

REGRESSÃO MULTIVARIADA POR OPLS E PLS DOS ESPECTROS DE RMN DE ¹H DE MISTURAS DIESEL/BIODIESEL DE MAFURRA PARA ESTIMATIVA DO TEOR DE BIODIESELAdemar D. V. Máquina^{a,b,*}, Baltazar V. Siteo^{a,b}, Maria T. C. Ferreira^a, Douglas Q. Santos^c e Waldomiro Borges Neto^a^aInstituto de Química, Universidade Federal de Uberlândia, 38408-100 Uberlândia-MG, Brasil^bFaculdade de Ciências Naturais e Matemática, Universidade Púnguè, 323 Chimoio – Manica, Moçambique^cEscola Técnica de Saúde, Universidade Federal de Uberlândia, 38400-902 Uberlândia – MG, Brasil

Recebido em 23/01/2020; aceito em 01/04/2020; publicado na web em 29/05/2020

OPLS AND PLS MULTIVARIATE REGRESSION OF ¹H NMR SPECTRA OF DIESEL/MAFURRA BIODIESEL BLENDS FOR ESTIMATION OF BIODIESEL CONTENT. Two methodologies were developed to monitor the biodiesel content of mafurra in mixtures with diesel using hydrogen nuclear magnetic resonance (¹H NMR) Spectroscopy combined with the multivariate regression by orthogonal projections to latent structure (OPLS) and partial least squares (PLS). The efficiency of these methodologies was analyzed based on the figures of merit and the fit of the models through the correlation of the measured and predicted values of the calibration and prediction sets. The results of the figures of merit in the OPLS model were better than in the PLS model. A high correlation between the measured and predicted values was evident in the OPLS model, with a correlation coefficient (R^2) greater than 0.99, demonstrating a better fit of the OPLS model in relation to the PLS model which presented a correlation coefficient (R^2) less than 0.98. The OPLS model is more robust and has good predictive capacity than the PLS model because it obtained a higher Q^2 value. The excellent results of the application of ¹H NMR spectroscopy combined with multivariate regression by OPLS suggest that this analytical methodology is ideal, feasible, efficient and suitable for use by inspection agencies to control the quality of this fuel.

Keywords: mafurra Biodiesel; ¹H NMR spectrometry; Monitoring; OPLS; PLS.

INTRODUÇÃO

O biodiesel é um combustível derivado de fontes renovável, comumente comercializado em mistura com diesel porque apresentam propriedades similares.^{1,2} Esta mistura pode ser feita em diferentes percentagens, de acordo com a legislação de cada país, mas no geral ela começou a ser feita em pequenas percentagens (2%) e devido ao seu sucesso evoluiu para médias percentagens (30%) em alguns países. Por exemplo, na França a mistura evoluiu de B5 (5% de biodiesel e 95% de diesel) para B30 (30% de biodiesel e 70% de diesel); nos EUA, de B2 para B20; em Moçambique, de B3 para B7,5; no Brasil, de B2 para B11, com perspectiva de evoluir para B15 em 2023.³⁻⁵ A variedade da matéria prima existente no Brasil contribui para o aumento da produção do biodiesel e consequentemente da evolução da sua percentagem adicionada ao diesel.⁶ Assim, pesquisas visando à introdução de outras matérias primas, como por exemplo, a mafurra na matriz brasileira torna-se importante.

Mafurra (*Trichilia emetica*) é uma árvore de origem africana, cultivada no sul de Moçambique para fins ornamentais e em formulações cosméticas tradicionais.⁷ Os frutos desta planta contêm sementes através do qual pode-se obter 55-65% de óleo, tornando essa planta como uma oleaginosa promissora para a produção de biodiesel e que pode ser introduzida na matriz brasileira.^{8,9}

A evolução da percentagem de biodiesel na mistura com diesel além de representar somente a esperança de haver uma substituição total do diesel pelo biodiesel a médio ou longo prazo, também representa preocupações em relação ao seu uso com teores fora do que é exigido pelas normas. O Brasil é um exemplo disso porque mesmo com a exigência de adição do biodiesel ao diesel em pequenas percentagens, a característica teor de biodiesel sempre representou mais de 40% das não conformidades.¹⁰ Com isso, o desenvolvimento de metodologias analíticas que sejam minimamente invasivas, fáceis

de interpretação e com altos valores de predição, é uma necessidade constante.

Existem pesquisas que relatam o uso de diversas técnicas analíticas associadas a algum tipo de análise estatística de dados, uni e/ou multivariada, como sendo bastante eficazes para predição. Como são os casos da espectroscopia FT-Raman associada à calibração multivariada por Quadrados Mínimos Parciais (PLS) para determinar simultaneamente aspartame, ciclamato, sacarina e acesulfame-K em adoçantes de mesa em pó;¹¹ da cromatografia gasosa de alta temperatura associada à calibração multivariada por PLS para determinar propriedades físico-químicas de petróleo bruto;¹² da espectroscopia ultravioleta e quimiometria para identificação de taninos vegetais.¹³ Para determinação do teor de biodiesel em mistura com o diesel, a norma brasileira ABNT NBR 15568 recomenda o uso da espectroscopia no infravermelho médio e calibração multivariada por PLS.¹⁴

A espectroscopia de ressonância magnética nuclear de hidrogênio (RMN de ¹H) combinada com calibração multivariada é bastante utilizada em diversas análises;¹⁵⁻¹⁸ devido às vantagens de fornecer informações sobre as fórmulas estruturais, moleculares e geométricas de compostos orgânicos; realizar simultaneamente análises qualitativas e quantitativas; não necessitar de isolamento do analito quando presente em misturas, permitindo que sejam realizadas simultaneamente análises de diferentes analitos em um espectro de forma rápida.^{19,20}

Os métodos quimiométricos PLS e Análise Discriminante por Quadrados Mínimos Parciais (PLS-DA) usados para previsões e classificações, respectivamente, foram estendidas para Projeções ortogonais à estrutura latente (OPLS) e Análise discriminante por projeções ortogonais à estrutura latente (OPLS-DA) com objectivo de melhorar os resultados de previsão e classificação, respectivamente. A aplicação destes métodos quimiométricos em dados de RMN de ¹H é exequível e se torna vantajosa porque esta técnica apresenta boa sensibilidade que a espectroscopia no infravermelho médio. Esta realidade é evidenciada em alguns trabalhos recentemente publicados pelos autores Silva *et al.*,²¹ Lin *et al.*²² e Souhi *et al.*²³

*e-mail: ademarmaquina@gmail.com

No entanto, não existem trabalhos publicados que relatam o uso da regressão multivariada por OPLS para previsões do teor de biodiesel nas misturas de diesel/biodiesel.

Diante do exposto, é proposta neste trabalho a utilização da espectroscopia de ressonância magnética nuclear de hidrogênio associada à regressão multivariada por OPLS e PLS para desenvolver uma metodologia analítica que fornece excelentes resultados de previsão do teor de biodiesel de mafurra nas misturas diesel/biodiesel.

MATERIAIS E MÉTODOS

Produção de biodiesel e preparação das amostras

O óleo de mafurra foi adquirido em Moçambique e o seu transporte para Brasil foi realizado sem restrições porque é um óleo comercial que foi transportado em pequena quantidade (1000 ml). Este óleo foi utilizado para a produção de biodiesel metílico de mafurra no Laboratório de Biocombustíveis do Instituto de Química da Universidade Federal de Uberlândia, seguindo os procedimentos descritos na referência.⁸

A Transpetro S/A (Brasil) forneceu o diesel puro usado para preparar 70 amostras de diesel/biodiesel, adicionando biodiesel B100 ao diesel em uma concentração que varia de 1,00 a 30,00% (v/v). Para a construção dos modelos OPLS e PLS, foram utilizadas 46 amostras no conjunto de calibração e as 24 amostras restantes foram usadas no conjunto de previsão. A seleção das amostras para cada conjunto foi realizada empregando o algoritmo Kennard-Stone.

Aquisição de dados espectrais

Todos os espectros de RMN de ¹H foram obtidos em um espectrômetro Bruker Avance III 400 MHz RMN, utilizando as seguintes condições experimentais: Programa de pulso, zg30; Largura espectral, de -3,84 a 16,19 ppm; Tamanho espectral, 65536 pontos; Pulso 90°, 11,75 µs; Atraso, 4 s e Número de varreduras, 16. Para obter os espectros, dissolveram-se 100 µL de cada amostra em 400 µL de CDCl₃ (clorofórmio deuterado), totalizando um volume de 500 µL de solução.

Análise quimiométrica

Software MATLAB, versão R2018b (Mathworks, Inc.); o PLS_Toolbox, versão 8.8 (Eigenvector Research) e o software SIMCA, versão 14.1, foram utilizados para pré-processar os dados e desenvolver os modelos OPLS e PLS. Para realizar os procedimentos multivariados, os dados foram organizados em uma matriz ordenada de linhas e colunas, constituindo uma matriz **X**, onde cada linha corresponde a uma amostra “*m*” e cada coluna corresponde a uma variável “*n*”, na qual *m* = 1,2,3, ..., 46 e *n* = 1,2,3, ..., 65536. Um vector **y** foi criada contendo valores de concentração (de 1,00 a 30,00% (v/v)) das amostras. Os espectros de RMN de ¹H foram alinhados usando uma rotina desenvolvida pelos autores, aplicando a função Interval-correlation-shift (icoshift)²⁴ no software MATLAB.

Na construção do modelo PLS, a matriz **X** é correlacionada com matriz **Y** através de operações matemáticas para se obter Variáveis Latentes (VLs) e Coeficientes de Regressão, usados para atingir a covariância máxima entre os espectros e as concentrações das espécies de interesse e determinar o valor da concentração de cada perfil espectral, respectivamente.¹ O objetivo deste processo é encontrar um pequeno número *A* de fatores relevantes que são preditivos para **Y** e que utilizam **X** com eficiência. Para tal, a matriz **X** é decomposta em um conjunto de fatores ortogonais que são usados para ajustar **Y**, de acordo com a Equação 1. A matriz **Y** é decomposta de acordo com a Equação 2:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E}_X = \sum t_h \mathbf{p}_h^T + \mathbf{E}_X \quad (1)$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{E}_Y = \sum u_h \mathbf{q}_h^T + \mathbf{E}_Y \quad (2)$$

Em que, **X** e **Y** são matrizes que contêm dados instrumentais de medição e resposta (concentração), respectivamente; **T** e **U** são os escores para as duas matrizes de dados; **P** e **Q** são os respectivos pesos, *h* é o número de variável latente (VL), **E_X** e **E_Y** são os respectivos resíduos. No entanto, se o bloco **Y** for univariado, o peso **Q** é definido como 1.

A relação linear entre as duas matrizes é estabelecida pela correlação dos escores de **X** e **Y** para cada VL, de acordo com a Equação 3:

$$\mathbf{U}_h = \mathbf{b}_h \mathbf{T}_h + \mathbf{E} \quad (3)$$

Em que, **U_h** é uma matriz que contém as propriedades de todas as amostras (concentração), **b_h** é um vetor que contém os parâmetros do modelo, **T_h** é uma matriz que contém os dados de medição instrumental (espectro) para as amostras de calibração e **E** é uma matriz que representa ruído.

Matrizes complexas, por exemplo, de espectros de RMN de ¹H de misturas de diesel/biodiesel, quando correlacionada com as informações da matriz **Y** apresentam forte variação sistemáticas ortogonais (não correlacionados) a **Y**. Estas variações desajustam o **Y**, no entanto, devem ser corrigidas para melhorar a capacidade preditiva do modelo. Para a correção destas variações, o modelo PLS é estendido para OPLS que integra o filtro Orthogonal Signal Correction (OSC).

O OPLS usa informações na matriz **Y** para decompor a matriz **X** em três partes distintas, conforme descrito na Equação 4, em que **T_p** é a matriz de escore preditiva para **X**, **P_p** é a matriz de carga preditiva para **X**, **T_o** é a matriz de escore Y-ortogonal correspondente, **P_o** é a matriz de carga dos componentes Y-ortogonais, e **E** é a matriz residual de **X**.^{25,26}

$$\mathbf{X} = \mathbf{T}_p \mathbf{P}_p^T + \mathbf{T}_o \mathbf{P}_o^T + \mathbf{E} \quad (4)$$

Para previsão, a variação ortogonal-Y é removida da matriz de dados **X**, de acordo com a Equação 5:

$$\mathbf{X}_p = \mathbf{X} - \mathbf{T}_o \mathbf{P}_o^T \quad (5)$$

Em seguida, a matriz **Y** prevista é estimada usando o **X_p** atualizado e os componentes preditivos do modelo OPLS, que são estimados a partir do conjunto de calibração.²⁵

Para o modelo PLS, os dados foram somente centrados na média enquanto que para o modelo OPLS foram centrados na média e pré-processados usando o filtro OSC, a fim de minimizar variações sistemáticas indesejáveis para produzir modelos menos complexos com melhor poder preditivo.^{27,28} A validação cruzada nos modelos OPLS e PLS foi realizada pelo critério da veneziana, com 14 blocos de divisão de dados e com duas amostras por bloco. O número de VLs com o menor RMSECV foi selecionado, seguindo os requisitos da ASTM E1655-55 (2012).²⁹

A comparação dos valores do RMSEP dos modelos OPLS e PLS em relação à significância estatística da diferença verificada foi realizada com base no Teste F ao nível de confiança de 95%, calculado de acordo com a Equação 6.³⁰

$$F_{calc} = \left(\frac{\text{RMSEP}_1}{\text{RMSEP}_2} \right)^2 \quad (6)$$

Em que, RMSEP₁ é do modelo PLS e RMSEP₂ é do modelo OPLS, sendo RMSEP₁ > RMSEP₂. O valor de *F_{calc}* é comparado com o valor da distribuição Fisher – Snedecor (*F_{tab}*) com graus de liberdade igual ao número de amostras de previsão e um nível de significância de 5%.

A validação dos modelos OPLS e PLS foi realizada com base na determinação das seguintes figuras de méritos: seletividade, sensibilidade, sensibilidade analítica, limite de detecção, limite de quantificação, teste para erros sistemáticos (*bias* e t_{bias}), de acordo com as equações apresentadas na Tabela 1.³¹⁻³⁴ A qualidade do modelo OPLS foi analisada com base no valor de R² e Q², que indicam a correlação entre os valores observados e previstos para a resposta estudada e a correlação entre a resposta prevista observada e validada cruzada, respectivamente.²³

Tabela 1. Equações usadas para calcular as figuras de mérito dos modelos*

Figura de mérito	Equação
Seletividade	$S\hat{E}L_i = \frac{n\hat{a}s_i}{x_i}$
Sensibilidade	$S\hat{E}N = \frac{1}{b_k}$
Sensibilidade analítica	$\gamma = \frac{S\hat{E}N}{\delta_x}$
Limite de detecção	$LQ = 10\delta_x \frac{1}{S\hat{E}N}$
Limite de quantificação	$LQ = 10\delta_x \frac{1}{S\hat{E}N}$
<i>bias</i>	$bias = \frac{\sum_{i=1}^{n_{val}} (y_i - \hat{y}_i)}{n_{val}}$
t_{bias}	$t_{bias} = \frac{ bias \sqrt{n_{val}}}{SDV}$
SDV	$SDV = \sqrt{\frac{\sum [(y_i - \hat{y}_i) - bias]^2}{n_{val} - 1}}$
Q ²	$Q^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_{icv})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$
R ²	$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_{ical})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$

*Onde \hat{y}_i é o valor previsto usando os modelos OPLS e PLS; y_i é o valor de referência para a amostra I; \hat{y}_{icv} e \hat{y}_{ical} são valores previstos para y_i na validação cruzada e no modelo final, respectivamente; \bar{y} é o valor médio de y_i ; n_{val} é o número de amostras de validação; $n\hat{a}s_i$ é a norma do vetor NAS e $\|x_i\|$ é a norma de cada espectro; b_k é o vetor dos coeficientes de regressão final, δ_x é o valor do desvio padrão de 9 amostras do branco (diesel).

RESULTADOS E DISCUSSÃO

Espectros de RMN de ¹H

A Figura 1 mostra 70 espectros de RMN de ¹H da mistura diesel/biodiesel de mafurra na faixa de concentração de 1,00 a 30,0%, na qual se observa os picos característicos do biodiesel e diesel. Os picos característicos do biodiesel são encontrados nas regiões de 3.7 e 5.4 ppm, atribuídos à prótons metoxi e hidrogênios olefínicos, respectivamente. Os picos que aparecem na região de 6.6 à 7.3 ppm, atribuídos à gama de hidrogênios aromáticos são característicos do diesel.^{19,35,36} A presença desses picos nos espectros demonstra que se trata de uma mistura de diesel/biodiesel, mas não permite monitorar visualmente o teor de biodiesel nesta mistura porque a semelhança espectral é grande.

Como alternativa ao monitoramento visual, os métodos quimio-métrico são práticos para quantificar o teor de biodiesel na mistura com diesel quando há variações entre os espectros que não é enxergado visualmente.

Análise quimiométrica

As Figuras 2 e 3 mostram a) os erros RMSEC e RMSECV, b) a variância acumulada capturada nos blocos **X** e **Y** obtidos em função da quantidade da variável latente (VL) escolhida para construção dos modelos OPLS e PLS, respectivamente. No modelo OPLS é otimizado o número de componentes ortogonais para cada variável preditora, no entanto, foram escolhidas 2 variáveis latentes devido ao pre-processamento OSC que separa a direção preditiva da direção ortogonal-**Y**, fornecendo um bom ajuste do modelo. Estas variáveis proporcionam menores erros e maiores variância capturada, isto é, representam 96,33% e 99,97% da variância explicada nos blocos **X** e **Y**, respectivamente. Enquanto que no modelo PLS é otimizado o número de variáveis latentes, razão pela qual as 4 variáveis latentes escolhidas proporcionam menores erros e maiores variância capturada, representando 93,31% e 98,77% da variância explicada nos blocos **X** e **Y**, respectivamente.

A Tabela 2 apresenta resultados dos parâmetros e figuras de mérito, calculados para os modelos OPLS e PLS, no qual se observa que os valores de Erro Quadrático Médio de Validação Cruzada (RMSECV), Erro Quadrático Médio de Calibração (RMSEC) e Erro Quadrático Médio de Previsão (RMSEP) do modelo OPLS são inferiores que do modelo PLS e estão próximos dos valores obtidos em alguns modelos de quantificação publicados na literatura.^{37,38} O valor do t_{bias} calculado foi menor que $t_{critico}$ nos dois modelos, o que equivale dizer que os modelos não apresentaram erros sistemáticos.

O valor do LQ obtido no modelo OPLS é superior ao LD, demonstrando uma boa concordância entre os parâmetros porque a concentração mais baixa da substância de interesse que é medida com uma incerteza máxima de 10% está dentro do que pode ser detectado. O modelo PLS também obteve valor de LQ superior ao LD. A sensibilidade expressa um aumento na fração do sinal quando a concentração do analito de interesse possui um valor alto para uma unidade, no entanto, o seu valor foi estimado em $5,73 \times 10^7$ e $3,77 \times 10^7$ para os modelos OPLS e PLS, respectivamente. A sensibilidade analítica do modelo OPLS foi maior que a do modelo PLS, mostrando que a influência dos resíduos na previsão de amostras desconhecidas é menor no modelo OPLS enquanto no modelo PLS é maior. O inverso do valor da sensibilidade analítica mostra que os modelos OPLS e PLS são capazes de distinguir diferenças entre amostras com uma variação de $4,03 \times 10^{-4}$ e $6,13 \times 10^{-4}$, respectivamente. O modelo OPLS apresentou valor da seletividade maior que o do modelo PLS devido ao processo de remoção das variações ortogonais a **Y**, ou seja, o modelo PLS apresenta uma sobreposição significativa do sinal de interferência com o analito.

A avaliação do ajuste dos modelos OPLS e PLS por meio da correlação dos valores medidos e previstos dos conjuntos de calibração e previsão é mostrada na Figura 4 (a) e (b), respectivamente. Nesta figura, o modelo OPLS apresentou melhor ajuste porque uma alta correlação entre os dois valores foi evidente, com o coeficiente de correlação (R²) superior a 0,99, enquanto que o modelo PLS apresentou um coeficiente de correlação (R²) inferior a 0,98. Apesar dos dois modelos apresentarem valor de Q² maior que 0,5, o modelo OPLS é mais robusto e com boa capacidade preditiva que o modelo PLS porque apresentou maior valor de Q² e boa capacidade de resistência a pequenas e deliberadas variações em relação à estabilidade das soluções analíticas preparadas.

A Figura 5 apresenta um gráfico de escores para o modelo OPLS

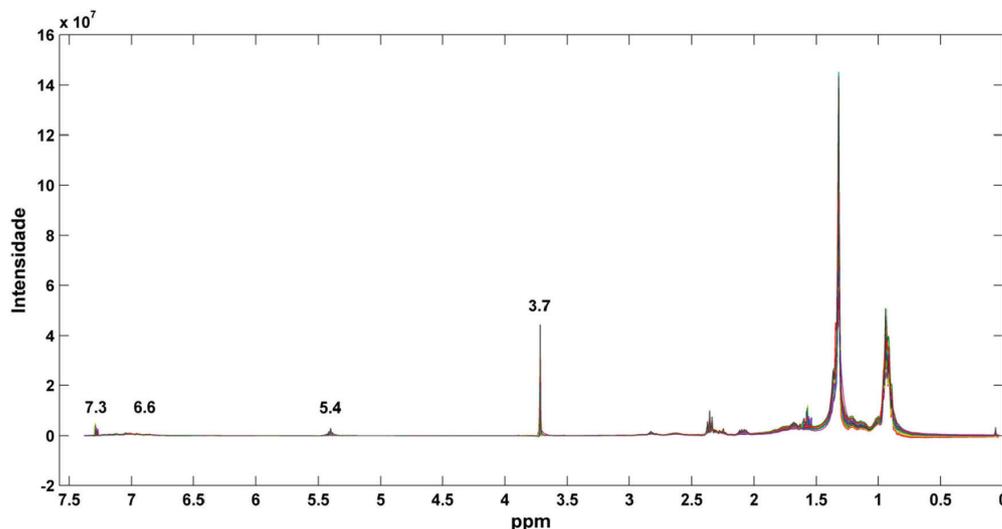


Figura 1. Espectros de RMN de ^1H da mistura diesel/biodiesel de mafurra

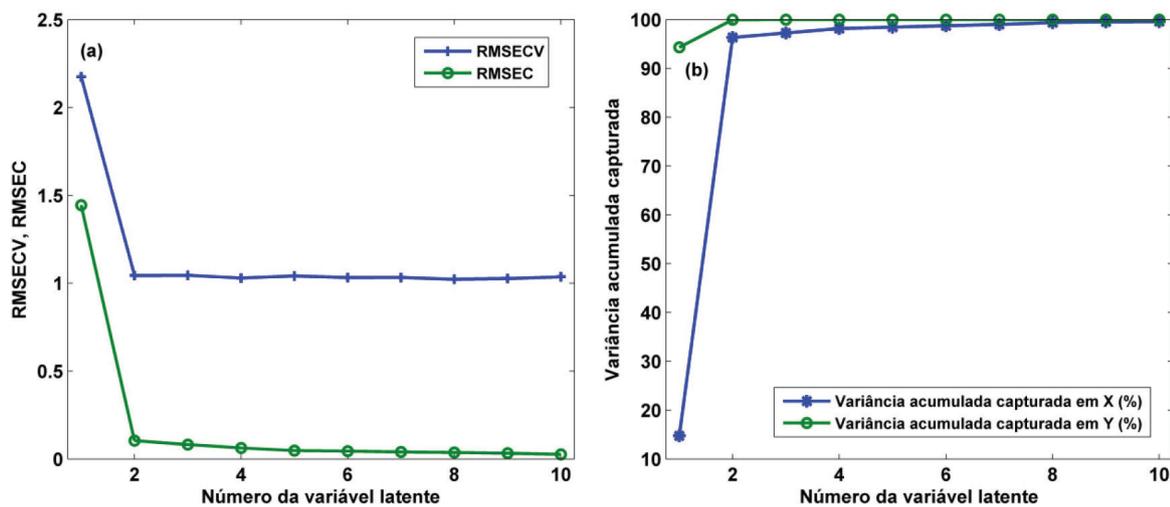


Figura 2. a) RMSEC e RMSECv, b) Variância acumulada capturada nos blocos X e Y versus variável latente do modelo OPLS

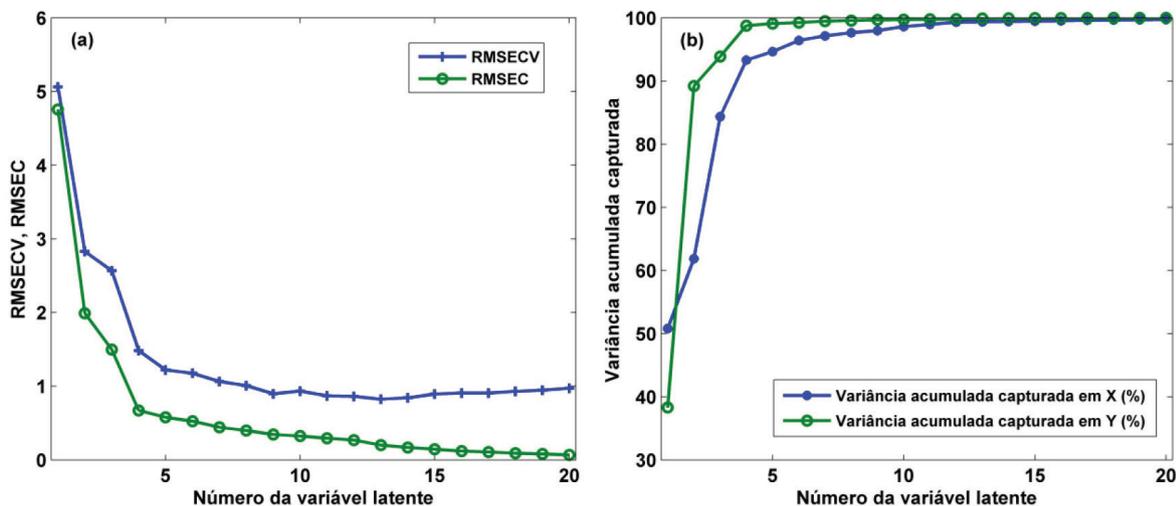


Figura 3. a) RMSEC e RMSECv, b) Variância acumulada capturada nos blocos X e Y versus variável latente do modelo PLS

no (a) conjunto de calibração e (b) conjunto previsão, que efetivamente separaram a direção preditiva em $t[1]$ da direção ortogonal-Y $t_0[1]$, fornecendo um bom ajuste do modelo.

A figura 6 mostra os pesos das variáveis latentes do modelo OPLS, na qual as variáveis da direção preditiva são descritas na VL1 com

81,59% da variância explicada das seguintes regiões espectrais: de 0.8 a 2.8 ppm atribuídos aos hidrogênios alifáticos; 3.7 e 5.4 ppm que são picos característicos do biodiesel, atribuídos à prótons metoxi e hidrogênios olefínicos, respectivamente^{35,39}. As demais variáveis descritas na VL2 com 14,74% são na sua maioria da direção ortogonal-Y. O

Tabela 2. Parâmetros e figuras de mérito obtidos nos modelos OPLS e PLS

Parâmetros/Figuras de mérito	Valores do modelo	
	OPLS	PLS
RMSECV / % (v/v)	1,04	1,48
RMSEC / % (v/v)	0,10	0,67
RMSEP / % (v/v)	0,78	1,32
Seletividade	0,70	0,18
Sensibilidade / % (v/v) ⁻¹	5,73×10 ⁷	3,77 ×10 ⁷
Sensibilidade analítica / % (v/v)	2,47 ×10 ³	1,63 ×10 ³
Inverso da sensibilidade analítica / (v/v) ⁻¹	4,03×10 ⁻⁴	6,13 ×10 ⁻⁴
Limite de detecção / % (v/v)	0,0013	0,0020
Limite de quantificação / % (v/v)	0,0040	0,0061
t _{bias}	0,32	0,41
t _{crítico}	2,0639	2,0639
R ²	0,995	0,977
Q ²	0,970	0,940

mesmo acontece com os pesos das variáveis latentes do modelo PLS, apresentados na figura 7, em que as 4 variáveis latentes (VL1, VL2, VL3 e VL4) descrevem as variações das regiões mais significativas para construção do modelo, que são os deslocamentos químicos esperados para o biodiesel.

A Tabela 3 apresenta os resultados dos testes de permutação realizados nos modelos OPLS e PLS para verificar o ajuste excessivo

dos modelos. Estes resultados mostram que os dois modelos não estão super ajustados porque os valores de Wilcoxon (teste de classificação assinado em pares Wilcoxon), Sign Test (teste de classificação assinada em pares) e Rand t-test (Teste-t de randomização) são menores que 0,05. No caso do modelo PLS, este resultado já era esperado porque a análise do ajuste do modelo por meio da correlação dos valores medidos e previstos dos conjuntos de calibração e previsão mostrou que o R² foi inferior a 0,98.

A Tabela 4 apresenta resultado do teste F aplicado para verificar se o valor de RMSEP do modelo OPLS é significativamente maior que o do modelo PLS com 95% de confiança. Nesta tabela se observa que o valor do F_{calculado} é maior que do F_{tabelado}, o que equivale dizer que existe diferença estatística significativa entre os valores de RMSEP dos modelos. Assim, considera-se que o valor de RMSEP do modelo OPLS é melhor que do modelo PLS.

CONCLUSÃO

A aplicação da regressão multivariada por OPLS e PLS nos espectros de RMN de ¹H de misturas diesel/biodiesel de mafurra permitiu o desenvolvimento de duas metodologias para quantificar o teor desse biodiesel na mistura com diesel. A eficiência destas metodologias foi analisada com base nas figuras de mérito e no ajuste dos modelos. Os resultados das figuras de mérito (a exatidão, os limites de detecção e quantificação e o inverso da sensibilidade analítica) do modelo OPLS foram concordantes com os requisitos estabelecidos na norma ASTM E1655-05 e melhores que do modelo PLS. O ajuste do modelo OPLS mostrou uma alta correlação existente entre valores de concentração real e previsto dos conjuntos de calibração e previsão,

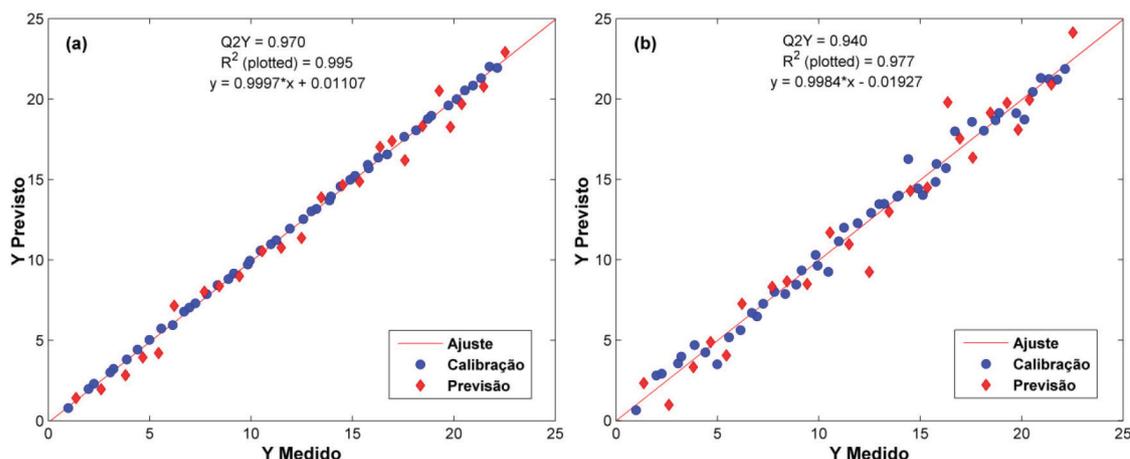


Figura 4. Correlação dos valores medidos e previstos dos conjuntos de calibração e previsão dos modelos (a) OPLS e (b) PLS

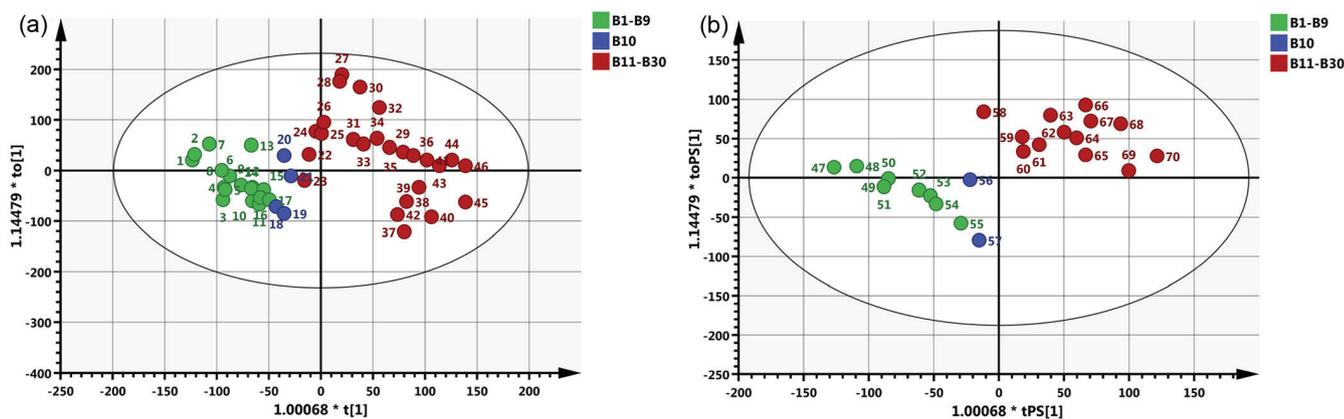


Figura 5. Escores para o modelo OPLS em (a) calibração e (b) previsão

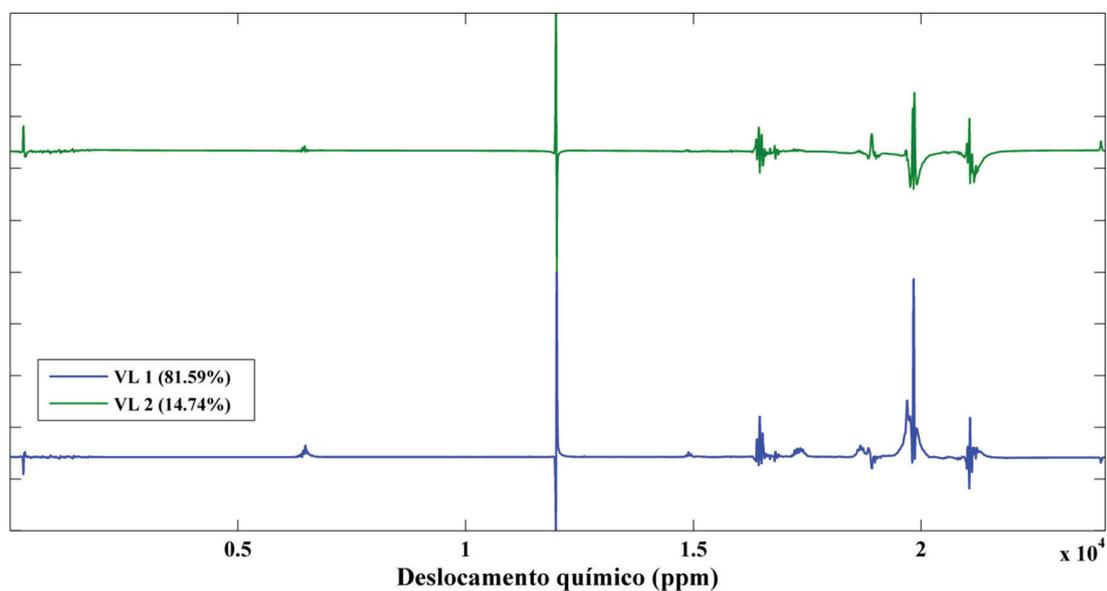


Figura 6. Pesos das variáveis latentes (VL1 e VL2) do modelo OPLS

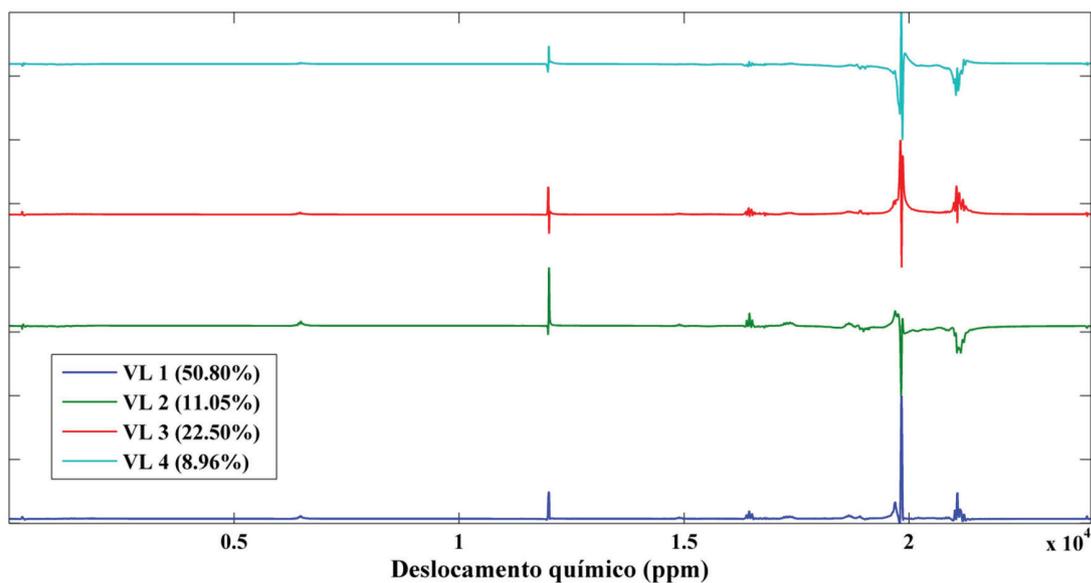


Figura 7. Pesos das variáveis latentes (VL1, VL2, VL3 e VL4) do modelo PLS

Tabela 3. Probabilidade de insignificância dos modelos OPLS e PLS vs. amostras permutadas para os modelos com 2 e 4 VL, respectivamente

	Wilcoxon		Sign Test		Rand t-test	
	OPLS	PLS	OPLS	PLS	OPLS	PLS
Y-column: 1						
Self-Pred (RMSEC)	0.000	0.000	0.000	0.000	0.005	0.005
Cross-Val (RMSECV)	0.000	0.000	0.000	0.000	0.005	0.005
Y-column: 2						
Self-Pred (RMSEC)	0.000	0.000	0.000	0.000	0.005	0.02
Cross-Val (RMSECV)	0.000	0.000	0.000	0.000	0.005	0.01

Tabela 4. Resultado do Teste F

	Modelo OPLS	Modelo PLS	F _{calculado}	F _{tabelado} (0,05,24,24)
RMSEP	0,776	1,319	2,889	2,027

com um coeficiente de correlação superior que 0,99 e com erros relativamente baixos que do modelo PLS.

Estes resultados demonstram que a metodologia desenvolvida com base na espectroscopia de RMN de ¹H associado à regressão multivariada por OPLS é ideal para ser usada pelas agências reguladoras no monitoramento do teor de biodiesel em mistura com diesel, pois apresenta excelente capacidade preditiva.

AGRADECIMENTOS

Os autores agradecem ao Programa PEC-PG (processo número 190025/2017-4) pelo apoio financeiro e à empresa Transpetro S/A (Brasil) pelo fornecimento de diesel puro.

REFERÊNCIAS E NOTA

- Máquina, A. D. V.; Siteo, B. V.; Buiatte, J. E.; Santos, D. Q.; Neto, W. B.; *Fuel* **2019**, *237*, 373. <https://doi.org/10.1016/j.fuel.2018.10.011>.
- Siteo, B. V.; Máquina, A. D. V.; de Santana, F. B.; Gontijo, L. C.; Santos, D. Q.; Borges Neto, W.; *Fuel* **2017**, *191*, 290.
- Mazivila, S. J.; de Santana, F. B.; Mitsutake, H.; Gontijo, L. C.; Santos, D. Q.; Neto, W. B.; *Fuel* **2015**, *142*, 222.
- Bambo, T. F.; *Tese de Mestrado*, Universidade de São Paulo, Brasil, 2014.
- República Federativa do Brasil *Lei no. 13.033, de 24 set. 2014: Dispõe sobre a adição obrigatória de biodiesel ao óleo diesel comercializado com o consumidor final; altera as Leis no 9.478, de 6 de agosto de 1997, e 8.723, de 28 de outubro de 1993; revoga dispositivos da Lei no 11.097, [s.d.]*.
- Brasil, R. F. do *Brazilian Energy Balance 2019: Year 2018* Copyright © 2019 EPE – Empresa de Pesquisa Energética: Rio de Janeiro, 2019.
- Máquina, A. D. V.; Souza, L. M. de; Gontijo, L. C.; Santos, D. Q.; Borges Neto, W.; *Anal. Lett.* **2017**, *50*, 2117. <https://doi.org/10.1080/00032719.2016.1267186>.
- Máquina, A. D. V.; Souza, L. M.; Buiatte, J. E.; Santos, D. Q.; Neto, W. B.; *Energy and Fuels* **2017**, *31*, 571. <https://doi.org/10.1021/acs.energyfuels.6b02079>.
- Grundy, I. M.; Campbell, B. M.; *Econ. Bot.* **1993**, *47*, 148.
- anp *Boletim de Monitoramento da Qualidade dos Combustíveis* in: PMQC. Rio de Janeiro, 2019.
- Duarte, L. M.; Paschoal, D.; Izumi, C. M. S.; Dolzan, M. D.; Alves, V. R.; Micke, G. A.; Dos Santos, H. F.; de Oliveira, M. A. L.; *Food Res. Int.* **2017**, *99*, 106. <https://doi.org/10.1016/j.foodres.2017.05.006>.
- Rodrigues, É. V. A.; Silva, S. R. C.; Romão, W.; Castro, E. V. R.; Filgueiras, P. R.; *Fuel* **2018**, *220*, 389. <https://doi.org/10.1016/j.fuel.2018.01.139>.
- Grasel, F. dos S.; Ferrão, M. F.; Wolf, C. R.; *Ind. Crops Prod.* **2016**, *91*, 279. <https://doi.org/10.1016/j.indcrop.2016.07.022>.
- ABNT NBR-15568. Biodiesel - Determination of biodiesel content in diesel fuel oil by mid infrared spectroscopy. In. ABNT NBR. Rio de Janeiro. **2008**.
- Vieira, A. P.; Portela, N. A.; Neto, Á. C.; Lacerda, V.; Romão, W.; Castro, E. V. R.; Filgueiras, P. R.; *Fuel* **2019**, *253*, 320. <https://doi.org/10.1016/j.fuel.2019.05.028>.
- Jović, O.; Pičuljan, K.; Hrenar, T.; Smolić, T.; Primožič, I.; *Chemom. Intell. Lab. Syst.* **2019**, *185*, 41. <https://doi.org/10.1016/j.chemolab.2018.12.012>.
- Monakhova, Y. B.; Diehl, B. W. K.; *J. Pharm. Biomed. Anal.* **2015**, *115*, 543. <https://doi.org/10.1016/j.jpba.2015.08.017>.
- Fan, S.; Zhong, Q.; Fahl-Hassek, C.; Pfister, M. K. H.; Horn, B.; Huang, Z.; *Food Control* **2018**, *88*, 113. <https://doi.org/10.1016/j.foodcont.2017.11.002>.
- Pavia, D. L.; Lampman, G. M.; Kriz, G. S.; Vyvyan, J. R. *Introdução à Espectroscopia. Tradução da 4ª edição norte-americana*, 4a CENGAGE Learning: São Paulo, 2010.
- Shimamoto, G. G.; Tubino, M.; *Fuel* **2016**, *175*, 99. <https://doi.org/10.1016/j.fuel.2016.02.019>.
- Silva, A. F. T.; Sarraguça, M. C.; Ribeiro, P. R.; Santos, A. O.; De Beer, T.; Lopes, J. A.; *Int. J. Pharm.* **2017**, *520*, 29. <https://doi.org/10.1016/j.ijpharm.2017.01.052>.
- Lin, W. S.; Yang, C. M.; Kuo, B. J.; *Chemom. Intell. Lab. Syst.* **2012**, *115*, 25. <https://doi.org/10.1016/j.chemolab.2012.04.005>.
- Souhi, N.; Lindegren, A.; Eriksson, L.; Trygg, J.; *Anal. Chim. Acta* **2015**, *857*, 28. <https://doi.org/10.1016/j.aca.2014.12.003>.
- Savorani, F.; Tomasi, G.; Engelsen, S. B.; *J. Magn. Reson.* **2010**, *202*, 190. <https://doi.org/10.1016/j.jmr.2009.11.012>.
- Max, B.; Mattias, R.; Cloarec, O.; Nicholson, J. K.; Holmes, E.; Trygg, J.; *J. Chemom.* **2006**, *20*, 341. <https://doi.org/10.1002/cem>.
- Trygg, J.; Wold, S.; *J. Chemom.* **2002**, *16*, 119. <https://doi.org/10.1002/cem.695>.
- Westerhuis, J. A.; de Jong, S.; Smilde, A. K.; *Chemom. Intell. Lab. Syst.* **2001**, *56*, 13. [https://doi.org/10.1016/S0169-7439\(01\)00102-2](https://doi.org/10.1016/S0169-7439(01)00102-2).
- Pers, T. H.; Martin, F. P.; Verdich, C.; Holst, C.; Johansen, J. V.; Astrup, A.; Polak, J.; Martinez, J. A.; Rezzi, S.; Blaak, E. E.; Saris, W. H. M.; Kochhar, S.; Macdonald, I. A.; Sørensen, T. I. A.; Ramadan, Z.; *Chemom. Intell. Lab. Syst.* **2008**, *93*, 34. <https://doi.org/10.1016/j.chemolab.2008.03.008>.
- ASTM E1655-05 Standard Practices for Infrared Multivariate Quantitative Analysis. **2012**, 1. <https://doi.org/10.1520/E1655-05R12>.
- Filgueiras, P. R.; Alves, J. C. L.; Poppi, R. J.; *Talanta* **2014**, *119*, 582. <https://doi.org/10.1016/j.talanta.2013.11.056>.
- Ferreira, M. H.; Braga, J. W. B.; Sena, M. M.; *Microchem. J.* **2013**, *109*, 158. <https://doi.org/10.1016/j.microc.2012.03.008>.
- Valderrama, P.; Braga, J. W. B.; Poppi, R. J.; *Quim. Nova* **2009**, *32*, 1278. <https://doi.org/10.1590/S0100-40422009000500034>.
- Silva, M. A. M.; Ferreira, M. H.; Braga, J. W. B.; Sena, M. M.; *Talanta* **2012**, *89*, 342. <https://doi.org/10.1016/j.talanta.2011.12.039>.
- De Carvalho Rocha, W. F.; Nogueira, R.; Vaz, B. G.; *J. Chemom.* **2012**, *26*, 456. <https://doi.org/10.1002/cem.2420>.
- Portela, N. A.; Oliveira, E. C. S.; Neto, A. C.; Rodrigues, R. R. T.; Silva, S. R. C.; Castro, E. V. R.; Filgueiras, P. R.; *Fuel* **2016**, *166*, 12. <https://doi.org/10.1016/j.fuel.2015.10.091>.
- Monteiro, M. R.; Ambrozín, A. R. P.; Lião, L. M.; Ferreira, A. G.; *Fuel* **2009**, *88*, 691. <https://doi.org/10.1016/j.fuel.2008.10.010>.
- Shi, T.; Zhu, M. T.; Zhou, X. Y.; Huo, X.; Long, Y.; Zeng, X. Z.; Chen, Y.; *Food Chem.* **2019**, *287*, 46. <https://doi.org/10.1016/j.foodchem.2019.02.072>.
- Singh, K.; Kumar, S. P.; Blümich, B.; *Fuel* **2019**, *243*, 192. <https://doi.org/10.1016/j.fuel.2019.01.084>.
- Máquina, A. D. V.; Siteo, B. V.; de Oliveira Cruz, W.; Santos, D. Q.; Neto, W. B.; *Talanta* **2019**, *120590*. <https://doi.org/10.1016/j.talanta.2019.120590>.