

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

GUSTAVO DAVID DIAS CRUZ

USO DE ESPECTROMETRIA NO INFRAVERMELHO MÉDIO E
SELEÇÃO DE VARIÁVEIS POR INTERVALOS NO CONTROLE DE
QUALIDADE DE BIODIESEL DO PINHÃO MANSO

UBERLÂNDIA

2018

GUSTAVO DAVID DIAS CRUZ

USO DE ESPECTROMETRIA NO INFRAVERMELHO MÉDIO E
SELEÇÃO DE VARIÁVEIS POR INTERVALOS NO CONTROLE DE
QUALIDADE DO BIODIESEL DE PINHÃO MANSO

Trabalho de Conclusão de Curso apresentado ao curso de
Graduação em Química Industrial, do Instituto de Química
da Universidade Federal de Uberlândia, como requisito
para conclusão do curso de Química Industrial.

Área de concentração: Química Analítica

Orientador: Prof. Dr. Waldomiro Borges Neto

UBERLÂNDIA

2018

GUSTAVO DAVID DIAS CRUZ

USO DE ESPECTROMETRIA NO INFRAVERMELHO MÉDIO E
SELEÇÃO DE VARIÁVEIS POR INTERVALOS NO CONTROLE DE
QUALIDADE DO BIODIESEL DE PINHÃO MANSO

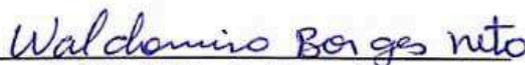
Trabalho de Conclusão de Curso apresentado ao curso de
Graduação em Química Industrial, do Instituto de Química
da Universidade Federal de Uberlândia, como requisito para
conclusão do curso de Química Industrial.

Área de concentração: Química Analítica

Orientador: Prof. Dr. Waldomiro Borges Neto

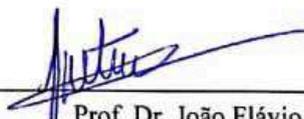
Uberlândia, 27 de Novembro de 2018.

BANCA EXAMINADORA



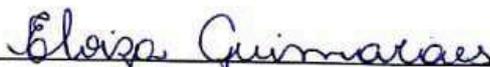
Prof. Dr. Waldomiro Borges Neto

Universidade Federal de Uberlândia - UFU



Prof. Dr. João Flávio da Silveira Petrucci

Universidade Federal de Uberlândia - UFU



Dra. Eloiza Guimarães

Universidade Federal de Uberlândia - UFU

AGRADECIMENTOS

“Somos a média das 5 pessoas que mais convivemos” Jim Rohn

Agradeço por TODAS as 5 pessoas que estiveram comigo em diversas fases da minha vida e que me moldaram a ser quem eu sou, por me conceder toda sabedoria e compreensão de mundo para continuar meus trabalhos e conseguir alcançar meus objetivos.

Ao meu orientador, prof. Dr. Waldomiro Borges Neto, por me mostrar uma área de conhecimento muito bela e por ser uma grande pessoa.

À Universidade Federal de Uberlândia e aos docentes do curso de Química Industrial por todo conhecimento, e contribuição para minha formação acadêmica.

Ao CNPq pela bolsa e a FAPEMIG pelo suporte financeiro.

Agradeço aos meus pais Sayonara e Cláudio e meus irmãos Leonardo e Gabriel pela união e pelos ensinamentos virtuosos que me proporcionaram. Serei muito grato por todo esforço que fizeram por mim para poder me dar o melhor estudo possível.

Agradeço aos meus amigos e principalmente a minha namorada Maria Gabriela, por me permitir sonhar grande, me ensinar valores de vida e por me ajudar no meu amadurecimento pessoal e profissional.

A todos os integrantes do LQT por todo compartilhamento de conhecimento e ajuda durante toda a minha iniciação científica.

RESUMO

O aumento da demanda energética é de fato uma constatação e a busca de energias renováveis no mundo cresce à medida que é necessário impulsionar alternativas menos prejudiciais ao meio ambiente. Analogamente, o biodiesel é apontado como uma boa opção de contribuição no desenvolvimento sustentável das áreas ambientais, sociais e econômicas.

O Pinhão Manso destaca-se das demais oleaginosas devido aos altos teores de óleos produzidos, além de ser bem resistente às terras secas e não ter finalidade alimentícia, como a soja. Simultaneamente a produção do biodiesel como fonte energética, as agências fiscalizadoras, como a ANP (Agência Nacional do Petróleo, Gás Natural e Biocombustíveis), se preocupam em assegurar a qualidade deste combustível no país, fazendo-se necessário o desenvolvimento de métodos para identificação e quantificação de teores em misturas biodiesel/diesel, a fim de verificar condições similares previstas pela lei nº 13.263/2016.

O presente trabalho teve como objetivo a utilização da Espectrometria no Infravermelho Médio (FT-MIR) aliada a métodos quimiométricos de calibração multivariada e seleção de variáveis por intervalos (iPLS) para quantificar o teor do biodiesel metílico de Pinhão Manso em misturas com diesel na faixa de concentração 0,75% a 30,00% (v/v). O iPLS permite a seleção de um único intervalo da região espectral, com a finalidade de melhorar significativamente o desempenho do método de calibração multivariada em relação a toda informação relacionada a região espectral e sua validação é baseada na norma ASTM E1655-05 e outras figuras de mérito relevantes.

Os espectros MIR foram adquiridos em quintuplicatas, no espectrofotômetro Spectrum Two (Perkin Elmer) com HATR, com cristal de ZnSe, na faixa de 4000 – 600 cm^{-1} resolução de 4 cm^{-1} e 16 varreduras. O programa utilizado para a construção dos modelos foi o Matlab versão R2015a juntamente com PLS_Toolbox 8.61 e iToolbox 1.0. Os modelos iPLS foram construídos dividindo-se a região espectral em 5, 10, 15, 20, 25, 30, 35, 40, 45 e 50 intervalos.

Portanto, o seguinte trabalho visou quantificar o teor de biodiesel metílico de Pinhão Manso na mistura com diesel assim como melhorar o modelo matemático utilizando a seleção de variáveis.

PALAVRAS-CHAVE: Biocombustível, FT-MIR, Quimiometria, Calibração Multivariada, iPLS, Pinhão Manso

ABSTRACT

The increase of energy demand indeed a fact and the pursue for renewable energies in the world grows as it is necessary to promote less harmful alternatives to the environment. Similarly, biodiesel is considered as a good alternative to contribute to the sustainable development of the environmental, social and economic areas.

Jatropha stands out from other oilseeds due to the high oil content produced, as well as being resistant to dry lands and not having a food purpose, such as soybeans. Simultaneously with the production of biodiesel as an energy source, inspection agencies, such as ANP (National Agency for Petroleum, Natural Gas and Biofuels), are concerned with ensuring the quality of this fuel in the country, making it necessary to develop methods for identification and quantification of contents in biodiesel / diesel blends, in order to verify similar conditions provided by Law n° 13.263 / 2016.

In this work, the objective was to use the MIR spectrometry (FT-MIR) allied to the chemometric methods of multivariate calibration and selection of variables by intervals (iPLS) to quantify the content of Jatropha's biodiesel in blends with diesel in the range 0.75% to 30.00% (v / v). The iPLS allows the selection of a single range of the spectral region in order to significantly improve the performance of the multivariate calibration method in relation to all spectral region related information and its validation is based on ASTM E1655-05 and other relevant figures of merit.

The MIR spectra were acquired in quintuplicate in Spectrum Two spectrophotometer (Perkin Elmer) with HATR, with ZnSe crystal, in the range of 4000 - 600 cm^{-1} resolution of 4 cm^{-1} and 16 scans. The program used to build the models was the Matlab version R2015a along with PLS_Toolbox 8.61 and iToolbox 1.0. The iPLS models were constructed by dividing the spectral region into 5, 10, 15, 20, 25, 30, 35, 40, 45 and 50 intervals.

Therefore, the following work was aimed at quantifying the methyl content of Jathopha in the mixture with diesel as well as improving the mathematical model using variable selection.

KEYWORDS: biofuel, FT-MIR, Chemometric, multivariate calibration, iPLS, Jatropha.

LISTA DE ILUSTRAÇÕES

Figura 1. Percentual da não conformidade do combustível diesel.....	12
Figura 2. O fruto encapsula três sementes que possuem amêndoas ricas em óleo.....	12
Figura 3. (a) Reflectância de radiação no interior do cristal de ZnSe; (b) Acessório de Reflectância total de ZnSe acoplado ao suporte FT-MIR SpectrumTwo (PerkinElmer)	13
Figura 4. Representação da construção da matriz X para modelagem multivariada	15
Figura 5. Representação Geométrica do NAS	17
Figura 6. Espectros MIR sobrepostos da mistura do diesel com BMPM.....	23
Figura 7. Espectros finais sobrepostos da mistura do diesel com BMPM	24
Figura 8. (a) Relação do número de variáveis latentes pelos erros RMSEC, RMSECV e RMSEP. (b) Relação do número de variáveis latentes pelo R^2 da validação cruzada. (c) Relação do número de variáveis latentes pelo erro sistemático de validação cruzada	25
Figura 9. Regressão dos valores reais pelos valores previstos pelo modelo	26
Figura 10. Região escolhida pelo algoritmo para a construção do novo PLS.....	27
Figura 11. (a) Escolha do número de variáveis latentes por comparação de erros RMSEC, RMSECV e RMSEP. (b) Escolha do número de variáveis latentes por avaliação do R^2 da validação cruzada. (c) Escolha do número de variáveis latentes devido ao erro sistemático de validação cruzada.....	28
Figura 12. Mensuração de amostras anômalas pelo <i>Leverage</i> e $Q_{residual}$ com uma confiabilidade de 95%.....	29
Figura 13. Linearidade dos valores reais frente aos valores previstos pelo modelo.....	29
Figura 14. (a) Elipse de Confiança do modelo (b) Avaliação dos erros absolutos.....	30

LISTA DE TABELAS

Tabela 1. Escolha da faixa a ser usada para a seleção de variáveis através do iPLS.....	27
Tabela 2. Comparação dos modelos pela estatística F.....	28
Tabela 3. Comparação dos erros do modelo global e do modelo de seleção de variáveis.....	30
Tabela 4. Figuras de mérito para o modelo de seleção de variáveis.....	31

LISTA DE ABREVIATURAS E SIGLAS

ANP - Agência Nacional do Petróleo, Gás Natural e Biocombustíveis

ASTM - *American Society for Testing and Materials*

ATR- Reflectância Total Atenuada (do inglês, *Attenuated Total Reflectance*)

BX - X% de biodiesel na mistura com diesel

B10 - 10% de biodiesel na mistura com diesel

BMPM - Biodiesel Metílico de Pinhão Manso

FT-MIR - Espectroscopia no Infravermelho Médio com Transformada de Fourier

iPLS - Quadrados Mínimos Parciais por Intervalos (do inglês, *Interval Partial Least Squares*)

NAS – Sinal Analítico Líquido (do inglês *Net Analyte Signal*)

PLS - Quadrados Mínimos Parciais (do inglês *Partial Least Squares*)

RMSEC - Erro Quadrático Médio da Calibração (do inglês, *Root Mean Square Error of Calibration*)

RMSECV - Erro Quadrático Médio da Validação Cruzada (do inglês, *Root Mean Square Error of Cross Validation*)

RMSEP - Erro Quadrático Médio da Previsão (do inglês, *Root Mean Square Error of Prediction*)

VL – Variável Latente

SUMÁRIO

1) INTRODUÇÃO E JUSTIFICATIVA	11
1.1. Óleo e Biodiesel Metílico de Pinhão Manso.....	12
1.2. Espectroscopia no Infravermelho Médio.....	13
1.3. Quimiometria.....	14
1.3.1. Quadrados Mínimos Parciais (PLS).....	15
1.3.2. Quadrados Mínimos Parciais por Intervalos (iPLS).....	16
1.3.3. Sinal Analítico Líquido – NAS.....	17
1.3.4. Validação do Método.....	17
2) OBJETIVOS.....	21
2.1. Geral	21
2.2. Específicos	21
3) METODOLOGIA.....	22
3.1. Preparo do B100	22
3.2. Preparo das Misturas biodiesel/diesel	22
3.3. Obtenção dos Espectros MIR.....	23
3.4. Modelos Quimiométricos	23
4) RESULTADOS E DISCUSSÃO	23
4.1. PLS	24
4.2. iPLS	26
5) CONCLUSÃO	32
6) REFERÊNCIAS	33

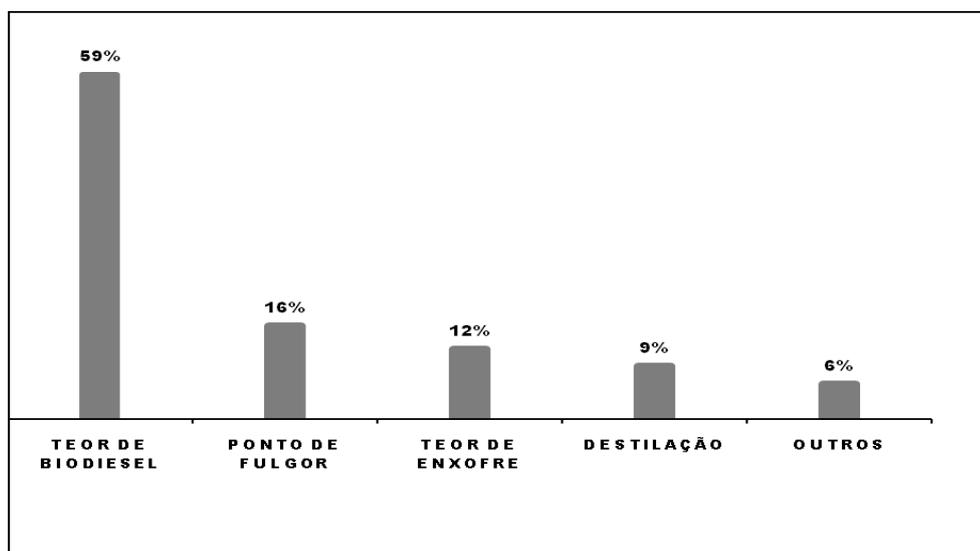
1. INTRODUÇÃO

A maioria do consumo de energia do mundo vem do petróleo, carvão e gás natural, mas estas fontes são limitadas por causa de sua origem fóssil além de causar danos ao meio ambiente. Estudos relatam que o uso de combustíveis alternativos, principalmente os derivados de fontes renováveis como o biodiesel, pode reduzir significativamente os danos ambientais causados, diminuir o déficit devido à importação de diesel, diversificar o número de matérias primas, melhora a renda social, pois favorece a produção em programas de agricultura familiar, dentre outras vantagens quando comparado ao diesel (PETERSON; HUSTRULD, 1998).

Na produção de biodiesel no país utiliza-se na maioria das vezes o óleo de soja, que por ser usado na alimentação pode levar a um aumento no custo de produção do biodiesel. Como o biodiesel pode ser produzido a partir de diferentes oleaginosas e gorduras tornam-se importante os estudos para produzir biodiesel, principalmente de oleaginosas que não são usadas na alimentação e que apresentem um teor de óleo igual ou superior ao da soja (PADULA *et al.*, 2012). Neste sentido, o Pinhão Manso, uma planta do cerrado se mostra como uma alternativa promissora como matéria prima para a produção de biodiesel. Como as propriedades físico-químicas do biodiesel são semelhantes às do óleo diesel, em quaisquer proporções, o biodiesel pode ser utilizado na sua forma pura ou em mistura com diesel em motores automotivos. O biodiesel usado na sua forma pura é denominado B100 e em misturas com o diesel BX, onde X corresponde à proporção percentual volumétrica do biodiesel na mistura (ABNT, 2008).

Assim, a comercialização é dada por uma mistura com o diesel estabelecido pela lei nº 13.263/2016, atualmente em vigência o B10 (BRASIL, 2016), sendo esta mistura a maior causa de não conformidade da qualidade de biocombustíveis, como mostra a figura 1. Deste modo, é de suma importância a padronização para controle da qualidade desses biocombustíveis, assim como uma análise rápida, reprodutível e de baixo custo. Portanto, a utilização de ferramentas quimiométricas é uma boa alternativa para o controle da mistura, uma vez que atendem os requisitos.

Figura 1. Percentual da não conformidade do combustível diesel.



Fonte: Boletim de Monitoramento da Qualidade de Biocombustíveis, 2018- ANP.

1.1. Óleo e Biodiesel Metílico de Pinhão Manso

O Pinhão Manso (*Jatropha curcas* L.) se destaca das demais oleaginosas de potencial crescimento por causa do alto teor de óleo, em especial do seu fruto que possui 3 amêndoas ricas em óleo, mostrado pela figura 2. Trata-se de uma planta perene com ciclo produtivo de até 40 anos, resistente à seca e se adapta a várias condições de clima e solo, porém apresenta como desvantagem não apresentar uma maturação dos frutos uniforme, o que poderia ser corrigido por alteração genética. Essas características aliadas à sua alta produtividade por área e rendimento de óleo, o torna uma cultura que pode ser desenvolvida em pequenas propriedades de agricultura familiar em regiões semiáridas, sendo propício ao território nacional que contém mais de 2.045.000 km² só de cerrado brasileiro (PADULA *et al.*, 2012).

Figura 2. O fruto encapsula três sementes que possuem amêndoas ricas em óleo.



Fonte: Ciclovivo, 2011

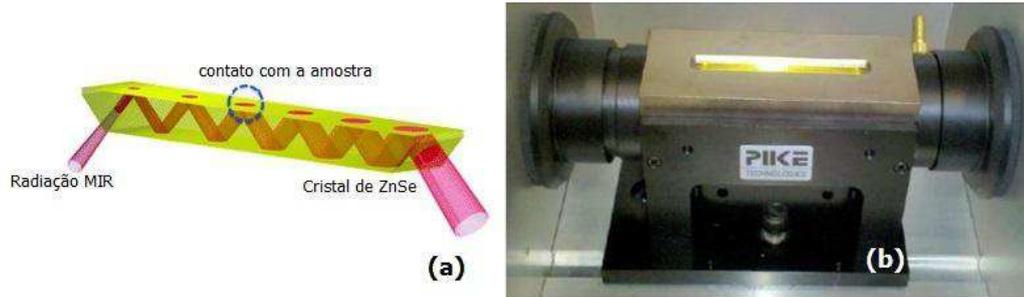
Além disso, quimicamente, possui um elevado teor de ácidos graxos saturados e um baixo teor de ácidos graxos insaturados, de forma que o biodiesel produzido a partir do Pinhão Manso apresenta boa estabilidade à oxidação (SARIN *et al.*, 2007). É comum se produzir dois tipos de biodieséis, o metílico que tem como reagente o metanol e o etílico que por sua vez é feito a partir do etanol. A preferência pela produção do biodiesel metílico de pinhão manso frente ao etílico é proporcionada pelas vantagens no rendimento de ésteres, consumo de álcool e eficiência energética (LÔBO; FERREIRA; CRUZ, 2009).

1.2. Espectroscopia no Infravermelho Médio

O espectro na região do Infravermelho Médio (MIR) é a faixa de frequência responsável pelas transições vibracionais e rotacionais das ligações das moléculas, faixa esta que compreende a região de 4.000 a 200 cm^{-1} . Com o aprimoramento da técnica, a Espectroscopia no Infravermelho Médio com a Transformada de Fourier (FT-MIR) foi desenvolvida com a finalidade de superar algumas limitações encontradas nos instrumentos dispersivos como, por exemplo, o longo processo de varredura. O diferencial é a formação da onda evanescente no seu interior, gerada a parte da reflexão da luz incidida e absorção por parte da amostra na superfície, que em seguida sofre uma atenuação do feixe, passando pelo detector assim gerando o espectro (MÜLLER *et al.*, 2011).

O seleneto de zinco (ZnSe) é o material de guia de onda do ATR mais utilizado, e pode analisar amostras em diversos estados físicos, tais como líquidos, pastas, polímeros, filmes, pós e sólidos, além de ser insolúvel em água e não absorver na região do MIR. No entanto, este cristal é atacado por ácidos e bases fortes, o que limita o pH das amostras a serem analisadas. A figura 3 destaca a irradiação sendo absorvida pelo contato com a amostra, assim como exhibe o acessório de ATR da Perkin Elmer.

Figura 3. (a) Reflectância de radiação no interior do cristal de ZnSe; (b) Acessório de ATR acoplado ao suporte FT-MIR SpectrumTwo (Perkin Elmer)



Fonte: PerkinElmer

Para que se atenda a demanda de quantificação do teor de biodiesel nas misturas com diesel, é necessário correlacionar as pequenas variações espectrais com as concentrações das amostras analisadas, por mais que as diferenças sejam muito sutis entre os espectros, a técnica de Espectrometria no Infravermelho Médio pode ser aplicada com ferramentas quimiométricas. Isso então permite uma eficiente quantificação, de maneira rápida, não destrutiva, e sem necessidade de grandes quantidades de amostras (ABNT, 2008).

Diferentemente do que a norma da NBR 15568 de março de 2008 propõe, este trabalho apresenta a construção de apenas uma única curva de faixa 0,75 – 30,00% de concentração, sem o uso de solventes, ou seja, uma análise direta e utiliza-se da seleção de variáveis, a quantidade de informação é reduzida, o que diminui os custos, uma vez que não seria mais necessário a utilização do infravermelho médio com alta varredura e com uma aparelhagem ótica muito sofisticada. Assim, o método aplicado apresenta como vantagem a portabilidade de um equipamento que ao emitir radiação na faixa selecionada do infravermelho, já correlacionará com a devida concentração do biocombustível através de ferramentas quimiométricas.

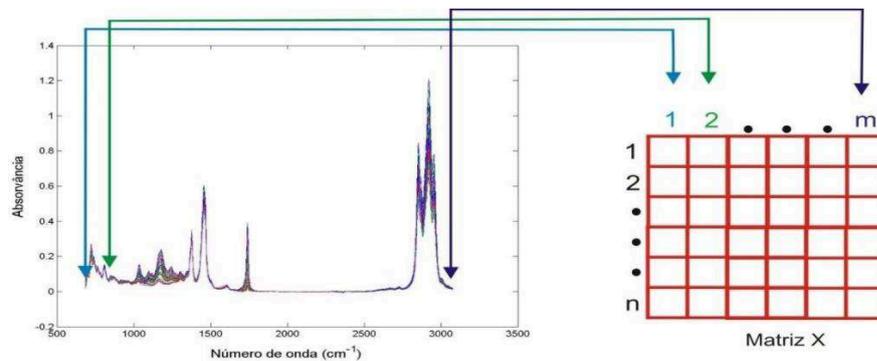
1.3. Quimiometria

A Quimiometria é definida como o uso de métodos matemáticos, estatísticos e de lógica formal para o tratamento de dados químicos (ALVES, 2010). Dentre todas as áreas da química, a área da química analítica tem sido a mais beneficiada em utilizar métodos quimiométricos, isso porque constantemente necessita de tratamento de dados químicos. Fato que se torna mais visível quando a *Chemometrics Society*, organização responsável pelo desenvolvimento de métodos quimiométricos, é composta quase totalmente por químicos analíticos.

A resposta instrumental, ou seja, os dados químicos processados foram os espectros MIR, sendo representada na forma de matriz \mathbf{X} para a construção dos modelos, enquanto que

a propriedade de interesse determinada pela metodologia padrão é representada por um vetor y , neste caso sendo a concentração BX (FERREIRA, 1999). A Figura 4 exhibe a construção da matriz X , de dimensão $n \times m$, onde n é a quantidade de objetos (espectros), e m a quantidade de variáveis (absorvâncias em determinado número de onda).

Figura 4. Representação da construção da matriz X para modelagem multivariada.



Fonte: Autor, 2018.

1.3.1. Quadrados Mínimos Parciais (PLS)

O Método dos Quadrados Mínimos Parciais (PLS) é um método supervisionado de regressão utilizado para a construção de modelos de calibração multivariada. Por ser multivariado pode se fazer a previsão de uma amostra contendo interferentes, uma vez que os mesmos também estiverem presentes no conjunto de calibração. No modelo PLS, os espectros, um para cada amostra, são organizados em uma matriz X ($n \times m$) de variáveis independentes, onde cada linha representa uma amostra e contém as respostas medidas para a mesma. O outro conjunto de dados é constituído das variáveis dependentes e organizado no vetor y que correspondem as concentrações das amostras (MACHADO, 2018).

O PLS também verifica se há amostras anômalas pelo cálculo residual de Q , sinalizando que a amostra está fora do limite de confiança de 95%, e pelo valor limite de Leverage, sendo três vezes o número de variáveis latentes dividido pelo número de amostras.

Para avaliar se o modelo de regressão está satisfatório, é feito o cálculo do erro médio quadrático (RMSE – do inglês, root mean square error). Esses cálculos avaliam a diferença entre o valor real de concentração e o valor previsto pelo modelo. O cálculo do erro quadrático médio pode ser feito com as amostras do conjunto de calibração, obtendo-se o valor de RMSEC (root mean square error calibration). Por outro lado, se o cálculo for feito com as amostras do conjunto de previsão ou validação externa, terá o valor do RMSEP (root mean square error prediction) para avaliar a qualidade do modelo (SOUZA, 2018).

$$RMSE = \sqrt{\frac{\sum (Y_{pred} - Y_{ref})^2}{N}} \quad (1)$$

1.3.2. Quadrados Mínimos Parciais por Intervalos (iPLS)

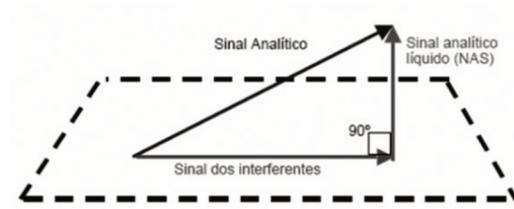
Ao se utilizar a Espectrometria no Infravermelho Médio alinhada à Quimiometria, depara-se com um grande número de variáveis, incluindo aquelas de potencial interferente e correlacionadas a não linearidades, que implicam desde problemas no desempenho computacional até nos modelos gerados (BALABIN, 2007). Para resolver este impasse são utilizados métodos de seleção de variáveis. Estes métodos são baseados no princípio da escolha de um pequeno número de variáveis, segregadas a partir do espectro original, onde a remoção de variáveis não informativas produzirá modelos mais robustos. A complexidade de um modelo pode ser reduzida através de uma pré-seleção criteriosa de números de onda. O simples fato de se dividir o espectro em regiões proporcionais em tamanho e de se analisar cada um separadamente, já caracteriza um método de seleção de variáveis.

Na prática, o método está baseado na identificação de um subconjunto dos dados inteiros que produzirão erros de modelagem mais baixos. Assim, em espectros vibracionais, por exemplo, os comprimentos de onda que apenas induzem a ruídos, informações irrelevantes ou não linearidades podem ser eliminados (GONTIJO, 2016).

1.3.3. Sinal Analítico Líquido

Em calibrações multivariadas, é de suma importância o conceito de sinal analítico líquido (NAS – *Net Analyte Signal*), pois tem a função de determinar algumas figuras de mérito, que por sua vez tem a finalidade de validar o método. A maioria dos artigos publicados com calibração multivariada atribui o método do cálculo do NAS para modelos multivariados ao trabalho de Lorber, em 1986. O método proposto por Lorber foi corrigido por Ferré, Brown e Rius para possibilitar o cálculo exato do NAS para modelos de calibração construídos a partir dos métodos de regressão por PLS. Tais métodos trazem em comum o cálculo de uma matriz de projeção ortogonal e a propriedade de ortogonalidade do NAS pode ser observada pela representação geométrica da Figura 5 (Ferré e Faber, 2003).

Figura 5. Representação Geométrica do NAS



Fonte: VALDERRAMA, 2009

A proposta do cálculo parte do princípio de trocar a matriz de dados e o vetor de interesse por uma combinação linear de variáveis latentes, cálculos mostrados pelas equações 2 e 3. Posteriormente, é construído a matriz que contém todas as informações exceto a espécie de interesse, conforme exibido pela equação 4.

$$X_a = T_A P_A^T + E \quad (2)$$

$$y_a = U_A Q_A^T + f \quad (3)$$

$$X_{a,-k} = (I - Y_{A,k} Y_{A,k}^+) X_a \quad (4)$$

Após estas etapas, como a matriz criada não possui nenhuma contribuição da espécie de interesse e o vetor do sinal analítico é uma combinação linear da matriz original, com uma nova projeção ortogonal é possível esboçar o NAS, ou seja, o sinal analítico que dita a correlação direta com a propriedade de interesse (VALDERRAMA *et al.*, 2009).

1.3.4. Validação do Método

A validação analítica avalia o desempenho do método com o objetivo de adequar às condições nas quais poderá ser aplicado. Ao se desenvolver um procedimento analítico, o processo de validação deve ser aplicado a fim de avaliar a performance e limitações do método proposto. Um dos parâmetros da validação analítica são as figuras de mérito que foram atribuídas a este trabalho e especificadas na norma ASTM E1655-05, além de parâmetros importantes no desenvolvimento de uma metodologia analítica. A norma dita o número mínimo de amostras pela quantidade de variáveis latentes pelo cálculo $6(k+1)$ para calibração e $6k$ para validação, onde k é o número de variáveis latentes. As figuras de mérito utilizadas foram:

linearidade, precisão, exatidão, seletividade, sensibilidade, limite de detecção, limite de quantificação e teste para erro sistemático (ASTM, 2012).

1.3.4.1. Linearidade

É representada pelo coeficiente de correlação R, que expressa a razão da covariância pelos desvios. Na prática, é o responsável por verificar a capacidade do método em fornecer resultados diretamente proporcionais à concentração da substância de interesse, dentro da faixa expressada. A linearidade pode ser inferida pela observação de parâmetros como o coeficiente de regressão “b”, o intercepto “a”. O R^2 expressa a variação explicada pela variação total, sendo de suma importância para entendimento de quantos por cento o modelo consegue captar da variância máxima (VALDERRAMA *et al.*, 2009).

1.3.4.2. Precisão

Indica o nível de conformidade dos resultados após uma certa quantidade de medidas para uma mesma amostra homogênea em condições determinadas. A repetição do número de amostras para uma mesma concentração, garantem uma precisão ao método, sendo aplicado no trabalho amostras em quintuplicatas (VALDERRAMA *et al.*, 2009).

1.3.4.3. Exatidão

Indica o nível de conformidade entre o valor estimado ou medido e o valor tido como verdadeiro ou de referência. Os processos mais utilizados para avaliar a exatidão de um método são: materiais de referência, comparação de métodos, ensaios de recuperação, adição de padrão. Para métodos multivariados, o valor de RMSE alinhado à Elipse de Confiança são procedimentos confiáveis (VALDERRAMA *et al.*, 2009).

1.3.4.4. Seletividade

É a predisposição da determinação da espécie de interesse frente a componentes que possam interferir com a sua determinação em uma amostra. Essa estimativa é realizada com base na razão entre a sensibilidade da propriedade de interesse e a sensibilidade dos interferentes, vista na equação 5 (VALDERRAMA *et al.*, 2009).

$$SEL_{k,j} = \frac{s_{k,j}^{as}}{\|s_{k,j}\|} \quad (5)$$

1.3.4.5. Sensibilidade

Indica a porção de sinal que é ampliada quando a concentração da espécie de interesse tem seu valor elevado a uma unidade. O cálculo multivariado é expresso pela equação 6 (VALDERRAMA *et al.*, 2009).

$$SEN = \|s_k^{as}\| \quad (6)$$

1.3.4.6. Limite de Detecção

Equivale à menor concentração da substância de interesse que pode ser detectada, mas não necessariamente quantificada. O cálculo multivariado é expresso pela equação 7 (VALDERRAMA *et al.*, 2009).

$$LD = 3,3\sigma_x \frac{1}{SEN} \quad (7)$$

1.3.4.7. Limite de Quantificação

Representa a menor concentração da substância de interesse que pode ser medida com uma incerteza máxima de 10%. O cálculo multivariado é expresso pela equação 8 (VALDERRAMA *et al.*, 2009).

$$LD = 10\sigma_x \frac{1}{SEN} \quad (8)$$

1.3.4.8. Teste para Erro Sistemático

O termo *bias* é atribuído a erros sistemáticos que são calculados pela diferença entre a média da população e o valor verdadeiro e são todas as componentes de erro que não são aleatórias. A norma E1655-00 da ASTM aborda a investigação desse parâmetro através de um teste-t para as amostras de validação no nível de 95% de confiança. A equação 9 e 10 exhibe seus respectivos cálculos (VALDERRAMA *et al.*, 2009).

$$bias = \frac{\sum_{i=1}^{n_v} (y_i - y_D)}{n_v} \quad (9)$$

$$t_{bias} = \frac{|bias| \sqrt{n_v}}{SDV} \quad (10)$$

2. OBJETIVOS

2.1. Geral

O projeto tem como objetivo geral, desenvolver métodos analíticos capazes de determinar o teor de biodiesel metílico de Pinhão Manso em mistura com diesel através da Espectrometria no Infravermelho Médio aliada a Calibração Multivariada e método de Seleção de Variáveis por Intervalos, que possam ser implementados em análises de rotina e controle de qualidade desse combustível.

2.2. Específicos

Obter espectros MIR dos biodieseis BX com adição na faixa de concentração de 0,75-30,00% (v/v);

Construir um modelo PLS para determinar o teor de biodiesel metílico de Pinhão Manso nas misturas com diesel;

Avaliar a utilização do método de Seleção de Variáveis por Intervalos (iPLS) para a quantificação do biodiesel metílico de Pinhão Manso em misturas com diesel, comparando com base nos valores de RMSEP e número de variáveis, sua eficiência com o respectivo modelo PLS global;

Validar o modelo obtido.

3. METODOLOGIA

3.1. Preparo do B100

As amostras de biodiesel foram produzidas no Laboratório do Instituto de Química da Universidade Federal de Uberlândia, e o diesel foi fornecido pela TRANSPETRO S/A. Os álcoois e o óleo de Pinhão Manso utilizados na produção de biodiesel encontram-se disponíveis no laboratório.

A primeira etapa da produção dos B100 foi à catálise ácida (esterificação), com a finalidade de formar os ácidos graxos livres integrantes nos óleos brutos e parte dos triglicerídeos em ésteres metílicos. Essa etapa consistiu no uso da proporção de 1:1 (m/m) do óleo em relação ao álcool e 3% (m/m) do ácido sulfúrico concentrado (usado como catalisador) em relação à massa do óleo. Após a reação de esterificação, a mistura heterogênea resultante foi transferida para um funil de decantação e permaneceu em repouso durante 1 hora para posterior separação das fases: inferior (mistura de ésteres) e superior (álcool). A fase inferior teve que ser submetida à catálise básica (transesterificação), com 100,0 g do óleo e 30,0 g de álcool na presença de 1,0 g de KOH. A solução resultante foi mantida em repouso durante 24 horas para separação de fases biodiesel (superior) e glicerol (inferior). O biodiesel obtido foi purificado por lavagens sucessivas com água destilada quente de aproximadamente 85°C para remover os resíduos, subprodutos e excedentes do álcool ou catalisador. O biodiesel puro obtido foi transferido para um frasco de vidro envolvido em papel de alumínio e mantido ao abrigo da luz à temperatura ambiente.

3.2. Preparo das Misturas biodiesel/diesel

As amostras foram preparadas à temperatura ambiente (25 ± 2 °C) utilizando-se uma balança analítica (Sartorius, modelo BP211D). Para esse procedimento, foi adicionada uma alíquota de biodiesel de massa conhecida em um frasco âmbar de 25,00 mL, seguida da adição do diesel S-10 e posterior homogeneização em Vórtex (Phoenix, modelo AP56). As amostras foram preparadas na faixa de concentração de 0,75% a 30,00% (v/v). Foram preparadas 69 amostras de misturas biodiesel/diesel, sendo 44 amostras para o conjunto de calibração e 25 amostras para o conjunto de previsão. Ambos os conjuntos de amostras foram preparados de forma bem distribuída ao longo da faixa de concentração avaliada. O registro dos espectros MIR das amostras foi realizado após 48 horas de repouso.

3.3. Obtenção dos espectros MIR

Os espectros MIR foram obtidos utilizando um Espectrômetro de Infravermelho da marca Perkin Elmer, modelo SpectrumTwo, com Transformada de Fourier, resolução de 4 cm^{-1} , 16 varreduras, na faixa de número de onda de 4000 a 600 cm^{-1} e dispositivo amostrador Horizontal por Reflectância Total Atenuada (H-ATR) com cristal de ZnSe. Foram obtidos cinco espectros para cada amostra analisada. Os espectros foram corrigidos pelo método de Correção de Linha de Base (Baseline) e feito um corte de informações irrelevantes, a sobrar a faixa de número de onda de 3099 a 679 cm^{-1} utilizando o programa Matlab.

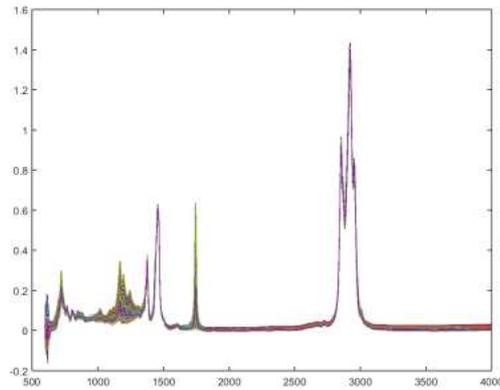
3.4. Modelos Quimiométricos

O programa utilizado para a construção dos modelos foi o Matlab versão R2015a juntamente com PLS_Toolbox 8.61 e iToolbox 1.0. Para construção do modelo de seleção de variáveis por intervalos foi utilizado o algoritmo iPLS empregando rotinas próprias modificadas a partir das demonstrações contidas no iToolbox 1.0. Neste algoritmo foi avaliada a divisão da faixa espectral completa em 5, 10, 15, 20, 25, 30, 35, 40, 45 e 50 intervalos. Para os resultados de saída do algoritmo iPLS foi selecionado e construído um modelo PLS, dentro de cada divisão, com o intervalo que apresentasse menor valor de RMSECV.

4. RESULTADOS E DISCUSSÃO

Ao observar a Figura 6, é notório que não é possível distinguir os espectros sobrepostos com relação às suas respectivas concentrações. Deste modo, é necessário aplicar métodos quimiométricos, uma vez que o espectro apresenta grande semelhança. O algoritmo que expressa uma boa correlação linear da matriz de dados (espectro) com a matriz de interesse (concentração) é a Regressão por Quadrados Mínimos Parciais (PLS).

Figura 6. Espectros sobrepostos da mistura do diesel com BMPM com variação de teor do biodiesel em diesel de 0,75% à 30%.



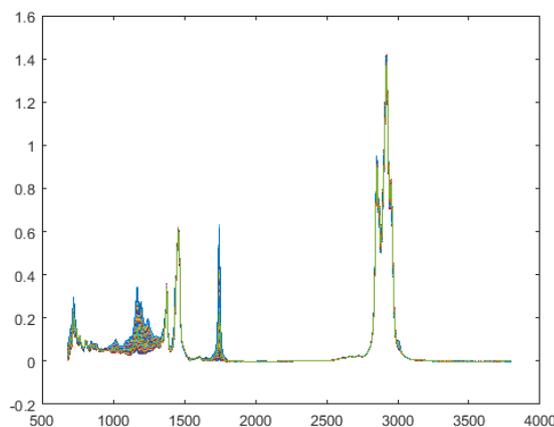
Fonte: Autor, 2018

4.1. PLS

Percebe-se que a região inicial dos espectros (menor número de onda) possui grande quantidade de ruídos que são inerentes ao equipamento. Para que esses ruídos não influenciam na construção do modelo é necessário que se faça um corte nos espectros nas faixas iniciais a fim de eliminar essas informações interferentes. Da mesma forma, na região final dos espectros (maiores números de onda) não há informações relevantes sobre a absorvância relativa aos números de onda. Sendo assim, essa região também é cortada minimizando assim possíveis erros na construção do modelo.

A região central dos espectros (números de onda intermediários) é uma região com absorvância aproximadamente zero. A fim de garantir essa aproximação, faz-se uma correção do espectro através do método de *baseline*. A Figura 7 mostra o resultado dos espectros gerados após os cortes e a correção por *baseline*.

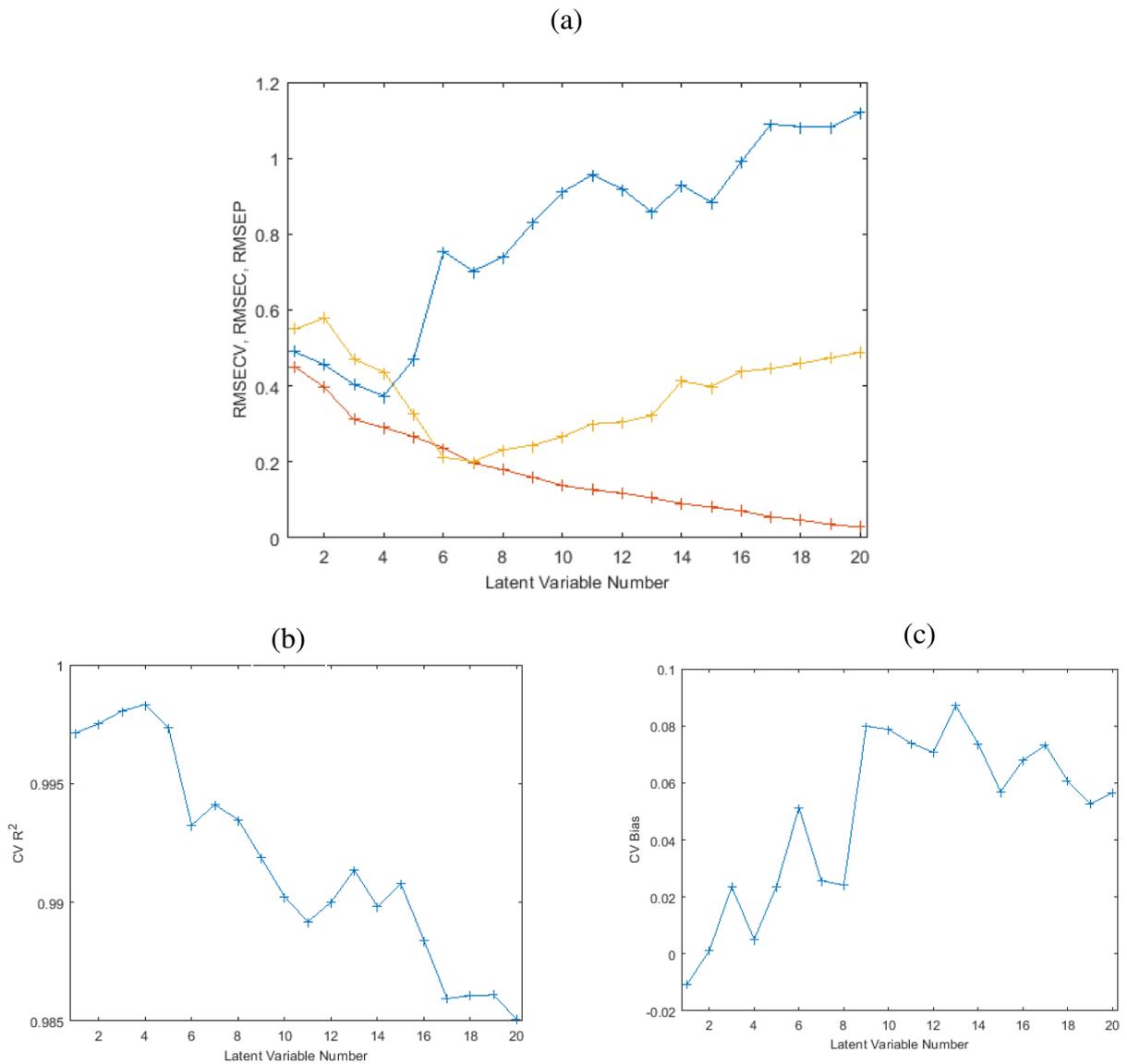
Figura 7. Espectros finais sobrepostos da mistura do diesel com BMPM.



Fonte: Autor, 2018

Para construção do modelo matemático, foi necessário avaliar o número de variáveis latentes, a escolha deve ser baseada de forma a minimizar os erros. A Figura 8 exibe três gráficos que fazem a avaliação do número a ser escolhido.

Figura 8. (a) Relação do número de variáveis latentes pelos erros RMSEC, RMSECV e RMSEP.(b) Relação do número de variáveis latentes pelo R^2 da validação cruzada. (c) Relação do número de variáveis latentes pelo erro sistemático de validação cruzada.

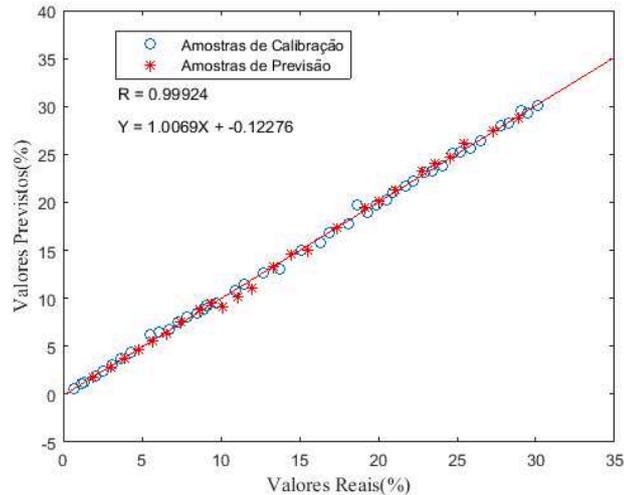


Fonte: Autor, 2018

Com base nestes critérios, a escolha foi de 4 variáveis latentes, uma vez que apresentou baixo valor de RMSECV no (a), teve o melhor R^2 de validação cruzada (b) e apresentou um dos menores erros sistemáticos de validação cruzada (c).

O modelo matemático é apresentado pela Figura 9 onde se compara os valores reais de concentração pelos valores previstos do modelo. Pode se constatar que o gráfico apresenta uma boa linearidade de modo que é possível quantificar os espectros em toda a faixa de concentração analisada.

Figura 9. Regressão dos valores reais pelos valores previstos pelo modelo.



Fonte: Autor, 2018

4.2. iPLS

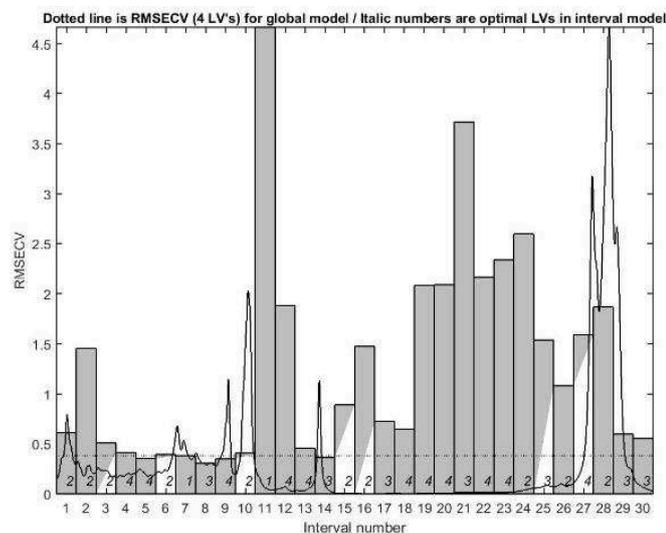
O modelo PLS global construído apresenta em sua estrutura regiões não informativas de linha de base e regiões que pode conter interferentes que prejudicam a robustez do método. Assim, se faz necessário a utilização de uma seleção destas variáveis de modo que diminua o erro de previsão, por isso a utilização o Regressão por Quadrados Mínimos Parciais por Intervalos (iPLS).

O espectro foi dividido em 5, 10, 15, 20, 25, 30, 35, 40, 45 e 50 faixas e selecionado em cada divisão a melhor região para construção do modelo. De acordo com a Tabela 1 a seleção que se mostrou mais eficiente foi a de 30 faixas.

Tabela 1. Escolha da faixa a ser usada para seleção de variáveis através do iPLS.

Nº de Faixas	V.L	Faixa	Variáveis Selecionadas (cm ⁻¹)	RMSECV	R ² CV	CV Bias
PLS Global	4	-	3099 - 679	0,3731	0,9980	0,0059
5	4	3	2132 - 1649	0,3578	0,9984	-0,0173
10	4	5	1890 - 1649	0,3489	0,9986	-0,0155
15	4	7	1812 - 1652	0,3440	0,9986	-0,0130
20	4	1	801 - 679	0,3071	0,9988	0,0030
25	4	1	776 - 679	0,2749	0,9992	-0,0129
30	3	8	1327 - 1247	0,3075	0,9988	0,0022
35	4	16	1789 - 1721	0,3498	0,9986	-0,0146
40	4	1	740 - 679	0,2839	0,9990	-0,0051
45	4	21	1814 - 1760	0,3018	0,9990	0,0051
50	3	12	1267 - 1219	0,3078	0,9986	-0,0012

O critério para a escolha do melhor modelo foi além do menor RMSECV que é amplamente descrito na literatura, foi levado em consideração a faixa, visto que alguns modelos selecionavam faixas de linhas de base ou baixa razão sinal/ruído, o erro sistemático da validação cruzada, o R² de validação cruzada, a existência de amostras anômalas e a diminuição do espaço N-Dimensional. A Figura 10 exibe a faixa selecionada para realização do iPLS.

Figura 10. Região escolhida pelo algoritmo para a construção do novo PLS.

Fonte: Autor, 2018

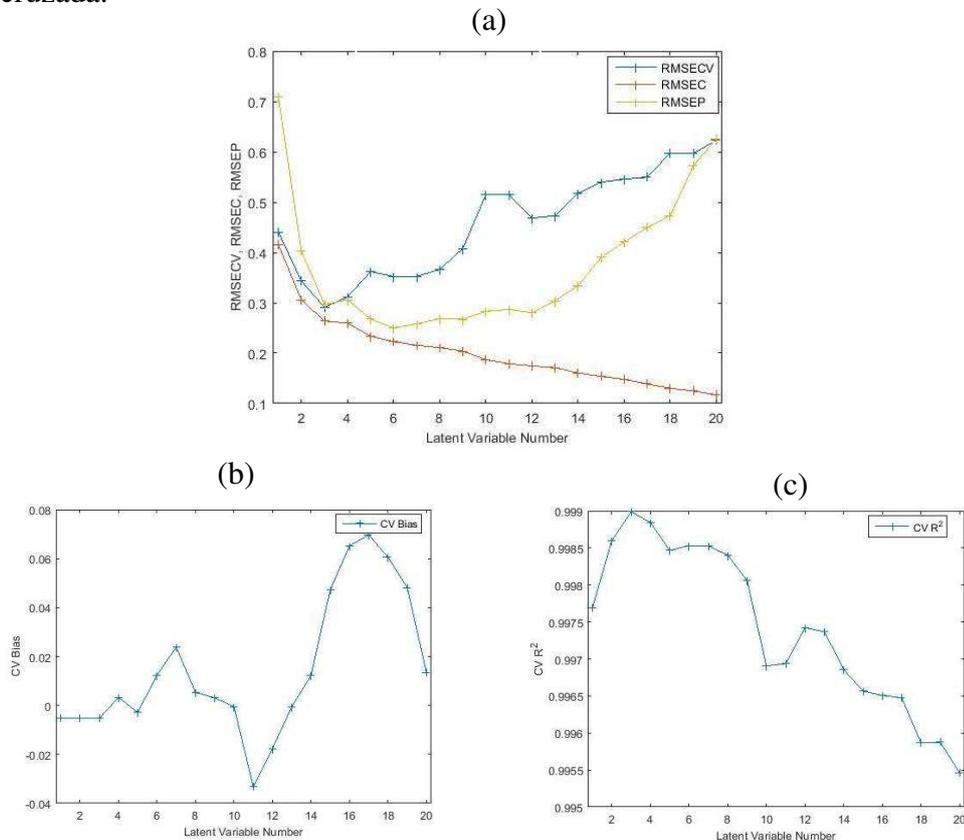
Sendo assim, a tabela 3 mostra a comparação mais aprofundada do modelo PLS Global frente ao modelo iPLS de 30 faixas.

Tabela 2. Comparação dos modelos pela estatística F.

Nº de Faixas	V.L	Faixa	Variáveis Seleccionadas (cm ⁻¹)	RMSEP	F _{tab} (0,05;24;24)	F _{cal}
PLS Global	4	-	3099 - 679	0,4339	1,98	2,14
iPLS_30	3	8	1327 - 1247	0,2969		

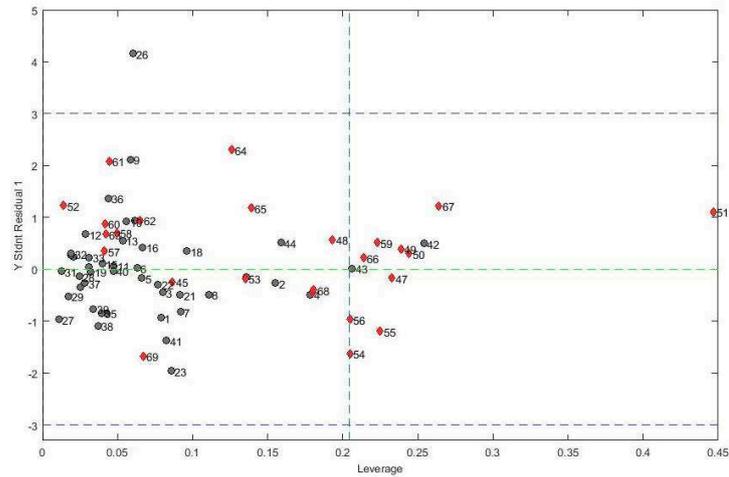
A escolha do número de variáveis latentes do novo modelo seguiu os mesmos critérios adotados pelo modelo global, sendo 3 o número de variáveis latentes a ser escolhido como mostra a Figura 11.

Figura 11. (a) Escolha do número de variáveis latentes por comparação de erros RMSEC, RMSECV e RMSEP. (b) Escolha do número de variáveis latentes por avaliação do R² da validação cruzada. (c) Escolha do número de variáveis latentes devido ao erro sistemático de validação cruzada.



A Figura 12 apresenta a relação das amostras anômalas do novo modelo, sendo que é possível verificar que não há nenhuma amostra sendo *outlier*.

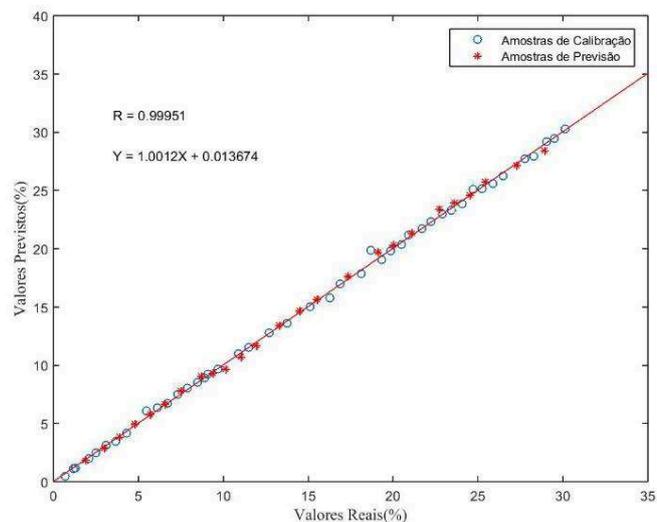
Figura 12. Mensuração de amostras anômalas pelo *Leverage* e Q_{residual} com uma confiabilidade de 95%.



Fonte: Autor, 2018.

A Figura 13 compara os valores reais de concentração pelos valores previstos do modelo por intervalos. O gráfico possui uma ótima linearidade de maneira que é quantificado corretamente os valores de concentração.

Figura 13. Linearidade dos valores reais frente aos valores previstos pelo modelo.



Fonte: Autor, 2018

Os modelos obtidos para quantificação do teor do biodiesel metílico de Pinhão Manso apresentaram excelente linearidade, com valores de R^2 igual a 0,999. Esses valores estão próximos de 1, o que indica que estes valores previstos pelos modelos estiveram muito próximos dos valores reais.

A Tabela 3 exibe os valores dos erros RMSEC, RMSECV e RMSEP em comparação com os valores do modelo global, o que demonstra significativa diminuição dos erros.

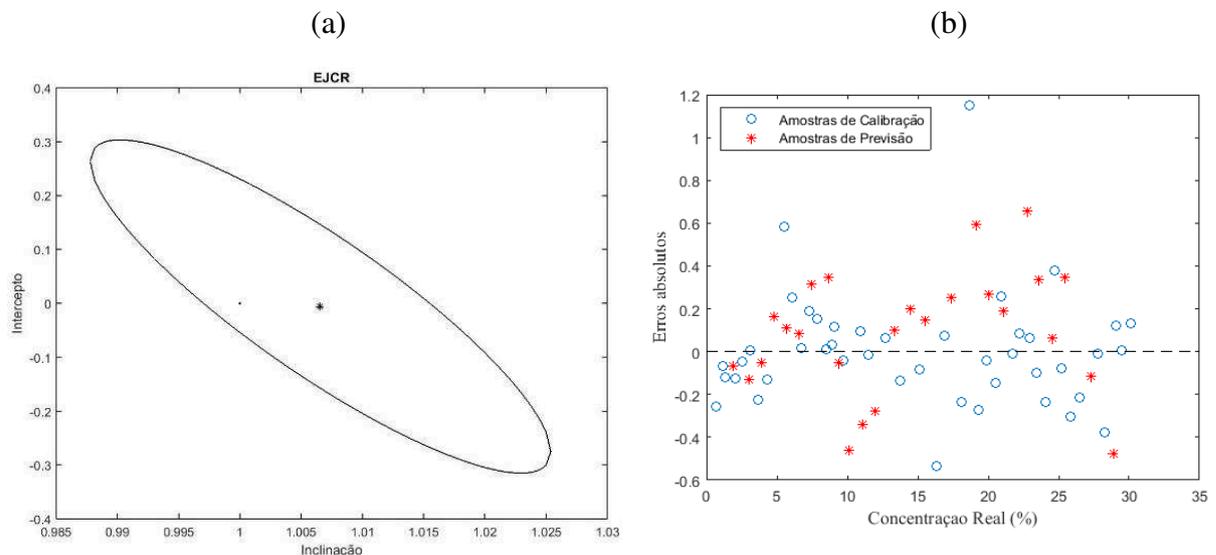
Tabela 3. Comparação dos erros do modelo global pelo modelo de seleção de variáveis.

Modelo	RMSEC	Variação (%)	RMSECV	Variação (%)	RMSEP	Variação (%)
PLS Global	0,2906	9,22	0,3731	28,18	0,4369	48,18
iPLS_30	0,2638		0,2912		0,2969	

Os valores de RMSEP e RMSEC obtidos apresentaram valores pequenos de forma que fornece maior credibilidade aos modelos. O valor de RMSEP foi menor que o de RMSEC, transmitindo uma confiabilidade na calibração, visto que os valores de uma calibração devem ser sempre menores do que a previsão, caso contrário, os modelos seriam tendenciosos.

Para validação do modelo, foram usadas as figuras de mérito assim como parâmetros da norma da ASTM. A Figura 14 se baseia na Elipse de Confiança por indicar que o modelo criado está dentro do limite de 95% de confiança e nos cálculos de erros absolutos que mostra uma aleatoriedade dos erros com variação de no máximo 1% se mostrando dentro de limites aceitáveis.

Figura 14. (a) Elipse de Confiança do modelo (b) Avaliação dos erros absolutos.



Fonte: Autor, 2018

A tabela 4 mostra os valores das figuras de mérito, onde é possível perceber que a técnica tem uma ótima sensibilidade, bons limites de detecção e quantificação, pois ambos os valores se apresentaram muito baixos.

Tabela 4. Figuras de mérito para o modelo de seleção de variáveis.

Seletividade	Sensibilidade (% v/v)	Limite de Detecção (% v/v)	Limite de Quantificação (% v/v)	Teste T_{bias} tabelado	Teste T_{bias} calculado
0,1021	0,0040	0,1738	0,5245	2,0639	1,2546

O valor de seletividade é baixo, porém o que caracteriza esta medida é a natureza dos dados e não o modelo em si, visto que a origem do FT-MIR era de finalidade ótica, sendo a ferramenta quimiométrica a grande responsável por propiciar outras práticas. Como o T_{bias} calculado é menor que o T_{bias} tabelado, o modelo não apresentou erros sistemáticos.

O modelo matemático com seleção de variáveis se mostrou melhor que o modelo PLS Global uma vez que após a estatística F apresentou uma diferença na variância do modelo através dos RMSEP's e indicou mais vantagens na validação por figuras de mérito.

5. CONCLUSÃO

Foi possível desenvolver métodos analíticos capazes de determinar o teor de biodiesel metílico de Pinhão Manso em mistura com diesel através da Espectrometria no Infravermelho Médio aliada a Calibração Multivariada e método de Seleção de Variáveis por Intervalos, de modo que viabiliza a implementação do modelo nas análises de rotina e controle de qualidade deste combustível, além de abrir perspectivas para realização de outras matrizes.

Foi obtido espectros MIR dos biodieseis BX com adição na faixa de concentração de 1,00 - 30,00% (v/v) não havendo problemas de linearidade dentro do limite estabelecido. Embora o modelo PLS foi satisfatório para a predição dos teores do biocombustível, a utilização do método de Seleção de Variáveis por Intervalos (iPLS) para a quantificação do biodiesel metílico de Pinhão Manso em misturas com diesel, obteve melhor valor de RMSEP e número de variáveis, o que comprova sua eficiência.

O modelo iPLS foi validado de acordo com a norma da ASTM E1655-05 e das figuras de mérito de maior notoriedade evidenciando-se que é reproduzível, com baixos erros, sem apresentar sistematicidade e com alta sensibilidade.

A formação como quimiometrista é um grande diferencial profissional, pois a implementação do modelo matemático, embora seja um modelo local, pode ser realizado independente da matriz e do equipamento analítico, a base teórica que é a fundamentação para o profissional da área estar preparado para realizar qualquer medição e extrair informações relevantes e de interesse.

Portanto, a análise de amostras de biodiesel metílico de pinhão manso a partir da técnica do infravermelho médio e da calibração multivariada por meio de métodos quimiométricos se mostrou muito eficiente na quantificação dos teores.

6. REFERÊNCIAS

ALVES, J. O. Espectrometria de Massas com Ionização Electrospray (ESI-MS) e Métodos Quimiométricos: Caracterização de Azeites de Oliva (Extra Virgem e Puro) e Outros Óleos Vegetais e Quantificação de Óleos Adulterantes em Azeite de Oliva Extra Virgem. 2010. 80 f. Dissertação (Mestrado) – **Instituto de Química**, Universidade Federal de Minas Gerais, Belo Horizonte, MG, 2010.

AMERICAN SOCIETY FOR TESTING AND MATERIALS – ASTM. E1655-05: Standard practices for infrared multivariate quantitative analysis. **West Conshohocken**, 2012.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS - ABNT. **NBR 15568**: Biodiesel = Determinação do teor de biodiesel em óleo diesel por espectroscopia na região do infravermelho médio. Rio de Janeiro, 2008.

BALABIN, R. M.; SYUNYAEV, R. Z.; KARPOV, S. A. Quantitative measurement of ethanol distribution over fractions of ethanol–gasoline fuel. **Energy & Fuels**, v. 21, p. 2460–2465, 2007.

BRASIL. Lei nº 13.263, de 23 de março de 2016. Altera a Lei nº 13.033, de 24 de setembro de 2014, para dispor sobre os percentuais de adição de biodiesel ao óleo diesel comercializado no território nacional. **Diário Oficial da União**, Brasília, DF, 24 mar. 2016

FERRÉ, J.; FABER, N. M. Net analyte signal calculation for multivariate calibration. **Chemometrics and Intelligent Laboratory Systems**, v. 69, n. 1-2, p. 123-136, 2003. ISSN 01697439.

FERREIRA, M. M. C.; ANTUNES, A. M.; MELGO, M. S.; VOLPE, P. L. O. Quimiometria I: calibração multivariada, um tutorial. **Química Nova**, São Paulo, v. 22, n. 5, p. 724-731, 1999.

GONTIJO L. C. Uso de espectrometria no infravermelho médio, calibração multivariada e seleção de variáveis por intervalos na quantificação de biodieseis em misturas com diesel. Tese (Doutorado) – **Instituto de Química**, Universidade Federal de Uberlândia, Uberlândia, MG, 2016.

LÔBO, I. P.; FERREIRA, S. L. C.; CRUZ, R. S. D. Biodiesel: parâmetros de qualidade e métodos analíticos. **Química Nova**, São Paulo, v. 32, n. 6, p. 1596-1608, 2009.

MACHADO C. C. S. Metodologia de detecção de adulterações em óleo de macadâmia usando espectroscopia no infravermelho médio e ferramentas quimiométricas. Tese (TCC) – **Instituto de Química**, Universidade Federal de Uberlândia, Uberlândia, MG, 2018.

MÜLLER, A. L. H. Attenuated total reflectance with Fourier transform infrared spectroscopy (ATR/FTIR) and different PLS Algorithms for simultaneous determination of clavulanic acid and amoxicillin in powder pharmaceutical formulation. **Journal of the Brazilian Chemical Society**, v. 22, n. 10, p. 1903-1912, 2011. ISSN 0103-5053.

PADULA, A. D.; SANTOS, M. S.; FERREIRA, L.; BORENSTEIN, D. The emergence of the biodiesel industry in Brazil: Current figures and future prospects. **Energy Policy**, Surrey, v.44, p. 395-405, 2012.

PETERSON, C. L.; HUSTRULID, T. Carbon Cycle for Rapeseed Oil Biodiesel Fuels. **Biomass and Bioenergy**, Oxford, v. 14, n. 2, p. 91-101, 1998.

SARIN, A.; ARORA, R.; SINGH, N. P.; SARIN, R.; MALHOTRA, R. K. Blends of biodiesels synthesized from non-edible and edible oils: Influence on the OS (oxidation stability). **Energy**, Oxford, v. 35, n. 8, p. 3449-3453, 2010

SOUZA, L. M. Controle de qualidade de óleos de linhaça e ginkgobiloba usando espectrometria no infravermelho médio e ferramentas quimiométricas. 2018. Tese (Doutorado) – **Instituto de Química**, Universidade Federal de Uberlândia, Uberlândia, MG, 2018.

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