

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DIRETORIA DE GRADUAÇÃO E EDUCAÇÃO PROFISSIONAL
COORDENAÇÃO DO CURSO DE TECNOLOGIA EM PROCESSOS QUÍMICOS

Lucas Wahl Da Silva

**Quantificação de teores de azeite de oliva em misturas binárias e ternárias,
por espectroscopia de infravermelho médio com uso de calibração
multivariada**

TRABALHO DE CONCLUSÃO DE CURSO

TOLEDO
2018

LUCAS WAHL DA SILVA

**Quantificação de teores de azeite de oliva em misturas binárias e ternárias,
por espectroscopia de infravermelho médio com uso de calibração
multivariada**

Trabalho de conclusão de Curso apresentado a Coordenação do Curso Superior de Tecnologia em Processos Químicos – COPEQ – da Universidade Tecnológica Federal do Paraná – UTFPR *Campus* Toledo, como requisito parcial para obtenção do título de Tecnólogo em Processos Químicos.

Orientador: Prof. Dr. Clayton Antunes Martin

TOLEDO
2018

**TERMO DE APROVAÇÃO
DO TRABALHO DE CONCLUSÃO DE CURSO**

LUCAS WAHL DA SILVA

**Quantificação de teores de azeite de oliva em misturas binárias e ternárias
por espectroscopia de infravermelho médio com uso de calibração
multivariada**

Trabalho apresentado como forma de avaliação para o Trabalho de Conclusão de Curso II do curso de Tecnologia em Processos Químicos da UTFPR, *Campus* Toledo, e aprovado pela banca examinadora abaixo.

Prof. Dr. Clayton Antunes Martin

Prof. Dr. Clóvis Bombardelli

Prof^a. Dr^a. Kelen Menzes Flores Rossi de Aguiar

Toledo, Junho de 2018

A Folha de Aprovação assinada encontra-se na Coordenação do Curso

RESUMO

O azeite de oliva é um alimento utilizado pelo homem a mais de 3000 a.C, e o seu valor agregado faz ser alvo de adulterações com outros óleos vegetais. A técnica de Infravermelho com Refletância Total Atenuada e Transformada de Fourier (ATR-FTIR) é um meio rápido e não destrutivo para análise de óleos vegetais, e exige o mínimo de preparação de amostras, tendo rapidez na obtenção dos espectros e boa resolução espectral (4 cm^{-1}). Pelo motivo da adulteração do azeite de oliva torna-se necessário o desenvolvimento de metodologias analíticas. Assim o trabalho teve por objetivo o desenvolvimento de uma metodologia de quantificação de azeite de oliva em mistura ternária com óleos de soja e canola, e mistura binária com óleo de girassol. Os espectros das misturas de óleos vegetais foram obtidos em espectrômetro infravermelho com módulo de refletância total atenuada. Os espectros foram normalizados e submetidos a correção do espalhamento multiplicativo (MSC), sendo centrados na média ou submetidos a primeira e segunda derivadas. Os espectros das amostras com seus respectivos teores foram divididos em conjuntos de calibração e validação pelo algoritmo de Kennard-Stone. Para a seleção das variáveis utilizadas na calibração por regressão linear múltipla (RLM) foi utilizado o algoritmo das projeções sucessivas (SPA), com o software MATLAB. Os valores da raiz quadrada do erro médio quadrático (RMSE) para os modelos RLM sem processamento e com correção MSC variaram de 2,84 a 3,03, e 2,31 a 2,86, respectivamente. Nos modelos RLM obtidos a partir da primeira e segunda derivadas o RMSE ficou em torno de 3,26 a 4,19, e 7,09 a 7,24, respectivamente. Os valores dos limites de detecção ficaram na faixa de $0,13\text{ g }100\text{ g}^{-1}$ até $5,53\text{ g }100\text{ g}^{-1}$, e os limites de quantificação variaram de $0,43\text{ g }100\text{ g}^{-1}$ a $18,25\text{ g }100\text{ g}^{-1}$. A exatidão foi calculada a partir do erro médio quadrático de previsão (RMSEP), com resultados para os modelos de misturas de oliva, soja e canola na faixa de 2,84 a 7,15, e para os modelos de misturas de azeite de oliva e óleo de girassol na faixa de 0,18 a 1,61. A partir do método desenvolvido é possível determinar o teor de azeite de oliva dentro da faixa de calibração (0 a 100%) para as misturas de azeite de oliva com óleos de soja, canola e girassol, com boa capacidade preditiva.

Palavras-chave: Figuras de mérito. Algoritmo de Projeções Sucessivas. *Outliers*. Óleos mistos. Quimiometria. Sinal Analítico Líquido.

ABSTRACT

Olive oil is a man-made food more than 3000 a. C, and its added value makes it the target of tampering with other vegetable oils. The infrared technique with total attenuated and transformed Fourier reflectance (ATR-FTIR) is a fast and non-destructive medium for plant oil analysis, and requires the minimum preparation of samples, taking speed in obtaining the spectra and good resolution Spectral (4 cm^{-1}). For the reason of adulteration of olive oil, it becomes necessary the development of analytical methodologies. Thus the work aimed to develop a methodology for quantification of olive oil in mixture ternary with soy and canola oils, and binary mixture with sunflower oil. The spectra of vegetable oil mixtures were obtained in the infrared spectrometer with a total attenuated reflectance module. The spectra were normalized and subjected to the multiplicative scatter correction (MSC), being mean centered or subjected to first and second derivatives. The spectra of the samples and your respectively concentrations of olive oil were divided into calibration sets and validation by the Kennard-Stone algorithm. For the selection of variables used in multi-linear regression calibration (RLM), The algorithm of successive projections (SPA) was used with MATLAB software. The square root values of the quadratic medium error (RMSE) for the RLM models without processing and with MSC correction ranged from 2.84 to 3.03, and 2.31 to 2.86, respectively. In the RLM models obtained from the first and second derivatives the RMSE was around 3.26 to 4,19, and 7.09 to 7.24, respectively. The values of the detection limits were in the range of $0.13\text{ g }100\text{ g}^{-1}$ to $5, 53\text{g }100\text{ g}^{-1}$, and the quantification limits ranged from $0.43\text{ g }100\text{g}^{-1}$ to $18.25\text{ g }100\text{ g}^{-1}$. The accuracy was calculated from the average quadratic prediction error (RMSEP), with results for the models of olive, soy and canola mixtures in the range of 2.84 to 7.15, and for the models of olive oil and sunflower oil mixtures in the range from 0.18 to 1.61. From the developed method it is possible to determine the olive oil content within the calibration range (0 to 100%) for the mixtures of olive oil with soy oils, canola, and sunflower, with good predictive capacity.

Key-Words: Figures of Merit. Successive projection algorithm. Outliers. Mixed oils. Chemometrics. Net analytical signal.

LISTA DE FIGURAS

Figura 1: Fluxograma de um processo de calibração	7
Figura 2: Projeções das variáveis no espaço ortogonal.	12
Figura 3: Representação geométrica das propriedades de ortogonalidade do sinal analítico líquido. ...	15
Figura 4 : Equipamento de Espectroscopia no Infravermelho Perkin Elmer Spectrum 65.	17
Figura 5: Espectros das misturas de azeite de oliva, soja e canola.....	21
Figura 6: Valores de RMSE pelo número de variáveis selecionadas.....	22
Figura 7: Números de onda selecionados pelo SPA para as misturas ternárias	22
Figura 8: <i>leverage versus</i> resíduos <i>studentizados</i> do modelo selecionado para a determinação do teor de azeite de oliva em mistura com óleos de soja e canola.....	24
Figura 9: Norma do NAS versus Teores de referência de azeite de oliva com óleos de soja e canola .	25
Figura 10: Norma do NAS versus teores de referência do teor de azeite do oliva em mistura com óleo de girassol.....	25

LISTA DE TABELAS

Tabela 1: Dados dos modelos obtidos a partir das misturas de azeite de oliva, soja e canola	23
Tabela 2: Dados dos modelos RLM para as misturas de azeite de oliva e óleo de girassol.....	23
Tabela 3: Dados dos modelos RLM sem <i>outliers</i> das misturas de azeite de oliva, e óleos de soja e canola	24
Tabela 4: Figuras de mérito para os modelos das misturas de azeite de oliva, com óleos de soja e canola	26
Tabela 5: Figuras de mérito para os modelos das misturas de azeite de oliva e óleo de girassol	26

LISTA DE SÍMBOLOS, SIGLAS E ABREVIATURAS

AO - Azeite de oliva

ATR-IR -Attenuated Total Reflectance–Infrared (Infravermelho Refletância Total Atenuada.)

ATR-FTIR - Infravermelho com Refletância Total Atenuada e Transformada de Fourier

IV - Infravermelho

FT-IR - Fourier Transform-InfraRed (Infravermelho com Transformada de Fourier)

g - Gramas

KS - Kennard – Stonne

LOD - Limit of detection (Limite de detecção)

NAS - Net Analite Signal (Sinal Analítico Líquido)

nm - Nanômetros

PLS - Partial Least Square (Mínimos quadrados parciais)

R - Coeficiente de correlação múltipla

RLM - Regressão Linear Múltipla

RMSE - Root Mean Square Error (Erro médio quadrático)

RMSEP - Root Mean Square Error of Prediction (Erro médio quadrático de previsão)

RMSEC - Root Mean Square Error of Calibration (Erro médio quadrático de calibração)

SIMCA - Multivariate Data Analysis Solution

SPA - Sucessive projections algorithm (Algoritmo de projeções sucessivas)

µm - Micrômetro

SUMARIO

DIRETORIA DE GRADUAÇÃO E EDUCAÇÃO PROFISSIONAL	1
1 INTRODUÇÃO	1
2 OBJETIVOS.....	3
2.1 OBJETIVO GERAL.....	3
2.2 OBJETIVOS ESPECÍFICOS.....	3
3 REVISÃO BIBLIOGRÁFICA	4
3.1 ESPECTROSCOPIA DE INFRAVERMELHO COM REFLEXÃO TOTAL INTERNA ATENUADA E PROCESSAMENTO ESPECTRAL.....	4
3.2 CALIBRAÇÃO MULTIVARIADA	6
3.3 O ALGORITMO DE KENNARD-STONE	7
3.4 REGRESSÃO LINEAR MÚLTIPLA	8
3.4.1 COEFICIENTE DE CORRELAÇÃO MÚLTIPLA (R^2).....	10
3.5 ALGORITMO DE PROJEÇÕES SUCESSIVAS (SPA)	11
3.6 DETECÇÃO DE <i>OUTLIERS</i>	12
3.7 AS FIGURAS DE MÉRITO.....	14
4 METODOLOGIA.....	16
4.1 MATERIAIS.....	16
4.2 – PREPARO DAS AMOSTRAS	16
4.3 INSTRUMENTOS E AQUISIÇÃO DOS ESPECTROS	17
4.4 SELEÇÃO DAS VARIÁVEIS.....	18
4.5 ANÁLISE DE <i>OUTLIERS</i>	18
4.6 FIGURAS DE MÉRITO.....	19
5 RESULTADOS E DISCUSSÕES.....	21
5.1 FIGURAS DE MÉRITO.....	25
6 CONCLUSÕES	27
REFERÊNCIAS	28

1 INTRODUÇÃO

O azeite de oliva é um óleo vegetal extraído através da planta *Olea europaea sativa* oriunda da região do Mediterrâneo, com valor cultural e uso na culinária. O azeite de oliva é classificado em quatro tipos de acordo com seu método de produção e algumas características, e são os seguintes: azeite de oliva virgem, azeite de oliva refinado, azeite de oliva extra virgem, e azeite de oliva comum (International Olive Council, 2018).

O azeite de oliva extra virgem deve possuir uma acidez livre, expressa em porcentagem de ácido oleico que não deve ser superior a 0,8 g 100 g⁻¹ de óleo (International Olive Council, 2018).

Devido as suas propriedades de sabor, odor e atividades culinárias, o azeite de oliva (AO) tem sido utilizado desde 3000 a.C, é um óleo obtido a partir dos frutos da *Olea europaea sativa*, sendo excluído qualquer processo de reesterificação ou mistura de outros óleos. A dieta com a presença de azeite de oliva é oriunda do mediterrâneo, e reconhecida desde 1960 para evitar doenças cardiovasculares e acidente vascular cerebral. No contexto nutricional o azeite de oliva é reconhecido pela grande quantidade de ácidos graxos monoinsaturados relacionado à diminuição da lipoproteína de baixa densidade (*low density lipoprotein cholesterol* LDL-C) do organismo humano (ALMEIDA et al., 2015).

Além do consumo fresco, o azeite é usado na culinária como um ingrediente em formulações industriais, condimentos, molhos, recheios, coberturas, massas, e produtos de padaria. Seu uso tem se tornado comum no desenvolvimento de filmes comestíveis (PEREDA et al., 2014).

Cerca de 80 % ou mais dos compostos fenólicos do AO são perdidos no processo de refino, assim, o seu teor é mais elevado em AO virgem. Os principais compostos fenólicos no AO são: fenóis simples (hidroxitiroso, tiroso) e polifenóis (oleuropeína glucósido) (FITO et al., 2007).

Este conjunto de propriedades funcionais e características tecnológicas, associadas a um preço relativamente elevado de venda faz do azeite de oliva um alvo de fraudes comerciais, e tornou-se uma prática comum onde é buscado o lucro rápido (MENDES et al., 2015). Os azeites de oliva são a classe de compostos químicos mais antiga utilizadas pelo homem, e estão sujeitos a adulterações há muito tempo. Na Espanha em 1981 um caso de adulteração causou danos a saúde dos consumidores e severas medidas passaram a ser tomadas para o controle de qualidade desses produtos (CRISTY et al, 2004).

Recentemente a instituição de defesa do consumidor Proteste avaliou 20 marcas de azeite de oliva extra virgem e reprovou 4 dessas, onde segundo a análise não eram azeite de oliva extra virgem (PROTESTE , 2016).

Ao longo dos últimos 15 anos, como resultado do desenvolvimento da Espectroscopia no Infravermelho com Transformada de Fourier (FT-IR), a aplicação desta técnica expandiu-se na pesquisa de alimentos e, particularmente, tornou-se uma ferramenta poderosa na área analítica em estudo de óleos e gorduras comestíveis (VLACHOS et al., 2006).

Muitos pesquisadores apresentaram resultados na literatura científica, ao utilizarem a técnica de espectroscopia no infravermelho médio para quantificação de azeite de oliva em misturas com outros óleos vegetais (MENDES, 2015) (LEMA-GARCIA, 2010).

Pelo motivo de adulteração em azeite de oliva é importante o desenvolvimento de métodos analíticos para a sua quantificação em misturas binárias e ternárias com óleos vegetais refinados e para o controle de qualidade.

Em uma situação ideal de previsão de valores pode-se obter uma fórmula teórica assim como na física é possível através dos valores de pressão, volume e massa conhecidas prever a temperatura. Mas na situação em questão não há uma fórmula teórica para calcular os teores de azeite de oliva em misturas por Infravermelho Médio, é nesse ponto que a calibração pode fornecer essa informação através de relações subjacentes.

A calibração multivariada abrange técnicas de predição estatística, e para obter melhores resultados em tempo hábil é importante considerar quais técnicas podem adequar-se a pesquisa e a rotina de laboratório, para tal a técnica de regressão linear múltipla desempenha um papel importante e especial interesse, pois, permite bons coeficientes de correlação.

É vantajoso poder exprimir os parâmetros do método analítico em função de dados numéricos padronizados, utilizando a precisão, exatidão, sensibilidade, limite de detecção (limit of detection - LOD) e limite de quantificação (limit of quantification - LOQ).

O trabalho visou desenvolver um método analítico de espectroscopia no infravermelho médio para fornecer dados quantitativos da presença de óleo de oliva em misturas com outros óleos vegetais, por calibração multivariada baseada na regressão linear múltipla.

2 OBJETIVOS

2.1 Objetivo Geral

Desenvolver uma metodologia analítica para a quantificação de teores de azeite de oliva em misturas com outros óleos vegetais, permitindo obter dados para detecção de adulterações de azeite de oliva comercial, utilizando a espectroscopia no infravermelho médio e a calibração multivariada baseada na regressão linear múltipla.

2.2 Objetivos específicos

Selecionar as variáveis utilizadas no modelo de calibração de regressão linear múltipla.

Avaliar a presença de *outliers* no modelo de calibração e validação.

Realizar a validação externa do modelo de calibração RLM.

Determinar as figuras de mérito: precisão, exatidão, sensibilidade, limite de detecção, limite de quantificação para o método desenvolvido.

3 REVISÃO BIBLIOGRÁFICA

3.1 Espectroscopia de Infravermelho com reflexão total interna atenuada e processamento espectral

A radiação infravermelha (IV) faz parte do espectro eletromagnético e abrange a gama de comprimentos de onda de 800 nm a 1.10^6 nm aproximadamente 0,8 μm a 1000 μm . A radiação IV possui menos energia que a radiação visível, e mais energia que as ondas de rádio (HIGSON, 2009).

A espectroscopia no infravermelho é uma técnica analítica usada para a caracterização de amostras de sólidos, líquidos e gases. A escolha da amostragem depende do objetivo da análise, identificação qualitativa, medidas quantitativas de analíticos específicos, tamanho e composição da amostra (ROBINSON et al., 2005).

As técnicas de reflexão interna total atenuada (ATR) permitem obter os espectros IV utilizando-se a reflexão de radiação quando esta passa de um meio para o outro que possua diferentes índices de refração. Um dos meios em que ocorre a refração é a amostra, enquanto o outro é um cristal que forma parte da instrumentação do espectrômetro. A fração da radiação refletida é determinada por vários fatores, incluindo o ângulo de incidência. A medida que o ângulo de incidência aumenta, o mesmo acontece com a porcentagem da radiação refletida, até ser atingido certo ângulo conhecido como ângulo crítico. Em ângulos maiores do que o crítico ocorre a reflexão total (HIGSON, 2009).

Grande parte dos equipamentos de ATR utilizam as transformadas de Fourier, sendo geralmente conhecida como ATR-FTIR, e possuem muitas vantagens, as mais importantes são a facilidade de preparação de amostras e a versatilidade com relação a ela. Os espectros gerados pelas técnicas de ATR-IR são basicamente semelhantes aqueles obtidos pelos métodos normais de transmissão, embora as intensidades relativas de algumas bandas possam ser diferentes (HIGSON, 2009).

A técnica de espectroscopia no infravermelho com transformada de Fourier (FTIR) é uma ferramenta analítica rápida e não destrutiva para o estudo de óleos e gorduras comestíveis, exigindo o mínimo de preparação de amostras (LEMA-GARCIA, 2010).

Quando a amostra líquida é pequena ou existe falta de solvente, é comum obter-se espectros de um líquido puro. Neste caso, apenas um filme muito fino pode fornecer um caminho óptico suficiente curto para produzir espectros satisfatórios (SKOOG et al., 2009).

Os pré-processamentos dos espectros permitem uma manipulação mais seletiva dos dados importantes para as calibrações reduzindo fontes de não-linearidade e variância causados por interferências e condições experimentais (Noord, 1993). O pré-processamento é definido como uma manipulação matemática antes das análises preliminares e tem como finalidade reduzir fontes de variações irrelevantes ou não informativas e tratar as matrizes de dados para tornar os cálculos melhor condicionados antes da modelagem para remover matematicamente fontes de variação indesejáveis (que não estão relacionadas ao parâmetro de interesse) que não serão removidas naturalmente durante a análise dos dados. O tratamento dos dados é aplicado antes da fase de calibração para melhorar a eficiência do método, produzindo um modelo de regressão com a melhor capacidade de previsão das amostras. Contudo, essa etapa de pré-tratamento deve ser realizada com cautela, uma vez que, informações espectrais relevantes, podem ser significativamente removidas (Noord, 1993).

Muitos softwares com ferramentas de trabalho para o processamento dos dados estão disponíveis, como o MATLAB, o R e o SIMCA. Softwares com suporte de ferramentas analíticas são utilizados e alguns métodos de processamento são disponíveis em caixas ferramentas de extensões dos softwares.

O problema da linha de base pode ser corrigido tomando-se a derivada dos espectros, sendo o algoritmo de Savitsky-Golay o mais utilizado. A correção multiplicativa de sinal é usada para corrigir efeitos do espalhamento de luz na espectroscopia por refletância, estes efeitos são eliminados fazendo-se a regressão linear das variáveis espectrais (X^T) nas variáveis do espectro médio (X_m).

$$X_j = a_i I + b_j X_m \quad (1)$$

Os coeficientes a_i e b_i da amostra i são calculados por quadrados mínimos fazendo-se a regressão de cada espectro no espectro médio (um conjunto a, b para cada amostra).

$$x_{i(msc)} = \frac{x_{ij} - a}{b} \quad (2)$$

O espectro corrigido $x_{i(msc)}$ é obtido subtraindo-se a absorvância de cada comprimento de onda do espectro original, x_i , de a_i e dividindo-a por b_i . O resultado final da regressão é uma matriz de coeficientes ($2 \times I$), sendo I igual ao número de amostras.

A centralização dos dados na média é usada para diminuir a flutuação do sinal analítico que podem ocorrer devido a efeitos físicos ou sistemáticos, após essa correção as flutuações são devidas as mudanças nas características químicas do analito de interesse. É um dos tratamentos mais simples para os dados espectrais (SILVA DE LIMA, 2013).

Primeiramente calcula-se a média de cada variável conforme equação (3):

$$X_{cj} = \frac{1}{n} \sum_{i=1}^n X_{ij} \quad (3)$$

Onde X_{cj} é a matriz dos dados X_{ij} contido em uma coluna (variáveis). Em seguida subtraem-se os dados originais pela média calculada.

$$X_{ijcm} = X_{ij} - X_{cj} \quad (4)$$

O autoescalamamento é definido por:

$$X_{ij(as)} = \frac{X_{ij} - \bar{X}_j}{S_j} \quad (5)$$

Onde S_j é igual a:

$$S_j^2 = \frac{1}{l-1} \sum_{i=1}^l (X_{ij} - \bar{X}_j)^2 \quad (6)$$

3.2 Calibração Multivariada

Em grande parte das análises químicas o objetivo é determinar a concentração de analitos, porém equipamentos muito utilizados para tal não informam diretamente a informação necessária, sendo necessários métodos de calibração. Na maior parte das técnicas instrumentais, a quantificação de um determinado analito deve ser feita após a calibração do método (SILVESTRE, 2014).

Infelizmente, as absorções moleculares da radiação IV são muito menores do que a radiação ultravioleta ou visível, o que normalmente impede que sejam determinados traços através de calibração clássica (HIGSON, 2009).

Através deste método é possível obter bons resultados em misturas binárias simples, mas, podem surgir problemas quando se trabalha com amostras reais, devido a interferências espectrais e desconhecimento da identidade dos compostos de interesse (FERREIRA et al., 1999).

Desde a origem da quimiometria como um ramo independente da química no início dos anos 1970, químicos tem desenvolvido métodos para a calibração multivariada (OLIVIERI et al., 2006).

Uma mensuração multivariada é definida como uma medida de múltiplas variáveis ou respostas para uma amostra. A predição é o processo de estimação das características das amostras desconhecidas pela aplicação de um modelo de calibração para a medida de um

vetor desconhecido. Diferentes características ou propriedades de sistemas químicos podem ser preditas usando a técnica de calibração multivariada (BEEBE, 1998).

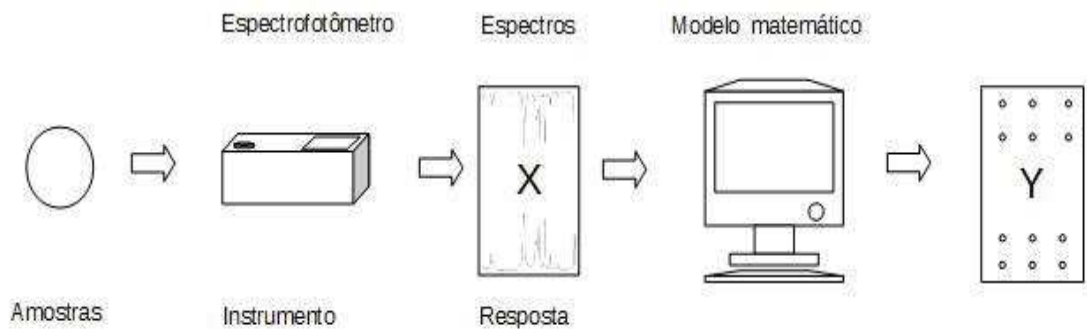
A calibração é um processo de construção de um modelo para o uso de predição de características ou propriedades de amostras desconhecidas, o modelo é construído a partir de dados de calibração medidos através de respostas multivariadas e correspondendo a amostras conhecidas de concentração, ou características físicas e químicas de interesse (BEEBE, 1998).

Os métodos multivariados são os mais adequados para a análise de dados, pois, permitem um estudo com várias espécies presentes ao mesmo tempo, não importando a existência ou ausência de diferenças espectrais marcantes entre elas nem a existência de alta correlação nos dados (FERREIRA et al., 1999).

Como uma poderosa extensão da calibração univariada, a calibração multivariada constrói um modelo preditivo com base em múltiplos valores do espectro (OLIVIERI et al., 2006).

Figura 1: Fluxograma de um processo de calibração

Fonte: autoria própria.



3.3 O algoritmo de Kennard-Stone

O algoritmo de Kennard-Stone (KS) seleciona as amostras com base em suas distâncias euclidianas, ele escolhe e separa um subconjunto significativo de um conjunto de n amostras. O algoritmo KS assegura a uniformidade de distribuição de cada subconjunto ao longo do espaço de dados X (respostas instrumentais) e segue um procedimento de *stepwise*

no qual a seleção ocorre em uma região do espaço longe das amostras já selecionadas (SOUZA, 2011).

A primeira amostra selecionada é a que apresenta a maior distância em relação a média das amostras, e a segunda amostra selecionada é que apresenta a maior distância da primeira, e as próximas amostras selecionadas serão as que tiverem a maior distância da última amostra selecionada, e assim consecutivamente até que um número mínimo de amostras sejam escolhidas (SOUZA, 2011).

A distância euclidiana entre os vetores X de cada par pode ser representado na equação (7):

$$d(x, p) = \sqrt{\sum_{j=1}^j [x_p(j) - x_q(j)]^2} \quad p, q \in [1, N] \quad (7)$$

Onde $x_p(j)$ e $x_q(j)$, são as respostas instrumentais do j -ésimo comprimento de onda para as amostras p e q , respectivamente, e j denota o número de comprimentos de ondas no espectro.

Normalmente, o algoritmo de Kennard-Stone é aplicado para a seleção de amostras do conjunto de calibração (SOUZA, 2011).

3.4 Regressão Linear Múltipla

Para Myers (2009), “na maioria dos problemas em que a análise de regressão é aplicada, necessitamos de mais de uma variável independente no modelo de regressão”. Para prevermos uma resposta que se ajuste a alta complexidade de muitos mecanismos científicos, precisamos de um modelo de regressão múltipla.

Quando o modelo de regressão é linear nos coeficientes, ele é chamado de modelo de regressão linear múltipla (MYERS et al., 2009).

Para o caso de k variáveis independentes x_1, x_2, \dots, x_k , a média de Y | x_1, x_2, \dots, x_k é dada pelo modelo de regressão linear múltipla representado na equação (8):

$$\mu Y | x_1, x_2, \dots, x_k = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (8)$$

E a resposta estimada é obtida da equação de regressão amostral (9):

$$\gamma = b_0 + b_1 x_1 + \dots + b_k x_k \quad (9)$$

Onde cada coeficiente de regressão β_i é estimado por b_i com base nos dados das amostras, usando o método dos mínimos quadrados (MYERS et al., 2009).

No ajuste de um modelo de regressão linear múltipla, é muito mais conveniente expressar as operações matemáticas usando a notação matricial. Suponha que haja k regressores e n observações $(x_{i1}, x_{i2}, \dots, x_{ik}, y_i)$, $i=1, 2, \dots, n$, e que o modelo relacionando os regressores à resposta seja:

$$y_i = b_0 + b_1 x_{i1} + \dots + b_k x_{ik} \quad (10)$$

$$i = 1, 2, \dots, n$$

Esse modelo é um sistema de n equações, que pode ser expresso na notação matricial como:

$$Y = X\beta + \epsilon \quad (11)$$

Em geral, y é um vetor ($n \times 1$) das observações, X é uma matriz ($n \times p$) dos níveis das variáveis independentes, β é um vetor ($p \times 1$) dos coeficientes de regressão e ϵ é um vetor ($n \times 1$) dos erros aleatórios (MONTGOMERY et al., 2008). Desejou-se encontrar o vetor dos estimadores de mínimos quadrados, $\hat{\beta}$, tal que :

$$X'X\hat{\beta} = X'y \quad (12)$$

A equação (12) é a equação normal de mínimos quadrados na forma matricial. Com o objetivo de resolver as equações normais, multiplica-se ambos os lados da equação (12) por $X'X$ (MONTGOMERY et al., 2008). Consequentemente, a estimativa de mínimos quadrados de β é :

$$\hat{\beta} = (X'X)^{-1} X'y \quad (13)$$

Nessa forma é possível ver que $X'X$ é uma matriz simétrica ($p \times p$) e $X'y$ é um vetor coluna ($p \times 1$), logo:

$$\hat{Y} = X(X'X)^{-1} X'y \quad (14)$$

3.4.1 Coeficiente de correlação Múltipla (R^2)

O coeficiente de correlação múltipla R^2 é definido como:

$$R^2 = \frac{SQr}{SQt} = 1 - \frac{SQe}{SQt} \quad (15)$$

Pode-se mostrar que :

$$SQe = \mathbf{y}'\mathbf{y} - \widehat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \quad (16)$$

$$SQr = \widehat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \quad (17)$$

$$SQt = SQr + SQe \quad (18)$$

Onde SQr é a soma quadrática da regressão, e SQt a soma quadrática total, e SQe é a soma quadrática dos erros. R^2 é uma medida da quantidade de redução na variabilidade de y , obtida pelo uso dos regressores $x_{i1}, x_{i2}, \dots, x_{ik}, y_i$, temos de ter $0 \leq R^2 \leq 1$ (MONTGOMERY et al., 2008).

3.4.2 – Seleção das variáveis na regressão linear múltipla e multicolinearidade

Um problema importante em muitas aplicações da análise de regressão envolve selecionar o conjunto de variáveis independentes ou regressores a serem usados para a construção do modelo (BEEBE, 1998). O problema consiste em selecionar um conjunto apropriado de regressores a partir de um conjunto que inclua todas as variáveis importantes. Em tais casos torna-se importante filtrar as variáveis candidatas para obter um modelo de regressão que contenha os “melhores” subconjuntos de regressores, para executar satisfatoriamente, isso requer interação do analista e seu julgamento (MONTGOMERY et al., 2008, p.255).

Em problemas de regressão múltipla, esperamos encontrar dependências entre a variável de resposta Y e os regressores x_j , no entanto também encontramos relação entre os regressores x_j . Em situações onde essas dependências forem fortes, dizemos que existem multicolinearidades, a multicolinearidade pode causar sérios efeitos nas estimativas dos

coeficientes de regressão e na aplicabilidade geral do modelo estimado (MONTGOMERY et al. 2008, p.264).

3.5 Algoritmo de projeções sucessivas (SPA)

Métodos analíticos modernos empregam técnicas instrumentais de análise de amostras de sólidos, líquidos e gases. E os instrumentos utilizados geram conjuntos de dados com um número considerado de variáveis.

A resposta instrumental apresenta forte correlação em diferentes canais analíticos, o que leva a redundância nos dados adquiridos, e também alguns canais não fornecem a informação necessária e seu uso compromete a precisão do resultados. Por estas razões, um método analítico pode se beneficiar de um subconjunto de canais, em vez de todo o conjunto de medições obtidas para cada amostra, facilitando a interpretação de resultados, e útil para a aplicação analítica do instrumento, na quimiometria isso é chamado de seleção de variáveis (SOARES et al. , 2013).

Em termos matemáticos, a matriz X é chamada colinear (ou multicolinear) se as colunas de X são linearmente dependentes. Em outras palavras, X é colinear se uma das variáveis X pode ter uma combinação linear com as outras variáveis. Assim, para a calibração RLM, a colinearidade em X pode ter um efeito prejudicial sobre a estabilidade dos coeficientes de β e torná-los inúteis para a interpretação (MARTENS; NAES, 2012) .

O Algoritmo de Projeções Sucessivas tem a finalidade de selecionar os comprimentos de onda cujo conteúdo de informação redundante seja mínima, a fim de resolver problemas de colinearidade (ARAÚJO et al., 2001).

Se dois vetores $X = (x_1+x_2+x_3)'$ e $Z = (z_1+z_2+z_3)'$ e um espaço tridimensional ($k=3$) são ortogonais entre eles, o ângulo entre eles será de 90° , algebricamente a soma do produto escalar é zero: $(x_1 z_1+x_2 z_2+x_3 z_3) = 0$.

A escolha das variáveis menos colineares entre si através do SPA consiste em três fases durante a execução do algoritmo. A primeira fase consiste na criação de cadeias de variáveis inicializada pela adição de uma primeira variável de referência. (SOARES et al., 2013).

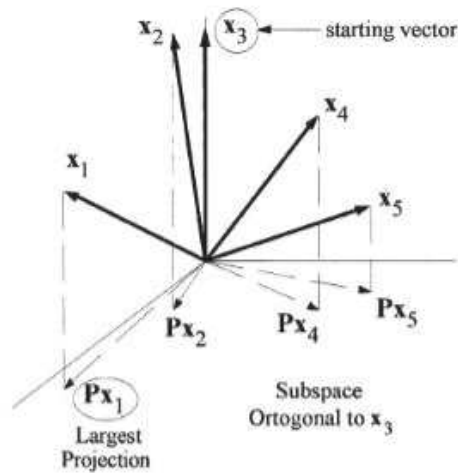
A segunda variável adicionada fornece o maior valor de projeção quando todas as variáveis não adicionadas na cadeia são projetadas no plano ortogonal a primeira variável

seleccionada. O processo repete-se até o número máximo de variáveis na cadeia, que deverá ser menor ou igual ao número de amostras de calibração. Quando todas as n variáveis do conjunto de calibração foram empregadas como variável de partida para a construção de n cadeias a primeira fase do algoritmo foi concluída (SOARES et al., 2013).

A segunda fase consiste na avaliação das cadeias de variáveis construídas em função da raiz quadrada do erro médio de validação (RMSEV), escolhendo-se a cadeia com menor valor de RMSEV. A última fase do algoritmo consiste na eliminação das variáveis não informativas para o modelo MLR na cadeia seleccionada (SOARES et al., 2013).

Figura 2: Projeções das variáveis no espaço ortogonal.

Fonte: SOARES et al., 2013.



3.6 Detecção de *Outliers*

Se uma amostra possuir um valor significativamente discrepante dos outros dados, há uma grande possibilidade de que seja errôneo, sendo originado possivelmente de um erro crasso, sendo necessário escolher entre manter ou rejeitar esse valor. Por outro lado, existe a possibilidade de que o valor seja válido, e assim, dessa forma a precisão do procedimento será menor do que o esperado.

Outliers em geral representam dados irrelevantes, erros grosseiros ou anormais, quando os dados são comparados entre si. Para manter o uso ótimo de um conjunto de dados é importante a identificação de *outliers*. Na química analítica convencional a detecção de *outliers* é um estágio valioso que não pode ser ignorado. Durante a calibração é importante detectar e se possível remover ou corrigir dados que possam diminuir a capacidade de predição dos coeficientes de calibração estimados (MARTENS, 2002).

É importante a detecção de *outliers* em todas as análises de dados. Os erros são a regra e não a exceção devido a por exemplo, erros triviais, erros de instrumentos e erros de amostragem. Se esses erros são suficientemente grandes, quer em quantidade ou qualidade, eles podem destruir qualquer resultado ou interpretação significativa. Pode parecer difícil de detectar valores atípicos quando complicados dados multivariados são utilizados, mas, na verdade, a detecção de outliers é muito maior em dados multivariados (BRO, 2003).

Os *outliers* podem ser identificados pelo *leverage* extremo, o qual representa a distância de uma amostra em relação a média do conjunto de dados, e tomando como bases espectrais, ele difere o espectro de uma amostra com relação as demais (VALDERRAMA, 2005 apud MARTENS, 1996).

A *leverage* pode ser interpretada geometricamente como a distância de uma amostra ao centróide do conjunto de dados e é calculada segundo a equação (19):

$$h_{ii} = \frac{1}{n} + (\mathbf{X}_i - \bar{\mathbf{X}})^t (\mathbf{X}^t \mathbf{X})^{-1} (\mathbf{X}_i - \bar{\mathbf{X}}) \quad (19)$$

Onde x_i é o espectro da i ésima amostra, $\bar{\mathbf{X}}$ é o espectro médio e $\mathbf{X}^t \mathbf{X}$ é a matriz de variância-covariância, n é o número de amostras do modelo. Pela expressão acima, vê-se claramente que a "*leverage*" é uma função da distância da amostra à média e das correlações entre as variáveis.

O trabalho também utilizou o cálculo dos resíduos de *student*, também chamados resíduos *studentizados*, onde amostras mal modeladas podem ser detectadas com valores altos de resíduos (FERREIRA et al., 1999), para saber a influência de cada amostra em particular o resíduo de *student* é calculado para a amostra i a partir das equações (20) e (21) :

$$Lresc_i = \sqrt{\frac{(c - c_i)^2}{(n-1)(\hat{1} - h_i)}} \quad (20)$$

$$Resíduos\ de\ student_i = \frac{(c - c_i)}{Lresc_i - \sqrt{\hat{1} - h_i}} \quad (21)$$

Onde C_i é a concentração obtida pelo modelo de calibração, C a concentração da amostra do modelo de validação, e h_i a *leverage* da respectiva amostra do modelo de calibração, e n o número de amostras no modelo.

3.7 As Figuras de mérito

Figuras de mérito para Valderrama et al. (2009) “são parâmetros que certificam que o procedimento proposto é confiável e atende a determinadas especificações impostas pela indústria ou por algum órgão de fiscalização.

A precisão é determinada pela variação entre vários resultados obtidos na medida de um determinado componente de uma mesma amostra. Para os dados de primeira ordem a precisão é semelhante aos métodos de calibração univariada, enquanto a sensibilidade precisa de uma abordagem específica para dados multivariados (VALDERRAMA et al., 2009).

A exatidão em calibração multivariada, normalmente é expressa pela raiz quadrada do erro médio quadrático de previsão, também chamado RMSEP (VALDERRAMA, 2005).

$$RMSEP = \sqrt{\sum_{i=1}^{I_v} \frac{(y_i - \hat{y}_i)^2}{I_v}} \quad (22)$$

Em que, I_v é o número de amostras do conjunto de validação, y_i e \hat{y}_i correspondem aos valores de referência e aos previstos pelo modelo, respectivamente.

O limite de detecção é o mínimo detectável de um procedimento químico. A capacidade de detecção é considerado pela IUPAC como performance fundamental de um processo químico de mensuração (OLIVIERI et al, 2006).

A sensibilidade analítica é um parâmetro que apresenta a sensibilidade do método em termos da unidade de concentração que é utilizada, e definida como a razão entre sensibilidade e o desvio padrão do sinal de referência (VALDERRAMA, 2005).

A *Union of Pure and Applied Chemistry* (IUPAC) adota a definição da *International Standardization Organization* (ISO) para a capacidade de detecção como sendo “a menor quantidade de uma substância para distinguir-se da ausência dessa substância em um limite confiável”.

O limite de detecção pode ser diretamente estimado através da calibração univariada como uma alternativa simples, em qual o nível detalhado é estimado a partir do nível do sinal da média e desvio padrão para medições repetidas de um branco das amostras e para amostras em concentrações próximas do limite de detecção (OLIVIERI, 2014).

Porém se tratando de dados multivariados a aplicação acima não é totalmente clara e alguns aspectos do âmbito de aplicação da norma ISO deve ser considerada, e na verdade ainda não há um aceite geral para esses parâmetros, e a principal dificuldade em estimar um

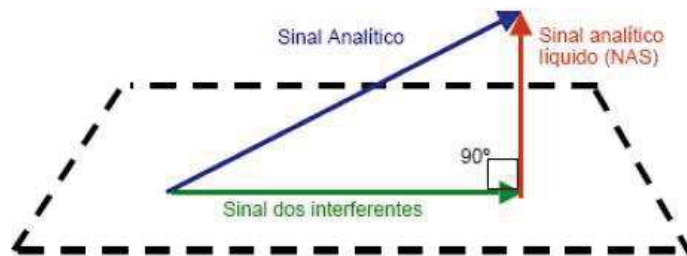
limite de detecção é que os sinais instrumentais não são específicos para um determinado analito (OLIVIERI, 2014).

O assunto tem especial interesse de muitos pesquisadores que descreveram métodos para estimar o limite de detecção em calibração multivariada, e destaca-se Lorber e colaboradores (1997) pela técnica de Sinal Analítico Líquido.

O Sinal Analítico Líquido é definido para uma propriedade de interesse k , como sendo a parte do sinal analítico ortogonal às contribuições de possíveis interferentes na amostra (VALDERRAMA, 2005).

Figura 3: Representação geométrica das propriedades de ortogonalidade do sinal analítico líquido.

Fonte: (VALDERRAMA,2005).



Farber apresentou como método para o cálculo do sinal analítico líquido a partir dos vetores de regressão do modelo (BRO et al., 2006), definido na equação (23):

$$x_{ki}^* = \mathbf{b} (\mathbf{b}^T \mathbf{b})^{-1} \mathbf{y}_i \quad (23)$$

Onde \mathbf{b} são os vetores de regressão do modelo RLM e \mathbf{y}_i os valores de concentração das amostras de calibração.

4 METODOLOGIA

4.1 Materiais

Para a realização do trabalho foram utilizados os óleos vegetais: Azeite de Oliva Extra-Virgem Andorinha, Lote: 071BL6005523. Óleo de Soja Coamo, Lote: L43042. Óleo de Canola Liza (Cargill), Lote: L08C. Óleo de Girassol Selecta, Lote: L101502:19GPR278941.

E os seguintes instrumentos de laboratório: Micropipeta (100 – 1000 μ L) Labmate HTL, Balança Analítica Shimadzu, Tubos de microcentrífuga, Agitador de soluções, ATR-FTIR Perkin Elmer Spectrum 65 e Álcool Isopropílico.

4.2 – Preparo das Amostras

Verificou-se a relação entre a massa e o volume de óleo de soja, pipetando-se diversos volumes de óleo de soja e pesando-se em balança analítica a massa de cada volume pipetado. Após isso foi construída uma curva de calibração, estabelecendo-se a quantidade de volume pipetado dos óleos vegetais para obter-se a massa necessária na proporção correta de cada óleo vegetal na mistura, através da equação da reta da curva de calibração, substituindo-se a variável x pela massa necessária.

Foram preparadas misturas de óleos de oliva, canola e girassol com óleos adquiridos no comércio da cidade de Toledo, em proporções de 0,00 a 100 % de azeite de oliva, totalizando 120 misturas. As misturas foram pesadas em balança analítica após cada adição de óleo vegetal, sucessivamente, com peso médio de 1,5000 g. Para as misturas onde a proporção de algum óleo vegetal foi muito pequena (< 5 %) a massa da mistura variou de 3,000 a 5,000 g. As amostras de 1,5000 g foram armazenadas em tubos de microcentrífuga, e as massas totais que foram de 3,0000 g e 5,0000 g armazenadas em frascos âmbar. As misturas foram armazenadas em freezer (-20 °C) até a realização da análise. Antes da análise, as misturas foram descongeladas e submetidas à homogeneização por meio de agitador vortex. Processo idêntico foi utilizado para o preparo das amostras de mistura de azeite de oliva e óleo de girassol, constituindo-se uma mistura binária.

4.3 Instrumentos e aquisição dos espectros

Os espectros das amostras foram obtidos em um espectrômetro de infravermelho com transformada de Fourier e módulo de refletância total atenuada (Perkin Elmer, Spectrum 65) (Figura 3). Foi utilizada a faixa de 3500 cm^{-1} a 650 cm^{-1} com resolução de 4 cm^{-1} e 40 varreduras, obtendo-se 4 espectros para cada amostra.

O cristal utilizado nessa técnica contém diamante na parte superior e seleneto de zinco no elemento focalizado. Os espectros foram normalizados em relação a banda 1744 cm^{-1} , que corresponde ao estiramento C=O de ésteres (TAVARES et al.,2012), com suas unidades de absorvância normalizadas entre 0 e 1, com o software Spectrum, versão 10.02, que também foi utilizado para a aquisição dos espectros, e posteriormente realizou-se a média entre os 4 espectros de cada amostra.

Figura 4 : Equipamento de Espectroscopia no Infravermelho Perkin Elmer Spectrum 65.

Fonte: autoria própria.



Os espectros foram submetidos a correção do espalhamento multiplicativo (MSC), sendo centrados na média ou submetidos a primeira e segunda derivadas. A seguir, os espectros pré-processados das misturas foram divididos em conjuntos de calibração ($n = 95$) e validação ($n = 33$) pelo algoritmo de Kennard-Stone.

4.4 Seleção das variáveis

O conjunto de espectros utilizados para a calibração foi proveniente de 128 amostras de misturas de óleo de oliva com óleos de soja e canola e 120 amostras de mistura de azeite de oliva com óleo de girassol. O conjunto de amostras então foi dividido em conjunto de calibração e validação através do algoritmo de Kennard – Stone, utilizando o software Matlab com o pacote SPA_GUI.

A seleção de números de onda utilizados no modelo foi obtida por meio do algoritmo de projeções sucessivas em um pacote do software MatLab (SPA – Successive Projections Algorithm), com um número mínimo de 5 variáveis, e máximo de 15 variáveis, aumentando progressivamente até o número máximo de 40 variáveis selecionadas.

Após cada seleção de variáveis realizada foi avaliado o valor de coeficiente de correlação múltipla (R), erro médio quadrático de previsão (RMSEP), erro médio quadrático de calibração (RMSEC) de cada um dos modelos de regressão linear múltipla obtidos com as variáveis selecionadas pelo algoritmo, a fim de constatar o melhor modelo de calibração para as determinações analíticas. O trabalho realizado verificou os coeficientes do modelo para os conjuntos de calibração e validação dos modelos de predição RLM, averiguando a qualidade desses modelos para as determinações analíticas experimentais.

4.5 Análise de *outliers*

Após as averiguações dos melhores modelos de regressão foi necessário avaliar a presença de amostras anômalas de tais modelos, a fim de aprimorar os dados como a precisão e a exatidão do método analítico. Os *outliers* foram analisados segundo a *leverage* e os resíduos *studentizados* para cada amostra, e as amostras fora dos padrões aceitáveis desses parâmetros excluídas dos modelos de calibração e validação.

Após a retirada das amostras anômalas os modelos foram recalculados com os mesmos números de variáveis e analisados conforme R, RMSEP e RMSEC.

4.6 Figuras de mérito

Após a calibração multivariada foram avaliadas as figuras de mérito: exatidão, precisão, sensibilidade, limite de detecção e quantificação.

A exatidão foi determinada a partir do cálculo do RMSEC para o conjunto de calibração e do RMSEP para o conjunto de validação. A repetibilidade foi calculada a partir de seis replicatas de uma amostra, obtendo-se seis espectros por replicata conforme método descrito pela ASTM E1655-0 (ASTM, 2005).

A precisão média foi determinada a partir de um número de amostras correspondente ao número de variáveis do modelo, com valores distribuídos em 95% da faixa de concentração do azeite de oliva, sendo obtidos seis espectros por amostra, com o cálculo da precisão realizado a partir da equação (24) (POPPI, 2010).

$$precisão = \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^m (\hat{y}_{ij} - \hat{y}_i)^2}{n(m-1)}} \quad (24)$$

Onde m é o número de replicatas feitas, n o número de amostras, e \hat{y}_{ij} é a média dos valores previstos de cada amostra \hat{y}_i .

A sensibilidade é definida como a fração de sinal responsável pelo acréscimo de uma unidade de concentração da propriedade de interesse, em modelos multivariados é definida pela equação (25) (POPPI, 2010) :

$$SEN = \frac{1}{\|b_k\|} \quad (25)$$

Onde b são os vetores de regressão para k componentes do modelo de regressão linear múltipla.

O limite de detecção é calculado a partir da expressão (26) (OLIVIERI, 2014):

$$LOD = 3\|\epsilon\| \|b_k\| \quad (26)$$

Em que ϵ equivale ao ruído instrumental registrado a partir de espectros de amostras do branco e calculado o desvio padrão da norma do NAS para essas amostras, onde foram utilizados espectros de placa metálica.

O limite de quantificação é equivalente a 3,3 vezes o limite de detecção, conforme expressão (27) (OLIVIERI, 2014).

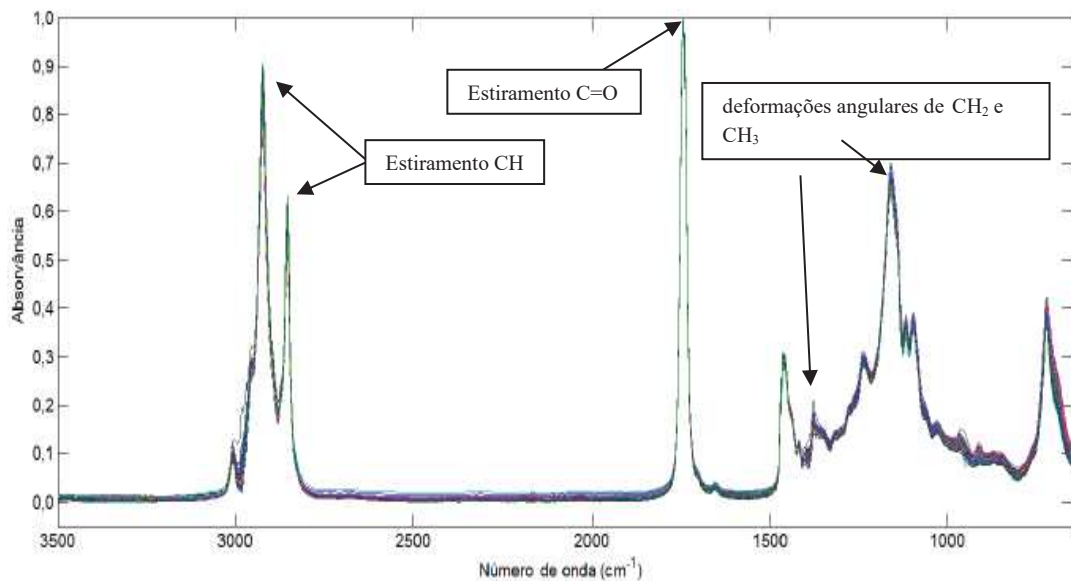
$$LOQ = 3,3 . LOD \quad (27)$$

5 RESULTADOS E DISCUSSÕES

A figura 5 apresenta os espectros das misturas de azeite de oliva, óleo de soja e óleo de canola na região do infravermelho médio e as principais bandas que ocorrem em: 1742 cm^{-1} e corresponde ao estiramento C=O do grupo éster dos triacilgliceróis; $2923\text{-}2853\text{ cm}^{-1}$ e estão relacionadas aos estiramentos C-H dos grupos metileno e metil terminais das cadeias de ácidos graxos. As bandas entre $1400\text{-}1200\text{ cm}^{-1}$ estão associadas com deformações angulares de CH_2 e CH_3 . Em 722 cm^{-1} ocorre a banda que está relacionada a deformação fora do plano (C-H) de hidrogênios olefinicos de duplas ligações *cis* que está sobreposta com o balanço da cadeia $(\text{CH}_2)_n$ (DE LA MATA et al., 2012).

Figura 5: Espectros das misturas de azeite de oliva, soja e canola

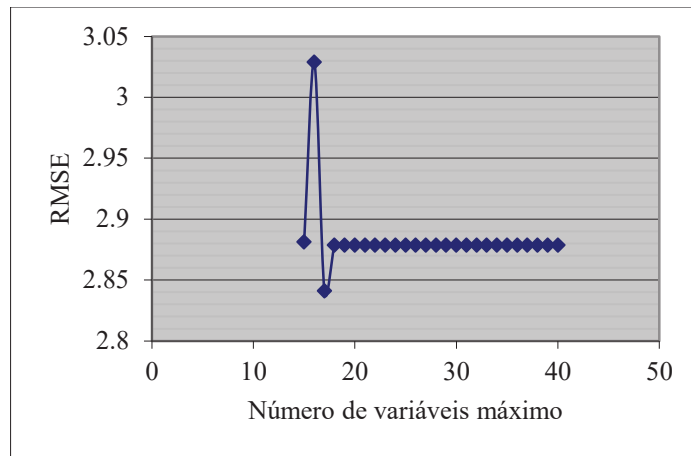
Fonte: autoria própria.



O número de variáveis selecionadas pelo método SPA para os modelos de regressão linear múltipla variou de 5 a 15. A figura 6 apresenta o gráfico das variáveis selecionadas para o modelo com um número máximo inicial de 15 variáveis até um número máximo final de 40 variáveis. Foram obtidos modelos para misturas de azeite de oliva, óleo de soja e óleo de canola, cujos espectros foram submetidos à correção do espalhamento multiplicativo (MSC - *multiplicative scatter correction*). A partir de 18 variáveis máximas, as variáveis selecionadas permaneceram constantes (15 variáveis).

Figura 6: Valores de RMSE pelo número de variáveis selecionadas

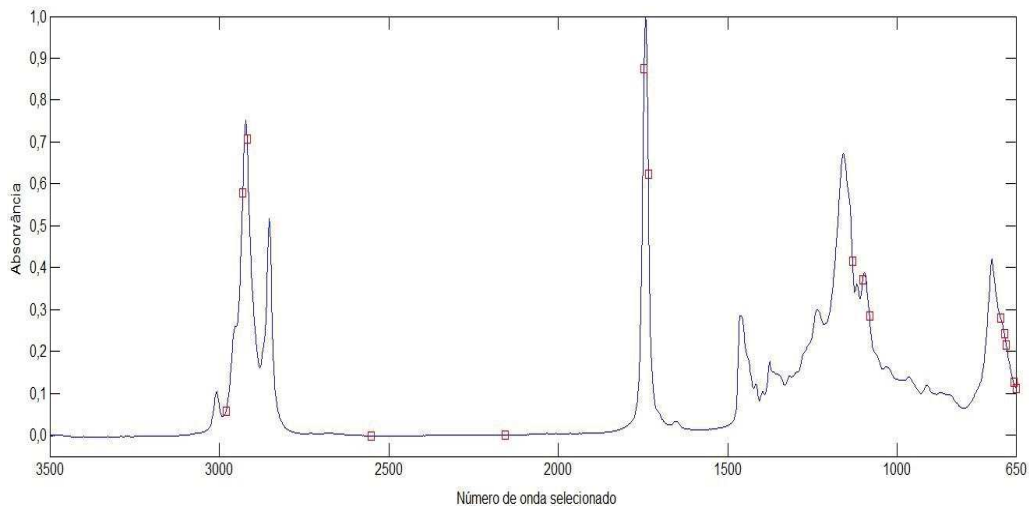
Fonte: Autoria própria.



É possível verificar que das 15 variáveis selecionadas para os modelos MSC das misturas ternárias, somente duas (2551 e 2156 cm^{-1}) apresentaram valores de absorvância muito baixos. Variáveis com absorvância baixa podem ser afetadas pelos ruídos instrumentais, abaixando a precisão do método.

Figura 7: Números de onda selecionados pelo SPA para as misturas ternárias

Fonte: autoria própria.



A tabela 1 apresenta os resultados obtidos para os maiores valores do coeficiente de correlação (R), determinados entre os valores de referência e os valores preditos, para os modelos sem pré-processamento e com pré-processamento, para os conjuntos de calibração e validação. Os valores de R variaram de 0,9507 a 0,9913 para o conjunto de calibração e de 0,9633 a 0,9973 para o conjunto de validação, das misturas de azeite de oliva com óleos de soja e canola.

Foram obtidos resultados semelhantes quando foram utilizados os espectros sem pré-processamento e utilizando a correção MSC, a centralização na média e o autoescalonamento, tanto para o conjunto de calibração como para o de validação. Entretanto, o autoescalonamento resultou nos menores valores de RMSEC e RMSEP, indicando que pode ser mais adequado empregar este pré-processamento, pela qualidade de predição dos teores da validação e baixo erro no cálculo através das variáveis selecionadas.

Tabela 1: Dados dos modelos obtidos a partir das misturas de azeite de oliva, soja e canola

Pré - Processamento	Número de variáveis	Calibração		Validação	
		R	RMSEC	R	RMSEP
-	12	0,9902	3,2958	0,9972	2,841
Correção MSC	15	0,9913	3,2725	0,9973	2,512
Centrados na média	12	0,9902	3,2958	0,9972	2,841
Autoescalonamento	14	0,9904	3,2632	0,9968	2,135
Primeira derivada	15	0,9855	3,6153	0,9944	3,503
Segunda derivada	6	0,9507	3,3038	0,9633	7,154

Para as misturas de azeite de oliva e óleo de girassol, os valores de R variaram de 0,9927 a 0,9993, para o conjunto de calibração e de 0,9982 a 0,9999, para o conjunto de validação. Ao analisar os valores de RMSEC e RMSEP foram obtidos valores muito próximos quando foram utilizados os espectros sem pré-processamento e com pré-processamento, tanto para o conjunto de calibração como para o de validação. A tabela 2 apresenta a quantidade de variáveis utilizadas nos modelos RLM. Os dados obtidos apresentaram boas capacidades preditivas com baixo número de *outliers*, tanto nos modelos de calibração e validação. A retirada das amostras *outliers* resultou na diminuição do coeficiente de correlação e RMSEP, e dessa forma não foi necessária a retirada dessas amostras.

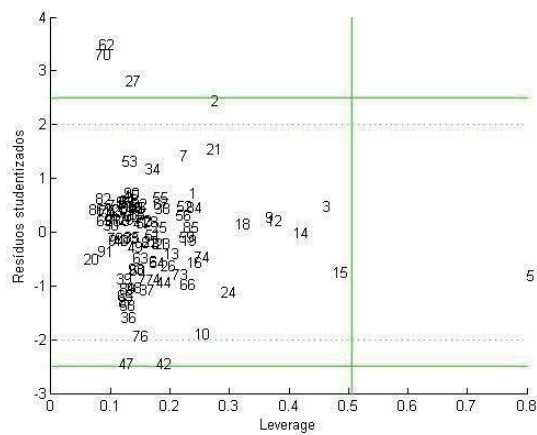
Tabela 2: Dados dos modelos RLM para as misturas de azeite de oliva e óleo de girassol

Pré -Processamento	Número de variáveis	Calibração		Validação	
		R	RMSEC	R	RMSEP
-	6	0,9993	1,1659	0,9984	1,614
Correção MSC	5	0,9981	1,8942	0,9982	1,5866
Centrados na média	11	0,9918	1,2607	0,9999	0,1885
Autoescalonamento	13	0,9985	1,7121	0,9999	0,5760
Primeira derivada	13	0,9927	3,6497	0,9997	0,2371
Segunda derivada	5	0,9983	1,8016	0,9978	1,8382

Foi realizada uma avaliação de amostras anômalas (*outliers*) dos conjuntos de calibração e validação das misturas de azeite de oliva, soja e canola utilizando os valores de *leverage* versus resíduos *studentizados*, seguido da obtenção de um novo modelo por regressão linear múltipla.

Figura 8: *leverage* versus resíduos *studentizados* do modelo selecionado para a determinação do teor de azeite de oliva em mistura com óleos de soja e canola

Fonte: Autoria própria.



No conjunto de validação para os espectros de oliva, soja e canola, com correção MSC encontraram-se 2 *outliers*. Nos espectros sem processamento também detectou-se 2 *outliers* no modelo de validação; para os modelos autoescalados detectou-se uma amostra fora do padrão. Os dados dos modelos em que removeram-se os *outliers* estão representados na tabela 3.

Tabela 3: Dados dos modelos RLM sem *outliers* das misturas de azeite de oliva, e óleos de soja e canola

Pré -Processamento	Número de variáveis	Calibração		Validação	
		R	RMSEC	R	RMSEP
-	12	0,9966	1,977	0,9977	2,6302
Correção MSC	15	0,9971	1,8928	0,9980	2,1387
Autoescalamento	14	0,9942	2,6353	0,9978	2,8856

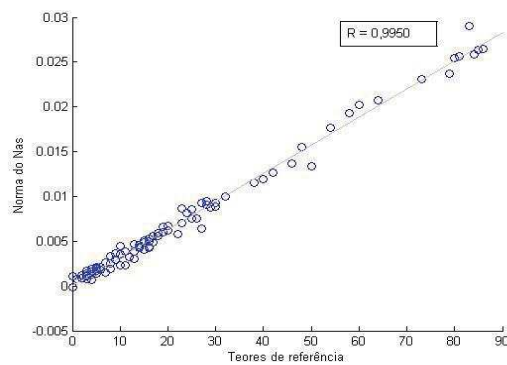
Aferiu-se que a retirada dos *outliers* melhorou a capacidade preditiva dos modelos para a mistura ternária de azeite de oliva com óleos de soja e canola, em relação a tabela 1, ao diminuir a raiz quadrada do erro médio de calibração (RMSEC).

5.1 Figuras de Mérito

Avaliou-se no trabalho desenvolvido os limite de detecção e quantificação, sensibilidade, precisão, exatidão e linearidade, como meio de estabelecer uma predição confiável para a quantificação de azeite de oliva em misturas com outros óleos vegetais, e assim concretizar o desenvolvimento do método.

Figura 9: Norma do NAS versus Teores de referência de azeite de oliva com óleos de soja e canola

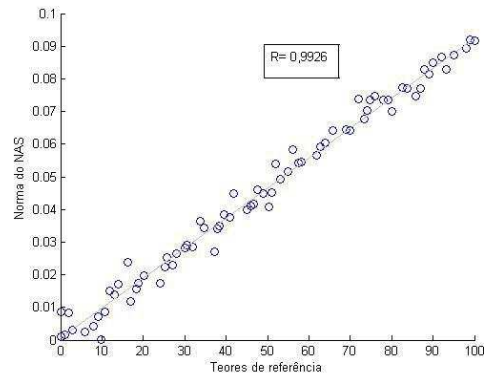
Fonte: Autoria própria.



Realizou-se o cálculo do NAS para os modelos de calibração MLR selecionados, das misturas de azeite de oliva, óleos de soja e canola e também para as misturas de azeite de oliva e óleo de girassol.

Figura 10: Norma do NAS versus teores de referência do teor de azeite do oliva em mistura com óleo de girassol

Fonte: Autoria Própria.



As figuras 9 e 10 representam a norma do sinal analítico líquido e os teores de referência do modelo de calibração para os espectros processados pela correção MSC. Em ambos os cálculos pode-se observar a alta correlação dos dados (R).

As figuras de mérito expressas nas tabelas 4 e 5 foram obtidas a partir do sinal analítico líquido calculado para os modelos de calibração, tanto para as misturas de azeite de oliva com óleos de soja e canola, quanto para azeite de oliva e óleo de girassol. Todos os dados apresentados são para modelos sem a presença de *outliers* no conjunto de calibração e validação.

Tabela 4: Figuras de mérito para os modelos das misturas de azeite de oliva, com óleos de soja e canola

Pré - Processamento	Número de variáveis	LOD (%m/m)	LOQ (%m/m)	Sensibilidade	Repetibilidade	Exatidão (RMSEP)
-	12	1,12	3,70	$3,06 \times 10^{-4}$	2,58	2,63
Correção MSC	15	1,49	4,93	$2,91 \times 10^{-4}$	1,23	2,14
Primeira derivada	15	5,56	18,36	$3,89 \times 10^{-5}$	1,33	2,88

Os valores de LOD e LOQ foram melhores para os espectros sem pré-processamento, porém a repetibilidade e a exatidão ficaram com melhores valores para a correção MSC, indicando a baixa variabilidade dos dados de calibração através desse processamento.

Tabela 5: Figuras de mérito para os modelos das misturas de azeite de oliva e óleo de girassol

Pré - Processamento	Número de variáveis	LOD (%m/m)	LOQ (%m/m)	Sensibilidade	Precisão	Repetibilidade	Exatidão (RMSEP)
-	6	0,28	0,93	$5,93 \times 10^{-4}$	1,93	7,23	1,61
Correção MSC	5	0,13	0,43	$9,12 \times 10^{-4}$	1,96	0,33	1,58
Primeira derivada	13	0,68	2,25	$8,29 \times 10^{-5}$	2,44	2,41	1,58

Para os modelos das misturas de azeite de oliva e óleo de girassol, a correção MSC apresentou menor LOD e LOQ, baixa variação expressa pela repetibilidade, e exatidão mais adequada, quando comparado ao modelo de 6 variáveis sem pré-processamento.

Christy et al. (2004) empregaram a espectroscopia NIR associada a calibração por PLS para detectar adulteração em azeite de oliva com diferentes tipos de óleos vegetais, dentre eles, o óleo de soja, onde obtiveram um erro médio de calibração e predição de 0,87 % e 0,96 %, respectivamente, em níveis de adulteração de 0-100 % (m/m).

Vasconcelos et al. (2015) obtiveram resultados de limite de detecção para misturas de azeite de oliva e óleo de algodão e colza, os valores ficaram em 1,4 e 1,32% v/v, respectivamente.

Mariani et al. (2016) também empregou a técnica de espectroscopia no infravermelho médio, e obteve o resultado de limite de detecção entre 1,6 a 3,1 g 100g⁻¹, e limite de quantificação entre 4,8 a 9,2 g 100g⁻¹.

6 CONCLUSÕES

O trabalho teve por objetivo desenvolver uma metodologia analítica para a quantificação de azeite de oliva em mistura com outros óleos vegetais refinados. Os resultados das figuras de mérito determinadas para o modelo MLR discutidos demonstram que modelos de calibração multivariada podem ser validados, de forma a atender às exigências impostas por órgãos de fiscalização e padronização.

Os métodos desenvolvidos para a quantificação de azeite de oliva em misturas ternárias com óleos de soja e canola e binárias com óleo de girassol apresentaram uma boa exatidão para a faixa de concentração avaliada (0 a 100,0 g 100 g⁻¹), indicando que estes métodos podem ser aplicados na quantificação de azeite de oliva adulterado e em óleos mistos, ambos contendo óleos de soja e canola, ou óleo de girassol.

A metodologia desenvolvida também mostrou ser sensível as mudanças espectrais causadas pelas variações dos teores de azeite de oliva em misturas com óleos de soja e canola, e óleo de girassol, com valores de sensibilidade variando entre $3,8917 \times 10^{-5}$ até $3,0575 \times 10^{-4}$ e $5,9324 \times 10^{-4}$ até $9,1255 \times 10^{-4}$, respectivamente.

Os valores do limite de detecção e limite de quantificação ficaram na faixa de 0,2093 g 100 g⁻¹ até 6,31g 100 g⁻¹, apresentando a validade do método de espectroscopia no infravermelho médio associada a regressão linear múltipla com a seleção das variáveis pelo SPA. Os resultados obtidos ficaram com baixa colinearidade, obtida pela seleção das variáveis pelo SPA e baixo índice de *outliers*.

O resultados dos parâmetros estimados foram consistentes e de boa concordância com os resultados do modelo e valores teoricamente esperados.

Referências

Almeida, Carlos Alberto Nogueira-de- et al. **Azeite de Oliva e suas propriedades em preparações quentes: revisão da literatura.** [S.l.], International Journal of Nutrology, v. 8, n. 2, p. 13-20, 2015.

Araújo, Mário C. U. et al. **The sucessive projections algorithm for variable selection in spectroscopic multicomponent analysis.** Brasil, Chemometrics and Intelligent Laboratory System, Vol. 57, p. 65-73, 2001.

ASTM International: **E1655-05 Standard Practices for Infrared Multivariate Quantitative Analysis**, 2005.

Beebe, R. Kenneth; Pell, Randy J.; Seasholtz, Beth Mary. **Chemometrics , A Practical Guide.** Vol.1. [s.l.]. Editora John Wiley. 1998.

Bro, Rasmus. **Multivariate calibration. What is in chemometrics for the analytical chemist?**. Analytica Chimica Acta. Dinamarca, V. 500 , n.1-2, p. 185-194, 2003.

Bro, Rasmus; Andersen, Charlotte M. **Theory of Net Analyte Signal Vectors in Inverse Regression.** Journal of Chemometrics, Dinamarca. V. 17, p.646-652, 2006.

Cristy, A. Alfred; Kasenmumram, S.; Du, Yiping; Yukihiro, Osaki. **The detection and quantification of adulteration in olive oil by Infrared Spectroscopy and Chemometrics.** Analytical Sciences, Japão. V. 20, n. 6, p. 935-939, 2004.

De la Mata, P.; Dominguez-Vidal, A.; Bosque-Sendra, J. M.; Ruiz-Medina, A.; Cuadros-Rodriguez, L.; Ayora-Canada, M. J. **Olive oil assessment in edible oil blends by means of ATR-FTIR and chemometrics.** Food Control, [s.l.] v. 23, n. 2, p. 449-455, 2012.

Ferreira, M.C Marcia et al. **Quimiometria I: calibração multivariada, um tutorial.** Brasil, Vol. 22 , N. 5, p. 724-731, 1999 .

Fito M, de la Torre R, Farre-Albaladejo M, Khymenetz O, Marrugat J, Covas MI. **Bioavailability and antioxidant effects of olive oil phenolic compounds in humans: a review.** Annali dell'Istituto superiore di sanita. [s.l] , 43(4):375-81, 2007.

Higson, Séamus P.J. **Química Analítica.** Vol. 1. N. 1. São Paulo. Editora Mac Graw Hill, 2009.

INSTITUTO TECNOLÓGICO DA AERONÁUTICA: **The sucessive algorithm projections.** Disponível em: <<http://www.ele.ita.br/~kawakami/spa/>>. Acesso em : 30 Nov. 2016.

INTERNATIONAL OLIVE COUNCIL: **Designations and definitions of olive oils.** 2018.

INTERNATIONAL ORGANIZATION OF STANDARDIZATION: **ISO 11843-1:1997.**

_Lerma-García, M.J. et al. **Authentication of extra virgin olive oils by Fourier-transform infrared spectroscopy.** United Kingdom, Food Chemistry, Vol. 118, p. 78-83, 2010.

Lorber, A. **Error propagation and figures of merit for quantification by solving matrix equations**. [s.l], Analytical Chemistry, Vol. 58, p. 1167-1172, 1986.

Lorber, A.; Faber, K.; Kowalsky, B. R. **Net Analyte Signal Calculation in Multivariate Calibration**. EUA, Anal. Chem, Vol. 69 (8), p. 1620–1626, 1997.

Mariani, Kelly; Silva, W.L; Martin, A. C. **Desenvolvimento de um método quimiométrico para quantificação de azeite de oliva em misturas ternárias com óleo de soja e canola**. ENDICT.IV.2016. Universidade Tecnológica Federal do Paraná Câmpus Toledo.

Martens, Harald; Naes, Tormod. **Multivariate Calibration**. 1º ed. [s.l]. Editora Wiley. 2002.

Mendes, Thiago O. et al. **Quantification of Extra-virgin Olive Oil Adulteration with Soybean Oil: a Comparative Study of NIR, MIR, and Raman Spectroscopy Associated with Chemometric Approaches**. Estados Unidos, Food Analytical Methods, Vol 8. N. 9, p. 2339- 2346, 2015.

Montgomery, Douglas C.; Runger, George C. **Estatística aplicada e probabilidade para engenheiros**. 2º ed. Rio de Janeiro. Editora LTC, 2008.

Noord, Onno E. **The influence of data pre-processing on the robustness and parsimony of multivariate calibration models**. Chemometrics and Intelligent Laboratory System. [s.l], Vol. 23, p. 65-70, 1994.

Olivieri, Alejandro C. et al. **Uncertainty estimation and figures of merit for multivariate calibration (IUPAC Technical Report)**. [s.l], Pure and applied Chemistry, The scientific journal of IUPAC. Jan. 2006. Vol. 78, N. 3, p. 633- 661.

Olivieri, Alejandro; Allegrini, Franco. **IUPAC-Consistent Approach to the Limit of Detection in Partial Least-Squares Calibration**. Analytical Chemistry. Jul. 2014. Vol. 86, N. 15, p. 7858 – 7866.

Pereda, Mariana et al. **Polyelectrolyte films based on chitosan/olive oil and reinforced with cellulose nanocrystals**. Carbohydrate Polymers, Argentina, Jan. 2014

Braga, Jez Willian B.; POPPI, Ronei Jesus. **Validação de modelos de calibração multivariada: uma aplicação na determinação de pureza polimórfica de carbamazepina por espectroscopia no infravermelho próximo**. São Paulo , Quím. Nova, v. 27, n. 6, p. 1004-1011, 2004 .

PROTESTE. **Quatro azeites adulterados em teste da PROTESTE**. Disponível em : <<http://www.proteste.org.br/azeite>>. Acesso em : 07 Set. 2016.

Robinsons, James W.; Frame Skelly Eileen M.; Frame II, George M. **Undergraduated instrumental analysis**. Vol. 1 .6º ed . [s.l]. Editora Dekker. 2005.

Silvestre, Daniel M. et al. **Experimento utilizando grãos para explorar a calibração em análises químicas**. Brasil, Química Nova, Jul. 2014. Vol. 37. N.9 p. 1545-1549.

Silva de Lima, Leomir Aires. **Estudo da potencialidade da espectroscopia de infravermelho próximo na análise de cabelo para diferenciar fumantes de não fumante**. 2013. 41 f. Dissertação (Mestrado). Universidade Federal do Rio Grande do Norte, 2013.

Skoog Douglas A.; Holler, F. James; Stanley R. Crouch. **Princípios de análise instrumental**. Vol. 1. Ed. 6ª. Brasil. Editora Bookman, 2009.

Soares, Sófacles F. C. et al. **The successive projections algorithm**. Trends in Analytical Chemistry. Jan. 2013. Vol. 42, p. 84-98.

Souza, Júlia Aparecida Lourenço. **Desenvolvimento de uma metodologia baseada na espectroscopia no infravermelho próximo para determinação dos produtos de degradação de captopril em comprimidos**. 2011. 49 f. Dissertação (Mestrado). Universidade Federal da Paraíba, 2011.

Tavares, Mansur K. et al. **Espectroscopia no infravermelho médio aplicada à detecção de adulteração de café torrado por adição de cascas de café**. São Paulo, Química Nova. Vol. 35, p. 1164-1168. 2012.

Valderrama, Patrícia et al. **Estado da arte de figuras de mérito em calibração multivariada. Estado da arte de figuras de mérito em calibração multivariada**. São Paulo, Química Nova, v. 32, n. 5, p. 1278-1287, 2009.

Valderrama, Patrícia. **Avaliação de figuras de mérito em calibração multivariada na determinação de parâmetros de controle de qualidade em indústria alcooleira por espectroscopia no infravermelho próximo**. 2005. 135 f. Dissertação (Mestrado). Universidade Estadual de Campinas, 2005.

Vlachos, N. et al. **Applications of Fourier transform-infrared spectroscopy to edible oils**. Grécia, Analytica Chimica Acta, 28 Jul. 2006. Vol. 573-574, p. 459-465.