a Data Fusion Perspective**. [s.l.] Elsevier, 2019. v. 31
- CHUNG, O. K. et al. Wheat Lipids. **Wheat: Chemistry and Technology: Fourth Edition**, p. 363–399, 1 jan. 2009.
- CORNELL, J. A. **A Primer on Experiments with Mixtures**. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2011.
- COZZOLINO, D. An overview of the use of infrared spectroscopy and chemometrics in authenticity and traceability of cereals. **Food Research International**, v. 60, p. 262–265, jun. 2014.
- DE AGUIAR, L. M. et al. Data fusion of middle-resolution NMR spectroscopy and low-field relaxometry using the Common Dimensions Analysis (ComDim) to monitor diesel fuel adulteration. **Talanta**, v. 236, n. August 2021, p. 122838, jan. 2022.
- DOS SANTOS, F. R. et al. EDXRF spectral data combined with PLSR to determine some soil fertility indicators. **Microchemical Journal**, v. 152, p. 104275, 2020.
- DOS SANTOS, F. R. et al. Evaluation of pre-processing and variable selection on energy dispersive X-ray fluorescence spectral data with partial least square regression: A case of study for soil organic carbon prediction. **Spectrochimica Acta Part B: Atomic Spectroscopy**, v. 175, n. June 2020, p. 106016, jan. 2021.
- FAO. **Methods of food analysis**. Disponível em:  
<<http://www.fao.org/3/y5022e/y5022e03.htm>>. Acesso em: 13 nov. 2020.
- FARRÉS, M. et al. Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation. **Journal of Chemometrics**, v. 29, n. 10, p. 528–536, 2015.
- FERREIRA, M. DO P.; TEIXEIRA TARLEY, C. R. Bioaccessibility estimation of metallic macro and micronutrients Ca, Mg, Zn, Fe, Cu and Mn in flours of oat and passion fruit peel. **LWT**, v. 150, n. February, p. 111880, out. 2021.

FERREIRA, M. H.; BRAGA, J. W. B.; SENA, M. M. Development and validation of a chemometric method for direct determination of hydrochlorothiazide in pharmaceutical samples by diffuse reflectance near infrared spectroscopy. **Microchemical Journal**, v. 109, p. 158–164, jul. 2013.

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

FIRMANI, P. et al. Multi-block classification of Italian semolina based on Near Infrared Spectroscopy (NIR) analysis and alveographic indices. **Food Chemistry**, v. 309, n. June 2019, p. 125677, 2020.

GALLARDO, H. et al. Possibilities of low-power X-ray fluorescence spectrometry methods for rapid multielemental analysis and imaging of vegetal foodstuffs. **Journal of Food Composition and Analysis**, v. 50, p. 1–9, 1 jul. 2016.

GALVAO, R. et al. A method for calibration and validation subset partitioning. **Talanta**, v. 67, n. 4, p. 736–740, 15 out. 2005.

GALVIN-KING, P.; HAUGHEY, S. A.; ELLIOTT, C. T. Herb and spice fraud; the drivers, challenges and detection. **Food Control**, v. 88, p. 85–97, jun. 2018.

GEBREMARIAM, M. M.; ZARNKOW, M.; BECKER, T. Teff (*Eragrostis tef*) as a raw material for malting, brewing and manufacturing of gluten-free foods and beverages: a review. **Journal of Food Science and Technology**, v. 51, n. 11, p. 2881–2895, 2 nov. 2014.

GELADI, P.; MACDOUGALL, D.; MARTENS, H. Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. **Applied Spectroscopy**, v. 39, n. 3, p. 491–500, 31 maio 1985.

GIROIL ALIMENTOS. **Teff: o grão etíope**. Disponível em: <<https://www.instagram.com/reel/CaXovZnJ-qj/?igshid=YmMyMTA2M2Y=>>>. Acesso em: 28 fev. 2022.

GRASSI, S.; ALAMPRESE, C. Advances in NIR spectroscopy applied to process analytical technology in food industries. **Current Opinion in Food Science**, v. 22, p. 17–21, 2018.

HAZRA, S. K. et al. Characterization of phytochemicals, minerals and in vitro medicinal activities of bael (*Aegle marmelos* L.) pulp and differently dried edible leathers. **Heliyon**, v. 6, n. 10, p. e05382, 1 out. 2020.

HE, M. et al. Advanced functional materials in solid phase extraction for ICP-MS determination of trace elements and their species - A review. **Analytica Chimica Acta**, v. 973, p. 1–24, jun. 2017.

HEALTHY SUPPLIES. **Organic Teff Grain, Gluten Free 500g (Sussex Wholefoods)**. Disponível em: <<https://www.healthysupplies.co.uk/organic-teff-grain-sussex-500g.html>>. Acesso em: 30 abr. 2022.

HERREROS-CHAVEZ, L. et al. Mineral profiles of legumes and fruits through partial least squares energy dispersive X-ray fluorescence. **Journal of Food Composition and Analysis**, v. 82, p. 103240, 1 set. 2019.

HESHE, G. G. et al. Effect of conventional milling on the nutritional value and antioxidant capacity of wheat types common in Ethiopia and a recovery attempt with bran supplementation in bread. **Food Science & Nutrition**, v. 4, n. 4, p. 534–543, 19 jul. 2016.

HUANG, F. et al. Detection of adulteration in Chinese honey using NIR and ATR-FTIR spectral data fusion. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 235, p. 118297, jul. 2020.

HUANG, H. et al. Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review. **Journal of Food Engineering**, v. 87, n. 3, p. 303–313, 2008.

IGNE, B.; HURBURGH, C. R. Local chemometrics for samples and variables: optimizing calibration and standardization processes. **Journal of Chemometrics**, v. 24, n. 2, p. n/a-n/a, 2010.

INSTITUTO ADOLFO LUTZ. **Métodos físico-químicos para análise de alimentos**. 1<sup>a</sup> edição ed. São Paulo: Instituto Adolfo Lutz, 2008.

JIANG, H.; LIU, T.; CHEN, Q. Quantitative detection of fatty acid value during storage of wheat flour based on a portable near-infrared (NIR) spectroscopy system. **Infrared Physics and Technology**, v. 109, n. March, p. 103423, 2020.

KAUR, M. et al. Some properties of rice grains, flour and starches: A comparison of organic and conventional modes of farming. **LWT - Food Science and Technology**, v. 61, n. 1, p. 152–157, abr. 2015.

KOTU, V.; DESHPANDE, B. Classification. **Data Science**, p. 65–163, 1 jan. 2019.

KUCHERYAVSKIY, S.; LOMBORG, C. J. Monitoring of whey quality with NIR spectroscopy—A feasibility study. **Food Chemistry**, v. 176, p. 271–277, 1 jun. 2015.

LASKOWSKI, W. et al. How Important are Cereals and Cereal Products in the Average Polish Diet? **Nutrients**, v. 11, n. 3, p. 679, 21 mar. 2019.

LI, F. et al. Review of energy-dispersive X-ray fluorescence on food elements detection. **X-Ray Spectrometry**, n. December 2021, p. 1–19, 20 jan. 2022.

LUYPAERT, J.; MASSART, D. L.; VANDER HEYDEN, Y. Near-infrared spectroscopy applications in pharmaceutical analysis. **Talanta**, v. 72, n. 3, p. 865–883, 15 maio 2007.

MAGALHÃES DE AGUIAR, L. et al. Application of low-field and medium-resolution <sup>1</sup>H NMR spectroscopy combined with chemometric methods for automotive gasoline quality control. **Fuel**, v. 282, p. 118684, 15 dez. 2020.

MAKIMORI, G. Y. F.; BONA, E. Commercial Instant Coffee Classification Using an Electronic Nose in Tandem with the ComDim-LDA Approach. **Food Analytical Methods**, v. 12, n. 5, p. 1067–1076, 24 maio 2019.

MAŁYJUREK, Z. et al. Authentication of honeybush and rooibos herbal teas based on their elemental composition. **Food Control**, v. 123, p. 107757, 1 maio 2021.

MANLEY, M. Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. **Chemical Society Reviews**, v. 43, n. 24, p. 8200–8214, 2014.

MARETTI SILVEIRA BUENO, M. I. et al. X-ray scattering processes and chemometrics for differentiating complex samples using conventional EDXRF equipment. **Chemometrics and Intelligent Laboratory Systems**, v. 78, n. 1–2, p. 96–102, jul. 2005.

MARQUETTI, I. et al. Partial least square with discriminant analysis and near infrared spectroscopy for evaluation of geographic and genotypic origin of arabica coffee. **Computers and Electronics in Agriculture**, v. 121, p. 313–319, 2016.

MARTENS, H.; STARK, E. Extended multiplicative signal correction and spectral interference subtraction: New preprocessing methods for near infrared spectroscopy. **Journal of Pharmaceutical and Biomedical Analysis**, v. 9, n. 8, p. 625–635, jan. 1991.

MASOTTI, F. et al. Analytical advances in the determination of calcium in bovine milk, dairy products and milk-based infant formulas. **Trends in Food Science & Technology**, v. 103, p. 348–360, set. 2020.

MCNEIL, B. et al. **Fermentation monitoring and control of microbial cultures for food ingredient manufacture**. [s.l.] Woodhead Publishing Limited, 2013.

MEDINA, S. et al. Food fingerprints – A valuable tool to monitor food authenticity and safety. **Food Chemistry**, v. 278, p. 144–162, abr. 2019.

MELQUIADES, F. L. et al. Quantification of metals in river water using a portable EDXRF system. **Applied Radiation and Isotopes**, v. 69, n. 2, p. 327–333, fev. 2011.

MELQUIADES, F. L. et al. Direct Determination of Sugar Cane Quality Parameters by X-ray Spectrometry and Multivariate Analysis. **Journal of Agricultural and Food Chemistry**, v. 60, n. 43, p. 10755–10761, 31 out. 2012.

MIR-MARQUÉS, A.; CERVERA, M. L.; DE LA GUARDIA, M. Mineral analysis of human diets by spectrometry methods. **TrAC Trends in Analytical Chemistry**, v. 82, p. 457–467, set. 2016.

MUTHAMILARASAN, M.; PRASAD, M. Small Millets for Enduring Food Security Amidst Pandemics. **Trends in Plant Science**, set. 2020.

NDLOVU, P. F. et al. Vis-NIR spectroscopic and chemometric models for detecting contamination of premium green banana flour with wheat by quantifying resistant starch

content. **Journal of Food Composition and Analysis**, v. 102, n. February, p. 104035, set. 2021.

NICOLAÏ, B. M. et al. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. **Postharvest Biology and Technology**, v. 46, n. 2, p. 99–118, nov. 2007.

OCEAN ROBBINS. **Teff: The Tiny but Mighty Grain Worth Knowing About**. Disponível em: <<https://foodrevolution.org/blog/food-and-the-environment/teff/>>. Acesso em: 28 fev. 2022.

OLIVIERI, A. C. et al. Uncertainty estimation and figures of merit for multivariate calibration (IUPAC Technical Report). **Pure and Applied Chemistry**, v. 78, n. 3, p. 633–661, 1 jan. 2006.

OLIVIERI, A. C. Analytical Figures of Merit: From Univariate to Multiway Calibration. **Chemical Reviews**, v. 114, n. 10, p. 5358–5378, 28 maio 2014.

OLIVIERI, A. C. Practical guidelines for reporting results in single- and multi-component analytical calibration: A tutorial. **Analytica Chimica Acta**, v. 868, p. 10–22, 2015.

PEREIRA, J. M. G. et al. Industrial Hydrogenation Process Monitoring Using Ultra-compact Near-Infrared Spectrometer and Chemometrics. **Food Analytical Methods**, v. 11, n. 1, p. 188–200, 11 jan. 2018.

PERUCHI, L. C. et al. Determination of inorganic nutrients in wheat flour by laser-induced breakdown spectroscopy and energy dispersive X-ray fluorescence spectrometry. **Spectrochimica Acta Part B: Atomic Spectroscopy**, v. 100, p. 129–136, out. 2014.

QUELAL-VÁSCONEZ, M. A. et al. Rapid fraud detection of cocoa powder with carob flour using near infrared spectroscopy. **Food Control**, v. 92, p. 183–189, out. 2018.

RODRIGUES, J. DE A. **Raios X: difração e espectroscopia**. São Carlos: EdUFSCar, 2005.

ROHANI, A.; TAKI, M.; ABDOLLAHPOUR, M. A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part: I). **Renewable Energy**, v. 115, p. 411–422, jan. 2018.

ROSEFL, C. M.; MARCO, C. Rice. 2008.

ROSENTRATER, K. A.; EVERS, A. D. Chemical components and nutrition. In: **Kent's Technology of Cereals**. [s.l.] Elsevier, 2018. p. 267–368.

SANTANA, E. J. et al. Improved prediction of soil properties with multi-target stacked generalisation on EDXRF spectra. **Chemometrics and Intelligent Laboratory Systems**, v. 209, n. January, 2021.

SANTOS, I. A. et al. NIR and MIR spectroscopy for quick detection of the adulteration of cocoa content in chocolates. **Food Chemistry**, v. 349, n. January, p. 129095, jul. 2021.

SHAN, P. et al. Stacked ensemble extreme learning machine coupled with Partial Least Squares-based weighting strategy for nonlinear multivariate calibration. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 215, p. 97–111, 15 maio 2019.

SIMOPOULOS, A. An Increase in the Omega-6/Omega-3 Fatty Acid Ratio Increases the Risk for Obesity. **Nutrients**, v. 8, n. 3, p. 128, 2 mar. 2016.

SMOLINSKA, A. et al. General Framing of Low-, Mid-, and High-Level Data Fusion With Examples in the Life Sciences. In: **Data Handling in Science and Technology**. [s.l: s.n.]. v. 31p. 51–79.

SPERANÇA, M. A. et al. Biofortification quality in bananas monitored by energy-dispersive X-ray fluorescence and chemometrics. **Food Chemistry**, v. 362, n. December 2020, p. 130172, nov. 2021.

SPERANÇA, M. A.; NASCIMENTO, P. A. M.; PEREIRA, F. M. V. Impurity in sugarcane juice as mineral content: A prospect for analysis using energy-dispersive X-ray fluorescence (EDXRF) and chemometrics. **Microchemical Journal**, v. 164, n. December 2020, p. 105951, 2021.

SPOLADORE, S. F.; BRÍGIDA DOS SANTOS SCHOLZ, M.; BONA, E. Genotypic classification of wheat using near-infrared spectroscopy and PLS-DA. **Applied Food Research**, v. 1, n. 2, p. 100019, dez. 2021.

THERMO SCIENTIFIC. **X-ray energy reference**. Disponível em: <<https://assets.thermofisher.com/TFS-Assets/CAD/posters/CAD-Niton-Periodictable-fxl.pdf>>. Acesso em: 11 set. 2021.

TOMAR, M. et al. Development of NIR spectroscopy based prediction models for nutritional profiling of pearl millet (*Pennisetum glaucum* (L.)) R.Br: A chemometrics approach. **Lwt**, v. 149, n. February, p. 111813, 2021.

TORBICA, A.; BELOVIĆ, M.; TOMIĆ, J. Novel breads of non-wheat flours. **Food Chemistry**, v. 282, n. January, p. 134–140, jun. 2019.

TORMENA, M. M. L. et al. Application of multi-block analysis and mixture design with process variable for development of chocolate cake containing yacon (*Smallanthus sonchifolius*) and maca (*Lepidium meyenii*). **Journal of the Science of Food and Agriculture**, v. 97, n. 11, p. 3559–3567, ago. 2017.

UNUVAR, A.; BOYACI, I. H.; KOKSEL, H. A novel approach for rapid discrimination of common and durum wheat flours using spectroscopic analyses combined with chemometrics. **Journal of Cereal Science**, v. 100, n. June, p. 103269, 2021.

VALDERRAMA, P.; BRAGA, J. W. B.; POPPI, R. J. Validation of multivariate calibration models in the determination of sugar cane quality parameters by near infrared spectroscopy. **Journal of the Brazilian Chemical Society**, v. 18, n. 2, p. 259–266, abr. 2007.

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

VIEIRA, T. F. et al. Chemometric Approach Using ComDim and PLS-DA for Discrimination and Classification of Commercial Yerba Mate (*Ilex paraguariensis* St. Hil.). **Food Analytical Methods**, v. 13, n. 1, p. 97–107, 22 jan. 2020.

VIELL, F. L. G. et al. Sensory characterization of gluten-free bread enriched with teff ( *Eragrostis tef* (Zucc.) Trotter) and yacon ( *Smallanthus sonchifolius* ) using flash profile and common dimension analysis. **Journal of Food Processing and Preservation**, v. 44, n. 2, 28 fev. 2020a.

VIELL, F. L. G. et al. Comparison between ultra-homogenisation and ultrasound for extraction of phenolic compounds from teff ( *Eragrostis tef* (Zucc. ) ). **International Journal of Food Science & Technology**, v. 55, n. 7, p. 2700–2709, 26 jul. 2020b.

WANG, L. et al. Quality analysis, classification, and authentication of liquid foods by near-infrared spectroscopy: A review of recent research developments. **Critical Reviews in Food Science and Nutrition**, v. 57, n. 7, p. 1524–1538, 3 maio 2017.

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

XUE, S. S. et al. Rapid, simultaneous and non-destructive determination of maize flour and soybean flour adulterated in quinoa flour by front-face synchronous fluorescence spectroscopy. **Food Control**, v. 130, n. May, p. 108329, 2021.

YU, J.; ZHAN, J.; HUANG, W. Identification of Wine According to Grape Variety Using Near-Infrared Spectroscopy Based on Radial Basis Function Neural Networks and Least-Squares Support Vector Machines. **Food Analytical Methods**, v. 10, n. 10, p. 3306–3311, 2017.

ZAREEF, M. et al. An Overview on the Applications of Typical Non-linear Algorithms Coupled With NIR Spectroscopy in Food Analysis. **Food Engineering Reviews**, v. 12, n. 2, p. 173–190, 10 jun. 2020.

ZAWISZA, B. et al. Non-destructive elemental analysis of herbal teas from South Africa. **Journal of Food Composition and Analysis**, v. 102, p. 104041, 1 set. 2021.

ZHOU, K. et al. Cereals and Legumes. In: **Biochemistry of Foods**. [s.l.] Elsevier, 2013. p. 3–48.

ZHU, F. Chemical composition and food uses of teff (*Eragrostis tef*). **Food Chemistry**, v. 239, p. 402–415, jan. 2018.

## **APÊNDICE A – Planejamento de misturas para a adulteração**



**Tabela A1 – Percentuais dos componentes presentes nas 70 misturas com teff A e 70 com teff B obtidos pelo planejamento de mistura {5,4} simplex-lattice**

Código	Teff (%)	Arroz (%)	Trigo integral (%)	Aveia (%)	Centeio (%)
1	65,00	0,00	0,00	0,00	35,00
2	65,00	0,00	0,00	8,75	26,25
3	65,00	0,00	0,00	17,50	17,50
4	65,00	0,00	0,00	26,25	8,75
5	65,00	0,00	0,00	35,00	0,00
6	65,00	0,00	8,75	0,00	26,25
7	65,00	0,00	8,75	8,75	17,50
8	65,00	0,00	8,75	17,50	8,75
9	65,00	0,00	8,75	26,25	0,00
10	65,00	0,00	17,50	0,00	17,50
11	65,00	0,00	17,50	8,75	8,75
12	65,00	0,00	17,50	17,50	0,00
13	65,00	0,00	26,25	0,00	8,75
14	65,00	0,00	26,25	8,75	0,00
15	65,00	0,00	35,00	0,00	0,00
16	65,00	8,75	0,00	0,00	26,25
17	65,00	8,75	0,00	8,75	17,50
18	65,00	8,75	0,00	17,50	8,75
19	65,00	8,75	0,00	26,25	0,00

---

20	65,00	8,75	8,75	0,00	17,50
21	65,00	8,75	8,75	8,75	8,75
22	65,00	8,75	8,75	17,50	0,00
23	65,00	8,75	17,50	0,00	8,75
24	65,00	8,75	17,50	8,75	0,00
25	65,00	8,75	26,25	0,00	0,00
26	65,00	17,50	0,00	0,00	17,50
27	65,00	17,50	0,00	8,75	8,75
28	65,00	17,50	0,00	17,50	0,00
29	65,00	17,50	8,75	0,00	8,75
30	65,00	17,50	8,75	8,75	0,00
31	65,00	17,50	17,50	0,00	0,00
32	65,00	26,25	0,00	0,00	8,75
33	65,00	26,25	0,00	8,75	0,00
34	65,00	26,25	8,75	0,00	0,00
35	65,00	35,00	0,00	0,00	0,00
36	73,75	0,00	0,00	0,00	26,25
37	73,75	0,00	0,00	8,75	17,50
38	73,75	0,00	0,00	17,50	8,75
39	73,75	0,00	0,00	26,25	0,00
40	73,75	0,00	8,75	0,00	17,50

---

---

41	73,75	0,00	8,75	8,75	8,75
42	73,75	0,00	8,75	17,50	0,00
43	73,75	0,00	17,50	0,00	8,75
44	73,75	0,00	17,50	8,75	0,00
45	73,75	0,00	26,25	0,00	0,00
46	73,75	8,75	0,00	0,00	17,50
47	73,75	8,75	0,00	8,75	8,75
48	73,75	8,75	0,00	17,50	0,00
49	73,75	8,75	8,75	0,00	8,75
50	73,75	8,75	8,75	8,75	0,00
51	73,75	8,75	17,50	0,00	0,00
52	73,75	17,50	0,00	0,00	8,75
53	73,75	17,50	0,00	8,75	0,00
54	73,75	17,50	8,75	0,00	0,00
55	73,75	26,25	0,00	0,00	0,00
56	82,50	0,00	0,00	0,00	17,50
57	82,50	0,00	0,00	8,75	8,75
58	82,50	0,00	0,00	17,50	0,00
59	82,50	0,00	8,75	0,00	8,75
60	82,50	0,00	8,75	8,75	0,00
61	82,50	0,00	17,50	0,00	0,00

---

---

62	82,50	8,75	0,00	0,00	8,75
63	82,50	8,75	0,00	8,75	0,00
64	82,50	8,75	8,75	0,00	0,00
65	82,50	17,50	0,00	0,00	0,00
66	91,25	0,00	0,00	0,00	8,75
67	91,25	0,00	0,00	8,75	0,00
68	91,25	0,00	8,75	0,00	0,00
69	91,25	8,75	0,00	0,00	0,00
70	100,00	0,00	0,00	0,00	0,00

---

**Fonte: Autoria própria (2022).**

## **APÊNDICE B – Bandas identificadas na espectroscopia NIR**

**Tabela B1 – Identificação das bandas mais importantes presentes nos espectros NIR das farinhas**

Grupo Funcional	Posição (nm)	Referência
C-H em lipídios	1115	(JIANG; LIU; CHEN, 2020)
2º sobretom de estiramento C-H	1204	(JIANG; LIU; CHEN, 2020)
2º sobretom de estiramento C-H	1351	(JIANG; LIU; CHEN, 2020)
Alongamento O-H em amilose; N-H em proteínas	1463	(BURNS; CURCZAK, 2021; TOMAR et al., 2021)
N-H em proteínas	1505	(BURNS; CURCZAK, 2021)
N-H em proteínas	1580	(BURNS; CURCZAK, 2021)
C-O em lipídios	1724	(TOMAR et al., 2021)
C-H em hidrocarbonetos	1770	(BURNS; CURCZAK, 2021)
Flexão e alongamento de O- H na amilose	1929	(TOMAR et al., 2021)
N-H de amida primária em proteínas	2065	(BURNS; CURCZAK, 2021)
O-H em carboidratos	2100	(BURNS; CURCZAK, 2021)
N-H, C-N e C=O em proteínas	2150	(BURNS; CURCZAK, 2021)
N-H em proteínas	2190	(BURNS; CURCZAK, 2021)
O-H e C-O em glicose	2280	(BURNS; CURCZAK, 2021)

---

Flexão de C-H em lipídios	2308	(BURNS; CURCZAK, 2021; TOMAR et al., 2021)
C-H em celulose	2336	(BURNS; CIURCZAK, 2008)
C-O em lipídios	2350	(BURNS; CURCZAK, 2021)

---

**Fonte: Aatoria própria (2022),**

**APÊNDICE C – Energias dos elementos químicos identificados na espectroscopia  
EDXRF**



**Tabela C1 – Energias raio-X de referência dos elementos químicos identificados nas amostras de farinhas\***

Elemento	Energia (keV)	
	K $\alpha$	K $\beta$
P	2,02	2,14
S	2,31	2,46
K	3,31	3,59
Ca	3,69	4,01
Mn	5,90	6,49
Fe	6,40	7,06
Cu	8,05	8,90
Zn	8,64	9,57
Rh	20,31	22,72

\*(THERMO SCIENTIFIC, 2018)

**Fonte: Autoria própria (2022).**