



**Universidade do Estado do Rio de Janeiro**  
Centro de Tecnologia e Ciências  
Instituto de Química

Camilla Lima Cunha

**Aplicação de métodos de calibração multivariada na previsão de propriedades de combustíveis fósseis e biocombustíveis usando dados da espectroscopia por infravermelho, médio e próximo, e espectroscopia Raman**

Rio de Janeiro  
2019

Camilla Lima Cunha

**Aplicação de métodos de calibração multivariada na previsão de propriedades de combustíveis fósseis e biocombustíveis usando dados da espectroscopia por infravermelho, médio e próximo, e espectroscopia Raman**

Tese apresentada, como requisito parcial para a obtenção do título de Doutor, ao Programa de Pós-Graduação em Engenharia Química do Instituto de Química, da Universidade do Estado do Rio de Janeiro. Área de Concentração: Processos Químicos, Petróleo e Meio Ambiente.

Orientadores: Prof. Dr. Aderval Severino Luna  
Prof. Dr. Alexandre Rodrigues Torres

Rio de Janeiro  
2019

CATALOGAÇÃO NA FONTE  
UERJ / REDE SIRIUS / BIBLIOTECA CTC/Q

C972 Cunha, Camilla Lima.

Aplicação de métodos de calibração multivariada na previsão de propriedades de combustíveis fósseis e biocombustíveis usando dados da espectroscopia por infravermelho, médio e próximo, e espectroscopia Raman. – 2019.

222 f.

Orientador: Aderval Severino Luna

Orientador: Alexandre Rodrigues Torres

Dissertação (Doutorado) – Universidade do Estado do Rio de Janeiro. Instituto de Química.

1. Calibração multivariada – Teses. 2. Combustíveis fósseis – Teses. 3. Espectroscopia de infravermelho – Teses. 4. Raman, Espectroscopia de. I. Luna, Aderval Severino. II. Torres, Alexandre Rodrigues. III. Universidade do Estado do Rio de Janeiro. Instituto de Química. IV. Título.

CDU 543.42

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação, desde que citada a fonte.

---

Assinatura

---

Data


Camilla Lima Cunha

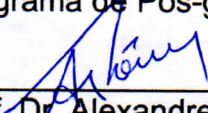
**Aplicação de métodos de calibração multivariada na previsão de propriedades de combustíveis fósseis e biocombustíveis usando dados da espectroscopia por infravermelho, médio e próximo, e espectroscopia Raman**

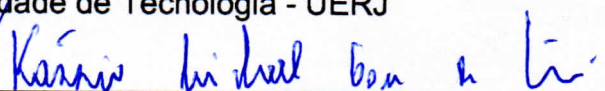
Tese apresentada, como requisito parcial para a obtenção do título de Doutor, ao Programa de Pós-Graduação em Engenharia Química do Instituto de Química, da Universidade do Estado do Rio de Janeiro. Área de Concentração: Processos Químicos, Petróleo e Meio Ambiente.

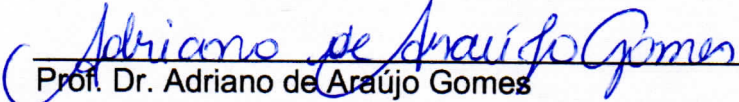
Aprovado em 31 de outubro de 2019.

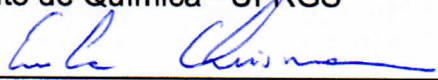
Banca Examinadora:


  
Prof. Dr. Aderval Severino Luna (Orientador)  
Programa de Pós-graduação em Engenharia Química - UERJ

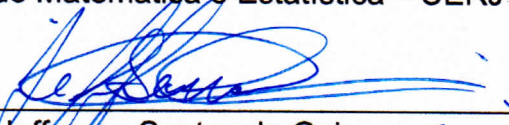
  
Prof. Dr. Alexandre Rodrigues Torres (Orientador)  
Faculdade de Tecnologia - UERJ

  
Prof. Dr. Kassio Michell Gomes de Lima  
Instituto de Química - UFRN

  
Prof. Dr. Adriano de Araújo Gomes  
Instituto de Química - UFRGS

  
Prof. Dr. Erika Christina Ashton Nunes Chrisman  
Escola de Química – UFRJ

  
Prof. Dr. Vinicius Layter Xavier  
Instituto de Matemática e Estatística – UERJ

  
Prof. Dr. Jefferson Santos de Gois  
Programa de Pós-graduação em Engenharia Química - UERJ

Rio de Janeiro

2019

## AGRADECIMENTOS

A Deus por ter me dado forças e iluminado meu caminho para que eu pudesse concluir mais essa realização na minha vida.

Aos meus familiares, em especial, a minha avó Maria Aparecida, minha mãe Sandra e a minha irmã Marianne, pelo apoio, incentivo e, exemplo de persistência e amor.

Aos meus orientadores Prof. Dr. Alexandre R. Torres e Prof. Dr. Aderval S. Luna pela paciência, amizade, dedicação ao meu aprendizado e à realização desse trabalho.

Ao meu namorado Rafael pelo carinho, apoio e companheirismo durante todo o doutorado.

Aos meus sogros, Selma e Romualdo, por todo o apoio, carinho e paciência, a minha cunhada e amiga Gabriela pela paciência, pelos ensinamentos compartilhados e, principalmente, pela amizade e apoio incondicional.

Aos meus amigos do laboratório da FAT-UERJ, Raquel, Luciana, Nilson, Wolney e Rafael pelas palavras de apoio e incentivo.

Ao meu grande amigo José Luís pela ajuda com os ensaios de viscosidade, apoio, carinho, palavras de incentivo e principalmente pelas piadas que alegraram meus dias no laboratório.

À Prof. Cristiane Henriques, Arnaldo e a Paula pela ajuda com os instrumentos e acessórios de infravermelho.

Aos professores da FAT-UERJ que muito me incentivaram com palavras de apoio e perseverança.

Aos meus tios, tias, primos e avós de coração, em especial, tia Carmita, tia Núbia, Marco, tia Rose, tio Raimundo, tia Erondina, tio José e Cecília que sempre me receberam com abraços carinhosos e me acolheram durante este período.

Aos amigos, Gesiene, Alanna, Silas, Érica, Marcela e Nadine que cada um, a sua maneira, foram e são pessoas importantes na minha vida, que mesmo longe sempre torceram e me apoiaram durante o doutorado.

Aos amigos que fiz no Programa de Pós-graduação em Engenharia Química da UERJ, em especial, Igor Lima, Arnaldo Peixoto, Denise Cândido e Wallace Antunes.

Ao corpo docente e aos funcionários do Programa de Pós-graduação em Engenharia Química da UERJ, especialmente a Renata e ao grande amigo Marco pelo suporte e apoio.

À CAPES (Coordenação de Aperfeiçoamento Pessoal de Nível Superior) pela assistência financeira.

## RESUMO

CUNHA, Camilla Lima. *Aplicação de métodos de calibração multivariada na previsão de propriedades de combustíveis fósseis e biocombustíveis usando dados da espectroscopia por infravermelho, médio e próximo, e espectroscopia Raman*. 2019. 222 f. Tese (Doutorado em Engenharia Química) – Instituto de Química – Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2019.

Com o desenvolvimento e o aumento do uso de combustíveis renováveis, surge também a necessidade de garantir sua produção e distribuição com a qualidade especificada. Desta forma, tornou-se imprescindível o estudo de metodologias eficazes para determinar as propriedades físico-químicas de forma rápida e não destrutiva, possibilitando a sua aplicação ao controle de qualidade no monitoramento on/in/at-line dos combustíveis em ambiente industrial. Na literatura, as espectroscopias de infravermelho, próximo (NIR) e médio (mid-IR), têm sido exaustivamente combinadas as ferramentas quimiométricas para determinação de propriedades de misturas diesel-biodiesel. Contudo, a inclusão da aplicação da espectroscopia Raman pode oferecer praticidade aos ensaios e contribuir na construção de instrumentos de baixo custo para controle de qualidade de combustíveis e biocombustíveis. Neste cenário, o presente estudo tem como objetivo apresentar modelos de previsão das propriedades de combustíveis (diesel) e biocombustíveis (biodiesel), puros e de suas misturas, baseados em técnicas espectroscópicas (NIR, mid-IR e Raman) acopladas com ferramentas quimiométricas, as quais se inclui a utilização de técnicas de seleção de variáveis combinadas com mínimos quadrados parciais (PLS) e a aplicação de florestas aleatórias (RF) e máquina de vetores de suporte (SVM) para as propriedades não lineares. Na primeira parte do estudo (banco 1), o índice de refração e a massa específica de 148 amostras de biodiesel foram determinados usando dados espectroscópicos. Na segunda parte do estudo (banco 2), o ponto de entupimento de filtro de frio (PEFF) e a viscosidade cinemática a 40°C de 333 amostras de biodiesel, diesel, diesel-biodiesel e misturas de diesel-biodiesel-etanol foram determinados usando dados espectroscópicos. Para ambos os estudos, 80 % das amostras formaram o conjunto de calibração e 20% formaram o conjunto de validação. Os resultados mostraram que o modelo PLS obtido com os espectros mid-IR e intervalos selecionados pelo algoritmo UVE forneceu a melhor previsão para o índice de refração, com 5 variáveis latentes, raiz quadrada do erro médio quadrático (RMSEP) igual a  $1,0 \cdot 10^{-4}$  e coeficiente de determinação da previsão ( $R^2_{pred}$ ) igual a 0,95. O melhor modelo de previsão para a massa específica foi obtido pelo método RF com os espectros de mid-IR, com RMSEP igual  $1,16 \cdot 10^{-4} \text{ g} \cdot \text{cm}^{-3}$  e  $R^2_{pred}$  igual a 0,98. Para a previsão do PEFF, o melhor modelo foi obtido com SVM em espectros mid-IR e intervalos selecionados pelo algoritmo RC, com RMSEP igual a 0,7°C e  $R^2_{pred}$  igual a 0,94. O melhor modelo de previsão da viscosidade cinemática a 40°C foi obtido pelo método RF com os espectros de mid-IR, com RMSEP igual a 0,0184  $\text{mm}^2 \cdot \text{s}^{-1}$  e  $R^2_{pred}$  igual a 0,98. Os resultados deste estudo mostraram que os espectros de mid-IR foram os mais adequados na previsão das propriedades estudadas. A figura de mérito Soma das probabilidades do teste de Wilcoxon (SWTP) desenvolvida neste estudo foi imprescindível para a escolha do melhor modelo.

Palavras-chave: PLS. SVM. Florestas aleatórias. Seleção de variáveis. NIR. mid-IR. Raman. Massa específica. Índice de refração. Ponto de entupimento de filtro a frio. Viscosidade.



## ABSTRACT

CUNHA, Camilla Lima. *Application of multivariate calibration methods to predict properties of fossil fuels and biofuels using middle and near infrared spectroscopy and Raman spectroscopy data*. 2019. 222 f. Tese (Doutorado em Engenharia Química) – Instituto de Química – Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2019.

With the development and increase in the use of renewable fuels, there is also need to guarantee its production and distribution with the specified quality. In this way, it became necessary to study effective methodologies to determine physicochemical properties in a fast and non-destructible way, allowing its application to the quality control on / in / at-line monitoring of fuels in an industrial environment. In literature, near and middle infrared spectroscopy have been exhaustively coupled with chemometric tools to determine the properties of diesel-biodiesel blends. However, the inclusion of Raman spectroscopy application can make testing easier and contribute to building of low-cost instruments for quality control of fuels and biofuels. In this scenario, the present study aims to present models for prediction of the properties of fuels (diesel) and biofuels (biodiesel), pure and its blends, based on spectroscopic techniques (NIR, mid-IR, and Raman) coupled with chemometric tools, which include the use of variable selection techniques combined with PLS and the application of random forests (RF) and support vector machine (SVM) for nonlinear properties. In the first part of the study (case 1), the refractive index and density of 148 biodiesel samples were determined using spectroscopic data. In the second part of the study (case 2), the cold filter plugging point (CFPP) and the kinematic viscosity at 40°C from 333 samples of biodiesel, diesel, diesel-biodiesel and blends diesel-biodiesel-ethanol were determined using spectroscopic data. In both cases, 80% of samples were used for the calibration set, and the remaining 20% was used as an external validation set. The results showed that PLS model with mid-IR spectra with intervals selected by UVE algorithm provided the best prediction for the refractive index with five latent variables, root-mean-square of error (RMSEP) equal to  $1.0 \cdot 10^{-4}$  and prediction coefficient ( $R^2_{\text{pred}}$ ) equal to 0.95. The best prediction model for density was obtained by RF method with mid-IR spectra, with RMSEP equal to  $1.16 \cdot 10^{-4} \text{ g.cm}^{-3}$  and  $R^2_{\text{pred}}$  equal to 0.98. For CFPP prediction model, the best model was obtained using SVM with mid-IR spectra and intervals selected by the RC algorithm, with RMSEP equal to 0.7°C and  $R^2_{\text{pred}}$  equal to 0.94. The best prediction model for kinematic viscosity at 40°C was obtained by RF method with mid-IR spectra, with RMSEP equal to  $0.0184 \text{ mm}^2.\text{s}^{-1}$  and  $R^2_{\text{pred}}$  equal to 0.98. The results in this work showed that mid-IR spectra were the most adequate in predicting the studied properties. The figure of merit sum of Wilcoxon test probability (SWTP) developed in this study was indispensable to choose the best model.

Keywords: PLS. SVM. Random forest. Variable selection. NIR. mid-IR. Raman.

Density. Refractive index. Cold filter plugging point. Viscosity.

## LISTA DE ILUSTRAÇÕES

Figura 1 - Esquema de decomposição das matrizes X e Y pela aplicação do PLS.....	40
Figura 2 - Classificação dos métodos de seleção de variáveis. ....	45
Figura 3 - Organograma das ferramentas e métodos quimiométricos utilizados. ..	57
Figura 4 - Fluxograma do estudo de previsão das propriedades. ....	57
Figura 5 - Reator contendo massa reacional durante a produção de biodiesel. ....	59
Figura 6 - Densímetro (DM40) acoplado ao refratômetro (RX40) com amostrador automático SC1.....	67
Figura 7 - Equipamento para ensaio do ponto de entupimento de filtro a frio, marca TANAKA modelo AFP-102. ....	68
Figura 8 - Viscosímetro utilizado nas medições de viscosidade cinemática à 40°C, modelo Ray-265D.....	69
Figura 9 - Espectrofotômetro de infravermelho com acessório ATR.....	70
Figura 10 - Espectrofotômetro de infravermelho com acessório NIRA.....	71
Figura 11 - Espectrofotômetro de Raman.....	72
Figura 12 - Espectros brutos de infravermelho médio (mid-IR) de amostras de biodiesel. ....	76
Figura 13 - Espectros de infravermelho médio (mid-IR) após a retirada da amostra anômala e remoção da região espectral do sinal do cristal de ZnSe.....	77
Figura 14 - Espectros brutos de infravermelho próximo (NIR) de amostras de biodiesel. ....	78
Figura 15 - Espectros de infravermelho próximo (NIR) de amostras de biodiesel após a correção de linha de base e remoção da amostra Am158.....	79
Figura 16 - Espectros de Raman (RAM) de amostras de biodiesel após a correção de linha de base, remoção da amostra Am158 e normalização dos espectros. ....	80
Figura 17 - Gráfico de autovalores <i>versus</i> número de componentes principais obtido pela PCA para os espectros NIR.....	81
Figura 18 - Gráfico de variância capturada <i>versus</i> número de componentes principais obtido pela PCA para os espectros NIR.....	82

Figura 19 - Gráfico dos resíduos <i>versus</i> escores (Q resíduos <i>versus</i> Hotelling T2) obtido pela PCA para os espectros NIR do Banco 1.....	82
Figura 20 - Gráfico da combinação de altos valores distância ortogonal e distância entre os escores obtido pelo ROBPCA para os dados de NIR do Banco 1 .....	83
Figura 21 - Gráfico de autovalores <i>versus</i> número de componentes principais obtido pela PCA para os espectros mid-IR .....	84
Figura 22 - Gráfico dos resíduos <i>versus</i> escores (Q resíduos <i>versus</i> Hotelling T2) obtido pela PCA para os espectros mid-IR do Banco 1 .....	85
Figura 23 - Gráfico da combinação de altos valores distância ortogonal e distância entre os escores obtido pelo ROBPCA para os dados de Mid-IR do Banco 1 .....	85
Figura 24 - Gráfico de autovalores <i>versus</i> número de componentes principais obtido pela PCA para os espectros Raman .....	86
Figura 25 - Gráfico dos resíduos <i>versus</i> escores (Q resíduos <i>versus</i> Hotelling T2) obtido pela PCA para os espectros Raman do Banco 1 .....	87
Figura 26 - Gráfico da combinação de altos valores distância ortogonal e distância entre os escores obtido pelo ROBPCA para os dados de Raman do Banco 1 .....	88
Figura 27 - Espectros de Raman do Banco 1 antes e após a transformação de sinal espacial (tss).....	88
Figura 28 - Fluxograma da divisão das amostras em conjunto de calibração e validação aplicando o algoritmo Kennard-Stone nos espectros de NIR .....	89
Figura 29 - Distribuição das amostras de calibração (+) e validação (•) no conjunto total de amostras do Banco 1, considerando os espectros NIR.....	90
Figura 30 - Intervalos espectrais do NIR selecionados pelos algoritmos UVE, BVE e VIP.....	94
Figura 31 - Intervalos espectrais do mid-IR selecionados pelos algoritmos VIP, BVE e IPW.....	94
Figura 32 - Intervalos espectrais de Raman selecionados pelos algoritmos BVE, GA e iPLS.....	95

Figura 33 - Modelos de previsão do índice de refração <i>versus</i> RMSEP (•) e RMSEC (•), obtidos a partir do conjunto espectros de NIR, a linha vertical indica o erro experimental.....	99
Figura 34 - Modelos de previsão do índice de refração <i>versus</i> RMSEP (•) e RMSEC (•), obtidos a partir do conjunto espectros de mid-IR, a linha vertical indica o erro experimental.....	100
Figura 35 - Modelos de previsão do índice de refração <i>versus</i> RMSEP (•) e RMSEC (•), obtidos a partir do conjunto espectros de Raman, a linha vertical indica o erro experimental.....	101
Figura 36 - Soma da probabilidade do teste de Wilcoxon (SWTP) <i>versus</i> RMSEP para todos os modelos de previsão do índice de refração, a linha vertical indica o erro experimental.....	102
Figura 37 - RMSEP <i>versus</i> intervalos espectrais para todos os modelos de previsão do índice de refração, a linha horizontal indica o erro experimental.....	103
Figura 38 - RMSEC e RMSEP <i>versus</i> dados espectrais (NIR, mid-IR e Raman) para todos os modelos de previsão do Índice de Refração, a linha horizontal indica o erro experimental.....	103
Figura 39 - Elipse da região conjunta de confiança (EJCR) do conjunto de validação para os modelos de previsão do Índice de refração: (-) MIR_UVE-PLS_IndRefra_drv; (-) NIR_VIP-PLS_IndRefra_snv; (- -) NIR_REP-PLS_IndRefra_snv; (- -) NIR_PLS_IndRefra_snv. ....	104
Figura 40 - Gráfico dos valores previstos <i>versus</i> medidos para o modelo MIR_UVE-PLS_IndRefra_drv.....	105
Figura 41 - Modelos de previsão da massa específica <i>versus</i> RMSEP (•) e RMSEC (•), em g·cm <sup>-3</sup> , obtidos a partir do conjunto espectros de NIR, a linha vertical indica o erro experimental.....	107
Figura 42 - Modelos de previsão da massa específica <i>versus</i> RMSEP (•) e RMSEC (•), em g·cm <sup>-3</sup> , obtidos a partir do conjunto espectros de mid-IR, a linha vertical indica o erro experimental.....	108
Figura 43 - Modelos de previsão da massa específica <i>versus</i> RMSEP (•) e RMSEC (•), em g·cm <sup>-3</sup> , obtidos a partir do conjunto espectros de Raman, a linha vertical indica o erro experimental.....	109

Figura 44 - Soma da probabilidade do teste de Wilcoxon (SWTP) <i>versus</i> RMSEP ( $\text{g}\cdot\text{cm}^{-3}$ ) para todos os modelos de previsão da massa específica, a linha vertical indica o erro experimental .....	110
Figura 45 - RMSEC ( $\text{g}\cdot\text{cm}^{-3}$ ) e RMSEP ( $\text{g}\cdot\text{cm}^{-3}$ ) <i>versus</i> dados espectrais (NIR, mid-IR e Raman) para todos os modelos de previsão da Massa Específica, a linha vermelha horizontal representa o valor do erro experimental .....	111
Figura 46 - Elipse da região conjunta de confiança (EJCR) do conjunto de validação para os modelos de previsão da Massa Específica: (-) MIR_RF_MasEsp_drv; (- -) MIR_RF_MasEsp_mc; (- -) NIR_RF_MasEsp_snv; (- -) MIR_RF_MasEsp_snv. ....	112
Figura 47 - Gráfico dos valores previstos <i>versus</i> medidos para o modelo MIR_RF_MasEsp_drv .....	113
Figura 48 - Espectros brutos de infravermelho médio (mid-IR) das amostras do Banco 2. ....	117
Figura 49 - Regiões dos espectros de mid-IR ricas em informações, a linha horizontal tracejada (--) indica a região do sinal de ZnSe. ....	118
Figura 50 - Espectros de infravermelho médio (mid-IR) após a remoção da região espectral do sinal do cristal de ZnSe e após correção de linha de base. ....	119
Figura 51 - Espectros de mid-IR para cada um dos grupos de amostras que compõem o Banco 2: Biodiesel (B), diesel (D), misturas diesel-biodiesel (M) e misturas diesel-biodiesel-etanol (T). ....	119
Figura 52 - Espectros brutos de infravermelho próximo (NIR) das amostras que compõem o Banco 2: biodiesel (B), diesel (D), diesel-biodiesel (M) e diesel-biodiesel-etanol (T). ....	120
Figura 53 - Espectros de infravermelho próximo (NIR) das amostras do Banco 2 após a correção de linha de base, as linhas verticais indicam intervalo a ser mantido para o estudo. ....	121
Figura 54 - Regiões dos espectros de NIR ricas em informações. ....	122
Figura 55 - Espectros de infravermelho próximo (NIR) após a etapa de normalização. ....	122

Figura 56 - Espectros de NIR para cada um dos grupos de amostras que compõem o Banco 2: Biodiesel (B), diesel (D), misturas diesel-biodiesel (M) e misturas diesel-biodiesel-etanol (T). .....	123
Figura 57 - Espectros de Raman (RAM) das amostras do Banco 2 após as etapas de correção de linha de base e de normalização. ....	124
Figura 58 - Espectros de Raman para cada um dos grupos de amostras que compõem o Banco 2: Biodiesel (B), diesel (D), misturas diesel-biodiesel (M) e misturas diesel-biodiesel-etanol (T). ....	124
Figura 59 - Gráfico de autovalores <i>versus</i> número de componentes principais obtido pela PCA para os espectros NIR (Banco 2) .....	126
Figura 60 - Gráfico dos resíduos <i>versus</i> escores (Q resíduos <i>versus</i> Hotelling T2) obtido pela PCA para os espectros NIR do Banco 2. ....	127
Figura 61 - Gráfico da combinação de altos valores distância ortogonal e distância entre os escores obtido pelo ROBPCA para os dados de NIR do Banco 2 .....	127
Figura 62 - Espectros de NIR do Banco 2 antes e após a transformação de sinal espacial (tss) .....	128
Figura 63 - Gráfico da combinação de altos valores distância ortogonal e distância entre os escores obtido pelo ROBPCA para os dados de NIR do Banco 2 após a transformação de sinal espacial .....	128
Figura 64 - Gráfico de autovalores <i>versus</i> número de componentes principais obtido pela PCA para os espectros mid-IR do Banco 2 .....	129
Figura 65 - Gráfico dos resíduos <i>versus</i> escores (Q resíduos <i>versus</i> Hotelling T2) obtido pela PCA para os espectros mid-IR do Banco 2 .....	130
Figura 66 - Gráfico da combinação de altos valores distância ortogonal e distância entre os escores obtido pelo ROBPCA para os dados de mid-IR do Banco 2 .....	131
Figura 67 - Gráfico de autovalores <i>versus</i> número de componentes principais obtido pela PCA para os espectros Raman do Banco 2 .....	132
Figura 68 - Gráfico dos resíduos <i>versus</i> escores (Q resíduos <i>versus</i> Hotelling T2) obtido pela PCA para os espectros Raman do Banco 2 .....	132
Figura 69 - Gráfico da combinação de altos valores distância ortogonal e distância entre os escores obtido pelo ROBPCA para os dados de Raman do Banco 2 .....	133

Figura 70 - Fluxograma da divisão das amostras em conjunto de calibração e validação aplicando o algoritmo Kennard-Stone nos espectros de NIR, mid-IR e Raman .....	134
Figura 71 - Distribuição das amostras de calibração (+) e validação (•) no conjunto total de amostras do Banco 2, considerando os espectros NIR.....	135
Figura 72 - Intervalos espectrais do NIR (Banco 2) selecionados pelos algoritmos REP, BVE e VIP .....	137
Figura 73 - Intervalos espectrais do mid-IR (Banco 2) selecionados pelos algoritmos VIP, BVE e IPW .....	137
Figura 74 - Intervalos espectrais do Raman (Banco 2) selecionados pelos algoritmos BVE, GA e REP .....	138
Figura 75 - Modelos de previsão do ponto de entupimento de filtro a frio (PEFF) <i>versus</i> RMSEP (•) e RMSEC (•), em °C, obtidos a partir do conjunto espectros de NIR, a linha vertical indica o erro experimental .....	142
Figura 76 - Modelos de previsão do ponto de entupimento de filtro a frio (PEFF) <i>versus</i> RMSEP (•) e RMSEC (•), em °C, obtidos a partir do conjunto espectros de mid-IR, a linha vertical indica o erro experimental.....	143
Figura 77 - Modelos de previsão do ponto de entupimento de filtro a frio (PEFF) <i>versus</i> RMSEP (•) e RMSEC (•), em °C, obtidos a partir do conjunto espectros de Raman, a linha vertical indica o erro experimental .....	144
Figura 78 - Soma da probabilidade do teste de Wilcoxon (SWTP) <i>versus</i> RMSEP (°C) para todos os modelos de previsão do PEFF, a linha vertical indica o erro experimental .....	145
Figura 79 - RMSEP (°C) <i>versus</i> intervalos espectrais para todos os modelos de previsão do PEFF, a linha horizontal tracejada indica o erro experimental.....	146
Figura 80 - RMSEP (°C) <i>versus</i> métodos PLS, RF e SVM para todos os modelos de previsão do PEFF, a linha horizontal tracejada indica o erro experimental.....	147
Figura 81 - RMSEP (°C) <i>versus</i> intervalos utilizados, considerando apenas os modelos SVM de previsão do PEFF, a linha horizontal tracejada indica o erro experimental .....	147

Figura 82 - RMSEC (°C) e RMSEP (°C) <i>versus</i> dados espectrais (NIR, mid-IR e Raman) para todos os modelos de previsão do PEFF, a linha horizontal tracejada indica o erro experimental.....	148
Figura 83 - Elipse da região conjunta de confiança (EJCR) do conjunto de validação para os modelos de previsão do PEFF: (–) MIR_RC-SVM_PEFF_mc; (- -) MIR_BVE-SVM_PEFF_drv; (- -) MIR_IPW-SVM_PEFF_mc; (- -) MIR_RC-SVM_PEFF_snv. ....	149
Figura 84 - Gráfico dos valores previstos (°C) <i>versus</i> medidos (°C) para o modelo MIR_RC-SVM_PEFF.....	149
Figura 85 - Modelos de previsão da viscosidade (VISCO) <i>versus</i> RMSEP (•) e RMSEC (•), em mm <sup>2</sup> ·s <sup>-1</sup> , obtidos a partir do conjunto espectros de NIR, a linha vertical indica o erro experimental .....	152
Figura 86 - Modelos de previsão da viscosidade (VISCO) <i>versus</i> RMSEP (•) e RMSEC (•), em mm <sup>2</sup> ·s <sup>-1</sup> , obtidos a partir do conjunto espectros de mid-IR, , a linha vertical indica o erro experimental.....	153
Figura 87 - Modelos de previsão da viscosidade (VISCO) <i>versus</i> RMSEP (•) e RMSEC (•), em mm <sup>2</sup> ·s <sup>-1</sup> , obtidos a partir do conjunto espectros de Raman (RAM), a linha vertical indica o erro experimental .....	154
Figura 88 - Soma da probabilidade do teste de Wilcoxon (SWTP) <i>versus</i> RMSEP (mm <sup>2</sup> ·s <sup>-1</sup> ) para todos os modelos de previsão da Viscosidade, a linha vertical indica o erro experimental.....	155
Figura 89 - RMSEC (mm <sup>2</sup> ·s <sup>-1</sup> ) e RMSEP (mm <sup>2</sup> ·s <sup>-1</sup> ) <i>versus</i> dados espectrais (NIR, mid-IR e Raman) para todos os modelos de previsão da viscosidade, a linha horizontal indica o erro experimental .....	156
Figura 90 - Elipse da região conjunta de confiança (EJCR) do conjunto de validação para os modelos de previsão da viscosidade: (–) MIR_RF-VISCO_snv, (–) MIR_VIP-SVM_VISCO_snv e (- -) MIR_REP-SVM_VISCO_snv. ....	157
Figura 91 - Gráfico dos valores previstos (mm <sup>2</sup> ·s <sup>-1</sup> ) <i>versus</i> medidos (mm <sup>2</sup> ·s <sup>-1</sup> ) para o modelo MIR_RF_VISCO_snv.....	158



## LISTA DE TABELAS

Tabela 1 - Composição das misturas ternárias de biodiesel.....	61
Tabela 2 - Composição das misturas quaternárias de biodiesel.....	61
Tabela 3 - Composição das misturas binárias de biodiesel. ....	62
Tabela 4 - Composição das misturas ternárias de diesel-biodiesel-etanol anidro ..	64
Tabela 5 - Composição das misturas binárias de diesel-biodiesel.....	65
Tabela 6 - Resultados das medidas de massa específica dos cinco tipos de biodiesel puro comparado com os resultados da literatura.....	74
Tabela 7 - Resultados das medidas de índice de refração dos cinco tipos de biodiesel puro. ....	74
Tabela 8 - Lista das ferramentas, comandos, pacotes e respectivas métricas utilizados neste trabalho.....	75
Tabela 9 - Identificação das regiões espectrais (mid-IR) das amostras de biodiesel. ....	77
Tabela 10 - Identificação das regiões espectrais (NIR) das amostras de biodiesel. ....	78
Tabela 11- Comparação autovalores e variância explicada do modelo PCA a partir dos espectros de NIR pré-processados com centralização na média.....	81
Tabela 12 - Comparação autovalores e variância explicada do modelo PCA a partir dos espectros de mid-IR pré-processados com centralização na média (Banco1).....	84
Tabela 13 - Comparação autovalores e variância explicada do modelo PCA a partir dos espectros de Raman pré-processados com centralização na média (Banco 1).....	86
Tabela 14 - Amostras selecionadas pelo algoritmo Kennard-Stone para compor os conjuntos de calibração e validação para os espectros de NIR, mid-IR e Raman do Banco 1.....	91
Tabela 15 - Quantidade de variáveis selecionadas por cada método de seleção de variáveis para cada um dos sensores (NIR, mid-IR e RAM), para o índice de refração e massa específica do Banco 1.....	96

Tabela 16 - Parâmetros de mérito dos 10 melhores modelos PLS com seleção de variáveis e Florestas Aleatórias (RF) obtidos a partir dos espectros de NIR, mid-IR e Raman para a previsão do índice de refração.....	98
Tabela 17 - Parâmetros de mérito dos 10 melhores modelos PLS com seleção de variáveis e Florestas Aleatórias (RF) obtidos a partir dos espectros de NIR, mid-IR e Raman para a previsão da Massa Específica .....	106
Tabela 18 - Resultados das medidas de PEFF de algumas amostras comparado com os resultados da literatura (obtidos por outros autores). .....	115
Tabela 19 - Resultados das medidas de Viscosidade cinemática a 40°C (VISCO) de algumas amostras comparado com os resultados da literatura (obtidos por outros autores). .....	116
Tabela 20 - Comparação autovalores e variância explicada do modelo PCA a partir dos espectros de NIR (Banco 2) pré-processados com centralização na média.....	125
Tabela 21 - Comparação autovalores e variância explicada do modelo PCA a partir dos espectros de mid-IR pré-processados com centralização na média (Banco 2) .....	129
Tabela 22 - Comparação autovalores e variância explicada do modelo PCA a partir dos espectros de Raman pré-processados com centralização na média (Banco 2) .....	131
Tabela 23 - Amostras selecionadas pelo algoritmo Kennard-Stone para compor os conjuntos de calibração e validação para os espectros de NIR, mid-IR e Raman do Banco 2.....	135
Tabela 24 - Quantidade de variáveis selecionadas por cada método de seleção de variáveis para cada um dos sensores (NIR, mid-IR e RAM), para o ponto de entupimento de filtro a frio (PEFF) e viscosidade cinemática do Banco 2 .....	139
Tabela 25 - Parâmetros de mérito dos 10 modelos com menor valor de SWTP obtidos a partir dos espectros de NIR, mid-IR e Raman para a previsão do Ponto de Entupimento de Filtro a Frio (PEFF) .....	141
Tabela 26 - Parâmetros de mérito dos 10 modelos com menor valor de SWTP, SVM com seleção de variáveis e Florestas Aleatórias (RF) obtidos a partir dos espectros de NIR para a previsão da Viscosidade Cinemática (VISCO).....	151

## LISTA DE ABREVIATURAS E SIGLAS

ANN	Artificial Neural Networks (Redes Neurais Artificiais)
ANP	Agência Nacional do Petróleo, Gás Natural e Biocombustíveis
ATR	Reflectância total atenuada
BVE	Backward variable elimination (Eliminação de variáveis para trás)
EM	Embedded Methods (Métodos integrados)
enPLS	Ensemble PLS
FM	Filter methods (Métodos de filtros)
FTIR	Espectrometria no infravermelho médio com transformada de Fourier
FTIR-ATR	Espectrometria de infravermelho por Transformada de Fourier com Reflexão Total Atenuada
GA	Genetic Algorithm (Algoritmo Genético)
iPLS	Interval Partial Least Squares (Mínimos Quadrados Parciais por intervalos)
IPW	Iterative predictor weighting (Ponderação iterativa do preditor)
KS	Algoritmo Kennard-Stone
LW	Loadings Weights (Pesos das cargas)
MC	Mean Center (Centralização dos dados na média)
mid-IR	Medium-Infrared (Infravermelho Médio)
MSC	Multiplicative Signal Correction (Correção de Sinal Multiplicativo)
NIR	Near-Infrared (Infravermelho Próximo)
NMR	Nuclear Magnetic Resonance (Ressonância Magnética Nuclear)
NMR-1H	Ressonância Magnética Nuclear de hidrogênio
PC	Principal Component (Componente Principal)
PCA	Principal Component Analysis (Análise por Componentes Principais)
PCR	Principal Component Regression (Regressão por Componentes Principais)
PEFF	Ponto de Entupimento de Filtro a Frio
PLS	Partial Least Squares (Mínimos Quadrados Parciais)
PLSR	Partial Least Squares Regression (Regressão por Mínimos Quadrados Parciais)
RBF	Radial Basis Function (Função de base radial)
RC	Regression Coefficients (Coeficientes de regressão)
REP	Regularized elimination procedure (Procedimento regularizado de eliminação)
RF	Random Forests (Florestas Aleatórias)
ROBPCA	Robust Principal Component Analysis (Análise por Componentes Principais Robusto)
R <sup>2</sup>	Coeficiente de Determinação

$R^2_{Cal}$	Coeficiente de Determinação da Calibração
$R^2_{CV}$	Coeficiente de Determinação da Validação Cruzada
$R^2_{Pred}$	Coeficiente de Determinação da Previsão
$R_{Cal}$	Coeficiente de Correlação da Calibração
$R_{CV}$	Coeficiente de Correlação da Validação Cruzada
$R_{Pred}$	Coeficiente de Correlação da Previsão
RMSE	Root Mean Square Error (Raiz Quadrada do Erro Médio Quadrático)
RMSEC	Root Mean Square Error Calibration (Raiz Quadrada do Erro Médio Quadrático de Calibração)
RMSECV	Root Mean Square Error Cross-Validation (Raiz Quadrada do Erro Médio Quadrático com Validação Cruzada)
RMSEP	Root Mean Square Error Prediction (Raiz Quadrada do Erro Médio Quadrático de Previsão)
SEV	Erro padrão de validação
SNV	Standard Normal Variate (Variação Normal Padronizada)
SPA	Successive Projections Algorithm (Algoritmo das Projeções Sucessivas)
SVM	Support Vector Machine (Máquinas de Vetores Suporte)
SVM-R	Support Vector Machine Regression (Regressão por Máquinas de Vetores Suporte)
SWTP	Sum of Wilcoxon Test Probability (Soma da Probabilidade do Teste de Wilcoxon)
UVE	Uninformative variable elimination (Eliminação de variáveis não informativas)
VIP	Variable importance in projection (Importância da variável na projeção)
VL	Variável latente
WM	Wrapper methods

## SUMÁRIO

	<b>INTRODUÇÃO</b> .....	22
1	<b>REVISÃO DA LITERATURA</b> .....	25
2	<b>ESPECTROSCOPIA VIBRACIONAL</b> .....	28
3	<b>QUIMIOMETRIA</b> .....	30
3.1	<b>Pré-processamentos e/ou transformações dos dados</b> .....	31
3.1.1	<u>Alisamento pelo método de Savitzky-Golay</u> .....	31
3.1.2	<u>Centralização ou centragem dos dados na média</u> .....	32
3.1.3	<u>Correção de linha de base</u> .....	32
3.1.4	<u>Derivadas</u> .....	33
3.1.5	<u>Padronização normal de sinal</u> .....	34
3.1.6	<u>Normalização</u> .....	34
3.1.7	<u>Transformação de sinais espaciais</u> .....	35
3.2	<b>Análise de componentes principais (PCA)</b> .....	36
3.3	<b>Métodos de regressão</b> .....	38
3.3.1	<u>Regressão por mínimos quadrados parciais (PLS)</u> .....	38
3.3.2	<u>Regressão por máquinas de vetores de suporte (SVM)</u> .....	40
3.3.3	<u>Florestas aleatórias</u> .....	44
3.4	<b>Seleção de variáveis</b> .....	44
3.4.1	<u>Métodos de filtro</u> .....	46
3.4.1.1	Pesos das cargas (LW) .....	46
3.4.1.2	Coeficientes de regressão (RC) .....	46
3.4.1.3	Importância da variável na projeção (VIP).....	47
3.4.2	<u>Métodos de wrapper</u> .....	47
3.4.2.1	Algoritmo genético (GA) .....	48
3.4.2.2	Eliminação de variáveis não-informativas por PLS (UVE-PLS).....	49
3.4.2.3	Eliminação de variáveis para trás por PLS (BVE-PLS) .....	49
3.4.2.4	Ponderação iterativa do preditor por PLS (IPW-PLS) .....	50
3.4.2.5	Análise de permutação de janelas secundárias por PLS (SPA-PLS).....	50
3.4.2.6	Procedimento regularizado de eliminação (REP-PLS).....	50
3.4.2.7	Mínimos quadrados parciais por intervalos (iPLS) .....	50

3.4.3	<u>Ensemble PLS</u> .....	51
3.5	<b>Algoritmo de seleção de amostras</b> .....	51
3.6	<b>Parâmetros de mérito</b> .....	52
3.7	<b>Soma das probabilidades do teste de Wilcoxon (SWTP)</b> .....	54
3.8	<b>Elipse da região conjunta de confiança</b> .....	55
3.9	<b>Resumo do estudo quimiométrico</b> .....	56
4	<b>OBJETIVO</b> .....	58
4.1	<b>Objetivos específicos</b> .....	58
5	<b>EXPERIMENTAL</b> .....	59
5.1	<b>Produção do biodiesel</b> .....	59
5.2	<b>BANCO 1: Misturas de biodiesel</b> .....	60
5.3	<b>BANCO 2: Misturas de biodiesel, diesel-biodiesel e diesel-biodiesel-etanol</b> .....	63
5.4	<b>Medição das propriedades</b> .....	66
5.4.1	<u>Índice de refração e massa específica</u> .....	66
5.4.2	<u>Ponto de entupimento de filtro a frio</u> .....	67
5.4.3	<u>Viscosidade cinemática a 40 °C</u> .....	68
5.5	<b>Medição dos espectros de infravermelho</b> .....	69
5.5.1	<u>Espectros de infravermelho médio</u> .....	69
5.5.2	<u>Espectros de infravermelho próximo</u> .....	70
5.5.3	<u>Espectros de Raman</u> .....	71
6	<b>RESULTADOS E DISCUSSÃO</b> .....	73
6.1	<b>BANCO 1: Previsão de massa específica e índice de refração em amostras de biodiesel</b> .....	73
6.1.1	<u>Caracterização do biodiesel</u> .....	73
6.1.2	<u>Análise dos dados</u> .....	74
6.1.3	<u>Análise de componentes principais (PCA)</u> .....	80
6.1.4	<u>Divisão do conjunto de amostras da calibração e validação</u> .....	89
6.1.5	<u>Modelos de regressão</u> .....	91
6.1.6	<u>Modelos de previsão para o índice de refração</u> .....	97
6.1.7	<u>Modelos de previsão para massa específica</u> .....	105
6.2	<b>BANCO 2: Misturas de biodiesel-biodiesel, diesel-biodiesel e diesel-biodiesel-etanol</b> .....	113
6.2.1	<u>Caracterização do diesel e das misturas</u> .....	114

6.2.2	<u>Análise dos dados</u> .....	116
6.2.3	<u>Análise de componentes principais (PCA)</u> .....	125
6.2.4	<u>Divisão do conjunto de amostras da calibração e validação</u> .....	133
6.2.5	<u>Modelos de regressão</u> .....	135
6.2.6	<u>Modelos de previsão para o ponto de entupimento de filtro a frio</u> .....	140
6.2.7	<u>Modelos de previsão para a viscosidade cinemática</u> .....	150
	<b>CONCLUSÃO</b> .....	159
	<b>REFERÊNCIAS</b> .....	162
	<b>ANEXO A</b> - Informação da composição e nomeação das amostras de biodiesel nos conjuntos espectrais de NIR, Mid-IR e Raman para o BANCO 1 .....	174
	<b>ANEXO B</b> - Informação da composição e nomeação das amostras de biodiesel nos conjuntos espectrais de NIR, Mid-IR e Raman para o BANCO 2 .....	180
	<b>ANEXO C</b> - Parâmetros de mérito dos modelos PLS com seleção de variáveis e Florestas Aleatórias (RF) obtidos a partir dos espectros de NIR, mid-IR e RAM para a previsão do índice de refração .....	187
	<b>ANEXO D</b> - Parâmetros de mérito dos modelos PLS com seleção de variáveis e Florestas Aleatórias (RF) obtidos a partir dos espectros de NIR, mid-IR e RAM para a previsão da massa específica .....	193
	<b>ANEXO E</b> - Parâmetros de mérito dos modelos PLS com seleção de variáveis e Florestas Aleatórias (RF) obtidos a partir dos espectros de NIR, mid-IR e RAM para a previsão do PEFF .....	199
	<b>ANEXO F</b> - Parâmetros de mérito dos modelos PLS com seleção de variáveis e Florestas Aleatórias (RF) obtidos a partir dos espectros de NIR, mid-IR e RAM para a previsão da viscosidade .....	211

## INTRODUÇÃO

A utilização dos recursos naturais foi intensificada na Revolução Industrial (século XVIII), quando houve a necessidade de atender uma gama de novas necessidades humanas, à medida que novos produtos e serviços eram criados e mais matéria-prima era demandada para alimentar os processos industriais.

Com o início da exploração não sustentável dos recursos naturais, os ecossistemas passaram a receber resíduos em quantidades cada vez maiores, inclusive com lançamento de novas substâncias da indústria química que passaram a estarem presentes no ar, solo, rios e mares, causando desestabilidade no equilíbrio natural. Por mais que existam técnicas eficientes para a redução e controle da poluição, a situação global é preocupante, florestas ainda são derrubadas, solos e rios poluídos, as cidades não param de crescer e com elas, a queima de combustíveis fósseis intensifica a poluição atmosférica interferindo no clima global e em inúmeras catástrofes ambientais.

Sem dúvida, o maior combustível responsável por este processo é o petróleo. O petróleo e seus derivados trouxeram uma série de benefícios para o desenvolvimento da população mundial, seja na geração de energia, transportes, agricultura, calefação ou nas necessidades corriqueiras do dia a dia. Assim, o petróleo e seus derivados favoreceram a industrialização, o desenvolvimento de mercados, o giro de capital e um grande desenvolvimento urbano e econômico para os países dotados de tecnologia de exploração e produção de petróleo.

Na década de 30, os problemas de abastecimento de petróleo no mercado mundial aliados à tentativa de países europeus de desenvolverem alternativas energéticas para as suas colônias levaram à busca por soluções viáveis para a substituição do combustível fóssil (SHAY, 1993). Hoje, o crescimento da demanda por produtos de petróleo e a proteção do meio ambiente, devem estar em equilíbrio, para mantermos a sustentabilidade e diminuir o impacto das atividades humanas. Neste contexto podemos citar o desenvolvimento e a utilização de combustíveis renováveis, produzidos com diversidade de matérias-primas, tais como o etanol e o biodiesel.

O biodiesel vem sendo empregado como uma fonte de energia renovável que contribui para a diminuição da demanda por diesel mineral, que é importado pelo



Brasil. Segundo Fangrui *et al.* (1999), o biodiesel apresenta algumas vantagens quando comparado ao diesel por ser biodegradável, praticamente isento de enxofre, e por sua combustão ser mais eficiente que a do diesel. Alguns estudos indicaram que o uso do biodiesel, além de reduzir a emissão de gases associados ao efeito estufa (CO<sub>2</sub> e CO), ocasiona uma redução de particulados e óxidos de enxofre (FANGRUI *et al.*, 1999; CANDEIA, 2008; SILVA e FREITAS, 2008; SUAREZ *et al.*, 2009; LÔBO *et al.*, 2009).

A Resolução N° 14 de 2012 da ANP (Agência Nacional de Petróleo, Gás Natural e Biocombustíveis) define o biodiesel como um combustível renovável, biodegradável e para uso em motores a combustão interna com ignição por compressão. O biodiesel tem a sua composição baseada em ésteres alquílicos de ácidos graxos, com cadeia longa, derivados de óleos vegetais ou de gorduras animais (ANP, 2012).

O etanol começou a ser utilizado como combustível no Brasil um ano após a queda do preço do açúcar em 1974, quando foi implantado o Programa Pró-Álcool para garantir o fornecimento de energia e também para apoiar a indústria açucareira com a diversificação da produção. A partir da implantação desse programa, foram construídas destilarias, as quais transformaram o excesso da produção de cana-de-açúcar em etanol anidro, usado como aditivo na gasolina, sem a necessidade de nenhuma modificação nos motores dos veículos. No início dos anos 1980, a produção dos motores de veículos brasileiros foi adaptada ao uso do etanol anidro (BORGES *et al.*, 1984). O etanol pode ser produzido de múltiplos tipos de biomassa, mas a cana-de-açúcar continua sendo a mais empregada atualmente.

Independente da natureza do combustível e do biocombustível, eles devem apresentar atributos de qualidade que devem ser monitorados para que se possa produzi-los e distribuí-los com a qualidade especificada. Deste modo, faz-se necessário o investimento de uma metodologia eficaz para determinar as propriedades físico-químicas estudadas de forma rápida e não destrutível, sendo possível utilizá-la para controle da qualidade no monitoramento on/in/at-line dos combustíveis em ambiente industrial, utilizando ferramentas quimiométricas.

De acordo com a literatura, os resultados obtidos utilizando as espectroscopias de infravermelho próximo (NIR) e médio (mid-IR) associadas às ferramentas quimiométricas para determinação de propriedades de misturas de biodiesel ou diesel-biodiesel (PIMENTEL *et al.*, 2006; OLIVEIRA *et al.*, 2007; BAPTISTA *et al.*,

2008; LIRA *et al.*, 2010; LUNA *et al.*, 2017; CUNHA *et al.*, 2017) apresentaram-se bastante promissoras. No entanto, a aplicação da espectroscopia Raman pode trazer informações químicas diferentes aos modelos, a qual pode ser útil na construção de instrumentos de baixo custo para controle de qualidade combustíveis e biocombustíveis (OLIVEIRA *et al.*, 2007; DEARING *et al.*, 2011; MARINOVIC *et al.*, 2012). A utilização da espectroscopia Raman é interessante principalmente pela praticidade das medidas para obtenção dos espectros.

Neste contexto, o presente trabalho tem como proposta apresentar modelos de previsão de propriedades de combustíveis (diesel) e biocombustíveis (etanol e biodiesel), puros e de suas misturas, baseados em técnicas espectroscópicas (NIR, mid-IR e Raman) acopladas com ferramentas quimiométricas, onde se inclui a utilização de técnicas de seleção de variáveis dos dados e aplicação de métodos não-lineares, como máquina de vetores suporte (SVM) e florestas aleatórias, para a previsão das propriedades com comportamento não linear. Estas técnicas instrumentais são não destrutivas, apresentam versatilidade e são capazes de fazer a previsão destas propriedades quando acopladas com técnicas quimiométricas.

## 1 REVISÃO DA LITERATURA

A literatura dispõe de estudos, como os de Pimentel *et al.* (2006), Oliveira *et al.* (2007), Baptista *et al.* (2008), Monteiro *et al.* (2009), Lira *et al.* (2010(a)) e Cunha *et al.* (2017), que utilizaram ferramentas quimiométricas em seus trabalhos para prever propriedades de biodiesel e misturas diesel-biodiesel.

Por exemplo, Pimentel *et al.* (2006) avaliaram a aplicação de modelos de calibração multivariada por mínimos quadrados parciais (PLS) baseados em espectros mid-IR (4000 a 650  $\text{cm}^{-1}$ ; refletância total atenuada (ATR) e NIR (12000 a 4000  $\text{cm}^{-1}$ ; caminho óptico = 1,0 cm) para prever o teor de biodiesel em misturas com óleo diesel considerando a presença de óleos vegetais. Os valores da raiz quadrada do erro médio quadrático da previsão (RMSEP) obtidos foram de 0,18% v/v (NIR) e 0,25% v/v (mid-IR). De acordo com o teste F (com nível de confiança de 95%), não houve diferença estatisticamente significativa entre os modelos construídos empregando as duas regiões espectrais.

Na comparação entre os modelos, é importante salientar que o teste F não deve ser empregado sem antes verificar se os resíduos apresentam distribuição normal. Em resumo, se os resíduos apresentam distribuição normal, os resíduos ao quadrado seguem a distribuição qui-quadrado e o teste F pode ser aplicado na comparação entre os modelos (MONTGOMERY e RUNGER, 2009, p. 278). Por outro lado, o teste de Wilcoxon por ser um teste não paramétrico é o mais indicado para comparação dos modelos, e será abordado ao longo deste estudo.

Oliveira *et al.* (2007) estudaram a adulteração de misturas de biodiesel com óleos vegetais a partir das espectroscopias vibracionais NIR (caminho óptico = 2,0 cm) e Raman (caminho óptico = 0,5 cm) com transformada de Fourier. Os autores analisaram o desempenho de modelos de calibração multivariada por PLS, regressão por componentes principais (PCR) e Redes Neurais Artificiais (ANN). Os resultados obtidos indicaram que entre os modelos de calibração estudados, aqueles que empregaram ANN e FT-Raman apresentaram uma melhor exatidão para amostras externas de validação.

Baptista *et al.* (2008) avaliaram a utilização de espectroscopia de infravermelho próximo (NIR) para a previsão de propriedades do biodiesel, como o índice de iodo, o

ponto de entupimento do filtro a frio (PEFF), a densidade a 15°C e a viscosidade cinemática à temperatura de 40°C. A priori, os autores utilizaram a análise dos componentes principais (PCA) para a realização de uma análise exploratória dos dados, e em seguida aplicaram a regressão por PLS para a previsão das propriedades. Baptista *et al.* (2008) prepararam 71 amostras de biodiesel produzidos a partir de misturas de óleo de soja, canola, palma e os óleos usados em frituras, sendo que 49 amostras foram destinadas ao conjunto de calibração e 22 amostras para validação. O modelo de previsão do PEFF do biodiesel foi construído utilizando o PLS com duas variáveis latentes, obtendo coeficiente de determinação da previsão ( $R^2_{\text{Pred}}$ ) de 0,951 e RMSEP 1,0°C para esta propriedade. O melhor modelo de previsão para a densidade a 15°C foi obtido por PLS e fez uso de seis variáveis latentes, com  $R^2_{\text{Pred}}$  de 0,999 e RMSEP de 0,9 kg·m<sup>-3</sup>.

Monteiro *et al.* (2009) tiveram como objetivo de seu trabalho a avaliação da qualidade das misturas de diesel-biodiesel e a previsão da concentração de óleo e biodiesel no diesel utilizando espectrometria de ressonância magnética nuclear de hidrogênio (NMR-1H). Os autores analisaram 46 amostras de mistura de óleo-diesel, diesel-biodiesel e óleo-diesel-biodiesel, sendo que 36 amostras pertenciam ao conjunto de calibração, restando as outras 10 amostras para o conjunto de validação. A análise exploratória dos dados foi realizada com aplicação da PCA e as amostras atípicas foram retiradas do conjunto de dados. Os modelos de calibração multivariada PCR e PLS foram construídos para previsão do teor de óleo biodiesel no diesel mineral. Os modelos obtidos por PCR e PLS foram mais eficientes para a previsão do teor de biodiesel em misturas de diesel-biodiesel, principalmente para a previsão de concentrações superiores a 2%. Para a determinação do teor de biodiesel o modelo PLS obteve coeficiente de correlação de validação ( $R_{cv}$ ) igual a 0,98 e erro padrão de validação (SEV) igual a 0,82 (% v/v).

Lira *et al.* (2010(a)) prepararam misturas utilizando ésteres metílicos de soja, óleo de rícino, óleo de sementes de algodão, óleo de canola e óleo de girassol, assim como amostras de diesel de diferentes regiões do Brasil. Foram preparadas um total de 161 amostras para o estudo da previsão de densidade, teor de enxofre e temperaturas de destilação, com suas respectivas medições espectrais no NIR e mid-IR. Os autores dividiram os dados em dois conjuntos, de calibração (80% das amostras) e de validação (20% das amostras). Os melhores modelos foram obtidos utilizando a regressão por PLS e com o emprego dos pré-processamentos da primeira

derivada após a suavização dos sinais espectrais. Para o modelo de previsão da densidade utilizando os dados espectrais de NIR, foram utilizadas oito variáveis latentes para geração do modelo PLS, com valores de  $R^2_{\text{Pred}}$  de 0,99 e RMSEP de  $0,56 \text{ kg}\cdot\text{m}^{-3}$ . Para realizar a previsão da densidade utilizando os dados espectrais de mid-IR, foram necessárias oito variáveis latentes para construir o modelo PLS, obtendo  $R^2_{\text{Pred}}$  igual a 0,97 e RMSEP de  $1,60 \text{ kg}\cdot\text{m}^{-3}$ .

Cunha *et al.* (2017) utilizaram regressão por PLS e SVM para modelar a relação entre dados espectroscópicos de FT-IR médio e a densidade, índice de refração e PEFF de amostras puras de biodiesel e suas misturas. Cento e quarenta e oito amostras foram preparadas usando biodiesel de diferentes fontes (canola, girassol, milho, soja e região sul do Brasil), sendo que 100 dessas amostras foram mantidas no conjunto de calibração e quarenta e oito amostras foram utilizadas para compor o conjunto de validação externa. O melhor modelo de previsão do ponto de entupimento de filtro a frio foi obtido pelo método de regressão SVM, o qual apresentou RMSEP igual a  $0,6^\circ\text{C}$ . O modelo PLS resultou na melhor previsão da densidade e índice de refração com valores de RMSEP iguais a  $0,2 \text{ kg}\cdot\text{m}^{-3}$  e 0,0001, respectivamente.

Com base no levantamento de informações e descrição do panorama da literatura, foi observado que em vários estudos as técnicas espectroscópicas mais utilizadas nos trabalhos disponíveis na literatura envolvendo quimiometria têm sido espectrometria no UV-visível, infravermelho (NIR e mid-IR), Raman, NMR, fluorescência, espectrometria de massa e cromatografia. Além disso, independente da natureza das amostras e dos sensores, esta área de estudo representa incontáveis possibilidades de combinações entre sensores e ferramentas quimiométricas, trazendo benefícios a modelagem. No entanto, vale ressaltar a importância de se aplicar e esgotar as técnicas convencionais mais simples (PLS, por exemplo) antes de se iniciar estudos com as ferramentas mais robustas, como técnicas não-lineares (Florestas Aleatórias e SVM).

## 2 ESPECTROSCOPIA VIBRACIONAL

A espectroscopia Raman é uma técnica analítica baseada na detecção de luz espalhada inelasticamente e está relacionada com as transições vibracionais e rotacionais das moléculas. Esta técnica é comumente utilizada para identificar substâncias desconhecidas, e pode ser considerada uma técnica complementar à espectroscopia no infravermelho, pois fornece informações complementares com relação a amostra analisada (RAMAN e KRISHNAN, 1928; DRAGO, 1992; SALA, 1996).

A espectroscopia Raman possui resultados superiores a espectroscopia infravermelho quando as amostras são aquosas, uma vez que no Raman, a água é um fraco espalhador e não causa interferências significativas nas bandas de deslocamento. Na espectroscopia de infravermelho, amostras aquosas tendem a apresentar absorções intensas em quase toda a faixa de interesse do infravermelho médio e a encobrir regiões contendo informações químicas de interesse (LOPES, 2005). Além disso, esta técnica permite o estudo de amostras de quase qualquer estado físico (líquidos, pós, pastas, filmes, e outros) com alguma ou sem nenhuma preparação da amostra.

De acordo com Freeman *et al.* (1974), na espectrometria Raman ao se incidir uma radiação monocromática na superfície de uma amostra, esta pode ser espalhada elasticamente (espalhamento Rayleigh) ou ser espalhada inelasticamente (espalhamento Raman). Os espalhamentos Raman são mensurados a partir da linha de excitação, que arbitrariamente é considerada como zero numa escala relativa de frequência. Independente da frequência da linha do laser usado como fonte de excitação, o espectrômetro irá medir o mesmo deslocamento.

Os fundamentos da espectroscopia no infravermelho se baseiam no fato de que as ligações químicas das substâncias possuem frequências de vibração específicas, que equivalem aos níveis de energia da molécula, conhecidos como níveis vibracionais. Essas frequências dependem de algumas características da molécula, tais como a forma da superfície de energia potencial, da geometria, das massas dos átomos e ocasionalmente do acoplamento vibrônico. Segundo Barbosa (2013), a radiação infravermelha promove uma alteração nos modos rotacionais e

vibracionais das moléculas. Assim, para que ocorra a vibração da ligação química e esta seja visível no espectro de infravermelho, a molécula precisa possuir uma variação no momento de dipolo, durante seu movimento rotacional ou vibracional (estiramentos e deformações).

De acordo com Costa (2000), a radiação NIR apresenta um nível baixo de energia que o mid-IR e a sua interação com a matéria produz apenas transições e excitação dentro do mesmo estado fundamental de vibração. Além disso, as moléculas diatômicas homonucleares não são detectáveis num espectro de NIR, enquanto compostos com ligações químicas heteronucleares (C-H, N-H, O-H e S-H) são, em geral, facilmente detectados na região espectral NIR.

As bandas no NIR são consideradas mais fracas que as vibrações fundamentais no mid-IR. No entanto, esta característica da absorção de NIR pode ser encarada como uma vantagem analítica na análise de algumas amostras, visto que facilita uma análise direta de matrizes com absorções com alta dispersão de luz (ARNOLD, 2002).

A região do mid-IR compreende a faixa de 4000 a 200  $\text{cm}^{-1}$  do espectro eletromagnético. Nessa região, os espectros mostram picos bem definidos correspondentes às transições fundamentais. As bandas observadas nesta região são intensas e os picos com valores de absorvância mais altos, o que favorece a interpretação dos espectros. A faixa espectral do mid-IR conhecida como região das impressões digitais (1200 a 600  $\text{cm}^{-1}$ ) é bastante rica de informações e portanto muito útil na identificação de compostos, já que pequenas diferenças na estrutura e na constituição das moléculas ocasionam em significativas mudanças no perfil e na distribuição dos picos de absorção. Assim sendo é provável identificar, por exemplo, uma substância orgânica comparando-se seu espectro com bancos de dados pré-existentes (PASQUINI, 2003; SKOOG *et al.*, 2009).

Atualmente, as técnicas espectrais tem sido muito aplicada no controle de processos industriais. Isso se deve principalmente pelo avanço nas análises instrumentais, o que permite a aquisição rápida e fácil de um grande número de dados de uma mesma amostra. No entanto, conforme observado por Costa Filho *et al.*, (2002), com o aumento do volume dos dados, o tratamento dos mesmos passou a exigir modelos mais complexos. Sendo assim, o uso de ferramentas quimiométricas se tornou imprescindível à análise de dados químicos de natureza multivariada (ESBENSEN, 2002).

### 3 QUIMIOMETRIA

A estatística multivariada pode ser utilizada como ferramenta de análise de dados em diversas áreas, como na economia (econometria), arqueologia (arqueometria), psicologia (psicometria), biologia (biometria) e química (quimiometria), entre outras.

A Quimiometria é uma ferramenta destinada à análise de dados químicos de natureza multivariada. Como descrito por Kowalski (1978), a quimiometria é a ferramenta que usa métodos matemáticos e estatísticos para planejar ou otimizar procedimentos experimentais e extrair o máximo de informação química relevante por meio da análise dos dados.

O início da história da quimiometria pode ser contado a partir da criação da palavra *Kemometri* (sueco) pelo Professor Svante Wold (UMEA University, Suécia) em 1971. Um ano depois, o Professor Svante Wold criou o primeiro grupo de quimiometria, chamado *Research Group of Chemometrics*. Grupo este, que escreveu o primeiro artigo com o termo *chemometrics* em 1973. A primeira definição do termo quimiometria foi mencionada por Wold em 1974: “A arte de extrair a informação química relevante de dados produzidos em experimentos químicos é chamada de quimiometria em analogia a biometria e econometria” (Mundo da Quimiometria, 2010; FERREIRA, 2015). Ainda em 1974, a Sociedade Internacional de Quimiometria foi fundada pelo Prof. Svante Wold (UMEA University) e Bruce R. Kowalski (University of Washington).

Em 1975, o segundo artigo contendo o termo *chemometrics* em seu título é publicado pelo Grupo de Pesquisa em Quimiometria do Prof. Bruce R. Kowalski, este artigo definiu a quimiometria como uma nova disciplina da Química. O primeiro Simpósio de Quimiometria, ‘*Chemometrics: Theory and Applications*’, ocorreu em 1976 na cidade de São Francisco (EUA). Em 1977, como resultado do primeiro Simpósio de quimiometria foi publicado o primeiro livro sobre quimiometria. O primeiro curso de quimiometria no Brasil foi ministrado pelo Prof. Kowalski, no Instituto de Química da Unicamp em 1980. Finalmente, em 1987 foram criadas as primeiras revistas dedicadas à área de quimiometria, *Chemometrics and Intelligent Laboratory Systems* e *Journal of Chemometrics* (Mundo da Quimiometria, 2010; FERREIRA,



2015). Desde então inúmeros estudos com abordagem quimiométrica têm sido desenvolvidos e aprimorados no Brasil e no mundo.

### 3.1 Pré-processamentos e/ou transformações dos dados

Segundo Ferreira (2015), o pré-processamento ou transformação dos dados deve ser realizado para diminuir variações indesejadas adquiridas durante a obtenção dos dados. Menezes *et al.* (2009), Zhao *et al.* (2010) e Ferreira (2015) descreveram que espectros podem indicar variações sistemáticas que não são explicadas pela concentração do analito, como por exemplo, espalhamento de luz e desvios da linha de base. Desta forma, o pré-tratamento se faz necessário para provocar melhorias na interpretação e simplificar o modelo, fazendo-o muitas vezes mais robusto e confiável contra essas variações indesejáveis.

#### 3.1.1 Alisamento pelo método de Savitzky-Golay

O filtro de Savitzky-Golay é um tipo de filtro descrito pela primeira vez em 1964 por Abraham Savitzky e Marcel J. E. Golay, sendo concebido para analisar os picos de absorção no espectro e remover os ruídos, mantendo detalhes importantes do espectro (SAVITZKY e GOLAY, 1964). Os espectros medidos geralmente podem apresentar ruídos aleatórios, tanto no início quanto no fim dos espectros. Para isto, aplica-se o conceito de média ponderada, no qual os pesos são obtidos por ajuste de um polinômio de grau  $n$ .

O método executa essencialmente uma regressão polinomial local, de grau  $n$ , relativa às respostas da primeira janela, de pelo menos  $(2m + 1)$  pontos, que são tratados como sendo igualmente espaçados na série  $(k = -m, -m + 1, \dots, -1, 0, 1, \dots, m - 1, m)$ , para determinar o valor do alisamento para cada ponto. O valor de  $k = 0$  representa o centro e a resposta no ponto central da janela é substituída pelo valor do polinômio ajustado. A janela se move por todo o espectro e o valor do polinômio é calculado no novo centro, sendo repetido inúmeras vezes até que todas as janelas tenham sido alisadas (MADDEN, 1978; BRERETON, 2003; FERREIRA, 2015). O polinômio de grau  $n$  a ser ajustado é apresentado na equação 1:

$$x_k = a_0 + a_1k + a_2k^2 + \dots + a_nk^n \quad (1)$$

onde  $m$  é o número de pontos da janela,  $n$  é o grau do polinômio ajustado e  $k = -m, \dots, 0, \dots, m$ .

A principal vantagem da suavização é que ela tende a conservar as características de distribuição, tais como larguras, máximos e mínimos relativos, os quais são geralmente "achatados" por outras técnicas de alisamento (MADDEN, 1978; BRERETON, 2003). No entanto é preciso ter compromisso entre a redução do ruído e a distorção do sinal, definidos principalmente ao grau do polinômio e a escolha do tamanho da janela (FERREIRA, 2015).

### 3.1.2 Centralização ou centragem dos dados na média

Para dados espectrais, a centralização dos dados na média é um dos pré-processamentos mais utilizados, e pode ser combinada com outros tipos de pré-processamentos.

Centralizar os dados na média é gerar um novo conjunto de dados ( $x_{ij(cm)}$ ), que exhibe o resultado do cálculo da diferença de cada dado ( $x_{ij}$ ) em relação à média ( $\bar{x}_j$ ) do conjunto original, conforme equação 2 (NAES *et al.*, 2002; FERREIRA, 2015). Esta operação de centralização da média remove a intensidade absoluta da informação de cada variável, tornando-se possível focar-se nas variações da resposta. A finalidade da remoção da média é eliminar dos dados o valor da intensidade de cada variável revelando flutuações dos dados em torno do valor médio (NAES *et al.*, 2002).

$$x_{ij(cm)} = x_{ij} - \bar{x}_j \quad (2)$$

### 3.1.3 Correção de linha de base

Antes de aplicar qualquer um dos métodos quimiométricos ou de utilizar outros pré-processamentos, é importante avaliar se há o deslocamento de linha de base, especialmente quando se utiliza instrumentos mais antigos (BRERETON, 2003). A ocorrência do deslocamento da linha de base é muito comum em cromatografia gasosa, em ressonância magnética nuclear e também na espectroscopia (FERREIRA, 2015).

Nas abordagens quimiométricas, a variação acima de uma linha de base é enxergada pelos métodos e por esse motivo pode influenciar a modelagem. Portanto,

se a correção de linha de base não for realizada, ruídos indesejáveis vão dificultar a análise dos espectros e a obtenção de modelos robustos. Segundo Brereton (2003), o ajuste da linha de base pode envolver simplesmente a média, ou então um ajuste linear ou polinomial, de modo a obter um modelo matemático simples, que em seguida que irá subtrair a linha de base a partir de toda a região de interesse (GAYDOU *et al.*, 2010).

Depois de corrigida a linha de base, um sinal com um pico bem definido é obtido, sem que ocorra a redução do número das variáveis originais (TEÓFILO, 2013). Esta correção é aplicada conforme demonstrado na equação 3.

$$f(x) = x - \min(X) \quad (3)$$

onde  $x$  é uma variável e  $X$  são todas as variáveis selecionadas para determinar a amostra.

#### 3.1.4 Derivadas

Para Ferreira (2015), o uso de derivadas, costuma ser o pré-processamento mais adequado para correções de linha de base espectrais, sendo que os cálculos não envolve a aplicação de funções paralelas e também não provoca um grande aumento da variância nos dados. No entanto, uma desvantagem do método é o decréscimo indesejável da razão sinal/ruído (S/R), mas se a razão se mantiver alta não deve ser considerada um problema.

O método de Savitsky-Golay, apresentado anteriormente no item 4.1.1, é o método mais utilizado para o cálculo de derivadas de um sinal. Como visto anteriormente, o método é um filtro de média móvel no qual ajusta-se um polinômio de grau  $n$  a todos os  $(2m + 1)$  pontos da janela móvel. Na equação 4, esta metodologia é aplicada para calcular a primeira derivada de  $x$  em relação a  $k$ , obtendo-se a derivada diretamente da expressão analítica do polinômio (FERREIRA, 2015).

$$\frac{dx}{dk} = a_1 + 2a_2k + \dots + na_nk^{n-1} \quad (4)$$

onde  $m$  é o número de pontos da janela,  $n$  é o grau do polinômio ajustado e  $k = -m, \dots, 0, \dots, m$ .

### 3.1.5 Padronização normal de sinal

A padronização normal de sinal (*Standard Normal Variate*, SNV) é utilizada em muitas aplicações espectrométricas, principalmente na remoção de interferências de espalhamento e de tamanho de partícula sólida (FERREIRA, 2015). Este pré-processamento realiza um ajuste aditivo e multiplicativo tal como a correção multiplicativa de sinal (MSC), embora os os parâmetros  $a_i$  (para o SNV,  $\bar{x}_i$ ) e  $b_i$  (para o SNV,  $s_i$ ) sejam determinados de modo diferente. No método SNV,  $\bar{x}_i$  é o valor médio das intensidades do  $i$ -ésimo espectro e o desvio padrão dos valores de todas as variáveis é  $s_i$ , como apresentado nas equações 5, 6 e 7.

$$\begin{bmatrix} \vdots \\ X_{i\ snv} \\ \vdots \\ \vdots \end{bmatrix} = \frac{1}{s_i} \left( \begin{bmatrix} \vdots \\ X_i \\ \vdots \\ \vdots \end{bmatrix} - \begin{bmatrix} \bar{x}_i \\ \vdots \\ \bar{x}_i \end{bmatrix} \right) \quad (5)$$

$$\bar{x}_i = \frac{1}{J} \sum_{j=1}^J x_{ij} \quad (6)$$

$$s_i = \sqrt{\sum_{j=1}^J (x_{ij} - \bar{x}_i)^2} \quad (7)$$

### 3.1.6 Normalização

O propósito da normalização dos dados originais é igualar a magnitude de cada observação (amostras), removendo assim as informações de distância de cada amostra com relação à origem, mas ao mesmo tempo preservando a direção. Ainda segundo Ferreira (2015), esse tratamento pode ser aplicado quando se quer reter a informação que qualitativamente difere uma amostra das outras e remover toda a informação que poderia diferenciar duas amostras de composição idênticas, mas com concentrações distintas.

Na normalização, os valores de cada uma das variáveis de uma dada amostra  $i$  são divididos por um fator de normalização, podendo ser pela norma  $\| X_i \|$  dessa

amostra, por exemplo. Desta forma, todas as amostras passam a apresentar uma escala fixa. Na equação 8, a expressão da normalização é apresentado para cada elemento de uma linha da matriz de dados (FERREIRA, 2015), com  $j = 1, 2, \dots, J$ .

$$x_{ij(norm)} = \frac{x_{ij}}{\|X_i\|} \quad (8)$$

Segundo Ferreira (2015), as normas mais utilizadas são: a norma *sup* ou ( $l_\infty$ ), norma um ( $l_1$ ) e a norma Euclideana ( $l_2$ ), apresentadas nas equações 9, 10 e 11, respectivamente.

$$\|X_i\|_\infty = \max_{1 \leq j \leq J} |x_{ij}| \quad (9)$$

$$\|X_i\|_1 = \sum_{j=1}^J |x_{ij}| \quad (10)$$

$$\|X_i\|_2 = \sqrt{\sum_{j=1}^J x_{ij}^2} \quad (11)$$

### 3.1.7 Transformação de sinais espaciais

Segundo Serneels, Nolf e Espen (2006) os estimadores baseados em sinais espaciais apresentam simples implementação, uma vez que são baseados em uma transformação dos dados em seus sinais espaciais, a partir dos quais o estimador clássico é calculado. O sinal espacial é uma extensão multivariada do conceito de sinal.

A função de sinal espacial, denotada por  $sgn(\mathbf{w})$ , é definida pela equação 12:

$$sgn(\mathbf{w}) = \begin{cases} \mathbf{w}/\|\mathbf{w}\| & \text{if } \mathbf{w} \neq 0 \\ \mathbf{0} & \text{if } \mathbf{w} = 0 \end{cases} \quad (12)$$

onde  $\|\mathbf{w}\|$  representa a norma euclidiana do argumento.

Geometricamente, a função de sinal espacial é uma projeção de qualquer ponto  $w$  na direção da origem em uma esfera unitária. Exceto pela origem, todos os pontos na linha são transformados em projeções em uma esfera unitária.

Para variáveis estocásticas, é necessário centralizar os dados antes que os sinais possam transportar uma quantidade considerável de informações. Se tivermos uma amostra finita que consiste em  $n$  observações de uma variável estocástica  $\mathbf{x}$ , a matriz de covariância de sinais espaciais é definida pela equação 13:

$$\Sigma = \frac{1}{n-1} \sum_{i=1}^n \text{sgn}(x_i - \hat{\mu}) \text{sgn}(x_i - \hat{\mu})^T \quad (13)$$

O cálculo da matriz de covariância de sinais espaciais se baseia na transformação dos dados em seus sinais espaciais (TSS), seguidos pelo cálculo de uma matriz de covariância normal (SERNEELS, NOLF e ESPEN, 2006).

### 3.2 Análise de componentes principais (PCA)

A análise por componentes principais é um método não supervisionado de análise exploratória, cujo o propósito é observar a estrutura dos dados, detectar amostras atípicas, encontrar semelhanças entre amostras e principalmente reduzir a dimensionalidade dos dados em poucos componentes principais (*Principal Components*, PC), que representam o conjunto de dados original utilizando uma combinação de variáveis que descrevem a tendência dos dados (FERRER-RIQUELME (2009); DE SOUZA e POPPI (2012); FERREIRA, 2015). Mesmo ao reduzir a dimensionalidade do conjunto de dados original, o PCA não interfere na relação entre as amostras (FERREIRA, 2015).

Ao representar os conjuntos de dados originais, os componentes principais podem ser definidos por um par de escores e pesos que explicam o máximo da variância. Os componentes principais são ortogonais e não correlacionados entre si. Sendo assim, o primeiro componente principal (PC1) é traçado no sentido da maior variação no conjunto de dados e o segundo (PC2) é traçado ortogonalmente ao primeiro, com a finalidade de explicar a maior porcentagem da variação não explicada pelo PC1 e mesmo para outros componentes principais que sejam necessários para explicar o comportamento dos dados (DE SOUZA e POPPI, 2012).

A decomposição da matriz  $X$  no PCA ocorre em um produto de duas matrizes, escores ( $T$ ) e pesos ( $P$ ), somado a uma matriz de Erros ( $E$ ), conforme apresentado na equação 14 (ESBENSEN e GELADI, 2009).

$$X = TP^t + E = t_1p_1^t + t_2p_2^t + \dots + t_Ap_A^t + E \quad (14)$$

onde  $X$  é a matriz dos dados de entrada ( $m \times n$ ),  $T$  é a matriz de escores ( $m \times$  número de PC),  $P^t$  é a matriz transposta dos pesos (número de PC  $\times n$ ),  $E$  é a matriz dos erros ou resíduos ( $m \times n$ ) e  $A$  é o  $a$ -ésimo componente principal.

O gráfico dos escores informam a presença de relações de similaridade entre as amostras, enquanto que o gráfico dos pesos informam quais variáveis contribuem mais significativamente para os agrupamentos observados no gráfico dos escores. Se analisados juntos, os gráficos de escores e pesos permitem a identificação das variáveis responsáveis pelas diferenças observadas entre as amostras. No modelo PCA, a quantidade de componentes principais pode ser definida pela percentagem da variância explicada, selecionando o número de componentes de modo a capturar a máxima variância presente no conjunto de dados originais (DE SOUZA e POPPI, 2012).

O PCA é uma ferramenta importante nos estudos estatísticos, mas sabe-se que o PCA clássico é sensível a presença de *outliers*. Segundo Ma e Aybat (2018), o PCA robusto (ROBPCA) pode ser usado para remover o efeito de erros brutos esparsos. Assim, para uma matriz de dados  $X$  ( $m \times n$ ), o ROBPCA decompõe essa matriz em duas matrizes, conforme mostrado na equação 15.

$$X = L + S \quad (15)$$

onde  $L$  é uma matriz de baixo nível e  $S$  é uma matriz esparsa. O PCA robusto considera que  $X$  é uma superposição de  $L$  e  $S$ . Assim, os erros grosseiros serão capturados pela matriz esparsa  $S$ , permitindo que a matriz de baixo nível  $L$  ainda possa se aproximar de  $X$ . Em resumo, o ROBPCA fornece uma aproximação de baixa dimensão que é robusta para valores extremos.

Neste estudo o PCA e ROBPCA estão sendo utilizados principalmente como um método exploratório que auxilia na investigação do comportamento dos dados

coletados, sendo capaz de separar as informações relevantes das redundantes e aleatórias, o que colabora na identificação de amostras atípicas.

### 3.3 Métodos de regressão

As ferramentas usadas neste trabalho para construir modelos de regressão podem ser divididas em dois tipos: métodos lineares, como a regressão por mínimos quadrados parciais, e métodos não lineares, como máquinas de vetores suporte e florestas aleatórias. Além disso, foram utilizadas ferramentas de seleção de variáveis na tentativa de melhorar os modelos de previsão.

#### 3.3.1 Regressão por mínimos quadrados parciais (PLS)

A modelagem por PLS é provavelmente uns dos métodos multivariados mais utilizados na química analítica (NADLER e COIFMAN, 2005), sendo proposto inicialmente por Herman Wold (WOLD, 1982) como uma técnica de análise de dados multivariados cuja o objetivo era correlacionar uma ou mais variáveis resposta ( $\mathbf{Y}$ ) com diversas variáveis independentes ( $\mathbf{X}$ ). O PLS considera a matriz  $\mathbf{X}$  ( $m \times n$ ) para o bloco  $x$  e o vetor  $\mathbf{y}$  ( $n \times 1$ ) no bloco  $y$  na construção do modelo. Em grande parte dos trabalhos, a matrix  $\mathbf{X}$  é composta por dados de origem instrumental, tais como conjunto de espectros de infravermelho (médio ou próximo), Raman, NMR, perfis cromatográficos e outros. O bloco  $\mathbf{y}$  pode ser constituído por mais de uma propriedade físico-química/característica, podendo ser um vetor  $\mathbf{y}$  ( $n \times 1$ ) para uma propriedade ou até mesmo uma matrix  $\mathbf{Y}$  ( $n \times p$ ) para duas ou mais propriedades. Para o caso de se prever simultaneamente múltiplas propriedades (variáveis respostas), ou seja, aplicar o PLS2, a situação se torna mais complicada do que para o algoritmo PLS padrão (conhecido como PLS1). De fato, a aplicação separada do algoritmo PLS1 em cada coluna da matrix  $\mathbf{Y}$  levaria à formação de diferentes conjuntos de escores para cada variável resposta. No PLS2, os escores separados são combinados em um único conjunto de escores, mas essa complicação adicional implica em um algoritmo mais complexo (AGUERSSIF *et al.*, 2008; ABDEL-RAHMAN *et al.*, 2017).

De um modo geral, o método PLS possibilita obter combinações lineares das variáveis independentes  $\mathbf{X}$ , que desempenham melhores resultados ao modelar as variáveis dependentes  $\mathbf{Y}$ . Além disso, de acordo com Teófilo (2013), o PLS pode ser aplicado para modelar conjuntos de dados onde há ruído aleatório e que possuam



variáveis altamente correlacionadas. Modelos PLS podem apresentar excelentes resultados, principalmente se as amostras a serem previstas possuírem características similares ou próximas aos dados originais. No entanto, não há como assegurar que o modelo apresentará um bom desempenho nos dados novos, visto que as características espectrais podem ser diferentes, os experimentos podem ter sido mal planejados ou até mesmo os dados experimentais podem ser inadequados.

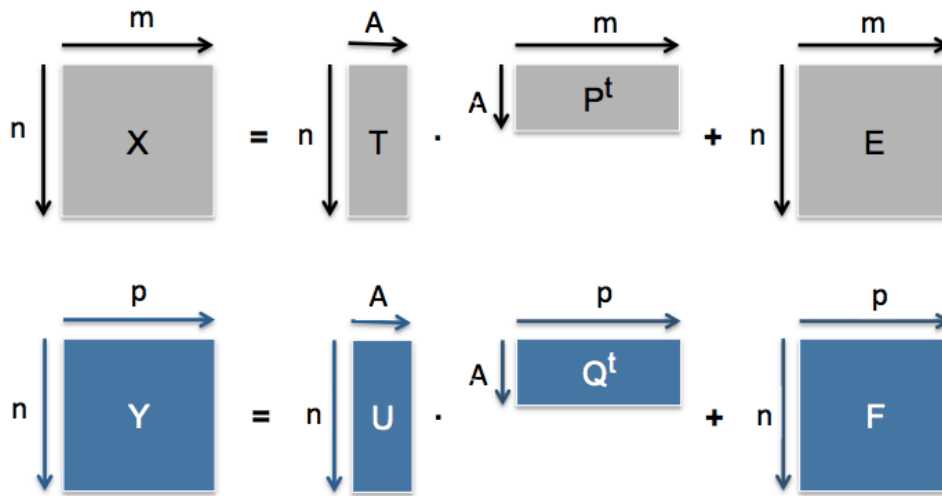
Uma característica importante sobre o PLS é que ele leva em conta os erros de ambas as matrizes, dos espectros (**X**) e das propriedades/características (**Y**). De um modo simplificado, o PLS assume que os erros estão igualmente distribuídos em ambas as matrizes, os quais são de igual importância, como mostrado nas equações 16 e 17 (WOLD *et al.*, 2001; VALDERRAMA *et al.*, 2007; PALERMO *et al.*, 2009; POMERANTSEV, 2014; FERREIRA, 2015). De acordo com Aleme (2008, p. 27) “No método PLS cada variável latente (VL) da matriz **X** é modificada de modo que a covariância entre ela e a matriz **Y** seja maximizado”.

$$\mathbf{X} = \mathbf{TP}^t + \mathbf{E} = \sum_{i=1}^A t_i \mathbf{p}_i^t + \mathbf{E} \quad (16)$$

$$\mathbf{Y} = \mathbf{UQ}^t + \mathbf{F} = \sum_{i=1}^A u_i \mathbf{q}_i^t + \mathbf{F} \quad (17)$$

onde, **X** e **Y** são as matrizes decompostas; **T** e **U** são as matrizes dos *escores* ( $m \times$  número de variável latente); **P<sup>t</sup>** e **Q<sup>t</sup>** são matrizes transpostas dos pesos (número de variável latente  $\times$   $n$ ); **E** e **F** são as matrizes dos erros das matrizes **X** e **Y**, ou seja, representam os resíduos que são a parte não modelada (WOLD *et al.*, 2001). A decomposição das matrizes **X** e **Y** pode ser observada no esquema da Figura 1:

Figura 1 - Esquema de decomposição das matrizes  $X$  e  $Y$  pela aplicação do PLS.



Fonte: Adaptado de Wold *et al.* (2001) e Luna *et al.* (2017)

### 3.3.2 Regressão por máquinas de vetores de suporte (SVM)

A regressão por Máquinas de Vetores de Suporte (*Support Vector Machines Regression*, SVM-R) constitui uma técnica de aprendizado de máquina que foi originalmente desenvolvida por Vapnik para solucionar problemas de reconhecimento de padrões (VAPNIK e CHERVONENKIS, 1971; VAPNIK, 1995).

Muitos autores têm utilizado a técnica de SVM para resolver problemas de classificação e regressão, principalmente para modelar processos ou propriedades não lineares (VAPNIK, 1999; BISHOP, 2006; STEINWART e CHRISTMANN, 2008(a); STEINWART e CHRISTMANN, 2008(b)). Isso é devido principalmente pela inclusão de funções de perda na estrutura da SVM, o que permitiu a ampliação de sua utilidade para aplicação nas áreas de regressão não linear e também em previsões de séries temporais (VAPNIK, 1995; MULLER *et al.*, 1997; LU e WANG, 2005; WANG *et al.*, 2008(b)).

Uma das etapas durante a construção de um modelo de regressão é a divisão do conjunto de dados estudado em dois conjuntos: calibração (treinamento) e validação (teste). O aprendizado na técnica SVM pode ser supervisionado ou não supervisionado. No aprendizado supervisionado, Haykin (1999) descreve de forma figurada, como se houvesse a presença de um professor, o qual expõe o conhecimento prévio por um conjunto de exemplos na forma de entrada e saída desejada. Desta forma, o modelo SVM é concebido a partir do conjunto de calibração, estando apto à produzir resultados para novos dados de entradas. Em contrapartida,

no aprendizado não supervisionado não há o conhecimento prévio do comportamento dos dados, sendo de responsabilidade do algoritmo aprender a agrupar as entradas segundo a similaridade entre os dados sem informações prévias sobre as amostras.

A validação cruzada K-dobra (K-fold) é largamente aplicada na técnica SVM. O objetivo dessa validação cruzada é dividir aleatoriamente as amostras do banco de dados em K sub-amostras, onde uma única amostra é mantida no conjunto de validação para testar o modelo e, as demais K-1 sub-amostras são mantidas no conjunto de calibração. Esse processo de validação cruzada deve ser repetido K vezes, com cada uma das K sub-amostras sendo utilizada uma única vez como informação de validação. Por fim, a média dos resultados das K-dobras é calculada para produzir uma única estimativa. Deste modo, todas as amostras são usadas tanto para o conjunto de calibração quanto para o conjunto de validação, o que pode ser considerado como uma vantagem deste método de sub-amostragem aleatória. Vale ressaltar que a validação cruzada K-fold mais utilizada é a com 10 dobras (PICARD e COOK, 1984; EFRON e TIBSHIRANI, 1997; SANCHEZ *et al.*, 2011).

Monard e Baranauskas (2003) ressaltaram em seu estudo que o modelo SVM pode estar sujeito a ocorrência de super-ajustamento (*overfitting*) ou sub-ajustamento (*underfitting*), mesmo com a utilização de conjuntos de dados diferentes para a calibração (treinamento) e validação (teste). O efeito de super-ajustamento (*overfitting*) ocorre quando o modelo se especializa nos dados do conjunto de treinamento, consequentemente apresentando baixa taxa de acerto quando utilizado na previsão de novos dados. Por outro lado, é também possível que o modelo apresente uma baixa taxa de acerto mesmo no subconjunto de treinamento, evidenciando um efeito de sub-ajustamento (*underfitting*), visto que provavelmente as amostras de calibração disponíveis são pouco representativas ou que o modelo construído é muito rudimentar.

De acordo com Wang *et al.* (2008(a)), a aproximação por regressão SVM direciona a um problema de estimação de uma função com base em dados de um determinado conjunto  $G = \{(x_i, y_i)\}_{i=1}^l$  (onde  $x_i \in R^n$  representa os vetores de entrada ou espectros,  $y_i \in R$  os valores desejados ou propriedades), que é produzido de  $\Phi$ . A regressão por SVM aproxima a função na forma da equação 18.

$$f(x) = \sum_{i=1}^l w_i \Phi_i(x) + b \quad (18)$$

Onde  $\{\Phi_i(x)\}_{i=1}^l = 1$  representa as variáveis de entrada,  $\{w_i\}_{i=1}^l = 1$  e  $b$  são os coeficientes. Os coeficientes podem ser estimados pela minimização da função risco apresentada na equação 19, onde  $L_\varepsilon(y, f(x))$  é descrito pela equação 20.

$$R(C) = C \frac{1}{l} \sum_{i=1}^l L_\varepsilon(y_i, f(x_i)) + \frac{1}{2} \|w\|^2 \quad (19)$$

$$L_\varepsilon(y_i, f(x_i)) = \begin{cases} |y - f(x)| - \varepsilon & |f(x) - y| \geq \varepsilon \\ 0 & \text{outros} \end{cases} \quad (20)$$

sabendo que  $\varepsilon$  é um parâmetro previsto e  $L_\varepsilon(y, f(x))$  é uma função perda  $\varepsilon$ -indiferente, que não penaliza erros menores que  $\varepsilon$ . O termo  $\frac{1}{2} \|w\|^2$  é aplicado como uma medida do nivelamento da função e  $C$  é a constante de regularização que define a troca entre o erro de calibração e o nivelamento do modelo. Ao introduzir no modelo as variáveis de folga  $(\zeta, \zeta^*)$ , têm-se as equações 21 e 22 a serem minimizadas, atendendo as restrições indicadas pelas equações 23 e 24 (WANG *et al.*, 2008(a)).

$$R(w, \zeta^*) = \frac{1}{2} \|w\|^2 + C^* \sum_{i=1}^n (\zeta_i + \zeta_i^*) \quad (21)$$

$$w\Phi(x_i) + b - y_i \leq \varepsilon + \zeta_i \quad (22)$$

$$y_i - w\Phi(x_i) - b - y_i \leq \varepsilon + \zeta_i \quad (23)$$

$$\zeta, \zeta^* \geq 0 \quad (24)$$

De acordo com Wang *et al.* (2008(a)), se a equação 18 for reescrita de modo explícita, será obtida a equação 25.

$$f(x, \alpha_i, \alpha_i^*) = \sum_{i=1}^l (\alpha_i - \alpha_i^*)K(x, x_i) + b \quad (25)$$

Na equação 26 podem ser observados os multiplicadores Lagrangianos,  $\alpha_i$  e  $\alpha_i^*$ , que satisfazem a igualdade:  $\alpha_i \cdot \alpha_i^* = 0$ ,  $\alpha_i \geq 0$ ,  $\alpha_i^* \geq 0$ ; e  $l$  é o número de vetores de suporte (WANG *et al.*, 2008(a); NAGUIB e DARWISH, 2012). Fazendo  $i = 1, \dots, l$  pode ser obtido pela maximização da forma dual da equação 26, considerando as restrições apresentadas nas equações 27 e 28, (VAPNIK, 1998; PLATT, 1998; YEGANEH *et al.* (2012)).

$$\Phi(\alpha, \alpha^*) = \sum_{i=1}^l y_i(\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(\alpha_i, \alpha_j) \quad (26)$$

$$\sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \quad (27)$$

$$\begin{aligned} 0 &\leq \alpha_i \leq C \\ 0 &\leq \alpha_i^* \leq C \end{aligned} \quad (28)$$

Sendo que  $K(x_i, x_j) = \Phi(x_i)^t \cdot \Phi(x_j)$  no espaço, conhecida como função Kernel. Segundo Vapnik (1998) e Schölkopf e Smola (2002), as funções *Kernels* mais aplicadas são as lineares, quadráticas e RBF (*Radial Basis Function*), sendo que para os dados não lineares o *Kernel* RBF é o mais indicado. Para o modelo de regressão, dois tipos de vetores suporte são amplamente utilizados, o Nu-SVR e Epsilon-SVR. O algoritmo SVM utiliza a validação cruzada para selecionar as melhores faixas de valores para estes parâmetros e, então o modelo é construído.

O algoritmo SVM possui importantes parâmetros que devem ser considerados, tais como os valores de *cost*, *gamma* e *epsilon*. O parâmetro *cost* retrata diretamente a penalidade associada a erros maiores que *epsilon*, visto que o aumento do valor do custo (*cost*) faz com que o ajuste do modelo seja mais próximo. O parâmetro *gamma* está relacionado ao aumento do número de vetores de suporte, devido a sua principal característica que é controlar a forma de separação do hiperplano. Para a função de

regressão SVM não há nenhuma penalidade associada aos pontos que estão previstos dentro do valor real da distância de *epsilon*, sendo que a diminuição do valor de epsilon faz com que o ajuste seja mais apropriado aos dados de calibração (WISE *et al.*, 2006).

Para este estudo, os modelos de SVM mais adequados serão os que proporcionarem os melhores valores dos parâmetros comparativos de ajuste do modelo (elevado  $R^2_{CV}$  e baixo RMSECV). Os melhores modelos SVM serão comparados com os modelos obtidos por PLS. Os pré-processamentos utilizados para as SVM serão praticamente os mesmos utilizados na realização do PLS, tendo em consideração que a validação cruzada utilizada no SVM é o K-dobras (K-fold) igual a 10.

### 3.3.3 Florestas aleatórias

O algoritmo de florestas aleatórias (*Random Forests*, RF) foi desenvolvido por Breiman (2001) e, desde então, tem sido amplamente aplicado em vários campos científicos, como medicina (LI *et al.*, 2012; HUANG *et al.*, 2013; DECONINCK *et al.*, 2014; NI *et al.*, 2017), ciências ambientais (ZHANG *et al.*, 2017), engenharia (SHAHBAZI, CHEHREH e MATIN, 2017) e alimentos (LIU *et al.*, 2018). Florestas aleatórias pode ser definida como um algoritmo de aprendizado supervisionado que combina várias árvores de decisão, de modo a construir uma floresta, melhorar a precisão, obter uma previsão estável e uma interpretação direta. Em resumo, o método adiciona aleatoriedade ao modelo e pesquisa o recurso mais crítico ao dividir um nó, e geralmente resulta em um modelo melhor (BREIMAN, 2001; WANG *et al.*, 2018; AMJAD *et al.*, 2018).

## 3.4 **Seleção de variáveis**

O uso de técnicas de seleção de variáveis permite a construção de um modelo robusto e de fácil interpretação, uma vez que a escolha de regiões específicas pode minimizar os erros do modelo multivariado.

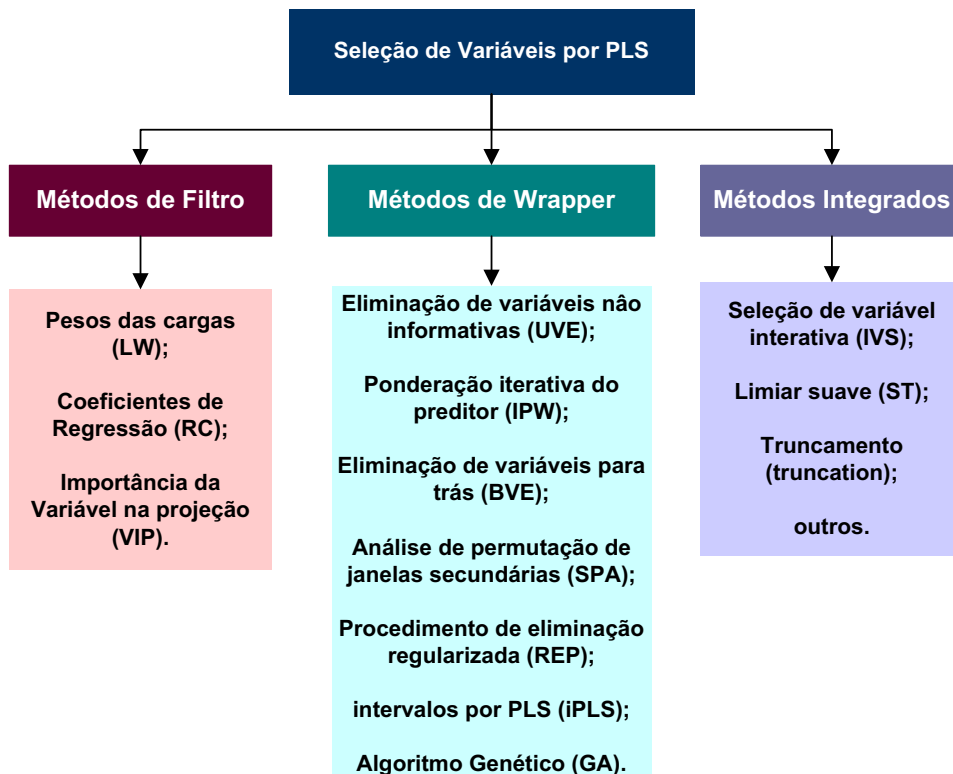
Os métodos de seleção de variáveis podem ser baseados em informações químicas ou em algoritmos. Neste trabalho, métodos baseados em informações químicas levam em consideração a identificação de grupos funcionais característicos do biodiesel em dados espectrais no infravermelho próximo e médio. Além do espectro

completo, as regiões espectrais com informações químicas sobre biodiesel serão consideradas para a construção dos modelos.

Mehmood *et al.* (2012) consideram que os métodos de seleção de variáveis em regressões por mínimos quadrados parciais (PLSR) podem ser divididos em três categorias principais: métodos de filtro (*filter methods*, FM), métodos wrapper (*wrapper methods*, WM) e métodos integrados (*embedded methods*, EM). Saeys *et al.* (2007) também consideraram essa classificação dos métodos de seleção de variáveis em seus estudos. Na Figura 2, são apresentados a classificação dos métodos de seleção de variáveis mais conhecidos.

Neste estudo foram considerados alguns métodos de seleção de variáveis: importância variável na projeção (VIP), pesos das cargas (LW), coeficientes de regressão (RC), *ensemble* PLS (enPLS), mínimos quadrados parciais por intervalos (iPLS), algoritmo genético (GA), eliminação de variável não informativa (UVE), eliminação de variáveis para trás (BVE), análise de permutação de janelas secundárias (SPA), procedimento regularizado de eliminação de variáveis (REP) e ponderação iterativa do preditor (IPW).

Figura 2 - Classificação dos métodos de seleção de variáveis.



Fonte: Adaptado de Mehmood *et al.* (2012)

### 3.4.1 Métodos de filtro

Segundo explicado por Mehmood *et al.* (2012), os métodos de filtro selecionam as variáveis em dois estágios, no primeiro o modelo de regress por PLS é ajustado aos dados e, no segundo, a seleção de variáveis é executada mediante a inserção de um limite (limiar) em alguma medida de relevância obtida a partir do modelo PLS ajustado. A adoção de um limite na medida do filtro é necessário para classificar variáveis como selecionadas ou não. Algumas vantagens dos métodos de filtro são a rapidez e a facilidade de se efetuar os cálculos. No entanto, esses métodos não oferecem informações quanto a relevância da previsão das variáveis selecionadas, e a escolha do limite é um fator crítico que afeta fortemente a seleção. De acordo com Mehmood *et al.* (2012), como exemplos de métodos de filtro, pode-se citar: *loadings weights* (LW), coeficientes de regressão (*Regression Coefficients, RC*) e importância da variável na projeção (*Variable Importance on Projection, VIP*).

#### 3.4.1.1 Pesos das cargas (LW)

Com o modelo PLS ajustado, aperfeiçoado para o mínimo número de componentes, o LW pode ser usado como medida de importância para selecionar variáveis (MARTENS, 1985; FRANK, 1987). Para cada um dos componentes são selecionadas as variáveis com um peso de carga superior a um determinado limite (limiar) em valor absoluto. Um exemplo disso, é quando os pesos das cargas (*loadings weights*) são extraídos do modelo ótimo e utilizados para a determinação dos comprimentos de onda (variáveis) mais correlacionados com a propriedade que permite obter melhores resultados com relação à previsão. O peso das cargas é um método de saída de filtro muito simples de ser aplicado a partir do modelo PLSR, principalmente se a implementação for no software R.

#### 3.4.1.2 Coeficientes de regressão (RC)

De acordo com Frenich *et al.* (1995), o método RC é uma medida única de associação entre cada variável e a resposta. Neste método, as variáveis que apresentarem baixos valores absolutos dos coeficientes de regressão podem ser eliminadas. Além disso, a determinação dos limites pode ser baseada nas considerações de significância de *jackknifing* ou *bootstrapping* (EFRON e



TIBSHIRANI, 1993) para considerar, além da observação única da medida do filtro, também a distribuição ou variação.

### 3.4.1.3 Importância da variável na projeção (VIP)

A importância da variável na projeção por PLS é um método que acumula a importância de cada  $j$ -ésima variável no modelo, que é projetada pelo vetor de pesos  $w$  de cada variável latente, conforme apresentado na equação 29 (WOLD *et al.*, 1993; ERIKSSON *et al.*, 2001; MEHMOOD *et al.*, 2012; GODINHO *et al.*, 2014).

$$v_j = \sqrt{\frac{p \sum_{a=1}^A \left[ SS_a \left( \frac{w_{aj}}{\|w_a\|^2} \right) \right]}{\sum_{a=1}^A (SS_a)}} \quad (29)$$

onde  $v_j$  é a medida VIP;  $SQ_a$  é a soma quadrática explicada pela  $a$ -ésima variável latente;  $W_a$  é a matriz de pesos (*loadings*);  $\|W_a\|$  representa a norma de  $W_a$ ;  $A$  é o total de variáveis latentes; e  $p$  simboliza as variáveis. A eliminação da  $j$ -ésima variável ocorre quando a mesma se encontra abaixo de um limite (limiar) de corte  $u$  (ou seja, se  $v_j < u$ ), onde  $u \in [0, \infty)$  e é definido pelo usuário. De acordo com Eriksson *et al.* (2001), Chong e Jun (2005), Tran *et al.* (2014) e Farrés *et al.* (2015), a variável  $j$  pode ser incluída no modelo de calibração se a mesma tiver a medida VIP maior que 1 ( $v_j > 1$ ).

### 3.4.2 Métodos de wrapper

Segundo Saeys *et al.* (2007), os métodos do tipo *Wrapper* usam principalmente os métodos de filtro de uma maneira iterativa e utilizam também alguma abordagem de aprendizagem supervisionada, onde a reconstrução do modelo é compreendido dentro de um algoritmo de busca de variável. O algoritmo de busca extrai o subconjunto de variáveis relevantes e avalia cada subconjunto ajustando um modelo ao subconjunto de variáveis. No entanto, vale ressaltar que todos os métodos de *wrapper* exigem maior tempo computacional do que os métodos de filtro. A seguir são descritos alguns dos métodos de *wrapper*.

### 3.4.2.1 Algoritmo genético (GA)

O algoritmo genético é um algoritmo adaptativo de busca heurística baseado nas idéias evolucionárias de seleção natural e genética, seguindo o princípio estabelecido por Darwin de sobrevivência do mais apto. Como tal, o algoritmo representa um método de inteligência artificial que tem sido muito aplicado na seleção de comprimentos ou números de ondas em aplicações espectrométricas quantitativas, e também muito aplicado para resolver problemas de otimização. Cartwright (1995) e Lacerda e Carvalho (1999) descreveram o algoritmo genético como uma ferramenta que atua modelando matematicamente a teoria Darwiniana. De modo análogo ao que ocorre no mundo biológico, o algoritmo otimiza as condições experimentais (cromossomos), que levam as melhores respostas (indivíduos mais adaptados ao ambiente) e tem maior chance de serem selecionados (sobreviver), sendo transmitidos às novas gerações pela reprodução. Sendo assim, a otimização das respostas (evolução da espécie) é obtida por meio de recombinação das variáveis (cruzamento de cromossomos) e de algumas modificações randômicas (mutações de genes).

Resumidamente, Lucas (2002) explica que o algoritmo genético gera uma população de possíveis respostas para o problema a ser tratado para depois submetê-la ao processo de evolução. Este processo é formado pelas etapas de avaliação, seleção, cruzamento, mutação, atualização e finalização. Durante a etapa de avaliação, as soluções do problema são avaliadas para descobrir o quão elas respondem ao problema proposto. A etapa de seleção compreende a escolha dos indivíduos para a reprodução, considerando que a probabilidade de uma dada solução ser selecionada é proporcional à sua aptidão. Lucas (2002) descreveu o cruzamento como uma etapa na qual as características das soluções escolhidas são recombinadas, gerando novos indivíduos. A mutação é uma etapa que inclui o acréscimo da variedade à população, em consequência às alterações das características dos indivíduos resultantes do processo de reprodução. Na etapa de atualização, os indivíduos criados nesta geração são inseridos na população. Por fim, a última etapa é a finalização, na qual há a verificação se as condições de encerramento da evolução foram alcançadas, retornando para a etapa de avaliação em caso negativo e encerrando a execução em caso positivo.

Wise *et al.* (2003) relataram a importância da técnica de seleção natural para selecionar variáveis em modelos calibração multivariada, descrevendo que para uma dada matriz  $\mathbf{X}$  (variáveis independentes) e um vetor ou matriz  $\mathbf{Y}$  (variáveis dependentes), o algoritmo determina um subconjunto de variáveis em  $\mathbf{X}$  e, usando um método de regressão multivariada e uma estratégia de validação cruzada, é determinado o valor da raiz quadrada do erro médio quadrático de validação cruzada (RMSECV) para o subconjunto de variáveis. O algoritmo genético aplica esta sequência de forma iterativa até alcançar os subconjuntos que proporcionam o menor valor de RMSECV.

#### 3.4.2.2 Eliminação de variáveis não-informativas por PLS (UVE-PLS)

No método de eliminação de variáveis não-informativas por PLS (*Uninformative variable elimination in PLS*, UVE-PLS), variáveis de ruído artificiais são adicionadas ao conjunto de previsão antes da aplicação do modelo de regressão por PLS (CENTNER *et al.*, 1996). Depois disso, todas as variáveis originais com menor "importância" do que as variáveis de ruídos são eliminadas antes do procedimento ser repetido até que um critério de parada seja atingido.

#### 3.4.2.3 Eliminação de variáveis para trás por PLS (BVE-PLS)

O método de eliminação de variável para trás (*Backward Variable Elimination*, BVE-PLS) tem como objetivo a eliminação de variáveis não informativas. Para a execução do método, as variáveis são primeiramente classificadas em relação a importância, sendo necessário aplicar uma das medidas de filtro descritas anteriormente. Depois, um limiar (threshold) é utilizado para eliminar o subconjunto das variáveis menos informativas. Em seguida, um modelo é ajustado novamente para as variáveis restantes e o desempenho é medido. Este procedimento é repetido inúmeras vezes até que o desempenho máximo do modelo seja alcançado (MEHMOOD *et al.*, 2012).

#### 3.4.2.4 Ponderação iterativa do preditor por PLS (IPW-PLS)

O método de ponderação iterativa do preditor por PLS (*Iterative predictor weighting* PLS, IPW-PLS) foi desenvolvido por Forina *et al.* (1999) como um processo iterativo para a eliminação de variáveis. Para o funcionamento do método, uma medida de importância preditora é calculada após a montagem de um modelo de regressão por PLS. A medida de importância obtida pelo modelo é usada para reescalonar as variáveis originais da matriz  $\mathbf{X}$  e para eliminar as variáveis menos importantes antes do reajuste do modelo subsequente.

#### 3.4.2.5 Análise de permutação de janelas secundárias por PLS (SPA-PLS)

Segundo Li *et al.* (2010), a análise de permutação de janelas secundárias (ou sub-janelas) acoplada com PLS (*Sub-window permutation analysis coupled with PLS*, SPA-PLS) fornece a influência de cada variável desconsiderando a influência do resto das variáveis. O uso de um subconjunto de variáveis torna o método SPA-PLS mais eficiente e rápido para grandes conjuntos de dados.

#### 3.4.2.6 Procedimento regularizado de eliminação (REP-PLS)

O método chamado procedimento regularizado de eliminação de variáveis (*Regularized elimination procedure in PLS*, REP-PLS) foi introduzido por Mehmood *et al.* (2011). No método, um procedimento de seleção de variáveis baseadas na estabilidade é adotado, onde as amostras são divididas aleatoriamente em um número predefinido de conjuntos de treinamento e teste.

#### 3.4.2.7 Mínimos quadrados parciais por intervalos (iPLS)

De acordo com Nørgaard *et al.* (2000), o método iPLS é uma extensão do PLS, no qual o algoritmo desfragmenta o conjunto de dados espectrais em um determinado número de intervalos de mesma largura, construindo um modelo PLS para cada intervalo, e expondo os resultados em um gráfico para simplificar a comparação com o modelo construído para todo o perfil espectral.

Segundo Silva *et al.* (2007) o método de regressão de mínimos quadrados parciais por intervalos indica o conjunto de variáveis espectrais baseando-se em

critérios objetivos, como por exemplo, avaliar a correlação dos intervalos selecionados com o método de referência a partir da raiz quadrada dos erros médios de validação cruzada (RMSECV). Deste modo, o método é capaz de oferecer uma visão completa dos dados, o que o torna bastante útil na interpretação de quais sinais do espectro são mais relevantes, contendo informações químicas importantes sobre a amostra, e devem ser incluídas no modelo de calibração (NØRGAARD *et al.*, 2000; PARISOTTO *et al.*, 2009; XIABO *et al.*, 2010).

### 3.4.3 Ensemble PLS

A aplicação do método enPLS pode ser usado em estudos de regressão e classificação multivariados, principalmente porque eles podem melhorar a precisão de preditores instáveis. Segundo alguns autores, o método enPLS busca desenvolver um algoritmo robusto para construir modelos baseados nos conceitos de aprendizagem por conjuntos e distribuição estatística. A distribuição estatística fornece algumas informações sobre o comportamento das variáveis. Os métodos de amostragem de Monte Carlo ou *bootstrap* são os métodos mais utilizados devido às suas propriedades assintóticas (MEVIK, SEGTMAN e NAES, 2005; HU *et al.*, 2012; BI *et al.*, 2015; CAO *et al.*, 2017)

## 3.5 **Algoritmo de seleção de amostras**

O algoritmo de Kennard-Stone (KS) é um dos mais conhecidos algoritmos para seleção de amostras. O algoritmo KS inicia os cálculos selecionando as duas amostras com a maior distância Euclidiana entre si no espaço X. Para cada uma das amostras que permaneceram, calcula-se a distância mínima com relação às amostras já selecionadas. A partir de então, a amostra com a maior distância mínima é retida, e o procedimento é repetido inúmeras vezes até que um determinado número de amostras seja selecionado (KENNARD e STONE, 1969).

O algoritmo de Kennard-Stone (KS) é utilizado na seleção das amostras do conjunto de validação e de calibração para o banco de dados espectrais. A partir do conjunto original dos dados, o algoritmo divide o total de amostras em dois conjuntos: treinamento que é responsável pela construção do modelo de regressão e o de teste que tem como objetivo avaliar a capacidade preditiva do modelo.

### 3.6 Parâmetros de mérito

Na calibração multivariada, os parâmetros estatísticos são muito importantes na avaliação e comparação entre modelos, tais como os valores de RMSEC, RMSECV, RMSEP,  $R^2$ ,  $R_{cal}$ ,  $R_{cv}$ ,  $R_{pred}$ , o viés (*bias*), SEP e SEC.

As amostras previstas durante a validação cruzada são comparadas com seus valores conhecidos e os parâmetros estatísticos são calculados para avaliar em que situação o modelo apresentou maior capacidade de previsão (TEÓFILO, 2013). Um exemplo disso é o coeficiente de determinação ( $R^2$ ), que é o parâmetro responsável por explicar o grau de ajuste do modelo. O coeficiente de determinação é dado pela razão da soma quadrática da regressão e da soma quadrática total, conforme apresentado na equação 30 (CHARNET *et al.*, 2008; FERREIRA, 2015).

$$R^2 = \frac{SQ_{regressão}}{SQ_{total}} = \frac{SQ_{total} - SQ_{resíduos}}{SQ_{total}} = 1 - \frac{SQ_{resíduos}}{SQ_{total}} \quad (30)$$

Um dos parâmetros considerados, a raiz quadrada do erro médio quadrático da calibração (RMSEC) é apresentada pela equação 31 e o coeficiente de correlação da calibração ( $R_{cal}$ ) apresentada pela equação 32. Ambos são obtidos a partir dos valores estimados pela validação cruzada e os valores conhecidos.

$$RMSEC = \sqrt{\frac{\sum_{i=1}^{I_c} (y_i - \hat{y})^2}{I_c - A}} \quad (31)$$

$$R_{cal} = \frac{\sum_{i=1}^{I_c} (\hat{y}_i - \bar{y})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{I_c} (\hat{y}_i - \bar{y})^2 (y_i - \bar{y})^2}} \quad (32)$$

onde  $\hat{y}_i$  é o valor estimado para a amostra  $i$  de calibração, não incluída na construção do modelo,  $\bar{y}$  é a média dos valores em  $y$ ,  $I_c$  é o número de amostras da calibração e  $A$  é o número de variáveis latentes.

A raiz quadrada do erro médio quadrático da validação cruzada (RMSECV) é apresentada pela equação 33 e o coeficiente de correlação de validação cruzada

( $R_{cv}$ ) apresentado pela equação 34. Ambos os parâmetros são obtidos a partir dos valores estimados pela validação cruzada (para este estudo, os blocos contíguos) e os valores conhecidos. Conforme descrito por Wise *et al.* (2006) e Ballabio e Consonni (2013), temos:

$$RMSECV = \sqrt{\frac{\sum_{i=1}^S (y_i - \bar{y})^2}{S}} \quad (33)$$

$$R_{cv} = \frac{\sum_{i=1}^S (\hat{y}_i - \bar{y})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^S (\hat{y}_i - \bar{y})^2 (y_i - \bar{y})^2}} \quad (34)$$

onde  $\hat{y}_i$  é o valor estimado para a amostra  $i$ , não incluída na construção do modelo,  $S$  é o número de segmentos ou divisões no conjunto de dados e  $\bar{y}$  é a média dos valores em  $y$ .

De acordo com Teófilo (2013), na etapa de validação externa, um conjunto de variáveis independentes é fornecido ao modelo com  $h$  componentes para verificar o seu poder preditivo. A habilidade de previsão do modelo é calculada usando a raiz quadrada do erro médio quadrático da previsão (RMSEP), conforme apresentado na equação 35 e o coeficiente de correlação de previsão ( $R_{val}$ ) apresentado na equação 36.

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{I_P} (y_i - \hat{y}_i)^2}{I_P}} \quad (35)$$

$$R_{val} = \frac{\sum_{i=1}^{I_P} (\hat{y}_i - \bar{y})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{I_P} (\hat{y}_i - \bar{y})^2 (y_i - \bar{y})^2}} \quad (36)$$

em que  $\hat{y}_i$  é o valor estimado pelo modelo final para a amostra  $i$ ,  $I_P$  é o número de amostra no conjunto de previsão e  $\bar{y}$  é a média dos valores em  $y$ . Após o modelo ter sido construído e validado, ele pode ser usado para previsões de novas amostras.

O erro padrão relativo de previsão (RSEP) é outro parâmetro utilizado para avaliar a performance do modelo construído, conforme equação 37 (FERREIRA, 2015).

$$RSEP(\%) = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i)^2}} \cdot 100\% \quad (37)$$

O viés (*bias*) é a tendência ou erro sistemático e pode ser descrito como a diferença média entre o valor predito e o valor medido, através da equação 38 (FERREIRA, 2015).

$$bias = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n} \quad (38)$$

onde  $n$  é o número de amostras do conjunto de teste,  $y_i$  é valor estimado pelo modelo,  $\hat{y}_i$  é o valor real ou medido da propriedade.

O parâmetro do erro-padrão de previsão (SEP) também pode ser utilizado para avaliar a qualidade do modelo e é apresentado pela equação 39 (FERREIRA, 2015).

$$SEP = \sqrt{\frac{\sum_{p=1}^P (y_p - \hat{y}_p - bias)^2}{P - 1}} \quad (39)$$

onde  $P$  é o número de amostras do conjunto de teste,  $y_p$  é valor estimado pelo modelo,  $\hat{y}_p$  é o valor real ou medido da propriedade.

### 3.7 Soma das probabilidades do teste de Wilcoxon (SWTP)

O teste de Wilcoxon foi utilizado para comparar e escolher o melhor modelo para cada propriedade. É um teste não paramétrico que compara dois grupos emparelhados (amostras ou modelos), calcula a diferença entre cada conjunto de pares e analisa essas diferenças. O teste assume que há informações nos sinais e nas magnitudes das diferenças nos valores de cada par (GEHAN, 1965).



O teste de Wilcoxon resulta em um valor de probabilidade quando aplicado para comparar apenas dois modelos de cada vez. Neste trabalho, o teste de Wilcoxon foi aplicado considerando uma nova abordagem para comparar todos os modelos. Os valores residuais de cada modelo de previsão foram utilizados como dados de entrada para o cálculo da probabilidade do teste de Wilcoxon (*Wilcoxon Test Probability*,  $WTP$ ). Caso o modelo  $i$  seja melhor que o modelo  $j$ , o  $WTP_{i,j}$  será um valor baixo. Então, é possível construir uma matriz  $N \times N$  aplicando o teste de Wilcoxon a um conjunto de modelos  $N$ , onde  $i = 1, \dots, N$  e  $j = 1, \dots, N$ . Considerando que as linhas dessa matriz, contém os resultados da comparação do modelo  $i$  com todos os modelos do conjunto. A soma de todas as probabilidades ( $SWTP_i$ ) da  $i$ -ésima linha da matriz, pode ser obtida com um valor que resume o comportamento do modelo. Este somatório é expresso na equação 40.

$$SWTP_i = \sum_{j=1}^N WTP_{i,j} - WTP_{i,i} \quad (40)$$

O menor valor contido neste vetor revelará o melhor modelo no conjunto de modelos estudados.

### 3.8 Elipse da região conjunta de confiança

A elipse da região conjunta de confiança (*elliptical joint confidence region*, EJCR) é aplicada para detectar a ocorrência de erros sistemáticos presentes nos modelos de calibração e validação. A elipse de região conjunta de confiança (com 95%) foi utilizada para comparar os valores experimentais (ou medidos) e os valores previstos, relativo à interceptação e inclinação obtidas a partir da regressão. Um gráfico bidimensional da região elíptica é construído e verifica-se este contém o ponto (1,0). Se o ponto (1,0) não estiver contido na elipse, isso indica que o método não é preciso e apresenta erros sistemáticos (VALDERRAMA, BRAGA e POPPI, 2007; DE SOUZA *et al.*, 2014).

### 3.9 Resumo do estudo quimiométrico

O estudo quimiométrico proposto a esta tese inclui o estudo de métodos de regressão para previsão das propriedades dos biocombustíveis e combustíveis, utilizando espectros de infravermelho próximo (NIR), médio (mid-IR) e Raman.

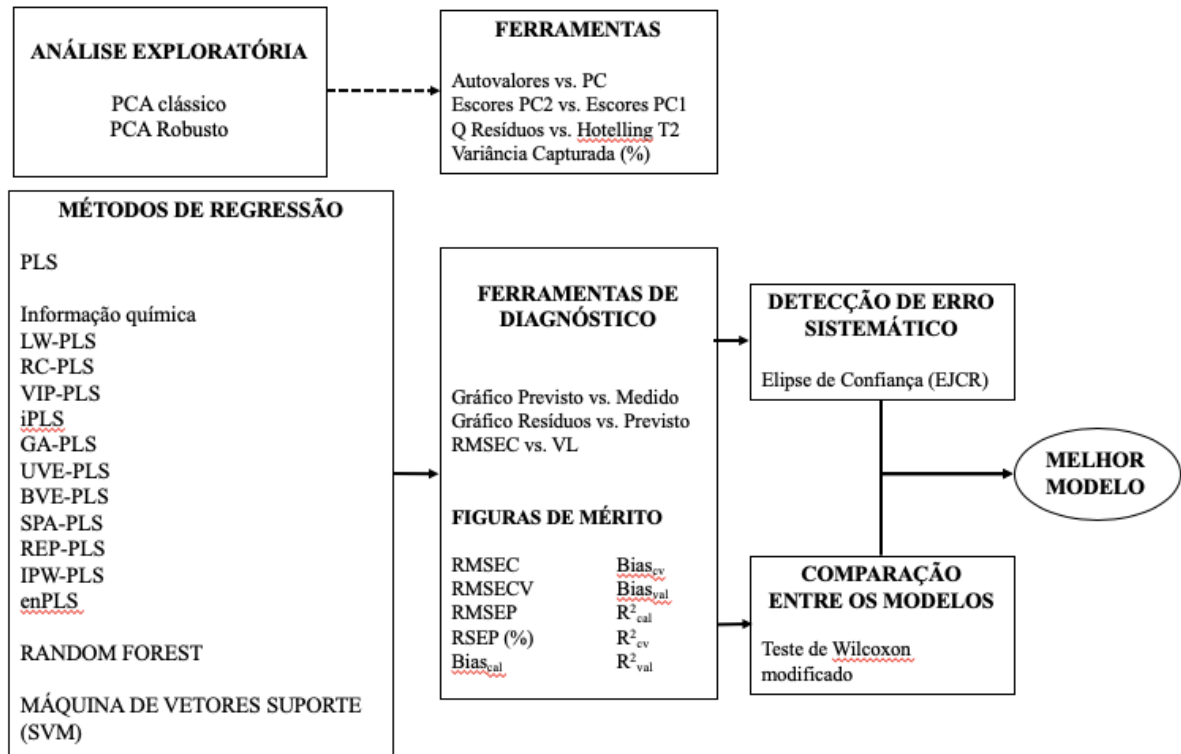
Na Figura 3 é apresentado um organograma dos métodos e das ferramentas aplicados a este estudo. A análise de componentes principais clássica (PCA) e robusta (ROBPCA) são utilizadas como a principal ferramenta de análise exploratória dos dados e detecção de possíveis amostras atípicas. Os dados espectrais (NIR, mid-IR e Raman) passaram por uma transformação de sinais espaciais, cuja o objetivo foi transformar os dados de modo a manter todas amostras viáveis para a construção dos modelos, inclusive aquelas indicadas como atípicas pelo PCA clássico e robusto.

O método de regressão linear a ser aplicado é a regressão por mínimos quadrados parciais (PLS) e os não-lineares são máquinas de vetores suporte (SVM) e Florestas Aleatórias (RF). Na aplicação do algoritmo PLS são considerados alguns métodos de seleção de variáveis na tentativa de melhorar o modelo antes de testar os métodos não-lineares (SVM e RF). Se após a aplicação dos métodos não-lineares os modelos ainda apresentarem resultados não satisfatórios, a técnica de fusão de dados de nível baixo será aplicada na construção de modelos mais satisfatórios.

Vale ressaltar que todos os modelos foram validados por um etapa de validação interna. Como método de validação interna foi aplicado a validação cruzada  $k$ -fold=10.

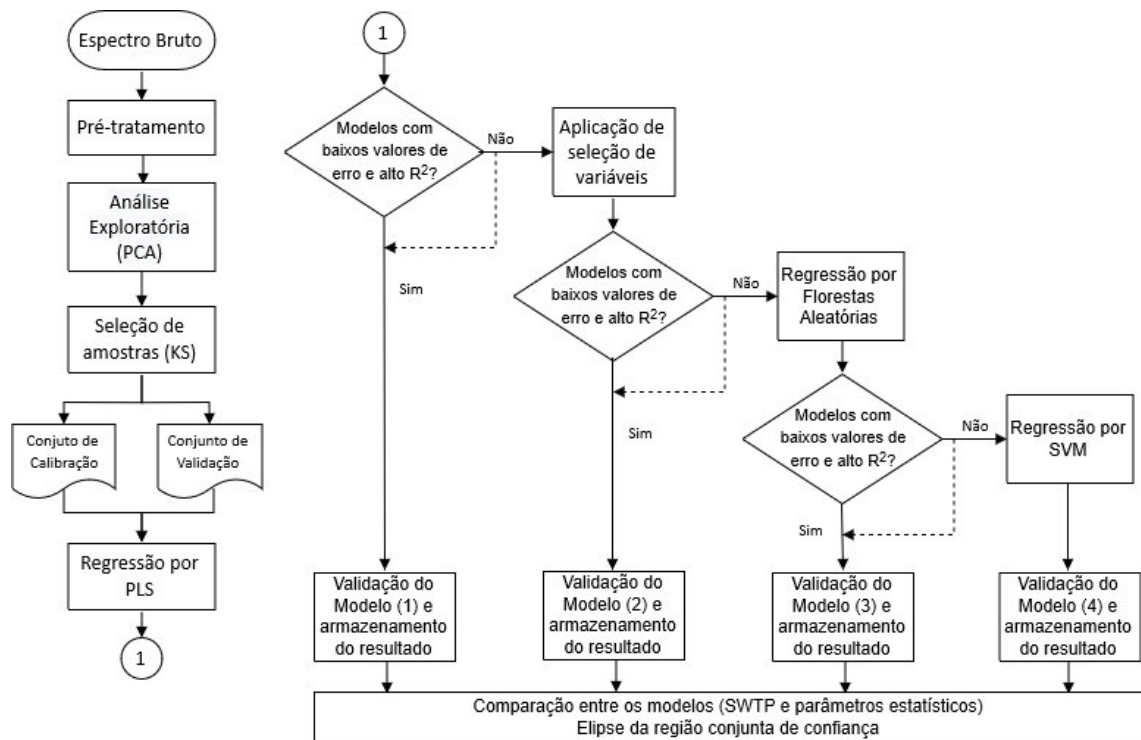
No fluxograma da Figura 4 é apresentado todas as etapas do estudo, no qual foi incluído desde a construção das matrizes de espectros brutos de NIR, mid-IR e Raman, o pré-tratamento desses espectros, análise exploratória utilizando PCA clássico e robusto, utilização de algoritmo de seleção de amostras (algoritmo KS) e algoritmos de seleção de variáveis, seguido pela construção de modelos preditivos para as propriedades estudadas, comparação entre os modelos utilizando as figuras de mérito e o teste Wilcoxon pareado, e por último a construção das elipses de região conjunta de confiança de cada modelo obtido. As determinações da norma ASTM E1655-05 (ASTM, 2005) foram atendidas na etapa de análise exploratória.

Figura 3 - Organograma das ferramentas e métodos quimiométricos utilizados.



Fonte: O autor, 2019.

Figura 4 - Fluxograma do estudo de previsão das propriedades.



Fonte: O autor, 2019.

## 4 OBJETIVO

O objetivo geral deste trabalho de pesquisa é a aplicação de ferramentas quimiométricas de calibração multivariada em dados espectrais de Raman e infravermelho (médio e próximo) provenientes de amostras de biocombustíveis e combustíveis, com o propósito de desenvolver metodologias analíticas alternativas de baixo custo, rápidas, não-destrutíveis, que possam ser aplicadas em equipamentos portáteis e remotamente.

### 4.1 Objetivos específicos

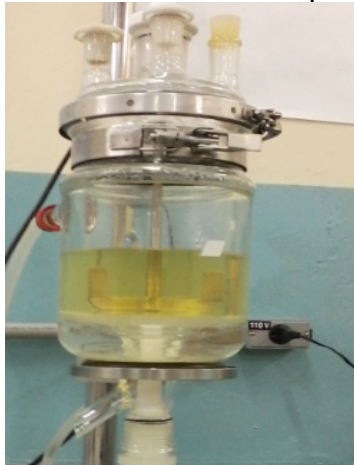
1. Investigar a possibilidade de prever o índice de refração e massa específica de amostras de biodiesel (Banco 1), utilizando ferramentas quimiométricas convencionais de regressão (PLS) e o emprego de seleção de variáveis (algoritmos e informações químicas), combinadas com dados de espectroscopia Raman. Comparar a robustez dos modelos obtidos com espectros Raman com os obtidos com infravermelho médio e próximo, motivado principalmente pela praticidade do método.
2. Verificar se as regiões químicas onde se encontram as mais evidentes bandas de absorção estão representadas nos intervalos que contribuem para a obtenção dos melhores modelos de previsão das propriedades físicas estudadas.
3. Comparar a aplicação de espectroscopia de infravermelho, médio e próximo, e espectroscopia Raman, combinadas com métodos de regressão multivariado de primeira ordem (PLS) e não-lineares, como Florestas Aleatórias (RF) e máquinas de vetores suporte (SVM) na previsão de propriedades físicas não-lineares, como viscosidade cinemática à 40°C e ponto de entupimento de filtro a frio, em amostras de biodiesel, diesel, diesel-biodiesel e diesel-biodiesel-etanol (Banco 2). E desenvolver uma ferramenta para facilitar a comparação entre os modelos obtidos.

## 5 EXPERIMENTAL

### 5.1 Produção do biodiesel

O biodiesel utilizado neste estudo foi obtido pela reação de transesterificação e posteriormente purificado. A síntese do biodiesel foi processada com razão molar de 6 mols de álcool metílico para 1 mol de óleo de soja, com 1% do catalisador básico (KOH). A reação de transesterificação foi realizada em um reator com capacidade volumétrica de 2 litros, com agitação mecânica (280 rpm) e temperatura controlada de 45°C. No reator foi adicionado o óleo refinado (soja, milho, canola ou girassol) e mantido sob agitação constante até atingir a temperatura de 45°C. Em seguida, adicionou-se a solução de metóxido de potássio (KOH + álcool metílico) e manteve-se a mistura reacional por 30 minutos a 45°C sob agitação (Figura 5).

Figura 5 - Reator contendo massa reacional durante a produção de biodiesel.



Fonte: O autor, 2019.

Após a transesterificação a solução foi mantida em repouso para permitir a decantação e separação das fases: biodiesel (superior) e inferior composta de glicerol, sabões, excesso de base e álcool. O biodiesel foi então submetido ao processo de sucessivas lavagens, com o objetivo de purificar os ésteres presentes, removendo impurezas e álcalis residuais.

A primeira lavagem foi feita neutralizando o meio com uma solução de HCl a 0,5%v/v, seguido de uma separação de fases. Na segunda, foi adicionado ao biodiesel uma solução saturada de cloreto de potássio (KCl), seguido de uma separação de fases. A adição de solução saturada de KCl no processo de purificação do biodiesel, tem como objetivo a quebra da emulsão formada entre o biodiesel e alguns resíduos

ainda presentes no biocombustível. A retirada de água também é facilitada neste processo, visto que com a adição do sal, aumenta-se a capacidade de solvatação com um eletrólito forte (no caso, o cloreto de potássio).

A terceira lavagem foi realizada com a adição de água destilada, seguida de uma separação de fases. A neutralização do catalisador foi confirmada medindo o pH da água de lavagem, que ficou entre 6 e 8. Por fim, foi adicionado sulfato de sódio anidro ( $\text{Na}_2\text{SO}_4$ ) ao biodiesel para remover qualquer traço de água presente no biodiesel, sucedida por uma etapa de filtração (duas camadas de algodão) e só então o éster foi armazenado em frascos âmbar adequadamente limpos.

Os resíduos gerados na produção do biodiesel, principalmente a glicerina, excesso de álcool metílico, o cloreto de potássio (KCl) e o sulfato de sódio ( $\text{Na}_2\text{SO}_4$ ), podem ser recuperados conforme descrito por Cunha (2014).

Todos os lotes de biodiesel de soja (10,5 litros) foram misturados formando um lote unificado de biodiesel de soja. Esta unificação dos lotes foi feita também para os demais tipos de biodiesel estudados, tais como o de milho (4 litros), canola (5 litros), girassol (4 litros) e misto (1 litro), sendo o biodiesel misto produzido com 25% v/v de óleo de soja, 25% v/v de óleo de canola, 25% v/v de girassol e 25% v/v de milho.

## 5.2 BANCO 1: Misturas de biodiesel

As amostras dos puros foram compostas por biodiesel puro de soja, canola, girassol, milho e biodiesel cedido por um distribuidor da região sul do Brasil, totalizando uma quantidade de 41 amostras.

As misturas contendo as quatro diferentes origens de biodiesel (soja, canola, girassol e milho) foram preparadas em triplicatas e em base volumétrica. Desta forma, como foram estabelecidas 36 misturas distintas, foi necessário o preparo de 108 misturas no total. Nas Tabelas 1, 2 e 3 são apresentadas as composições das misturas ternárias, quaternárias e binárias entre os óleos de biodiesel, em percentagem volumétrica. Para compor as misturas foram utilizados 2 litros do biodiesel cedido por uma distribuidora da região sul. O biodiesel de canola, por apresentar valor de ponto de entupimento mais distinto entre os tipos de biodiesel deste estudo, foi o eleito para compor a mistura binária com o biodiesel da região sul.

Tabela 1 - Composição das misturas ternárias de biodiesel.

Misturas Ternárias					
Biodiesel_1	%v/v_1	Biodiesel_2	%v/v_2	Biodiesel_3	%v/v_3
Soja	50	Milho	40	Canola	10
Soja	40	Milho	30	Canola	30
Soja	50	Milho	10	Canola	40
Soja	50	Milho	40	Girassol	10
Soja	40	Milho	30	Girassol	30
Soja	50	Milho	10	Girassol	40
Soja	50	Canola	40	Girassol	10
Soja	40	Canola	30	Girassol	30
Soja	50	Canola	10	Girassol	40

Em resumo, o Banco 1 é constituído por 40 amostras de puros, 61 amostras binárias, 27 amostras ternárias e 21 amostras quaternárias, totalizando 149 amostras, conforme apresentado no ANEXO A.

Tabela 2 - Composição das misturas quaternárias de biodiesel.

Misturas Quaternárias							
Biodiesel 1	%v/v 1	Biodiesel 2	%v/v 2	Biodiesel 3	%v/v3	Biodiesel 4	%v/v 4
Soja	25	Milho	25	Canola	25	Girassol	25
Soja	40	Milho	40	Canola	10	Girassol	10
Soja	40	Milho	10	Canola	40	Girassol	10
Soja	40	Milho	10	Canola	10	Girassol	40
Soja	10	Milho	40	Canola	40	Girassol	10
Soja	10	Milho	40	Canola	10	Girassol	40
Soja	10	Milho	10	Canola	40	Girassol	40

Tabela 3 - Composição das misturas binárias de biodiesel.

Misturas Binárias			
Biodiesel 1	%v/v 1	Biodiesel 2	%v/v 2
Soja	10	Milho	90
Soja	30	Milho	70
Soja	50	Milho	50
Soja	70	Milho	30
Soja	90	Milho	10
Soja	10	Girassol	90
Soja	30	Girassol	70
Soja	50	Girassol	50
Soja	70	Girassol	30
Soja	90	Girassol	10
Soja	10	Canola	90
Soja	30	Canola	70
Soja	50	Canola	50
Soja	70	Canola	30
Soja	90	Canola	10
Canola	10	Região Sul	90
Canola	30	Região Sul	70
Canola	50	Região Sul	50
Canola	70	Região Sul	30
Canola	90	Região Sul	10

As amostras do Banco 1 foram preparadas e mantidas em geladeira (5 a 10°C) desde 2014. Cada uma das amostras foi segregada em frascos com o volume requerido para cada ensaio. A obtenção dos espectros NIR e mid-IR, bem como os ensaios de índice de refração, massa específica e PEFf foram obtidos em 2014, conforme Cunha (2014).

Após um período de 4 anos sob refrigeração, algumas amostras escolhidas aleatoriamente foram retiradas da geladeira, mantidas em temperatura ambiente e novamente foram obtidos os espectros mid-IR. Os novos espectros de mid-IR foram então comparados com os espectros obtidos em 2014 para garantir que as amostras não sofreram deterioração durante o período sob refrigeração. Confirmado a



integridade das amostras, estas foram consideradas aptas a serem utilizadas neste estudo.

### 5.3 BANCO 2: Misturas de biodiesel, diesel-biodiesel e diesel-biodiesel-etanol

Todas as amostras de biodiesel e suas misturas preparadas no Banco 1, estão presentes no Banco 2. As misturas de diesel-biodiesel e diesel-biodiesel-etanol foram preparadas em triplicata e em base volumétrica. As amostras dos puros utilizados na preparação das misturas dessa etapa também foram incluídas no banco, totalizando 33 amostras de puros (Diesel S-10, Diesel S-500, Diesel padrão puro, biodiesel de canola, biodiesel de soja, biodiesel de girassol, biodiesel de milho, biodiesel da região sul e biodiesel de um distribuidor do RJ). Para compor as misturas foram utilizados 2 litros de biodiesel de soja, 1 litro de biodiesel de girassol, 1 litro de biodiesel de canola, 1 litro de biodiesel de milho (produzidos conforme item 6.1), 1 litro de biodiesel da região sul, 0,5 litro de biodiesel do RJ, 12 litros de diesel S-10, 5 litros de diesel S-500, 1,5 litros de diesel padrão puro (sem adição de biodiesel) e 0,5 litro de etanol anidro.

Como representante do diesel nas misturas ternárias, o Diesel S-10 foi o escolhido por apresentar maior número de cetano (48) que o S-500 (42) e por oferecer a qualquer veículo a diesel, mesmo os fabricados antes de 2012, uma melhor conservação do motor e redução dos custos de manutenção. Para compor as misturas ternárias, o biodiesel de soja foi eleito como representante do biodiesel nas misturas, por ser uma fonte oleaginosa muito comum na produção desse biocombustível. Na Tabela 4 são apresentadas as composições (% v/v) das 16 misturas ternárias entre o Diesel S-10, biodiesel de soja e etanol anidro.

Tabela 4 - Composição das misturas ternárias de diesel-biodiesel-etanol anidro

Diesel (%)	Biodiesel (%)	Etanol anidro (%)
88,0	10,0	2,0
85,0	10,0	5,0
83,0	10,0	7,0
80,0	10,0	10,0
83,0	15,0	2,0
80,0	15,0	5,0
78,0	15,0	7,0
75,0	15,0	10,0
78,0	20,0	2,0
75,0	20,0	5,0
73,0	20,0	7,0
70,0	20,0	10,0
73,0	25,0	2,0
70,0	25,0	5,0
68,0	25,0	7,0
65,0	25,0	10,0

Diesel S-10 (10 ppm de enxofre), S-500 (500 ppm de enxofre) e diesel padrão (puro) foram utilizados na composição das misturas. Na Tabela 5 são apresentadas as 34 misturas binárias diesel-biodiesel que foram preparadas em triplicata, resultando em um total de 102 amostras.

Tabela 5 - Composição das misturas binárias de diesel-biodiesel

Diesel (%)			Biodiesel (%)					
S-10	S-500	Padrão	Soja	Canola	Girassol	Milho	Região Sul	Rio
90,0	-	-	10,0	-	-	-	-	-
85,0	-	-	15,0	-	-	-	-	-
80,0	-	-	20,0	-	-	-	-	-
75,0	-	-	25,0	-	-	-	-	-
90,0	-	-	-	10,0	-	-	-	-
85,0	-	-	-	15,0	-	-	-	-
80,0	-	-	-	20,0	-	-	-	-
75,0	-	-	-	25,0	-	-	-	-
90,0	-	-	-	-	10,0	-	-	-
85,0	-	-	-	-	15,0	-	-	-
80,0	-	-	-	-	20,0	-	-	-
75,0	-	-	-	-	25,0	-	-	-
90,0	-	-	-	-	-	10,0	-	-
85,0	-	-	-	-	-	15,0	-	-
80,0	-	-	-	-	-	20,0	-	-
75,0	-	-	-	-	-	25,0	-	-
90,0	-	-	-	-	-	-	10,0	-
85,0	-	-	-	-	-	-	15,0	-
80,0	-	-	-	-	-	-	20,0	-
75,0	-	-	-	-	-	-	25,0	-
-	90,0	-	10,0	-	-	-	-	-
-	85,0	-	15,0	-	-	-	-	-
-	90,0	-	-	10,0	-	-	-	-
-	85,0	-	-	15,0	-	-	-	-
-	90,0	-	-	-	10,0	-	-	-
-	85,0	-	-	-	15,0	-	-	-
-	90,0	-	-	-	-	10,0	-	-
-	85,0	-	-	-	-	15,0	-	-
-	-	90,0	10,0	-	-	-	-	-
-	-	85,0	-	-	-	-	-	15,0
80,0	-	-	-	-	-	-	-	20,0
70,0	-	-	-	-	-	-	-	30,0

Os ANEXOS A e B deste trabalho apresentam todas as amostras e suas respectivas composições, bem como as propriedades medidas. No total, o Banco 2 é constituído por 334 amostras, sendo 149 amostras de biodiesel do Banco 1, mais 17 amostras de biodiesel, 18 amostras de diesel, 102 amostras de diesel-biodiesel e 48 amostras de diesel-biodiesel-etanol.

## 5.4 Medição das propriedades

As propriedades físico-químicas medidas no presente estudo são massa específica, índice de refração, ponto de entupimento de filtro a frio (PEFF) e viscosidade cinemática à 40°C.

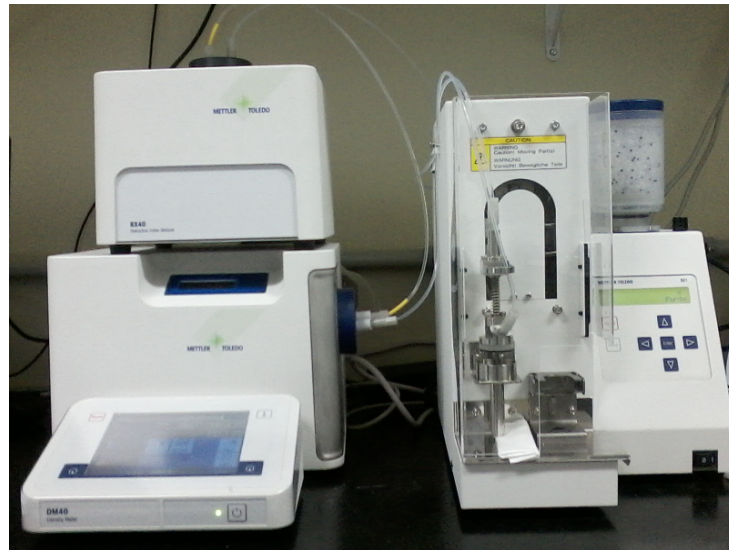
Índice de refração e massa específica são propriedades que apresentam contribuições lineares com as variáveis espectrais e por este motivo podem ser normalmente previstas utilizando regressão a partir de métodos lineares, como o PLS. Por outro lado, as propriedades PEFF e viscosidade são propriedades físicas que apresentam comportamento não linear e por este motivo são melhores modeladas utilizando ferramentas não lineares, como SVM e florestas aleatórias. Este comportamento é observado no estudo de Cunha *et al.* (2017).

### 5.4.1 Índice de refração e massa específica

Os ensaios de massa específica e índice de refração foram realizados simultaneamente no densímetro DM40 acoplado ao refratômetro RX40 com amostrador automático SC1, equipamentos da marca Mettler Toledo conforme apresentado na Figura 6. O procedimento de análise baseia-se na adição de aproximadamente 10 mL da amostra no frasco que acompanha o equipamento, que deve ser acomodado adequadamente no amostrador automático. Seleciona-se o método de análise e iniciam-se as medições. O método de análise foi construído com as seguintes características: temperatura de amostragem de 20°C; com configuração do amostrador SC1 ajustado para operar com velocidade automática, taxa de enchimento de 120% e máxima duração do enchimento de 30 segundos; a limpeza da linha foi ajustada para realizar a lavagem com dois solventes (álcool e acetona).

Para ambas as propriedades (massa específica e índice de refração) os resultados foram apresentados com 4 casas decimais, sendo que a unidade resposta para a massa específica foi  $\text{g}\cdot\text{cm}^{-3}$ .

Figura 6 - Densímetro (DM40) acoplado ao refratômetro (RX40) com amostrador automático SC1.



Fonte: O autor, 2019.

#### 5.4.2 Ponto de entupimento de filtro a frio

O ponto de entupimento de filtro a frio (PEFF) é a temperatura, em °C, na qual um determinado volume de amostra deixa de passar através de um filtro metálico padrão, num período de tempo específico, quando resfriado sob certas condições.

Os ensaios de PEFF foram realizados no equipamento da marca TANAKA, modelo AFP-102, conforme apresentado na Figura 7. O método baseia-se em resfriar, com uma taxa de  $1^{\circ}\text{C}\cdot\text{min}^{-1}$ , um volume de 45 ml de amostra, que é succionado para uma pipeta através de um filtro de malha metálico padronizado, sob vácuo controlado. Esse procedimento é repetido tantas vezes quanto possível até que a quantidade de cristais que se separam da solução seja suficiente para interromper ou reduzir a circulação da amostra pelo filtro. Alternativamente, este procedimento é repetido se o tempo necessário para preencher a pipeta ultrapasse 60 segundos ou ainda que a amostra deixe de retornar completamente para o recipiente de ensaio, antes de ser resfriado mais  $1^{\circ}\text{C}$ . A temperatura na qual a última filtração foi iniciada é o ponto de entupimento de filtro a frio.

Figura 7 - Equipamento para ensaio do ponto de entupimento de filtro a frio, marca TANAKA modelo AFP-102.



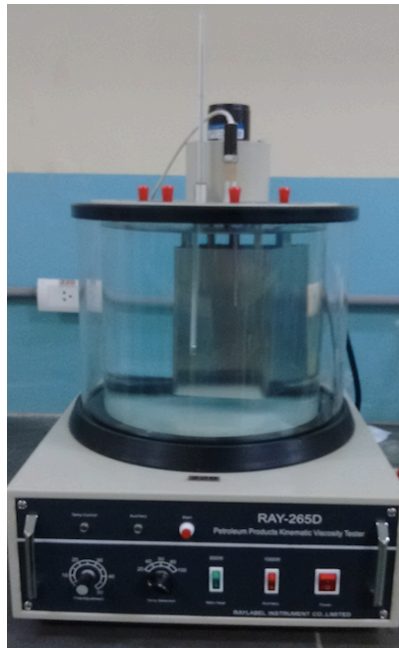
Fonte: O autor, 2019.

#### 5.4.3 Viscosidade cinemática a 40°C

A viscosidade cinemática é uma propriedade que mede a resistência ao escoamento sob ação da gravidade de uma certa massa de fluido em relação ao seu volume, que em outras palavras pode ser entendido como sendo a razão entre a viscosidade dinâmica e a massa específica do fluido.

Durante as medidas de viscosidade, a amostra foi homogeneizada no frasco original e, um alíquota de 10 ml foi filtrada e transferida via filtro-seringa (PTFE 0,2 $\mu$ m e 25mm) para o viscosímetro capilar do tipo Cannon-Fenske. Com o auxílio de uma pêra, a amostra foi succionada acima da linha superior, e o tempo gasto para o menisco superior passar sucessivamente pelas duas marcas de calibração foi anotado. Foram realizadas 10 determinações desse tempo. Na Figura 8 podemos observar o equipamento onde foram realizadas as medições de viscosidade cinemática à 40°C.

Figura 8 - Viscosímetro utilizado nas medições de viscosidade cinemática à 40°C, modelo Ray-265D.



Fonte: O autor, 2019.

## 5.5 Medição dos espectros de infravermelho

Os espectros de infravermelho médio, próximo e Raman dos combustíveis e bicomcombustíveis puros e de suas misturas suas misturas serão medidos.

### 5.5.1 Espectros de infravermelho médio

Os espectros de infravermelho médio (Mid-IR) foram adquiridos na faixa de 7800 a 450  $\text{cm}^{-1}$ , com intervalos de dados de 1  $\text{cm}^{-1}$  e 20 varreduras, utilizando um acessório de reflectância total atenuada (ATR) horizontal com cristal de seleneto de zinco (ZnSe) da PIKE Technologies acoplado a um espectrofotômetro de infravermelho da marca Perkin Elmer modelo FT-IR/NIR Spectrometer Frontier (Figura 9).

As amostras das misturas dos diferentes óleos de biodiesel foram analisadas colocando-se 0,2 mL de cada mistura no porta-amostra (cristal). Foram obtidos os espectros de infravermelho médio das misturas e do biodiesel puro.

Para a limpeza da superfície do cristal foi usado lenços de papel umedecido em água destilada, seguido por lenço de papel umedecido em álcool etílico e após a

evaporação do álcool etílico foi adquirido o espectro do branco para certificar-se da ausência de resíduos e ou de contaminantes no cristal.

Figura 9 - Espectrofotômetro de infravermelho com acessório ATR.



Fonte: O autor, 2019.

### 5.5.2 Espectros de infravermelho próximo

Os espectros de NIR foram adquiridos na faixa de  $10000$  a  $4000\text{ cm}^{-1}$ , com intervalos de dados de  $1\text{ cm}^{-1}$ , 20 varreduras e espessura do filme da amostra de  $0,224\text{ mm}$ . Utilizou-se um acessório para transfectância no infravermelho próximo (NIRA) com cristal de seleneto de zinco (ZnSe) da Perkin-Elmer acoplado a um espectrofotômetro de infravermelho, modelo FT-IR/NIR Spectrometer Frontier, do mesmo fabricante (Figura 10).

O porta-amostra utilizado nas medições consiste de uma pequena placa de Petri e um limitador da espessura do filme da amostra. A limpeza do porta-amostra foi realizada com uma lavagem com água e detergente, seguida de um enxague com água destilada, álcool etílico e colocado em estufa para secar.



Figura 10 - Espectrofotômetro de infravermelho com acessório NIRA.



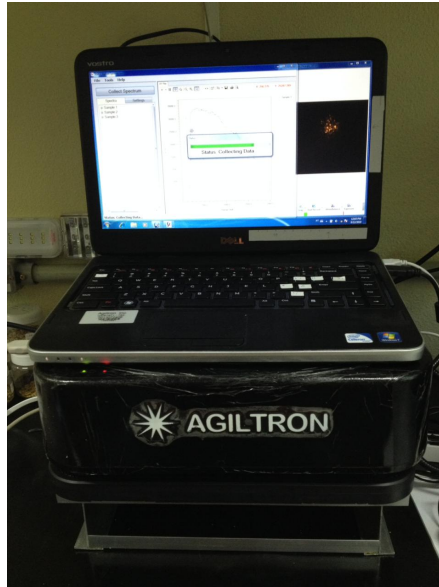
Fonte: O autor, 2019.

### 5.5.3 Espectros de Raman

Os espectros de Raman foram adquiridos em um espectrômetro dispersivo com janela espectral de 180 a 2000  $\text{cm}^{-1}$ , com intervalo de dados de 2  $\text{cm}^{-1}$ , comprimento de onda de excitação do laser em 785 nm e potência do laser em 20 mW, do fabricante Agiltron modelo PeakSeeker PRO-785 (Figura 11).

As amostras foram acondicionadas em tubos de eppendorf translúcidos, de fundo chato e capacidade de 2 mL. O acessório de amostragem utilizado nas medições foi um porta-frasco para líquidos e pós, que consiste em um suporte que permite o posicionamento preciso da sonda de fibra óptica sobre amostras.

Figura 11 - Espectrofotômetro de Raman



Fonte: O autor, 2019.

## 6 RESULTADOS E DISCUSSÃO

### 6.1 BANCO 1: Previsão de massa específica e índice de refração em amostras de biodiesel

A previsão da massa específica e índice de refração foram realizadas com os dados espectrais de NIR, mid-IR e Raman obtidos das amostras que compõe o Banco 1. O Banco 1 é constituído por amostras de biodiesel produzidas a partir de 4 oleaginosas diferentes (canola, girassol, milho e soja) e um biodiesel da região sul. No total, para o Banco 1, são 40 amostras de biodiesel puro, 61 amostras binárias, 27 amostras ternárias e 21 amostras quaternárias, totalizando 149 amostras, conforme apresentado no ANEXO A.

#### 6.1.1 Caracterização do biodiesel

Para a caracterização do biodiesel foram realizadas medições das propriedades: massa específica e índice de refração. Os espectros de NIR e mid-IR foram obtidos em um estudo anterior (CUNHA, 2014) e os espectros Raman das amostras puras e misturas dos biocombustíveis foram obtidos nessa etapa do estudo.

Os resultados das medidas de massa específica a 20°C, para os cinco tipos de biodiesel puro, foram comparados com os resultados relatados na literatura e são mostrados na Tabela 6.

Os resultados para as misturas binárias, ternárias e quaternárias são apresentados em tabela no ANEXO A. A repetibilidade do método de medição da massa específica foi de  $1,6 \cdot 10^{-4} \text{ g} \cdot \text{cm}^{-3}$ , para um nível de confiança de 95%, e a reprodutibilidade de  $5,2 \cdot 10^{-4} \text{ g} \cdot \text{cm}^{-3}$ . O intervalo de variação observado para o conjunto de dados em consideração foi de 0,8776 a 0,8821  $\text{g} \cdot \text{cm}^{-3}$  com erro de medição igual a  $3,1 \cdot 10^{-4} \text{ g} \cdot \text{cm}^{-3}$ .

Os resultados das medidas de índice de refração a 20°C, para os cinco tipos de biodiesel puro são exibidos na Tabela 7. Para as misturas binárias, ternárias e quaternárias os resultados são apresentados no material ANEXO A. A repetibilidade do método de medição do índice de refração foi de  $5,0 \cdot 10^{-5}$  e a reprodutibilidade foi de  $1,0 \cdot 10^{-4}$  para um nível de confiança de 95%. O intervalo de variação observado para o conjunto de dados em consideração foi de 1,4534 a 1,4575 com erro de medição igual a  $2,6 \cdot 10^{-4}$ .

Tabela 6 - Resultados das medidas de massa específica dos cinco tipos de biodiesel puro comparado com os resultados da literatura.

Experimental		Literatura	
Descrição	Média $\rho(\text{g}\cdot\text{cm}^{-3})$	$\rho(\text{g}\cdot\text{cm}^{-3})$	Referência
Biodiesel de Soja	0,8815	0,8820	Hoekman <i>et al.</i> , 2012
Biodiesel de Canola	0,8791	0,8830	Hoekman <i>et al.</i> , 2012
Biodiesel de Milho	0,8798	0,8830	Hoekman <i>et al.</i> , 2012
Biodiesel de Girassol	0,8814	0,8780	Hoekman <i>et al.</i> , 2012
Biodiesel Região Sul	0,8777	-	-

Tabela 7 - Resultados das medidas de índice de refração dos cinco tipos de biodiesel puro.

Experimental	
Descrição	Média do Índice Refração
Biodiesel de Soja	1,4570
Biodiesel de Canola	1,4555
Biodiesel de Milho	1,4559
Biodiesel de Girassol	1,4573
Biodiesel Região Sul	1,4535

### 6.1.2 Análise dos dados

O plataforma estatística R (R Core Team), versão 3.6.1, e R Studio (R Core Team), versão 1.2.1335, incluindo os pacotes *{stats}*, *{base}*, *{mvoutliers}*, *{pls}*, *{mdatools}*, *{prospectr}*, *{e1071}*, *{randomForest}*, *{ellipse}* e outros, foram aplicados para tratamentos dos dados, análises estatísticas, identificação de valores anômalos, divisão do conjunto de dados em treinamento, análise exploratória por PCA, métodos de seleção de variáveis, e métodos de regressão lineares (PLS, enPLS) e não-lineares (SVM, Florestas Aleatórias). Todos os comandos, pacotes e respectivas métricas utilizados neste trabalho estão listados na Tabela 8.

Neste trabalho, espectros no infravermelho, próximo (NIR) e médio (mid-IR), e Raman de biodiesel puro de cinco fontes distintas e suas misturas (binária, ternária e

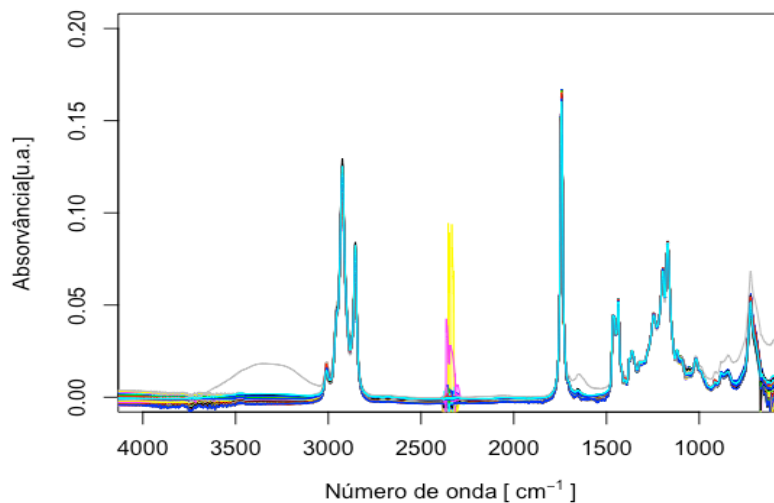
quaternária) foram obtidos na região espectral de 10000 a 4000  $\text{cm}^{-1}$ , 4000 a 680  $\text{cm}^{-1}$ , e 201 a 1800  $\text{cm}^{-1}$ , respectivamente.

Tabela 8 - Lista das ferramentas, comandos, pacotes e respectivas métricas utilizados neste trabalho.

Ferramentas	Tipo de Ferramenta	Comando	Pacote	Métrica	Referência
Espectros	Regiões com informação	surveySpectra	{ChemoSpec}	-	Hanson, 2019
SNV	Pré-processamento	standardNormal Variate	{prospectr}	-	Stevens e Ramirez-Lopez, 2013
Correção de linha de base	Pré-processamento	baseline	{baseline}	method="irls"	Liland e Mevik, 2015
Centralização na média	Pré-processamento	scale	{base}	scale=FALSE; center=TRUE	R core team, 2019
PCA	Análise exploratória	pca	{mdatools}	scale = FALSE; alpha=0.05; method="svd"; cv=10	Kucheryavskiy, 2015
1ª Derivada	Pré-processamento	diff	{base}	differences = 1	R core team, 2019
PLS	Regressão de 1ª ordem	mvr	{pls}	method="simpls"; ncomp=10	Mevik et al., 2015
PCA Robusto	Análise exploratória	PcaHubert	{rrcov}	-	Todorov e Filzmoser, 2009
Kennard-Stone	Seleção de amostras	kenStone	{prospectr}	metric="mahal; k=100	Stevens e Ramirez-Lopez, 2013
Validação cruzada	Validação cruzada	crossval	{pls}	segments = 10; segment.type = "interleaved"	Mevik et al., 2015
RMSEC/ RMSECV/ RMSEP	Parâmetros	RMSEP	{pls}	-	Mevik et al., 2015
$R^2_{cal} / R^2_{cv} / R^2_{val}$	Parâmetros	R2	{pls}	-	Mevik et al., 2015
LW-PLS/ RC-PLS/ VIP-PLS	Seleção de variáveis	filterPLSR	{plsVarSel}	ncomp=10; LW.threshold=0.1; RC.threshold=1; VIP.threshold=1	Mehmood et al., 2012
BVE-PLS	Seleção de variáveis	bve_pls	{plsVarSel}	ncomp=10; ratio=0.75; VIP.threshold=1	Mehmood et al., 2012
UVE-PLS	Seleção de variáveis	mcuve_pls	{plsVarSel}	N=3; ncomp=10; ratio=0.75; MCVUE.threshold=1	Mehmood et al., 2012
GA-PLS	Seleção de variáveis	ga_pls	{plsVarSel}	iter=5; popsize=100; GA.threshold=10	Mehmood et al., 2012
IPW-PLS	Seleção de variáveis	ipw_pls	{plsVarSel}	ncomp=10; no.iter=10; IPW.threshold=0.01	Mehmood et al., 2012
REP-PLS	Seleção de variáveis	rep_pls	{plsVarSel}	ncomp=10; ratio=0.75; N=3, REP.threshold=1	Mehmood et al., 2012
SPA-PLS	Seleção de variáveis	spa_pls	{plsVarSel}	ncomp=10; ratio=0.8; N=3, REP.threshold=0.01; Qv=50	Mehmood et al., 2012
iPLS	Seleção de variáveis	ipls	{mdatools}	ncomp=10; cv = 10; method='forward'	Kucheryavskiy, 2015
Elipse de confiança	deteção de erro sistemático	ellipse	{ellipse}	-	Murdoch e Chow, 2013
SVM	Regressão	tune	{e1071}	kernel=c('linear', 'radial', 'polynomial', 'sigmoid')	Meyer et al., 2018
Florestas Aleatórias	Regressão	randomForest	{randomForest}	-	Liaw e Wiener, 2002
enPLS	Regressão	enpls	{enpls}	-	Xiao et al., 2018

Para a espectroscopia de infravermelho médio por transformada de Fourier com reflectância total atenuada (FT-Mid-IR-ATR), os espectros foram obtidos para as 149 amostras no intervalo de 4000 a 680  $\text{cm}^{-1}$ , conforme pode ser observado na Figura 12. Dessa forma, inicialmente o tamanho da matriz de espectros mid-IR apresentou dimensão de 149 x 3321 (amostras *versus* variáveis).

Figura 12 - Espectros brutos de infravermelho médio (mid-IR) de amostras de biodiesel.



Fonte: O autor, 2019.

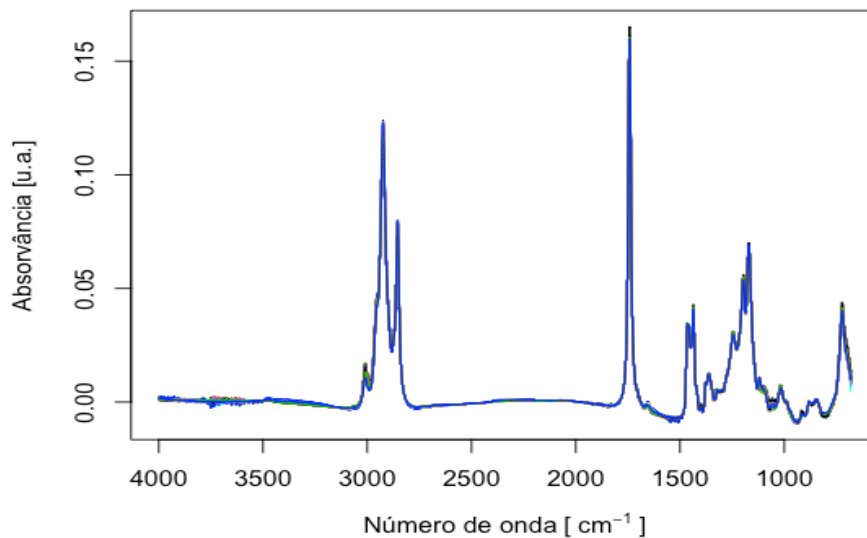
Como pode ser observado na Figura 12, há uma amostra (Am158) que apresenta comportamento distinto das amostras similares que compõem o conjunto de amostras triplicatas, por esse motivo foi necessário retirá-la do conjunto de dados. Este comportamento distinto foi atribuído como absorção de água, que provavelmente se deve a um erro na limpeza do equipamento entre as medidas das amostras. Para a identificação dessa amostra foi aplicado a ferramenta de detecção de amostras anômalas do pacote `{mvoutliers}` (FILZMOSEER E GSCHWANDTNER, 2018).

Além disso, o sinal observado na região espectral entre 2450 a 2200  $\text{cm}^{-1}$ , observado na Figura 12, foi identificado como sendo sinal de resposta do cristal de seleneto de zinco (ZnSe) do acessório de reflectância total atenuada (ATR) e não de bandas espectrais das amostras de biodiesel. Por este motivo, optou-se pela remoção desta região do espectro mid-IR. Com a remoção da amostra Am158 e a remoção da região espectral 2450 a 2200  $\text{cm}^{-1}$ , o conjunto de dados espectrais de mid-IR foi mantido com 148 amostras e 3072 variáveis.

A amostra Am158 foi removida do conjunto de dados espectrais de mid-IR e deve ser também retirada dos conjuntos espectrais de NIR e Raman, para que todos os modelos obtidos possuam as mesmas amostras e possam ser comparados pelo teste de Wilcoxon e elipse de confiança.

Na Figura 13 são apresentados os mesmos espectros mid-IR após a remoção da amostra Am158, a remoção da região espectral entre 2450 a 2200  $\text{cm}^{-1}$  e após a correção de linha de base.

Figura 13 - Espectros de infravermelho médio (mid-IR) após a retirada da amostra anômala e remoção da região espectral do sinal do cristal de ZnSe.



Fonte: O autor, 2019.

As regiões espectrais com os picos (Figura 13) foram identificadas na Tabela 9, evidenciando a presença de bandas de deformações axiais e angulares da ligação C-H, e as deformações axiais das ligações C=O e C-O, respectivamente.

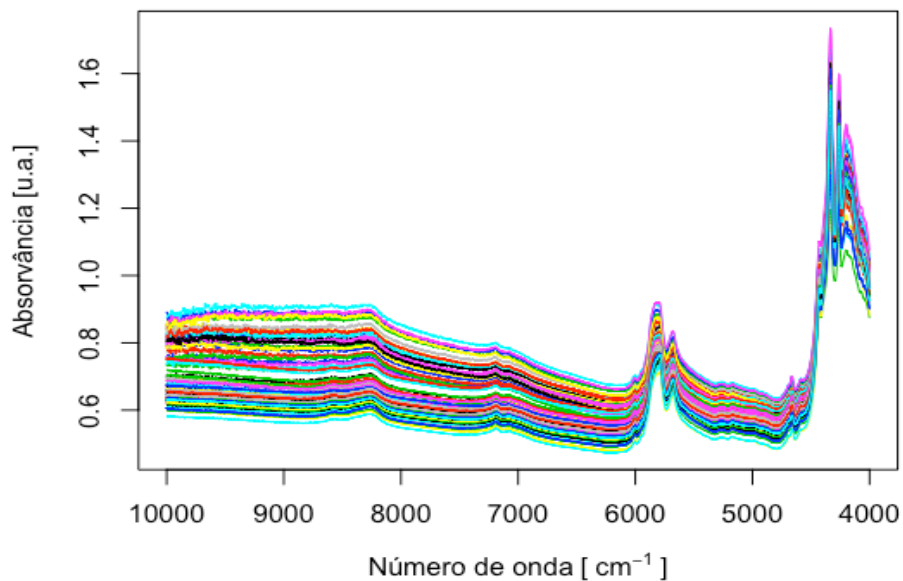
Tabela 9 - Identificação das regiões espectrais (mid-IR) das amostras de biodiesel.

Região	Nº de onda ( $\text{cm}^{-1}$ )*	Grupo provável	Classe dos Compostos
1	3000-2840	$\nu\text{C-H}$	Alcanos
2	1750-1730	$\nu\text{C=O}$	Éster
3	1500-1400	$\delta\text{C-H}$	Alcanos
4	1300-1000	$\nu\text{C-O}$	Ácidos carboxílicos, ésteres
5	750-700	$\delta\text{C-H}$	$(\text{CH}_2)_n$ ; $n \geq 4$

\*(Silverstein e Webster, 2000)

Para a espectroscopia de infravermelho de próximo foi utilizada a faixa de número de onda de 10000 a 4000  $\text{cm}^{-1}$ , com intervalo espectral de 1  $\text{cm}^{-1}$ . Na Figura 14, é apresentado o perfil espectral bruto obtido a partir de medidas de infravermelho próximo (NIR).

Figura 14 - Espectros brutos de infravermelho próximo (NIR) de amostras de biodiesel.



Fonte: O autor, 2019.

As regiões espectrais apresentadas na Figura 14 foram identificadas conforme mostrado na Tabela 10, evidenciando a presença combinação de bandas de deformações axiais das ligações C-H, O-H, além de deformações axiais da ligação C=O de derivados carbonilados.

Tabela 10 - Identificação das regiões espectrais (NIR) das amostras de biodiesel.

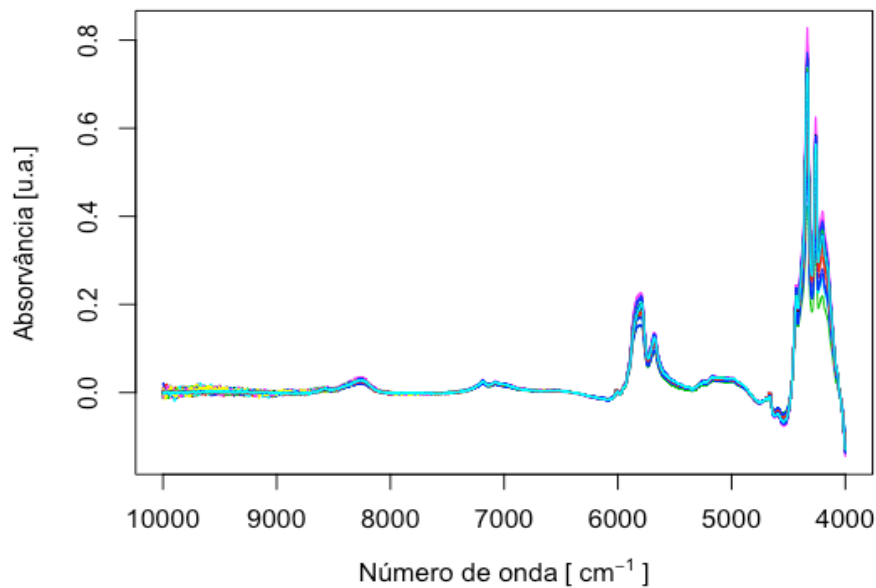
Banda	Nº de onda ( $\text{cm}^{-1}$ )*	Grupo provável	Classe das substâncias
1	9000-8000	3 $\nu$ C-H	Alcanos ( $\text{CH}_3/\text{CH}_2$ )
2	7500-6150	2 $\nu$ O-H	Ácido Carboxílico, álcoois (O-H)
3	5750-5700	2 $\nu$ C-H	Alcanos ( $\text{CH}_3/\text{CH}_2$ )
4	5350-4900	3 $\nu$ C=O	Derivados carbonilados
5	4370-4260	1ª Região de Combinação de C-H	Alcanos ( $\text{CH}_3/\text{CH}_2$ )

\*(WORKMAN, 2008)



Na Figura 15 são apresentados os mesmos espectros após o ajuste de linha de base e a remoção da amostra Am158. A matriz com os espectros NIR apresenta o sinal em absorvância de medidas de 6001 números de onda (10000 a 4000  $\text{cm}^{-1}$ ) para 148 amostras, formando uma matriz espectral de 148 x 6001 (amostras x variáveis).

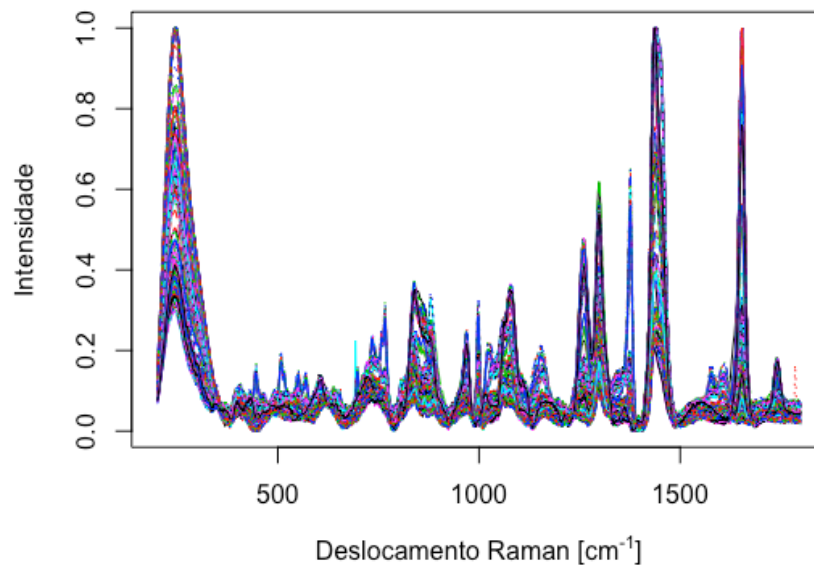
Figura 15 - Espectros de infravermelho próximo (NIR) de amostras de biodiesel após a correção de linha de base e remoção da amostra Am158.



Fonte: O autor, 2019.

Os espectros de Raman foram obtidos na faixa de número de onda de 201 a 1800  $\text{cm}^{-1}$ , com intervalo espectral de 2,03  $\text{cm}^{-1}$ . Na Figura 16, é apresentado o perfil espectral Raman após o ajuste de linha de base, a remoção da amostra Am158 e uma etapa de normalização. A matriz com os espectros Raman apresenta o sinal em 786 números de onda (201 a 1800  $\text{cm}^{-1}$ ) para 148 amostras, formando uma matriz espectral de 148 x 786 (amostras x variáveis).

Figura 16 - Espectros de Raman (RAM) de amostras de biodiesel após a correção de linha de base, remoção da amostra Am158 e normalização dos espectros.



Fonte: O autor, 2019.

### 6.1.3 Análise de componentes principais (PCA)

O estudo envolvendo a análise de componentes principais foi realizado para o banco de dados contendo os espectros de infravermelho próximo (NIR), médio (mid-IR) e Raman com o objetivo de realizar uma análise exploratória dos dados.

Neste estudo, diferentes pré-processamentos foram testados na matriz espectral durante a construção dos modelos, como centralização na média, derivada de 1ª ordem, padronização normal de sinal, além da transformação do sinal espacial (tss).

Para o banco de dados de espectros de infravermelho próximo, médio e Raman, o estudo foi realizado utilizando combinações de pré-processamentos, além do já aplicado correção de linha de base. Os pré-processamentos testados para a matriz dos espectros de NIR foram centralização dos dados na média (mc), 1ª derivada (drv) e padronização normal de sinal (snv). Na etapa de validação do modelo, utilizou-se o procedimento da validação cruzada randômica com a seleção de blocos com 10 divisões na métrica da ferramenta PCA clássico e PCA robusto (ROBPCA), o qual se mostrou adequado para o tamanho dos bancos.

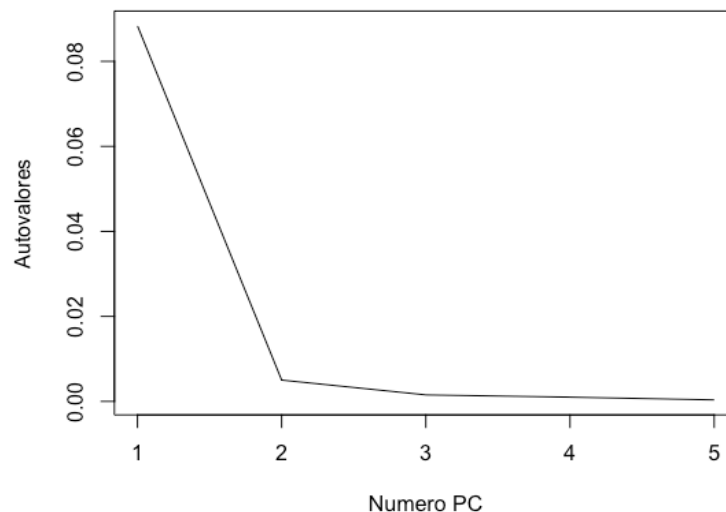
O modelo PCA para os dados espectrais de NIR foi construído usando a correção da linha de base (base) combinada com a centralização média (mc). A

Tabela 11 foi construída para comparar os autovalores e os valores de variância capturada (%) obtidos no modelo PCA para facilitar a escolha do número de componentes principais mais adequado para o modelo. Na Tabela 11 e Figura 17, pode-se observar que dois PCs são suficientes na construção do modelo PCA.

Tabela 11- Comparação autovalores e variância explicada do modelo PCA a partir dos espectros de NIR pré-processados com centralização na média

PC	Autovalores	Variância Capturada (%)	Variância Acumulada (%)
1	0.08887	87.86	87.86
<b>2</b>	<b>0.00504</b>	<b>4.98</b>	<b>92.84</b>
3	0.00153	1.51	94.35
4	0.00100	0.99	95.34
5	0.00035	0.34	95.68

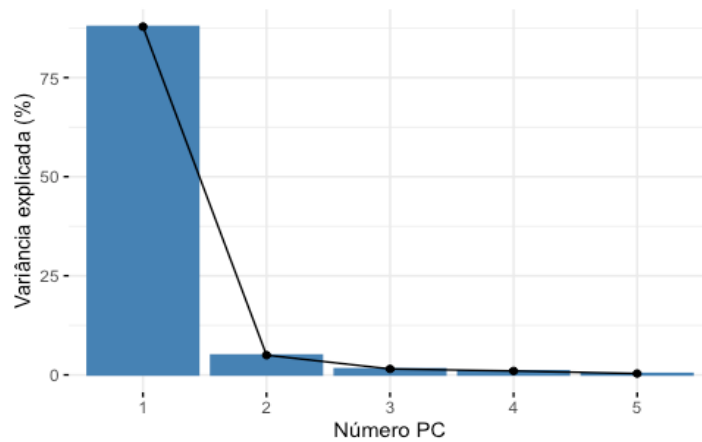
Figura 17 - Gráfico de autovalores *versus* número de componentes principais obtido pela PCA para os espectros NIR



Fonte: O autor, 2019.

A escolha de dois PCs obtida pelo gráfico autovalores *versus* número PC é confirmada com o gráfico *screeplot*, apresentado na Figura 18, que é uma ferramenta muito útil na escolha do número de componentes principais de modelos PCA.

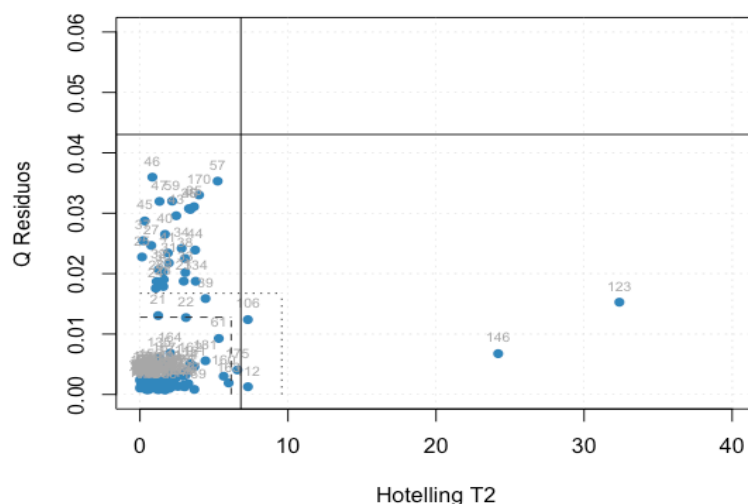
Figura 18 - Gráfico de variância capturada versus número de componentes principais obtido pela PCA para os espectros NIR



Fonte: O autor, 2019.

A Figura 19 apresenta o gráfico dos resíduos *versus* escores (Q Resíduos *versus* Hotelling T2), na qual observa-se que todas as amostras que apresentaram elevados valores de resíduos não indicaram elevada importância no modelo, e as amostras que indicaram elevada importância no modelo apresentaram baixos valores de resíduos. Desta forma, todas as amostras podem ser mantidas no modelo sem comprometer a robustez do mesmo e não será necessário aplicar a transformação de sinal espacial para mantê-las no modelo.

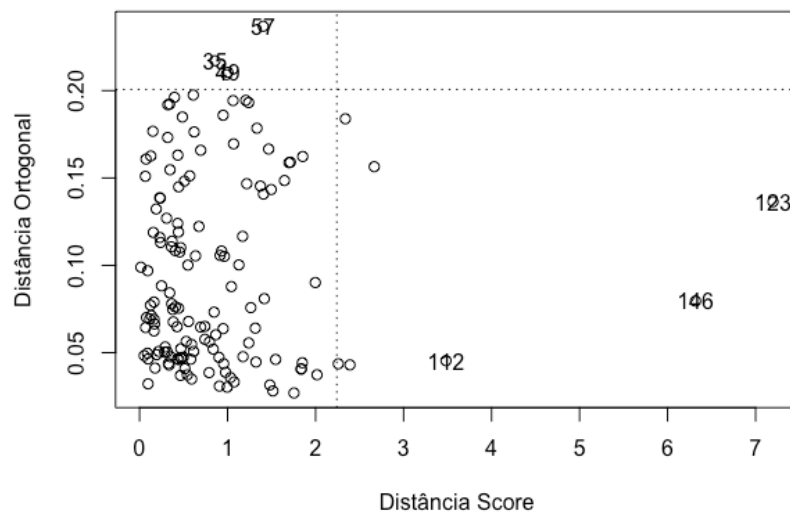
Figura 19 - Gráfico dos resíduos *versus* escores (Q resíduos *versus* Hotelling T2) obtido pela PCA para os espectros NIR do Banco 1



Fonte: O autor, 2019.

A Análise de Componentes Principais Robusta (ROBPCA) foi aplicado como uma ferramenta para garantir e corroborar o resultado obtido pelo PCA clássico. Na Figura 20 pode ser observado que nenhuma das amostras apresentaram uma combinação de altos valores distância ortogonal e distância entre os escores.

Figura 20 - Gráfico da combinação de altos valores distância ortogonal e distância entre os escores obtido pelo ROBPCA para os dados de NIR do Banco 1



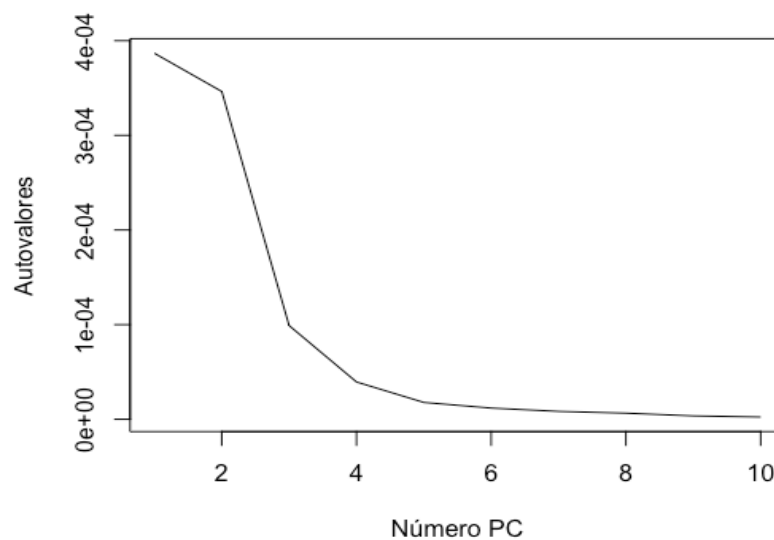
Fonte: O autor, 2019.

O modelo PCA para os dados espectrais de mid-IR foi construído usando a correção da linha de base (base) combinada com a centralização média (mc). A Tabela 12 foi construída para comparar os autovalores e os valores de variância capturada (%) obtidos no modelo PCA para facilitar a escolha do número de componentes principais mais adequado para o modelo. Para a escolha do número de componentes principais foi adotado como critério os autovalores e o gráfico *screeplot*. Na Tabela 12 e Figura 21, é apresentado que com quatro PCs são necessários na construção do modelo PCA. Esta quantidade de PCs foi confirmada pelo gráfico *screeplot*.

Tabela 12 - Comparação autovalores e variância explicada do modelo PCA a partir dos espectros de mid-IR pré-processados com centralização na média (Banco1)

PC	Autovalores	Variância Capturada (%)	Variância Acumulada (%)
1	3.87E-04	41.67	41.67
2	3.46E-04	37.31	78.98
3	9.90E-05	10.67	89.65
<b>4</b>	<b>3.94E-05</b>	<b>4.24</b>	<b>93.89</b>
5	1.78E-05	1.92	95.81
6	1.18E-05	1.27	97.08

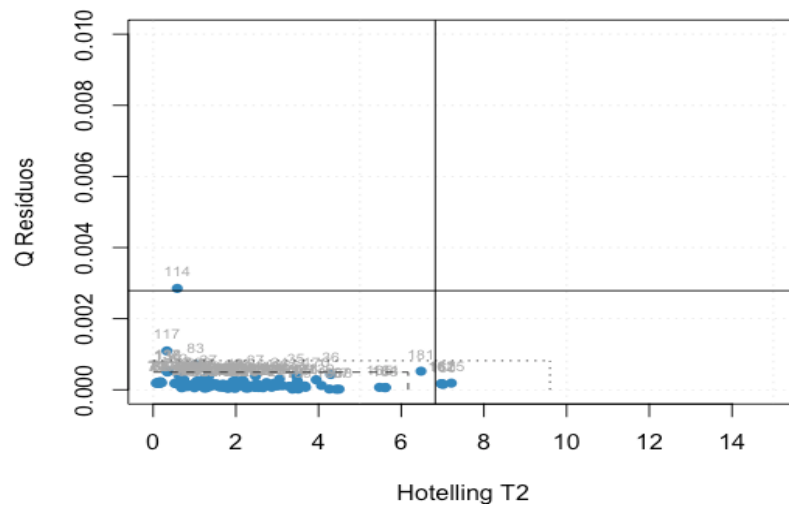
Figura 21 - Gráfico de autovalores *versus* número de componentes principais obtido pela PCA para os espectros mid-IR



Fonte: O autor, 2019.

A Figura 22 demonstra a ausência de amostras anômalas, na qual a maioria das amostras apresentaram baixos valores de resíduos, e as amostras que indicaram elevada importância no modelo apresentaram baixos valores de resíduos. Desta forma, todas as amostras foram mantidas no modelo.

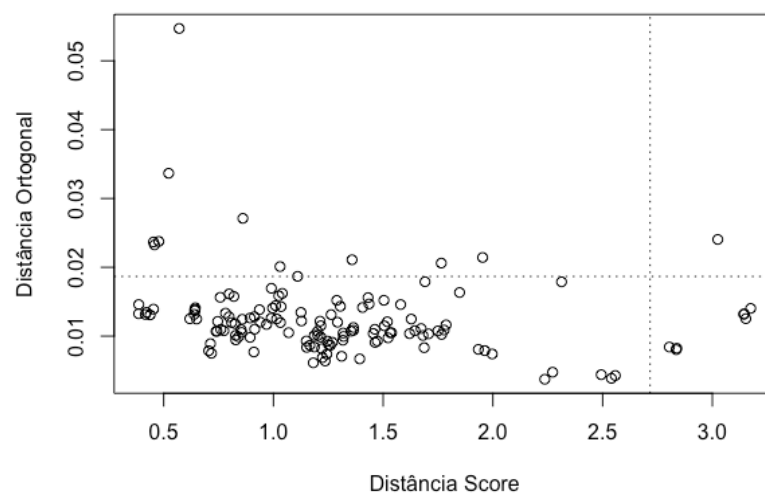
Figura 22 - Gráfico dos resíduos *versus* escores (Q resíduos *versus* Hotelling T2) obtido pela PCA para os espectros mid-IR do Banco 1



Fonte: O autor, 2019.

De modo a garantir o resultado obtido pelo PCA clássico, foi aplicado o algoritmo ROBPCA e o resultado é apresentado na Figura 23, onde as amostras apresentaram uma combinação de baixos valores distância ortogonal e distância escore, confirmando o resultado obtido pelo PCA clássico.

Figura 23 - Gráfico da combinação de altos valores distância ortogonal e distância entre os escores obtido pelo ROBPCA para os dados de Mid-IR do Banco 1



Fonte: O autor, 2019.

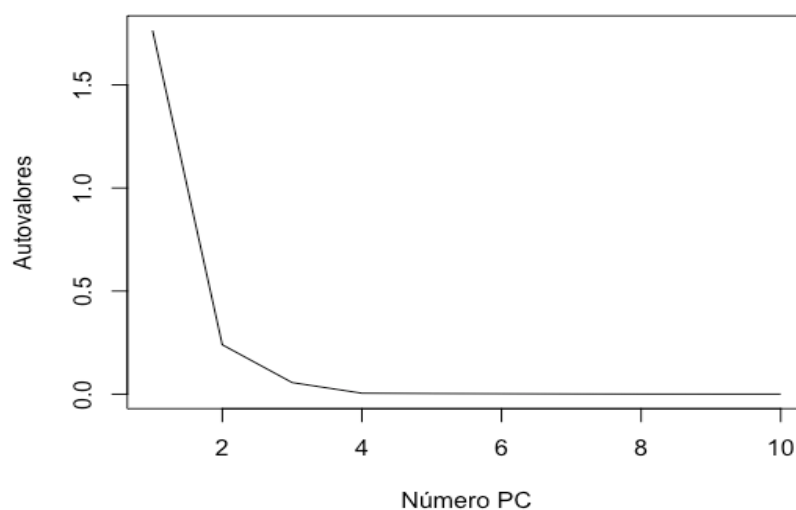
Para os espectros de Raman (RAM), o modelo PCA foi construído usando a correção da linha de base (base) combinada com a centralização média (mc). A

Tabela 13 foi construída para comparar os autovalores e os valores de variância capturada (%) obtidos no modelo PCA para facilitar a escolha do número de componentes principais mais adequado para o modelo. Para a escolha do número de componentes principais foi adotado como critério os autovalores. No caso deste estudo são necessários dois componentes principais conforme mostrado na Tabela 13 e no gráfico de autovalores *versus* número de componentes principais, Figura 24. A aplicação do gráfico *screeplot* confirmou que dois PCs são suficientes para o modelos PCA.

Tabela 13 - Comparação autovalores e variância explicada do modelo PCA a partir dos espectros de Raman pré-processados com centralização na média (Banco 1)

PC	Autovalores	Variância Capturada (%)	Variância Acumulada (%)
1	1.7645	84.76	84.76
<b>2</b>	<b>0.2398</b>	<b>11.52</b>	<b>96.28</b>
3	0.0564	2.71	98.99
4	0.0047	0.22	99.21
5	0.0032	0.15	99.36

Figura 24 - Gráfico de autovalores *versus* número de componentes principais obtido pela PCA para os espectros Raman

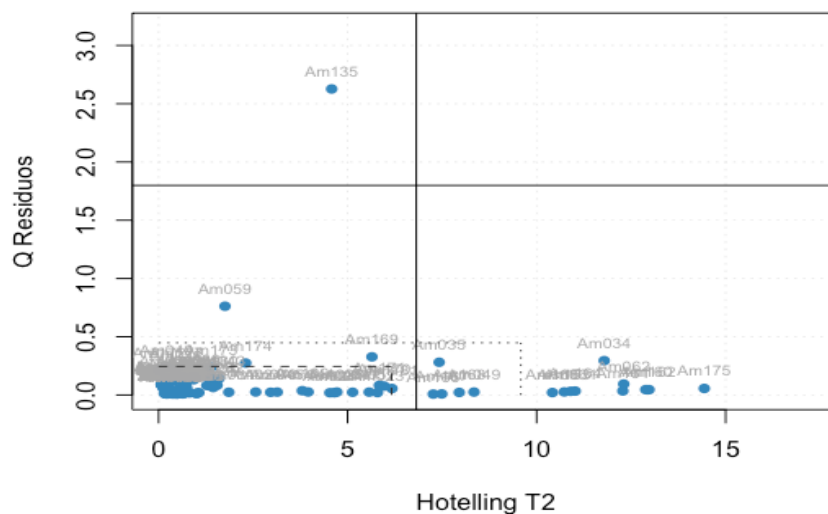


Fonte: O autor, 2019.



Na Figura 25 é observado a ausência de amostras anômalas, visto que a amostra que apresentou alto valor de resíduos possui baixa importância no modelo, e as amostras que apresentaram elevada importância no modelo apresentaram baixos valores de resíduos. Desta forma, nenhuma das amostras apresentaram comportamento anômalo segundo o PCA clássico.

Figura 25 - Gráfico dos resíduos *versus* escores (Q resíduos *versus* Hotelling T2) obtido pela PCA para os espectros Raman do Banco 1

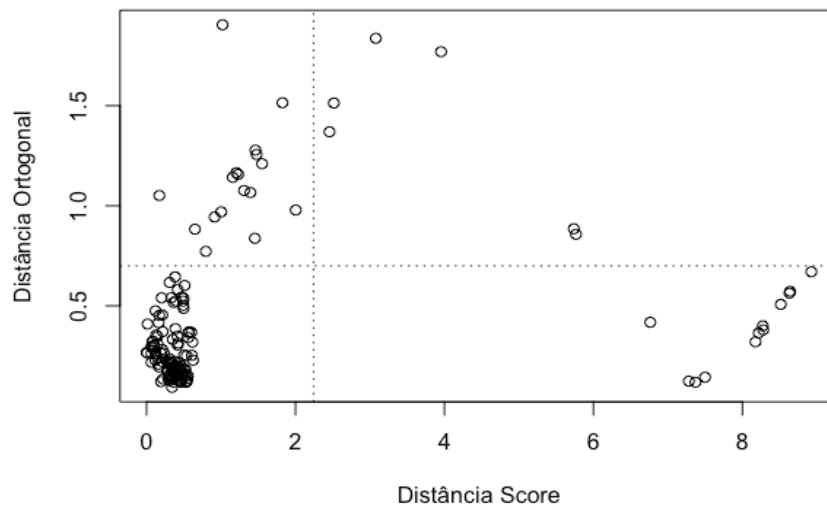


Fonte: O autor, 2019.

A PCA robusta foi aplicada para garantir o resultado obtido pelo PCA clássico, conforme Figura 26. No entanto, pode ser observado que algumas amostras apresentaram uma combinação de altos valores de distância ortogonal e distância entre os escores. Sendo assim, para os espectros de Raman do Banco 1 foi necessário realizar a transformação de sinal espacial (tss) para deslocar essas amostras para dentro conjunto de dados sem a necessidade de removê-las.

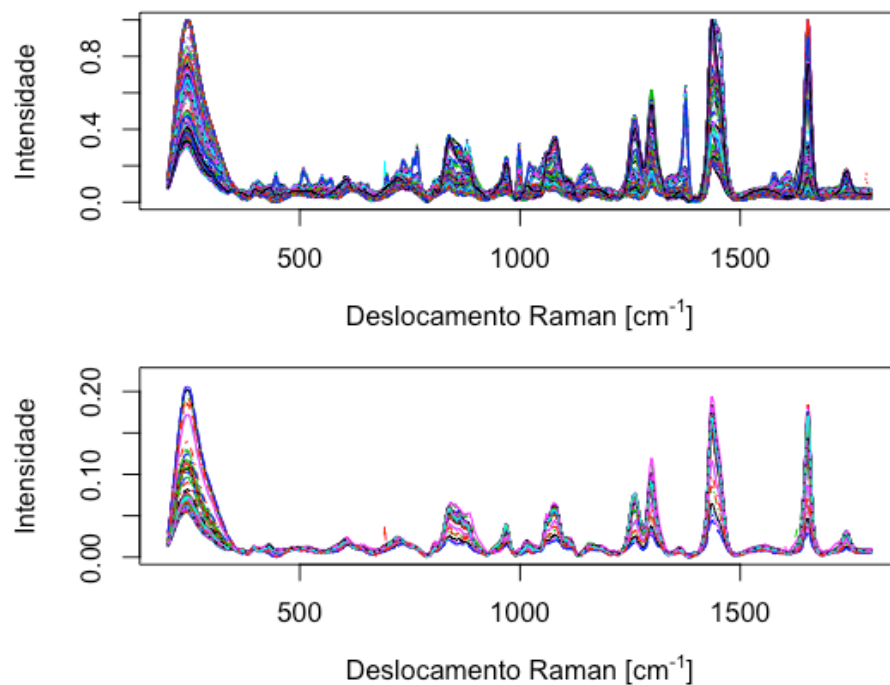
Os espectros de Raman antes e após a transformação de sinal espacial são apresentadas na Figura 27. Todos os modelos de regressão de previsão da massa específica e índice de refração contruídos a partir do conjunto espectral de Raman serão considerados após a transformação do sinal espacial.

Figura 26 - Gráfico da combinação de altos valores distância ortogonal e distância entre os escores obtido pelo ROBPCA para os dados de Raman do Banco 1



Fonte: O autor, 2019.

Figura 27 - Espectros de Raman do Banco 1 antes e após a transformação de sinal espacial (tss)



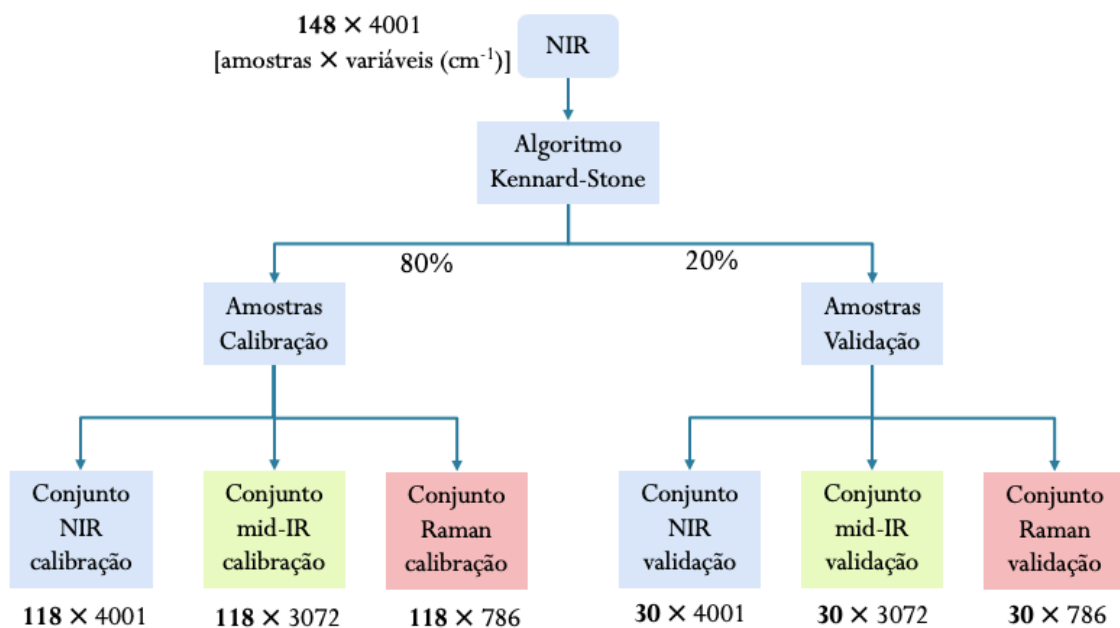
Fonte: O autor, 2019.

#### 6.1.4 Divisão do conjunto de amostras da calibração e validação

Após a PCA, as 148 amostras que compõe o Banco 1 foram divididas em dois conjuntos, a calibração (80% das amostras) e a validação (20% das amostras), utilizando o algoritmo de seleção Kennard-Stone (KS).

O algoritmo KS foi aplicado para escolher o conjunto de amostras de calibração e validação para o banco de dados. A partir do conjunto de dados espectrais de NIR, o algoritmo dividiu a amostra total no conjunto de calibração, responsável pela construção do modelo de regressão, e o conjunto de validação que visa avaliar a capacidade preditiva do modelo. As amostras foram divididas nos dois conjuntos de dados de acordo com esse algoritmo, 118 amostras para o conjunto de calibração (80%) e as 30 amostras restantes foram usadas para o conjunto de validação externa (20%). Conforme demonstrado no fluxograma da Figura 28, essa divisão das amostras foi aplicada aos dados espectrais de mid-IR e Raman, de modo a manter a equidade das amostras de calibração e validação e facilitar a comparação dos modelos.

Figura 28 - Fluxograma da divisão das amostras em conjunto de calibração e validação aplicando o algoritmo Kennard-Stone nos espectros de NIR

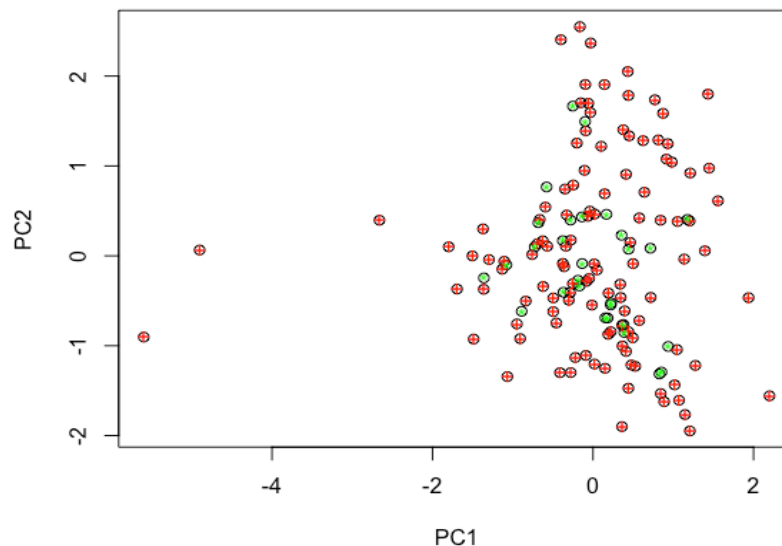


Fonte: O autor, 2019.

Após a divisão das amostras em calibração e validação, os conjunto de dados de calibração de NIR, mid-IR e Raman, passaram a ser matrizes de tamanho  $118 \times 6001$ ,  $118 \times 3072$  e  $118 \times 786$  (amostras  $\times$  variáveis), respectivamente. Os conjuntos de dados de validação de NIR, mid-IR e Raman passaram a ser matrizes de  $30 \times 6001$ ,  $30 \times 3072$  e  $30 \times 786$  (amostras  $\times$  variáveis), respectivamente.

Na Figura 29 pode ser observado como ficou a distribuição das amostras de calibração (vermelho) e validação (verde) no conjunto total de amostras do Banco 1, considerando os espectros NIR e o algoritmo de Kennard Stone.

Figura 29 - Distribuição das amostras de calibração (+) e validação (•) no conjunto total de amostras do Banco 1, considerando os espectros NIR



Fonte: O autor, 2019.

O conjunto de calibração e validação dos espectros (NIR, mid-IR e Raman) formados apresentaram 118 amostras e 30 amostras, respectivamente. Na Tabela 14 são mostradas quais amostras foram selecionadas para o conjunto de calibração e validação.

Tabela 14 - Amostras selecionadas pelo algoritmo Kennard-Stone para compor os conjuntos de calibração e validação para os espectros de NIR, mid-IR e Raman do Banco 1

Calibração	Validação
Am021 a Am029	Am030
Am031 a Am040	Am041
Am042 a Am064	Am066
Am067, Am068	Am069
Am070	Am071
Am072 a Am078	Am079
Am080, Am081	Am082
Am084, Am087	Am083
Am088, Am091	Am085
Am093 a Am104	Am086, Am089
Am106	Am090, Am092
Am109 a Am111	Am105
Am113, Am114	Am107, Am108
Am117 a Am127	Am112
Am129 a Am133	Am115, Am116
Am135 a Am137	Am134, Am138
Am139, Am140	Am141, Am142
Am143, Am145	Am144, Am147
Am146	Am161, Am167
Am148 a Am160	Am176, Am177
Am162 a Am166	Am180
Am168 a Am175	
Am181	

#### 6.1.5 Modelos de regressão

Um total de 294 modelos foram construídos, sendo 147 modelos para a propriedade índice de refração e 147 modelos para a propriedade massa específica. Todos os 294 modelos foram avaliados através da validação cruzada randômica com a seleção de blocos com 10 divisões.

Neste estudo, a qualidade dos modelos foram avaliados pelos parâmetros estatísticos: raiz quadrada do erro médio quadrático de calibração (RMSEC), de validação cruzada (RMSECV) e validação (RMSEP); erro-padrão relativo de validação (RSEP(%)); os coeficientes de determinação de calibração ( $R^2_{cal}$ ), de validação

cruzada ( $R^2_{cv}$ ) e de validação ( $R^2_{val}$ ); o viés de calibração ( $bias_{cal}$ ), de validação cruzada ( $bias_{cv}$ ) e de validação ( $bias_{val}$ ). Segundo Pomerantsev (2014), os modelos devem apresentar baixos valores de RMSEC e RMSEP, e com valores de bias muito próximos a zero.

Além disso, a comparação entre os modelos foi realizada com a aplicação do teste pareado de Wilcoxon. Devido a elevada quantidade de modelos de regressão obtidos para cada uma das propriedades, foi necessário modificar o teste de Wilcoxon de modo a torná-lo útil na comparação entre um número ilimitado de modelos. O teste pareado de Wilcoxon compara os resíduos entre dois modelos ( $M_i$  e  $M_j$ ) e retorna como resposta um valor de probabilidade entre 0 e 1, valor este que identifica qual dos modelos é melhor. Por exemplo, ao comparar o modelo  $M_i$  com o  $M_j$ , se o valor de probabilidade for perto de 0, significa que o primeiro modelo (neste caso,  $M_i$ ) é melhor que o segundo modelo ( $M_j$ ). Para valores próximos a 1, significa que o segundo modelo ( $M_j$ ) é melhor que o primeiro ( $M_i$ ).

Para este estudo, em uma nova abordagem do teste de Wilcoxon, uma matriz ( $n \times n$ ) com os valores de probabilidade resultantes do teste de Wilcoxon foi construída. Foi obtido um vetor contendo a Soma da Probabilidade do Teste de Wilcoxon (SWTP) de cada linha e o valor mínimo desse vetor revelou o melhor modelo para cada propriedade.

Antes de iniciar a construção dos modelos de calibração multivariada, as amostras foram separadas em conjuntos de calibração e validação. O algoritmo de Kennard-Stone foi utilizado para realizar a divisão dessas amostras, conforme descrito no tópico 6.1.4.

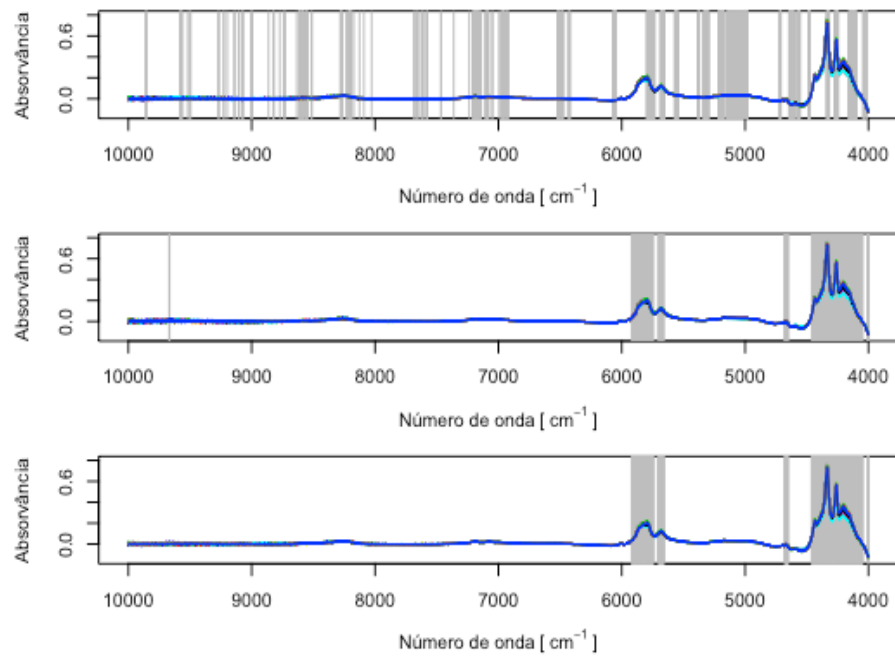
Depois da separação das amostras em calibração e validação, para os conjuntos de dados espectrais de NIR, mid-IR e Raman, foram adotados algumas estratégias para a construção dos modelos de previsão das propriedades, tais como a construção de 12 modelos de calibração multivariada de primeira ordem (PLS): i) utilizando todo o perfil espectral - PLS; ii) com seleção de variáveis por LW-PLS; iii) com seleção de variáveis por RC-PLS; iv) com seleção de variáveis por VIP-PLS; v) com seleção de variáveis por iPLS; vi) com seleção de variáveis por GA-PLS; vii) com seleção de variáveis por UVE-PLS; viii) com seleção de variáveis por BVE-PLS; ix) com seleção de variáveis por SPA-PLS; x) com seleção de variáveis por REP-PLS; xi) com seleção de variáveis por IPW-PLS; e xii) com seleção de variáveis pelo enPLS seguido pelo PLS.

Para o conjunto de NIR, os intervalos selecionados com base no conhecimento prévio das informações químicas, basearam-se nas cinco regiões espectrais apresentadas na Tabela 10: (1º) 9000-8000  $\text{cm}^{-1}$ , (2º) 7500-6150  $\text{cm}^{-1}$ , (3º) 5750-5700  $\text{cm}^{-1}$ , (4º) 5350-4900  $\text{cm}^{-1}$  e (5º) 4370-4260  $\text{cm}^{-1}$ , pois apresentaram informações sobre grupos funcionais característicos do biodiesel. Os intervalos selecionados para o conjunto mid-IR foram baseados pelas regiões espectrais mostradas na Tabela 9: (1º) 3000-2840  $\text{cm}^{-1}$ , (2º) 1750-1730  $\text{cm}^{-1}$ , (3º) 1500-1400  $\text{cm}^{-1}$ , (4º) 1300-1000  $\text{cm}^{-1}$  e (5º) 750-700  $\text{cm}^{-1}$ , pois apresentaram informações sobre grupos funcionais característicos do biodiesel. Para o conjunto de Raman não foram considerados os intervalos definidos por informações químicas.

A etapa de seleção de variáveis utilizando algoritmos foi realizada com o uso dos métodos: LW, RC, VIP (WOLD *et al.*, 1993; ERIKSSON *et al.*, 2001; MEHMOOD *et al.*, 2012; GODINHO *et al.*, 2014), BVE (FRANK, 1987), GA (HASEGAWA *et al.*, 1997), IPW (FORINA *et al.*, 1999), UVE (CENTNER *et al.*, 1996), REP (MEHMOOD *et al.*, 2011) e SPA (LI *et al.*, 2010). No emprego dos métodos VIP e BVE, o valor de limiar (VIP.threshold) adotado foi igual a 1, conforme indicado por Cassoti e Grisoni (2014). Deste modo todas as variáveis com escores maiores que o valor adotado do limiar são importantes para o modelo.

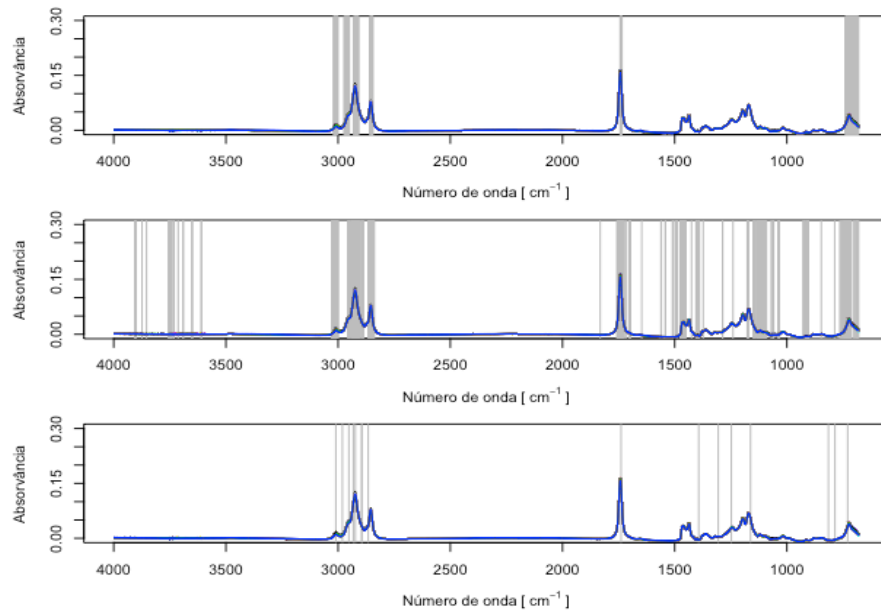
Nas Figuras 30, 31 e 32 estão representados alguns dos intervalos selecionados pelos algoritmos estudados para os espectros NIR, mid-IR e Raman, respectivamente. Como pode ser observado nas Figuras 30, 31 e 32, os intervalos selecionados pelos algoritmos VIP, BVE, IPW e iPLS estão distribuídos pelas regiões onde há bandas de absorção. Por outro lado, os algoritmos UVE e GA selecionaram variáveis distribuídos por todo o espectro.

Figura 30 - Intervalos espectrais do NIR selecionados pelos algoritmos UVE, BVE e VIP



Fonte: O autor, 2019.

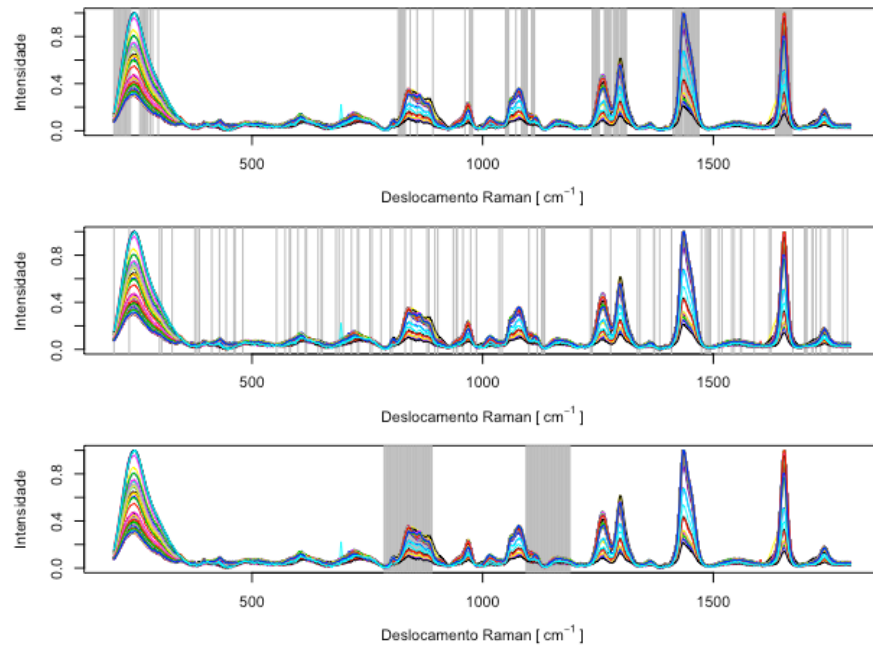
Figura 31 - Intervalos espectrais do mid-IR selecionados pelos algoritmos VIP, BVE e IPW



Fonte: O autor, 2019.



Figura 32 - Intervalos espectrais de Raman selecionados pelos algoritmos BVE, GA e iPLS



Fonte: O autor, 2019.

A Tabela 15 apresenta o número de variáveis selecionadas em cada método, para o conjunto de dados espectrais de NIR, Mid-IR e Raman. Alguns dos métodos empregados (RC, VIP, iPLS, GA, BVE, IPW e REP) conseguiram reduzir significativamente o número de variáveis dos três conjuntos de dados NIR, Mid-IR e Raman das propriedades índice de refração (IndRefra) e massa específica (MasEsp). No entanto, o método SPA não foi capaz de selecionar um número tão reduzido de variáveis para o conjunto de dados de MIR para ambas as propriedades (79,4%).

De acordo com a Tabela 15, para o conjunto de dados de Raman, o método LW não foi capaz de selecionar de selecionar uma quantidade reduzida de variáveis para ambas as propriedades índice de refração e massa específica, 79,5% e 76,8%, respectivamente.

Tabela 15 - Quantidade de variáveis selecionadas por cada método de seleção de variáveis para cada um dos sensores (NIR, mid-IR e RAM), para o índice de refração e massa específica do Banco 1

Propriedade	Sensor	Todas	LW	RC	VIP	iPLS	GA	UVE	BVE	SPA	REP	IPW
IndRefra	NIR	6001	1026	10	663	100	549	851	660	953	684	10
		100%	17,9%	0,2%	11,0%	1,7%	9,2%	14,2%	11,0%	15,9%	11,4%	0,2%
MaseEsp	NIR	6001	830	10	641	200	599	948	624	2115	645	28
		100%	13,8%	0,2%	10,7%	3,3%	10,0%	15,8%	10,4%	35,2%	10,7%	0,5%
IndRefra	Mid-IR	3072	792	10	148	51	342	742	358	2439	420	16
		100%	25,8%	0,3%	4,8%	1,7%	11,1%	24,2%	11,7%	79,4%	13,7%	0,5%
MaseEsp	Mid-IR	3072	415	10	115	100	317	505	366	2442	414	48
		100%	13,5%	0,3%	3,7%	3,3%	10,3%	16,4%	11,9%	79,5%	13,5%	1,6%
IndRefra	RAM	786	625	10	45	98	89	46	142	201	157	166
		100%	79,5%	1,3%	5,7%	12,5%	11,3%	5,9%	18,1%	25,6%	20,0%	21,1%
MaseEsp	RAM	786	604	10	40	98	77	497	146	201	159	207
		100%	76,8%	1,3%	5,1%	12,5%	9,8%	63,2%	18,6%	25,6%	20,2%	26,3%

#### 6.1.6 Modelos de previsão para o índice de refração

Os modelos PLS foram construídos considerando espectros completos e os intervalos selecionados pela seleção de variáveis com base em algoritmos e informações químicas. O critério R de Wold foi aplicado na escolha do número apropriado de variáveis latentes a serem incluídas em cada modelo de previsão. De acordo com Stone e Wold, o critério é baseado na validação cruzada, e sugere que uma variável latente (VL) adicional não será incluída no modelos PLS, a menos que ofereça melhorias significativas no modelo de previsão (STONE, 1974; WOLD, 1978; KRZANOWSKI, 1987).

Na Tabela 16 são apresentados o número de variáveis latentes e os parâmetros de mérito para todos os modelos de previsão do índice de refração construídos a partir dos conjuntos espectrais de NIR, mid-IR (MIR) e Raman (RAM), respectivamente.

A priori, os modelos que apresentaram RMSEP menor que o erro experimental das propriedades estudadas podem ser considerados adequados para a previsão dessas propriedades. Nas Figuras 33, 34 e 35, a linha vertical indica o erro experimental do índice de refração dos modelos de NIR, mid-IR e Raman, respectivamente. Assim, os modelos que estão à esquerda da linha vertical apresentam RMSEP menor que o erro experimental,  $2,6 \cdot 10^{-4}$  para índice de refração.

Entre os modelos com valor de RMSEP menor que o erro experimental, é necessário comparar os modelos e identificar o modelo com o menor valor residual entre os modelos estudados.

O teste de Wilcoxon pareado foi escolhido para comparar os valores residuais entre os modelos. Em uma nova abordagem, uma matriz ( $147 \times 147$ ) com os valores de probabilidade resultantes do teste de Wilcoxon foi construída. Foi obtido um vetor contendo todas as linhas de SWTP e, o valor mínimo desse vetor revelou o melhor modelo para índice de refração.

Tabela 16 - Parâmetros de mérito dos 10 melhores modelos PLS com seleção de variáveis e Florestas Aleatórias (RF) obtidos a partir dos espectros de NIR, mid-IR e Raman para a previsão do índice de refração

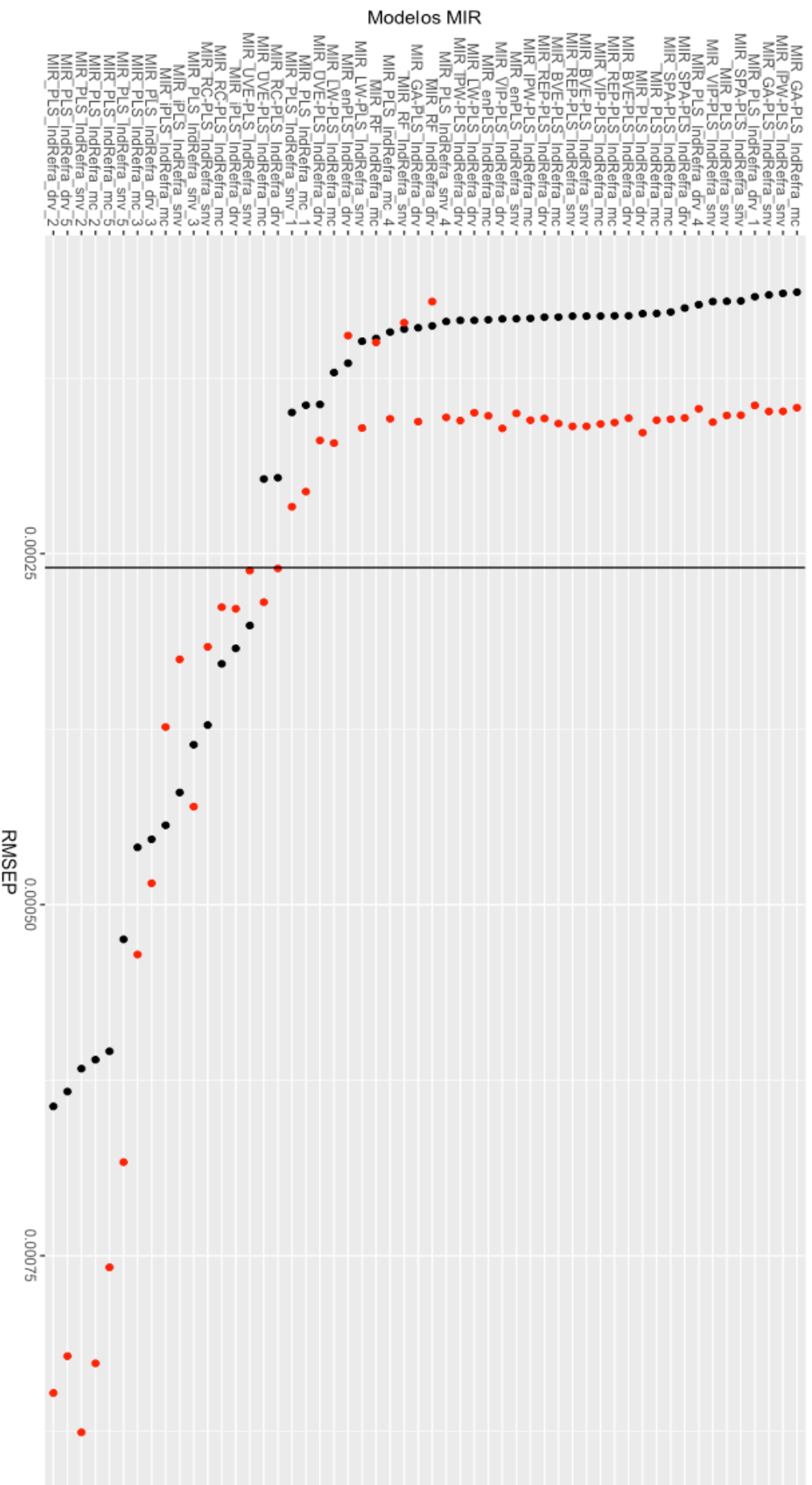
Modelos	Espectros	Intervalos	PPx	VL	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
<b>MIR_UVE-PLUS_IndRefra_drv</b>	<b>MIR</b>	<b>UVE</b>	<b>drv</b>	<b>5</b>	<b>1.69E-04</b>	<b>0.97</b>	<b>1.83E-04</b>	<b>0.97</b>	<b>1.44E-04</b>	<b>0.95</b>	<b>-3.33E-20</b>	<b>8.90E-08</b>	<b>6.46E-09</b>	<b>1.02E-03</b>	<b>1.00</b>
NIR_VIP-PLUS_IndRefra_snv	NIR	VIP	snv	3	1.63E-04	0.97	1.70E-04	0.97	8.01E-05	0.98	-3.26E-20	2.47E-08	1.52E-09	4.96E-04	4.80
NIR_REP-PLUS_IndRefra_snv	NIR	REP	snv	3	1.63E-04	0.97	1.71E-04	0.97	8.02E-05	0.98	1.66E-20	8.85E-10	1.40E-09	4.77E-04	4.83
NIR_PLS_IndRefra_snv	NIR	Todo	snv	4	1.54E-04	0.98	1.71E-04	0.97	7.18E-05	0.99	3.17E-20	2.49E-08	1.23E-10	1.41E-04	4.95
NIR_BVE-PLUS_IndRefra_snv	NIR	BVE	snv	3	1.63E-04	0.97	1.71E-04	0.97	8.04E-05	0.98	8.29E-21	6.27E-10	1.30E-09	4.60E-04	5.85
MIR_SPA-PLUS_IndRefra_drv	MIR	SPA	drv	4	1.53E-04	0.98	1.65E-04	0.97	7.51E-05	0.99	2.02E-20	1.15E-08	7.78E-10	3.56E-04	5.90
NIR_VIP-PLUS_IndRefra_mc	NIR	VIP	mc	4	1.63E-04	0.97	1.72E-04	0.97	8.30E-05	0.98	-1.49E-20	6.96E-08	2.74E-10	2.11E-04	9.53
MIR_PLS_IndRefra_drv_1	MIR	1	drv	4	1.44E-04	0.98	1.57E-04	0.97	6.71E-05	0.99	1.62E-21	4.96E-08	5.51E-10	2.99E-04	9.69
MIR_VIP-PLUS_IndRefra_snv	MIR	VIP	snv	4	1.56E-04	0.97	1.65E-04	0.97	7.05E-05	0.99	1.09E-20	1.37E-08	1.23E-09	4.47E-04	10.04
NIR_REP-PLUS_IndRefra_mc	NIR	REP	mc	4	1.62E-04	0.97	1.72E-04	0.97	8.29E-05	0.98	-3.91E-21	8.41E-10	6.38E-10	3.22E-04	11.06

Figura 33 - Modelos de previsão do índice de refração *versus* RMSEP (•) e RMSEC (•), obtidos a partir do conjunto espectros de NIR, a linha vertical indica o erro experimental



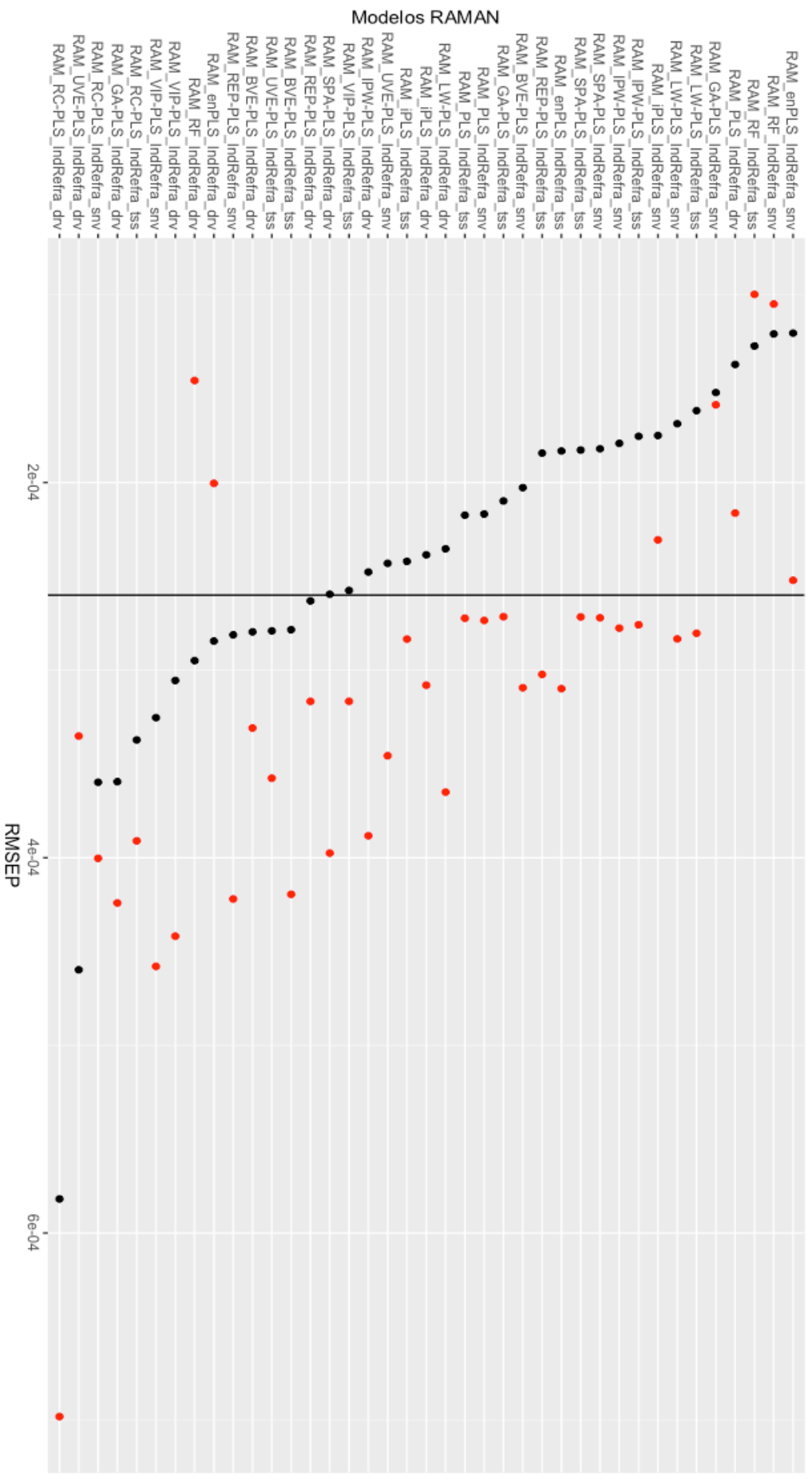
Fonte: O autor, 2019.

Figura 34 - Modelos de previsão do índice de refração *versus* RMSEP (•) e RMSEC (◦), obtidos a partir do conjunto espectros de mid-IR, a linha vertical indica o erro experimental



Fonte: O autor, 2019.

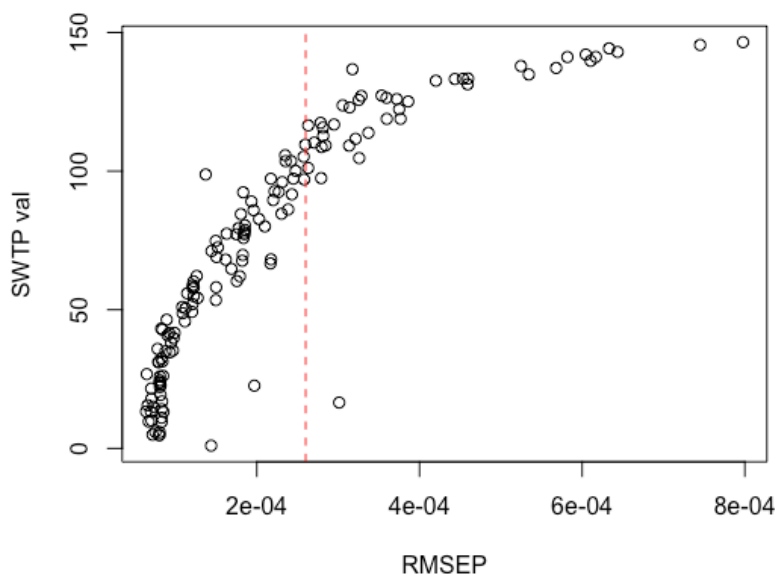
Figura 35 - Modelos de previsão do índice de refração *versus* RMSEP (•) e RMSEC (•), obtidos a partir do conjunto espectros de Raman, a linha vertical indica o erro experimental



Fonte: O autor, 2019.

A Figura 36 mostra o comportamento dos modelos ao comparar os valores  $SWTP_{val}$  com os valores RMSEP para o índice de refração.

Figura 36 - Soma da probabilidade do teste de Wilcoxon ( $SWTP$ ) *versus* RMSEP para todos os modelos de previsão do índice de refração, a linha vertical indica o erro experimental



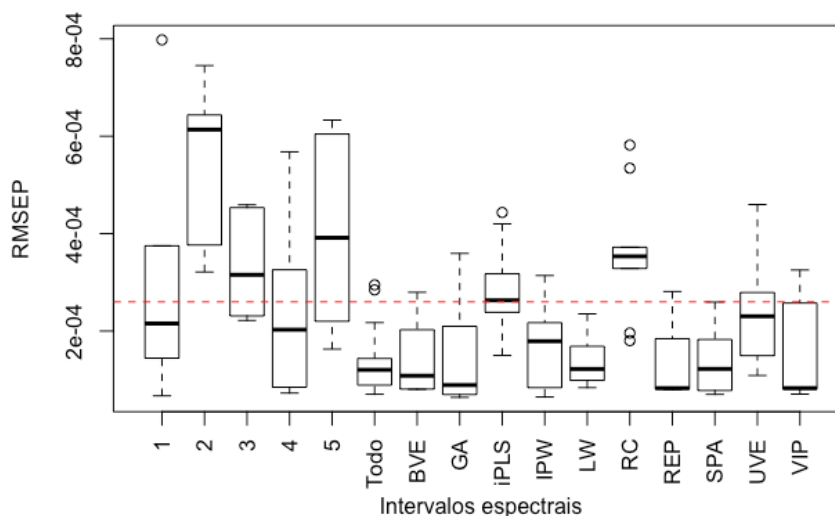
Fonte: O autor, 2019.

Os valores  $SWTP$  da validação para os 10 melhores modelos estão na Tabela 16, e o valor mínimo indica o melhor modelo entre os modelos estudados. No caso da previsão do índice de refração, o modelo PLS, utilizando os espectros de MIR e os intervalos obtidos pelo algoritmo UVE e pré-processamento da primeira derivada (MIR\_UVE-PLS\_IndRefra\_drv), apresentou o menor valor  $SWTP_{val}$  (1,00) entre os modelos avaliados.

Os próximos três modelos com baixos valores de  $SWTP_{val}$  são: NIR\_VIP-PLS\_IndRefra\_snv ( $SWTP_{val} = 4,80$ ) e NIR\_REP-PLS\_IndRefra\_snv ( $SWTP_{val} = 4,83$ ), ambos com 3 variáveis latentes, RMSEP igual a  $1,0 \cdot 10^{-4}$  e  $R^2_{pred}$  igual a 0,98; e NIR\_PLS\_IndRefra\_snv ( $SWTP_{val} = 4,95$ ) com 4 variáveis latentes, RMSEP igual a  $1,0 \cdot 10^{-4}$  e  $R^2_{pred}$  igual a 0,99. Os melhores modelos de previsão do índice de refração utilizando espectros de NIR foram obtidos com a aplicação de seleção de variáveis (VIP e REP) e com os espectros completos ( $10000$  a  $4000 \text{ cm}^{-1}$ ), o que corrobora com os algoritmos que apresentaram menores valores de RMSEP ao serem aplicados, conforme mostrado na Figura 37.



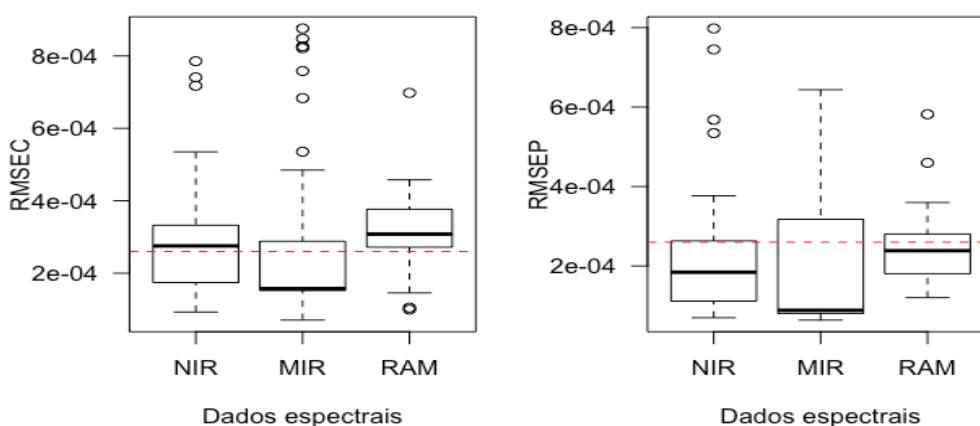
Figura 37 - RMSEP *versus* intervalos espectrais para todos os modelos de previsão do índice de refração, a linha horizontal indica o erro experimental



Fonte: O autor, 2019.

O modelo Raman com menor  $SWTP_{val}$  aparece na 49ª posição dos melhores modelos para a previsão do índice de refração. Sabendo disso, concluímos que a utilização dos espectros Raman (RAM) para prever o índice de refração não trouxe melhorias aos resultados obtidos por NIR e mid-IR (MIR). De fato, este fenômeno pode ser confirmado com a Figura 38, onde é observado que os modelos utilizando espectros NIR e mid-IR apresentam menores valores de RMSEC e RMSEP, embora os modelos com mid-IR tenham apresentado maior variabilidade.

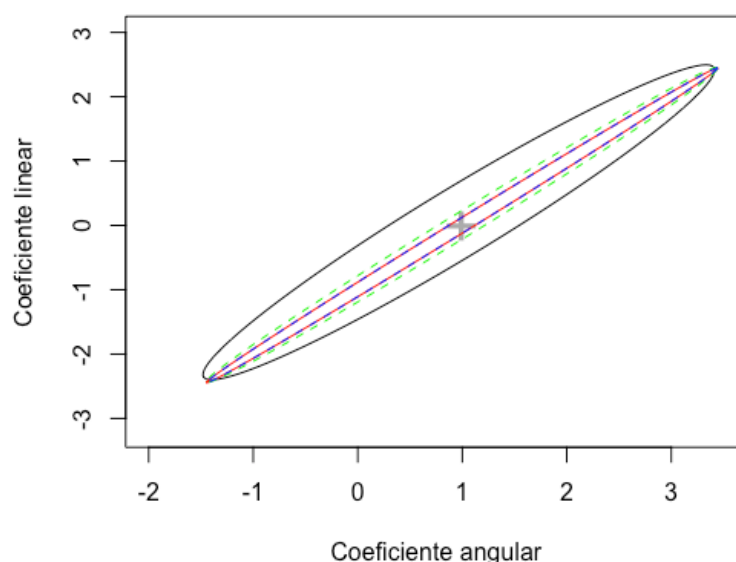
Figura 38 - RMSEC e RMSEP *versus* dados espectrais (NIR, mid-IR e Raman) para todos os modelos de previsão do Índice de Refração, a linha horizontal indica o erro experimental



Fonte: O autor, 2019.

A elipse da região conjunta de confiança (EJCR) foi empregada nos 4 melhores modelos de previsão do índice de refração, para avaliar a existência de erros sistemáticos e a precisão do método. Conforme mostrado na Figura 39, é possível observar que os modelos de previsão selecionados não apresentaram erro sistemático e que o modelo NIR\_VIP-PLS\_IndRefra\_snv é o mais preciso.

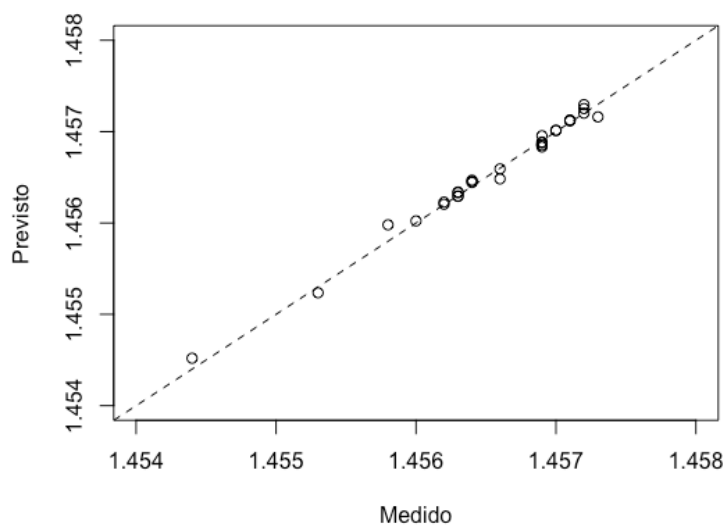
Figura 39 - Elipse da região conjunta de confiança (EJCR) do conjunto de validação para os modelos de previsão do Índice de refração: (-) MIR\_UVE-PLS\_IndRefra\_drv; (-) NIR\_VIP-PLS\_IndRefra\_snv; (- -) NIR\_REP-PLS\_IndRefra\_snv; (- -) NIR\_PLS\_IndRefra\_snv.



Fonte: O autor, 2019.

O modelo de previsão de índice de refração, ao comparar os modelos por SWTP obtidos e por não apresentar erro sistemático na EJCR, o modelo MIR\_UVE-PLS\_IndRefra\_drv foi selecionado como o melhor modelo, apresentando RMSEP igual a  $1,0 \cdot 10^{-4}$  e  $R^2_{pred}$  0,95. Na Tabela 16 são apresentados os 10 melhores resultados segundo a classificação por valor de SWTP. No ANEXO C foram disponibilizados todos os resultados obtidos. O gráfico dos valores preditos *versus* medidos para o melhor modelo do índice de refração é mostrado na Figura 40.

Figura 40 - Gráfico dos valores previstos *versus* medidos para o modelo MIR\_UVE-PLS\_IndRefra\_drv



Fonte: O autor, 2019.

Esse resultado pode ser comparado com um estudo anterior deste grupo de pesquisa. Cunha *et al.* (2017) estudaram uma metodologia para determinar índice de refração de misturas de biodiesel usando espectroscopia no infravermelho médio ( $4000-680\text{ cm}^{-1}$ ) e modelo PLS com três variáveis latentes e RMSEP igual a 0,0001. Este modelo foi reproduzido neste estudo, com o modelo MIR\_PLS\_IndRefra\_mc, e os parâmetros de mérito alcançados foram similares. No entanto, de acordo com o teste de SWTP este modelo foi considerado o 32º melhor.

#### 6.1.7 Modelos de previsão para massa específica

Os parâmetros de mérito dos 10 melhores modelos de previsão da massa específica obtidos a partir dos conjuntos espectrais de NIR, mid-IR (MIR) e Raman (RAM) são apresentados nas Tabela 17, respectivamente.

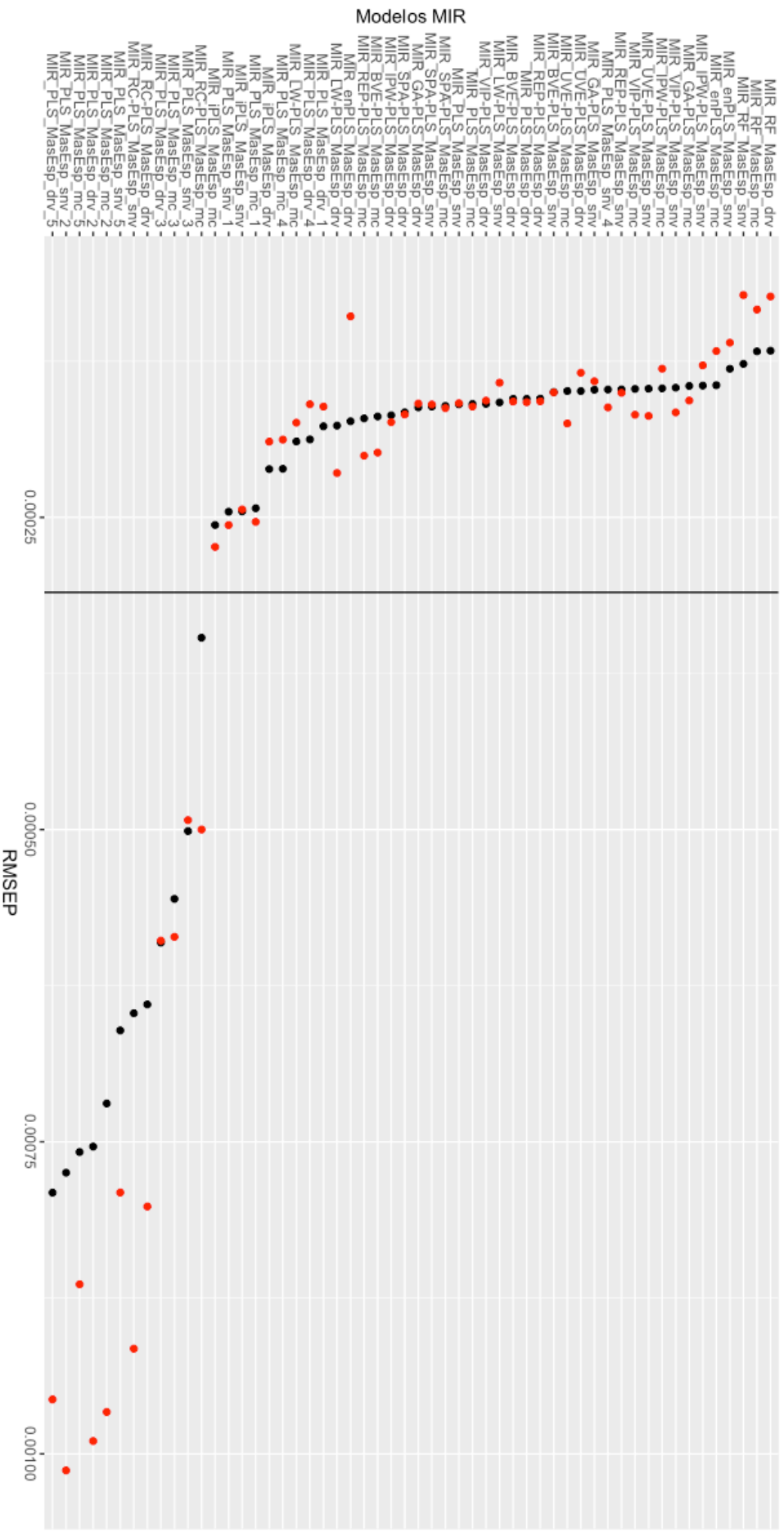
A princípio, os modelos que apresentaram RMSEP menor que o erro experimental das propriedades estudadas podem ser considerados adequados para a previsão da massa específica. Nas Figuras 41, 42 e 43, a linha vertical indica o erro experimental da massa específica dos modelos obtidos a partir dos espectros de NIR, mid-IR e Raman. Deste modo, os modelos que estão à esquerda da linha vertical apresentam RMSEP menor que o erro experimental de medição da massa específica, que é  $3,1 \cdot 10^{-4}\text{ g}\cdot\text{cm}^{-3}$ .

Tabela 17 - Parâmetros de mérito dos 10 melhores modelos PLS com seleção de variáveis e Florestas Aleatórias (RF) obtidos a partir dos espectros de NIR, mid-IR e Raman para a previsão da Massa Específica

Modelos	Espectros	Intervalos	PPx	VL	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
MIR_RF_MasEsp_drv	MIR	Todo	drv	-	7.3E-05	0.99	1.8E-04	0.98	1.2E-04	0.98	8.228E-10	8.24E-09	3.52E-08	4.0E-03	2.06
MIR_RF_MasEsp_mc	MIR	Todo	mc	-	8.4E-05	0.99	2.1E-04	0.97	1.2E-04	0.98	5.581E-09	7.43E-09	3.28E-08	3.8E-03	2.44
NIR_RF_MasEsp_snv	NIR	Todo	snv	-	8.9E-05	0.99	2.3E-04	0.96	1.3E-04	0.97	3.945E-09	4.45E-08	2.16E-09	9.8E-04	2.57
MIR_RF_MasEsp_snv	MIR	Todo	snv	-	7.2E-05	0.99	1.8E-04	0.98	1.3E-04	0.97	3.237E-09	1.11E-08	7.30E-08	5.7E-03	5.29
MIR_enPLS_MasEsp_snv	MIR	Todo	snv	9	1.1E-04	0.99	1.2E-04	0.99	1.3E-04	0.97	2.136E-09	3.04E-09	5.15E-08	4.8E-03	5.57
NIR_GA-PLS_MasEsp_mc	NIR	GA	base	3	1.7E-04	0.98	1.9E-04	0.97	1.3E-04	0.97	5.023E-17	5.98E-08	1.67E-08	2.7E-03	6.23
RAM_RF_MasEsp_snv	RAM	Todo	snv	-	1.1E-04	0.99	3.1E-04	0.93	1.7E-04	0.96	5.777E-10	1.21E-08	1.67E-08	2.7E-03	8.79
NIR_RF_MasEsp_drv	NIR	Todo	drv	-	8.1E-05	0.99	2.4E-04	0.96	1.4E-04	0.97	2.911E-10	8.29E-09	8.31E-09	1.9E-03	9.62
MIR_VIP-PLS_MasEsp_snv	MIR	VIP	snv	2	1.7E-04	0.98	1.7E-04	0.98	1.5E-04	0.97	5.017E-17	7.76E-08	1.07E-09	6.9E-04	9.82
MIR_IPW-PLS_MasEsp_snv	MIR	IPW	snv	4	1.3E-04	0.99	1.4E-04	0.98	1.4E-04	0.97	5.020E-17	1.21E-07	5.85E-08	5.1E-03	10.96

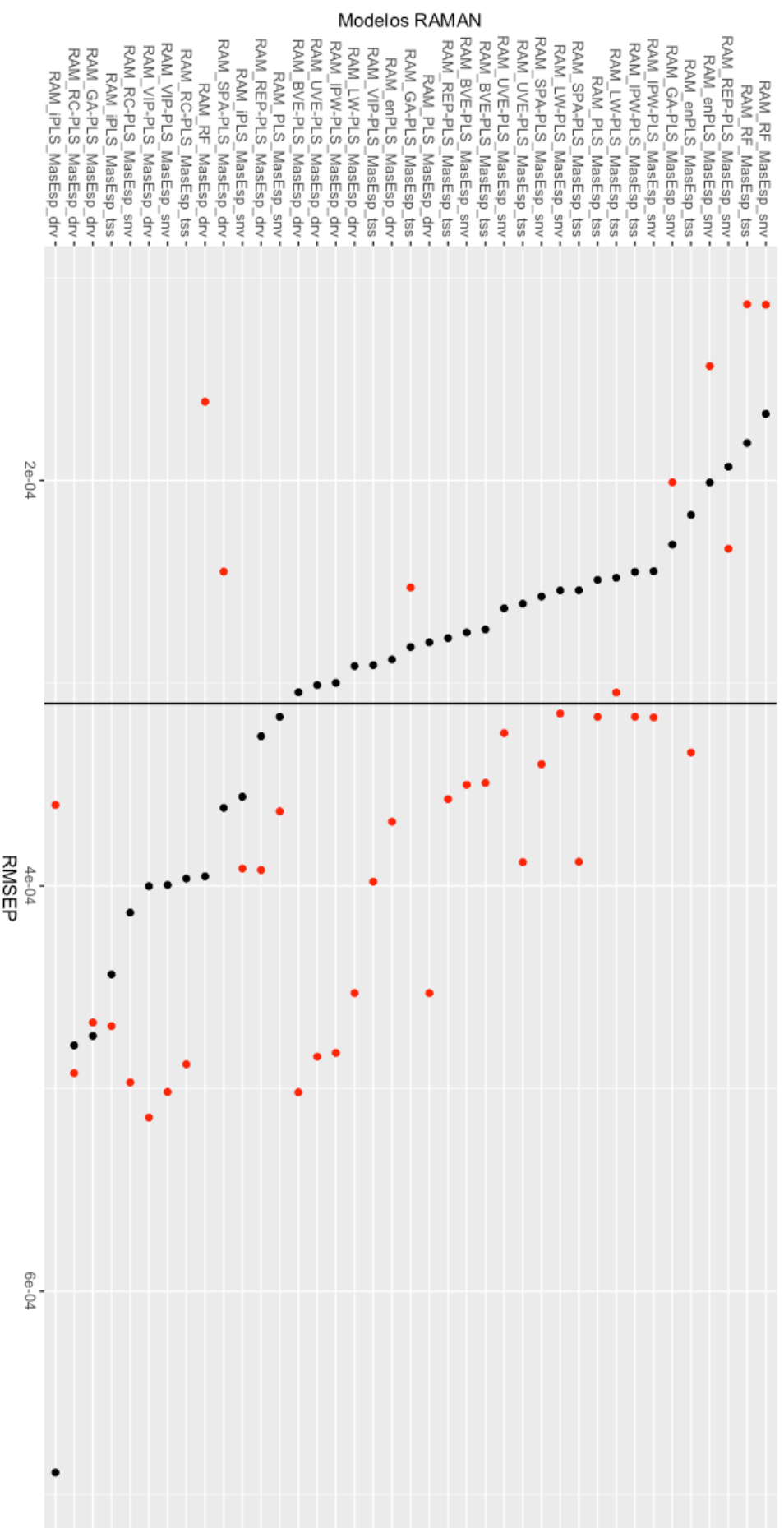


Figura 42 - Modelos de previsão da massa específica versus RMSEP (•) e RMSEC (•), em  $g \cdot cm^{-3}$ , obtidos a partir do conjunto espectros de mid-IR, a linha vertical indica o erro experimental



Fonte: O autor, 2019.

Figura 43 - Modelos de previsão da massa específica versus RMSEP (•) e RMSEC (•), em  $\text{g}\cdot\text{cm}^{-3}$ , obtidos a partir do conjunto espectros de Raman, a linha vertical indica o erro experimental

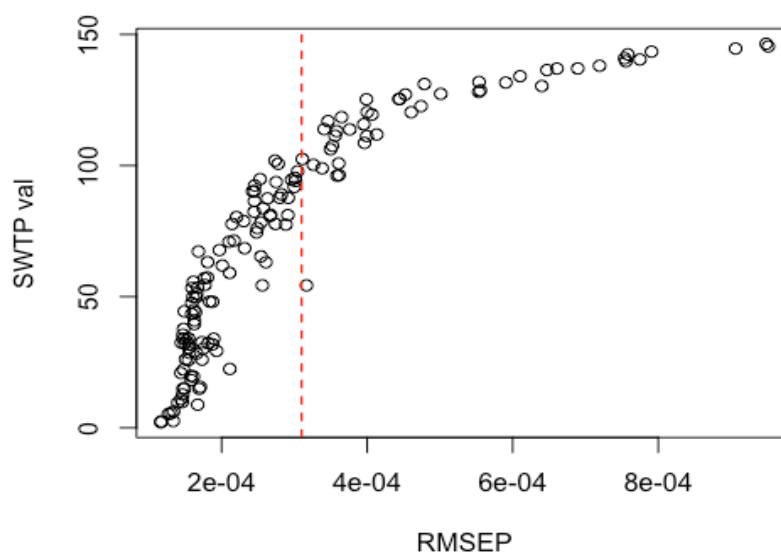


Fonte: O autor, 2019.

Além de considerar os modelos que apresentam valor de RMSEP menor que o erro experimental, é necessário comparar os modelos e verificar o modelo com o menor valor residual entre os modelos estudados. O teste de Wilcoxon pareado foi escolhido para comparar os valores residuais entre os modelos e identificar o modelo mais adequado dentre os obtidos. Uma matriz (147 × 147) com os valores de probabilidade resultantes do teste de Wilcoxon foi construída e um vetor contendo todas as linhas da Soma da Probabilidade do Teste de Wilcoxon (SWTP) foi obtido e, o valor mínimo desse vetor revelou o melhor modelo para a massa específica.

A Figura 44 mostra o comportamento dos modelos ao comparar os valores  $SWTP_{val}$  com os valores RMSEP (em  $g \cdot cm^{-3}$ ) para a massa específica.

Figura 44 - Soma da probabilidade do teste de Wilcoxon (SWTP) *versus* RMSEP ( $g \cdot cm^{-3}$ ) para todos os modelos de previsão da massa específica, a linha vertical indica o erro experimental



Fonte: O autor, 2019.

Para cada um dos 10 melhores modelos de previsão da massa específica são apresentados os valores de SWTP da validação, conforme Tabela 17, e o valor mínimo indica o melhor modelo dentre os modelos estudados. No caso da previsão da massa específica, o modelo das florestas aleatórias (RF), utilizando os espectros de mid-IR e o pré-processamento da primeira derivada (MIR\_RF\_MasEsp\_drv), apresentou o menor valor  $SWTP_{val}$  (2,06) com o maior valor de  $R^2_{pred}$  (0,98) e o menor

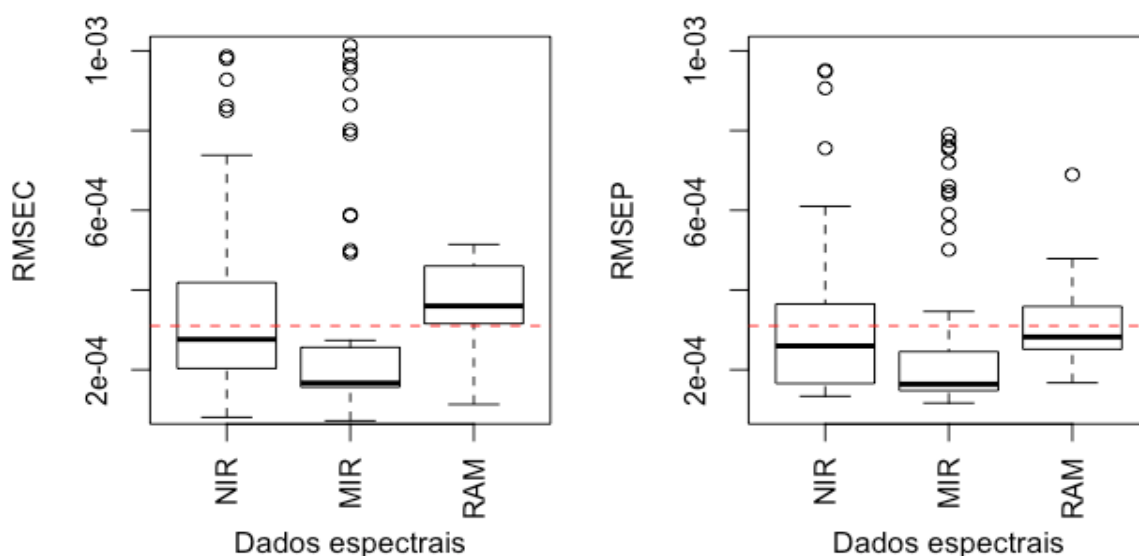


valor de RMSEP ( $1,16 \cdot 10^{-4} \text{ g} \cdot \text{cm}^{-3}$ ) dentre os modelos avaliados. No ANEXO D estão disponibilizados todos os resultados (parâmetros) obtidos.

Os próximos três modelos com baixos valores de  $\text{SWTP}_{\text{val}}$  são: MIR\_RF\_MasEsp\_mc ( $\text{SWTP}_{\text{val}} = 2,44$ ) com RMSEP igual  $1,17 \cdot 10^{-4} \text{ g} \cdot \text{cm}^{-3}$  e  $R^2_{\text{pred}}$  igual a 0,98; NIR\_RF\_MasEsp\_snv ( $\text{SWTP}_{\text{val}} = 2,57$ ) com RMSEP igual a  $1,34 \cdot 10^{-4} \text{ g} \cdot \text{cm}^{-3}$  e  $R^2_{\text{pred}}$  igual a 0,97; e MIR\_RF\_MasEsp\_snv ( $\text{SWTP}_{\text{val}} = 5,29$ ) com RMSEP igual a  $1,27 \cdot 10^{-4}$  e  $R^2_{\text{pred}}$  igual a 0,97.

Os melhores modelos de previsão do índice de refração utilizando espectros de mid-IR e NIR foram obtidos com a aplicação do método das florestas aleatórias (RF). O modelo Raman com menor  $\text{SWTP}_{\text{val}}$  aparece na 7ª posição dos melhores modelos para a previsão da massa específica. Com isso, pode-se concluir que a utilização dos espectros Raman para prever a massa específica, embora tenha apresentado uma superioridade na capacidade preditiva em comparação com os modelos de Raman para a previsão do índice de refração, não conseguiu alcançar resultados melhores que os modelos obtidos por NIR e mid-IR. De fato, este fenômeno pode ser confirmado com a Figura 45, onde é observado que os modelos utilizando espectros NIR e mid-IR apresentam menores valores de RMSEC ( $\text{g} \cdot \text{cm}^{-3}$ ) e RMSEP ( $\text{g} \cdot \text{cm}^{-3}$ ).

Figura 45 - RMSEC ( $\text{g} \cdot \text{cm}^{-3}$ ) e RMSEP ( $\text{g} \cdot \text{cm}^{-3}$ ) versus dados espectrais (NIR, mid-IR e Raman) para todos os modelos de previsão da Massa Específica, a linha vermelha horizontal representa o valor do erro experimental

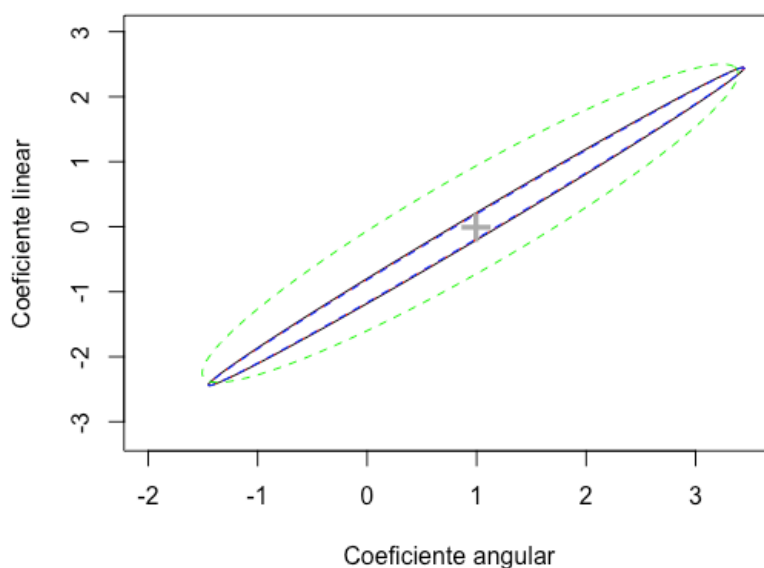


Fonte: O autor, 2019.

A elipse da região conjunta de confiança (EJCR) foi aplicada nos 4 melhores modelos de previsão da massa específica, para avaliar a existência de erros sistemáticos e garantir a precisão do método. De acordo com a Figura 46, é possível observar que os modelos de previsão selecionados não apresentaram erro sistemático.

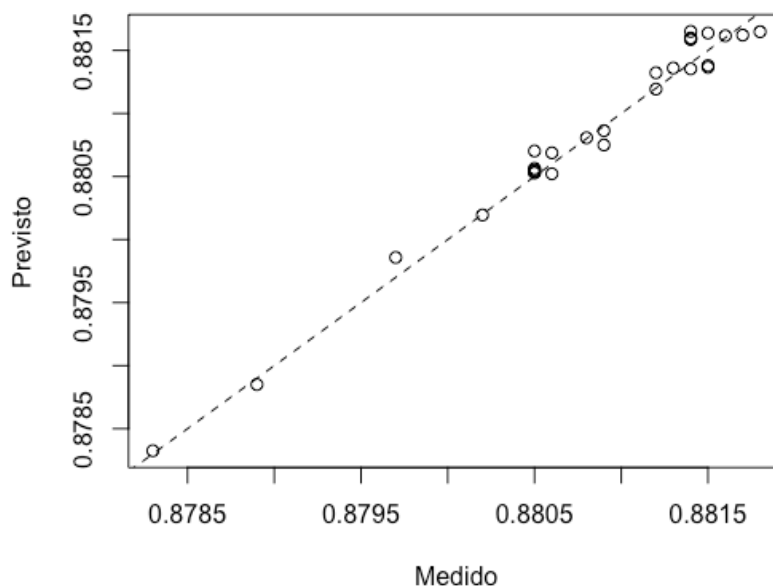
O modelo de previsão da massa específica MIR\_RF\_MasEsp\_drv foi selecionado como o melhor modelo, ao comparar os modelos por SWTP obtidos e por não apresentar erro sistemático e boa precisão na EJCR, apresentando RMSEP igual a  $1,16 \cdot 10^{-4} \text{ g} \cdot \text{cm}^{-3}$  e  $R^2_{\text{pred}} 0,98$  (Tabela 17). O gráfico dos valores previstos *versus* medidos para o melhor modelo é apresentado na Figura 47.

Figura 46 - Elipse da região conjunta de confiança (EJCR) do conjunto de validação para os modelos de previsão da Massa Específica: (-) MIR\_RF\_MasEsp\_drv; (- -) MIR\_RF\_MasEsp\_mc; (- -) NIR\_RF\_MasEsp\_snv; (- -) MIR\_RF\_MasEsp\_snv.



Fonte: O autor, 2019.

Figura 47 - Gráfico dos valores previstos *versus* medidos para o modelo MIR\_RF\_MasEsp\_drv



Fonte: O autor, 2019.

Esse resultado pode ser comparado com estudos anteriores da literatura. Baptista *et al.* (2008) desenvolveram uma metodologia para determinar a massa específica de misturas de biodiesel usando espectros de NIR na faixa de 9000-4500  $\text{cm}^{-1}$  com 6 variáveis latentes, obtendo um modelo com RMSEP igual a  $9,0 \cdot 10^{-4} \text{ g} \cdot \text{cm}^{-3}$ . Ferrão *et al.* (2011) também estudaram a previsão de massa específica por PLS e propuseram um modelo com 8 variáveis latentes, RMSEP igual a  $7,3 \cdot 10^{-4} \text{ g} \cdot \text{cm}^{-3}$ . Cunha *et al.* (2017) determinaram a massa específica de misturas de biodiesel usando espectroscopia no infravermelho médio ( $4000\text{-}680 \text{ cm}^{-1}$ ) com um modelo PLS com três variáveis latentes e RMSEP igual a  $2,0 \cdot 10^{-4} \text{ g} \cdot \text{cm}^{-3}$ . Todos estes autores obtiveram modelos com valores de RMSEP superiores quando comparados ao melhor modelo obtido neste estudo, com o modelo MIR\_RF\_MasEsp\_drv (RMSEP= $1,16 \cdot 10^{-4} \text{ g} \cdot \text{cm}^{-3}$ ).

## 6.2 BANCO 2: Misturas de biodiesel-biodiesel, diesel-biodiesel e diesel-biodiesel-etanol

A previsão do ponto de entupimento de filtro a frio (PEFF) e da viscosidade cinemática a  $40^\circ\text{C}$  (VISCO) foram realizadas com os dados espectrais de NIR, mid-IR e Raman obtidos das amostras que compõe o Banco 2. O Banco 2 é constituído pelas 148 amostras de biodiesel (B) do Banco 1 mais 17 amostras de biodiesel (B),

18 amostras de diesel (D), 102 misturas binárias diesel-biodiesel (M) e 48 amostras ternárias diesel-biodiesel-etanol (T) . No total, para o Banco 2, são 333 amostras cuja a composição podem ser consultadas no ANEXO A e ANEXO B.

Vale ressaltar que a amostra Am158, removida do conjunto de dados Banco 1, foi desconsiderada também dos conjuntos espectrais de NIR, mid-IR e Raman do Banco 2.

### 6.2.1 Caracterização do diesel e das misturas

Para a caracterização do diesel foram realizadas medições das propriedades: ponto de entupimento de filtro a frio e viscosidade. Os espectros de NIR, mid-IR e Raman das amostras puras e misturas dos combustíveis foram obtidos nessa etapa do estudo.

Os resultados das medidas de PEFF ( $^{\circ}\text{C}$ ) e viscosidade cinemática ( $\text{mm}^2\cdot\text{s}^{-1}$ ) para as 333 amostras que compõem o Banco 2 estão disponíveis nos ANEXO A e B.

A repetibilidade do método de medição da viscosidade cinemática foi de  $0,0155 \text{ mm}^2\cdot\text{s}^{-1}$ , para um nível de confiança de 95%, e a reprodutibilidade de  $0,0279 \text{ mm}^2\cdot\text{s}^{-1}$ . O intervalo de variação observado para o conjunto de dados em consideração foi de  $2.6629$  a  $4.8524 \text{ mm}^2\cdot\text{s}^{-1}$  com erro de medição igual a  $0,0267 \text{ mm}^2\cdot\text{s}^{-1}$ .

Para o PEFF ( $^{\circ}\text{C}$ ), a repetibilidade do método de medição foi de  $1,8^{\circ}\text{C}$  e a reprodutibilidade foi de  $2,0^{\circ}\text{C}$  para um nível de confiança de 95%. O intervalo de variação observado para o conjunto de dados em consideração foi de  $-26,0^{\circ}\text{C}$  a  $7,0^{\circ}\text{C}$  com erro de medição igual a  $1,6^{\circ}\text{C}$ .

As Tabela 18 e 19 mostram os resultados das medidas de ponto de entupimento (PEFF) e viscosidade cinemática para alguns grupos de amostras e compara com os resultados encontrados na literatura. A diferença apresentada nas medidas de PEFF podem ser explicadas pelas diferenças na síntese e nas etapas de purificação do biodiesel, além da qualidade/especificações da matéria-prima.

Tabela 18 - Resultados das medidas de PEFF de algumas amostras comparado com os resultados da literatura (obtidos por outros autores).

Experimental		Literatura	
Descrição	Média PEFF (°C)	PEFF (°C)	Referência
Biodiesel de Soja	-5,0	-4,0	Sanford <i>et al.</i> , 2009
Biodiesel de Canola	-6,6	-8,8	Giakoumis, 2013
Biodiesel de Milho	-2,6	-3,0	Sanford <i>et al.</i> , 2009
Biodiesel de Girassol	-1,6	-3,0	Sanford <i>et al.</i> , 2009
Biodiesel Região Sul	6,0	-	-
Biodiesel médio	-3,95	-13,0	Kalogeras <i>et al.</i> , 2010
B10	-0,45	-6,76	Kalogeras <i>et al.</i> , 2010
B20	-5,6	-7,48	Kalogeras <i>et al.</i> , 2010
B15	-0,7	-	-
B25	-3,5	-	-
B30	-12,0	-	-
Diesel médio	-2,22	-6,00	Kalogeras <i>et al.</i> , 2010
Diesel S500	4,9	-	-
Diesel S10	-1,17	-	-
Diesel Padrão	-26,0	-	-

Tabela 19 - Resultados das medidas de Viscosidade cinemática a 40°C (VISCO) de algumas amostras comparado com os resultados da literatura (obtidos por outros autores).

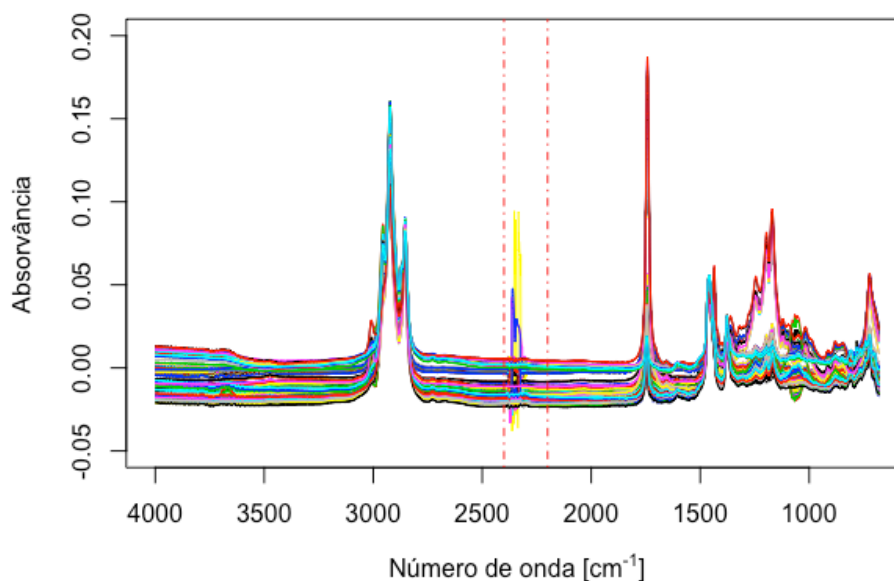
Experimental		Literatura	
Descrição	Viscosidade média (mm <sup>2</sup> ·s <sup>-1</sup> )	Viscosidade (mm <sup>2</sup> ·s <sup>-1</sup> )	Referência
Biodiesel de Soja	4,2267	-	-
Biodiesel de Canola	4,5915	-	-
Biodiesel de Milho	4,3729	-	-
Biodiesel de Girassol	4,3555	-	-
Biodiesel Região Sul	4,4569	-	-
Biodiesel médio	4,4637	4,4700	Kalogeras <i>et al.</i> , 2010
B10	3,6667	2,8868	Kalogeras <i>et al.</i> , 2010
B20	3,4186	3,0627	Kalogeras <i>et al.</i> , 2010
B15	3,6788	-	-
B25	3,5710	-	-
B30	3,0161	-	-
Diesel médio	3,5597	2,7109	Kalogeras <i>et al.</i> , 2010
Diesel S500	3,4984	-	-
Diesel S10	3,8944	-	-
Diesel Padrão	2,6787	-	-

### 6.2.2 Análise dos dados

Neste estudo, espectros NIR, mid-IR e Raman de amostras de biodiesel, diesel, diesel-biodiesel e diesel-biodiesel-etanol foram obtidos na região espectral de 10000 - 4000 cm<sup>-1</sup>, 4000 a 680 cm<sup>-1</sup>, e 201 a 1800 cm<sup>-1</sup>, respectivamente.

Para a espectroscopia de infravermelho médio por transformada de Fourier com reflectância total atenuada (FT-mid-IR-ATR), os espectros foram obtidos para as 333 amostras no intervalo de 4000 a 680 cm<sup>-1</sup>, conforme pode ser observado na Figura 48. Dessa forma, inicialmente o tamanho da matriz de espectros mid-IR apresentou dimensão de 333 x 3321 (amostras *versus* variáveis).

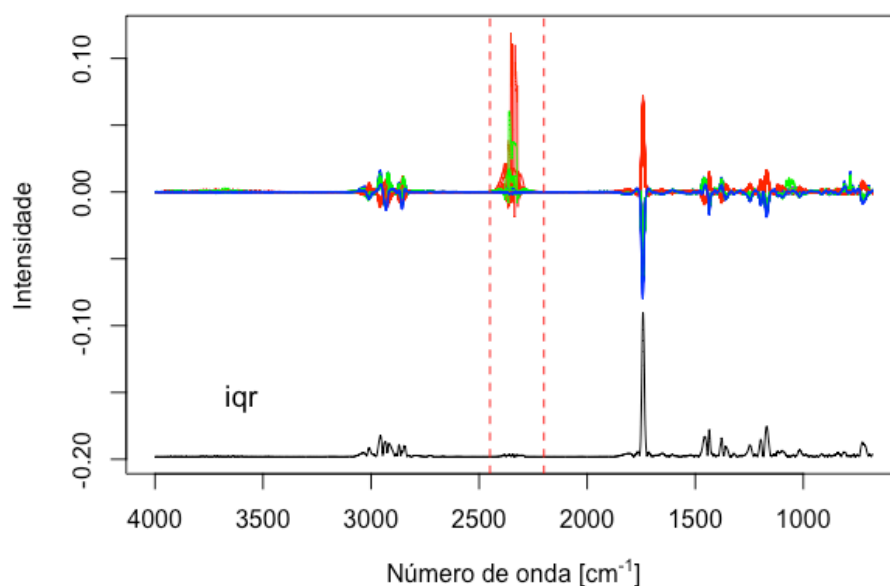
Figura 48 - Espectros brutos de infravermelho médio (mid-IR) das amostras do Banco 2.



Fonte: O autor, 2019.

O sinal de resposta do cristal de seleneto de zinco (ZnSe) do acessório de reflectância total atenuada (ATR) foi observado na região espectral entre 2450 a 2200  $\text{cm}^{-1}$ , conforme indicado na Figura 49. Antes de remover esta região, foi avaliado se a região em questão contém informações relevantes ou apenas ruído oriundo do sinal do cristal do equipamento. Para isto foi aplicado a função *surveySpectra*, baseado na amplitude interquartil, do pacote {ChemSpec} (HANSON, 2019) e foi gerada a Figura 49, que comprova que a região espectral de 2450 a 2200  $\text{cm}^{-1}$  não é uma região com informações relevantes. Por este motivo, optou-se pela remoção desta região do espectro mid-IR.

Figura 49 - Regiões dos espectros de mid-IR ricas em informações, a linha horizontal tracejada (--) indica a região do sinal de ZnSe.



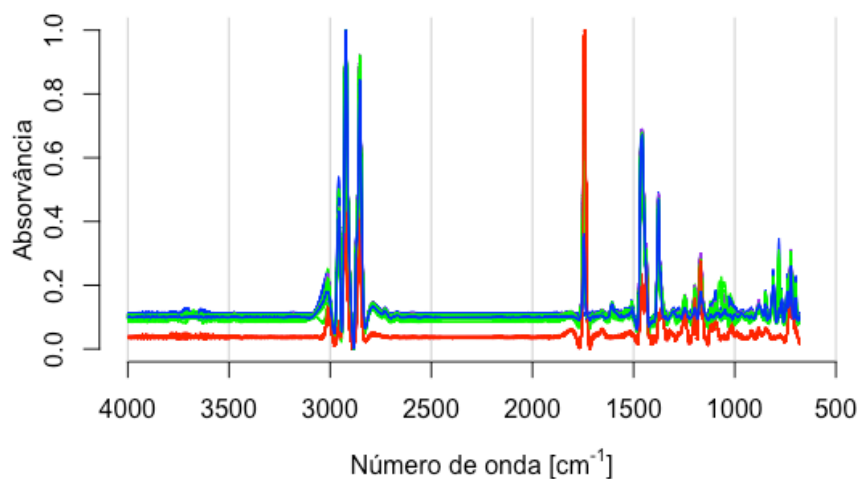
Fonte: O autor, 2019.

Na Figura 50 são apresentados os espectros mid-IR após a correção da linha de base e da remoção da região espectral entre 2450 a 2200  $\text{cm}^{-1}$ . Os quatro principais grupos (ou classes) presentes no Banco 2 estão representados na Figura 50 em cores distintas, tais como: 166 amostras de biodiesel (**B**) em vermelho, 18 amostras de diesel (**D**) em azul, 102 amostras de misturas binárias entre diesel-biodiesel (**M**) em verde; e 48 amostras de misturas ternárias entre diesel-biodiesel-etanol (**T**) em roxo.

Os espectros apresentados na Figura 50 podem ser avaliados separadamente considerando as 4 diferentes classes presentes no Banco 2, conforme mostrado na Figura 51. Com o auxílio da Figura 51 é possível avaliar as regiões espectrais (1500-1400  $\text{cm}^{-1}$ ; 3000-2900  $\text{cm}^{-1}$ ) que possibilitam a diferenciação das amostras de biodiesel das amostras que contém diesel, na forma de puros ou em misturas).

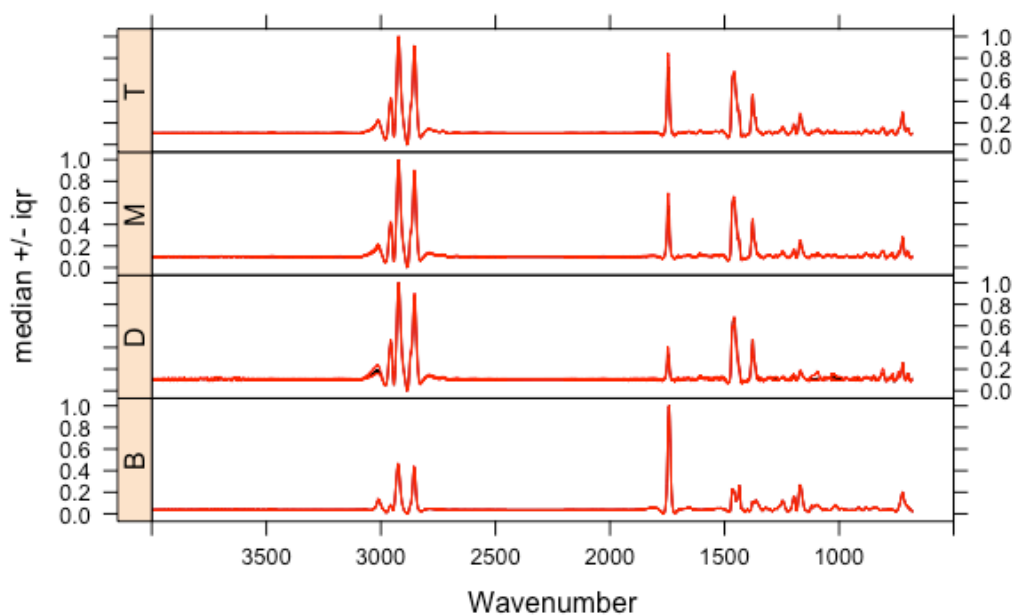


Figura 50 - Espectros de infravermelho médio (mid-IR) após a remoção da região espectral do sinal do cristal de ZnSe e após correção de linha de base.



Fonte: O autor, 2019.

Figura 51 - Espectros de mid-IR para cada um dos grupos de amostras que compõem o Banco 2: Biodiesel (B), diesel (D), misturas diesel-biodiesel (M) e misturas diesel-biodiesel-etanol (T).

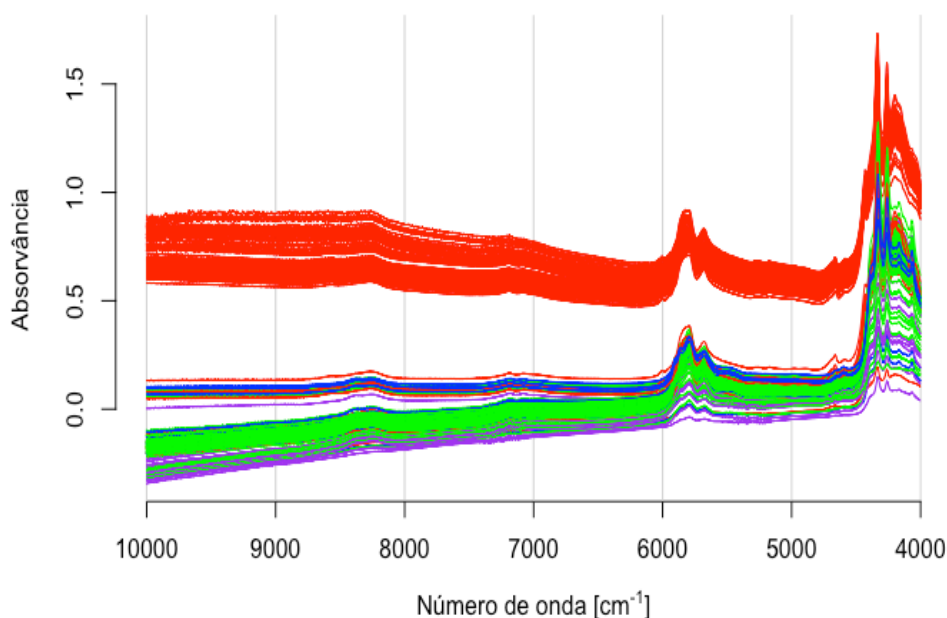


Fonte: O autor, 2019.

Com a remoção da amostra Am158 e a remoção da região espectral 2450 a 2200  $\text{cm}^{-1}$ , o conjunto de dados espectrais de mid-IR ficou com 333 amostras e 3072 variáveis (4000-2450  $\text{cm}^{-1}$  e 2200-680  $\text{cm}^{-1}$ ).

Para a espectroscopia de infravermelho próximo foi utilizada a faixa de número de onda de 10000 a 4000  $\text{cm}^{-1}$ , com intervalo espectral de 1  $\text{cm}^{-1}$ . Na Figura 52, são apresentadas o perfil espectral bruto obtido a partir de medidas de infravermelho próximo (NIR).

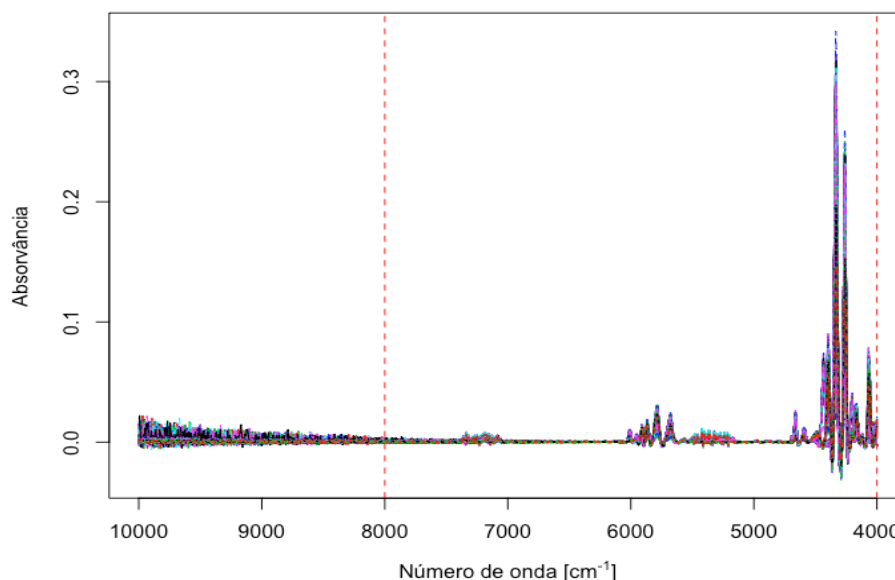
Figura 52 - Espectros brutos de infravermelho próximo (NIR) das amostras que compõem o Banco 2: biodiesel (**B**), diesel (**D**), diesel-biodiesel (**M**) e diesel-biodiesel-etanol (**T**).



Fonte: O autor, 2019.

Na Figura 53 são apresentados os mesmos espectros após o ajuste de linha de base. A matriz com os espectros NIR apresentam o sinal em absorvância de medidas de 6001 números de onda (10000 a 4000  $\text{cm}^{-1}$ ) para 333 amostras, formando uma matriz espectral de 333 x 6001 (amostras x variáveis).

Figura 53 - Espectros de infravermelho próximo (NIR) das amostras do Banco 2 após a correção de linha de base, as linhas verticais indicam intervalo a ser mantido para o estudo.

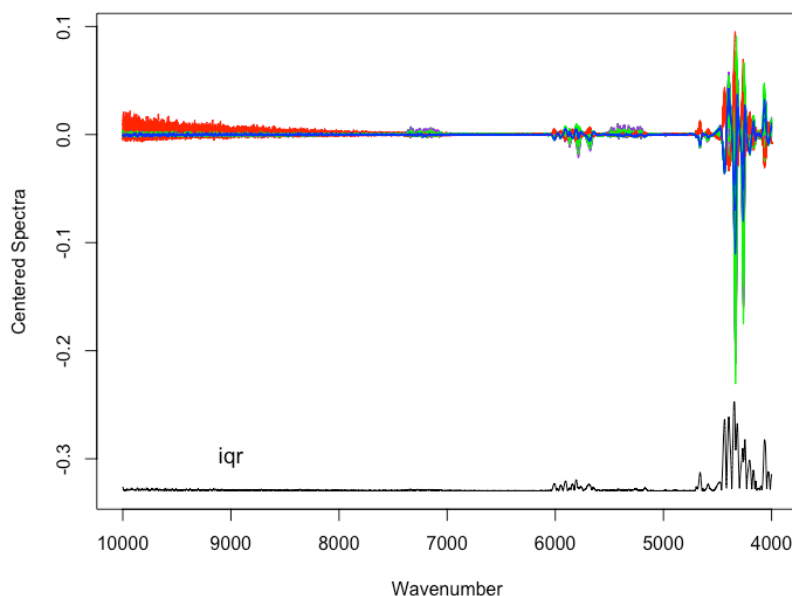


Fonte: O autor, 2019.

Na Figura 53, é possível observar que há uma região (10000 a 8000  $\text{cm}^{-1}$ ) que apresenta ruídos nos espectros de NIR do Banco 2. Antes de remover esta região, foi avaliado se a região em questão contém informações relevantes ou apenas ruído. A função *surveySpectra* do pacote {ChemSpec} foi aplicada e a Figura 54 foi obtida, comprovando que a região espectral de 10000 a 8000  $\text{cm}^{-1}$  não é uma região importante, uma vez que a amplitude interquartil (iqr) para essa região é próxima de zero. Por este motivo, optou-se pela remoção desta região do espectro NIR.

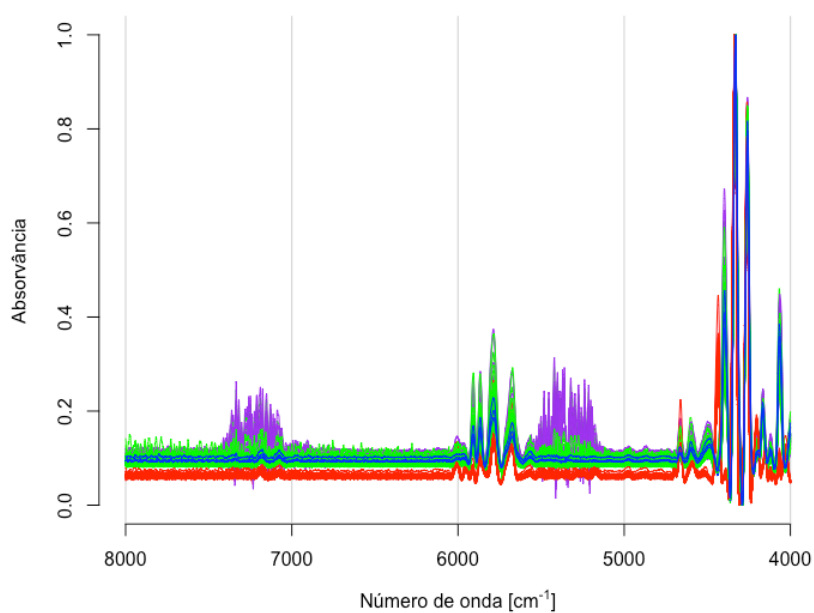
Na Figura 55 são apresentados os espectros NIR após a correção de linha de base, seguido da remoção da região 10000 a 8000  $\text{cm}^{-1}$  e a normalização dos espectros.

Figura 54 - Regiões dos espectros de NIR ricas em informações.



Fonte: O autor, 2019.

Figura 55 - Espectros de infravermelho próximo (NIR) após a etapa de normalização.

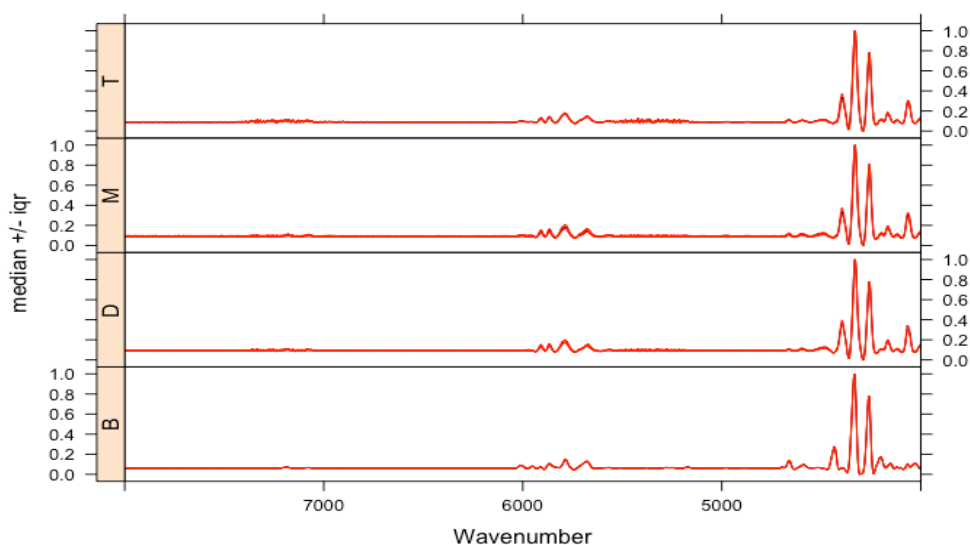


Fonte: O autor, 2019.

Os espectros apresentados na Figura 55 podem ser avaliados separadamente considerando as 4 diferentes classes presentes no Banco 2, conforme pode ser visto na Figura 56. Com o auxílio da Figura 56 é possível avaliar as regiões espectrais ( $6050\text{-}6000\text{ cm}^{-1}$ ;  $4660\text{-}4630\text{ cm}^{-1}$ ) que possibilitam a diferenciação das amostras de biodiesel das amostras que contém diesel, na forma de puros ou em misturas.

Após a etapa de tratamento dos espectros, a matriz com os espectros NIR ficaram com 4001 números de onda ( $8000$  a  $4000\text{ cm}^{-1}$ ) e intervalo de  $1\text{ cm}^{-1}$  para 333 amostras, formando uma matriz espectral de  $333 \times 4001$  (amostras x variáveis).

Figura 56 - Espectros de NIR para cada um dos grupos de amostras que compõem o Banco 2: Biodiesel (B), diesel (D), misturas diesel-biodiesel (M) e misturas diesel-biodiesel-etanol (T).

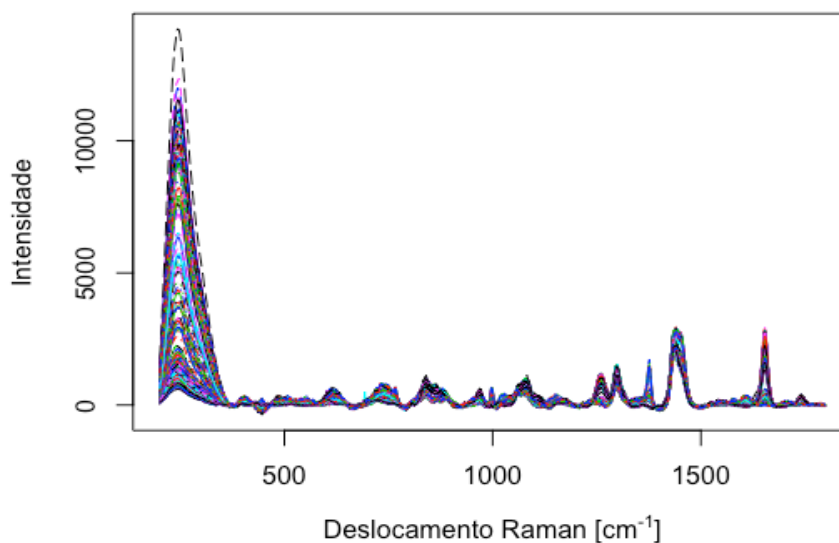


Fonte: O autor, 2019.

Os espectros de Raman foram obtidos na faixa de número de onda de  $201$  a  $1800\text{ cm}^{-1}$ , com intervalo espectral de  $2,04\text{ cm}^{-1}$ . Na Figura 57, é apresentado o perfil espectral de Raman após as etapas de correção de linha de base e normalização. A matriz com os espectros Raman apresentaram 786 números de onda ( $201$  a  $1800\text{ cm}^{-1}$ ) e intervalo de  $2,04\text{ cm}^{-1}$  para 333 amostras, formando uma matriz espectral de  $333 \times 786$  (amostras x variáveis).

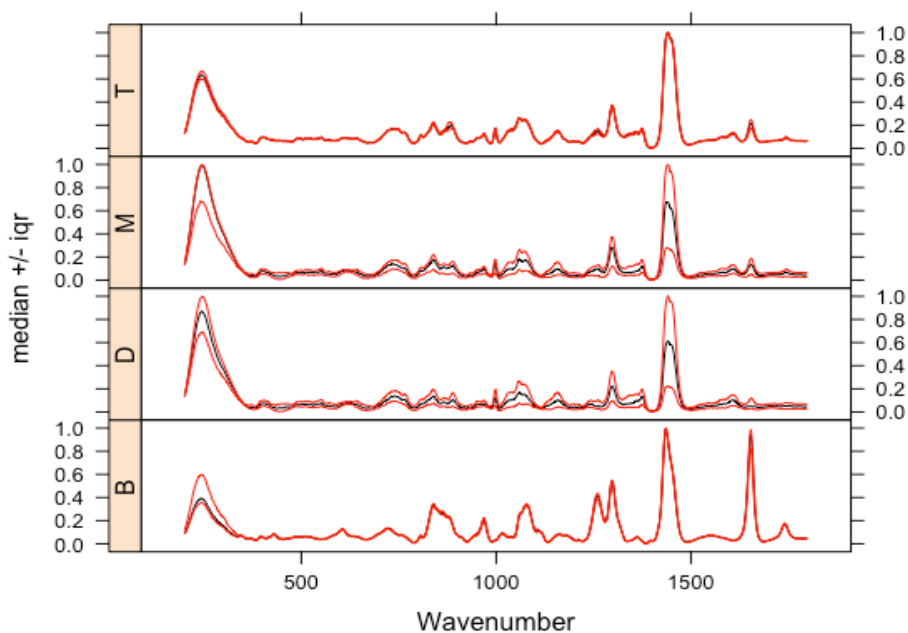
Os quatro principais grupos (ou classes) de amostras presentes no Banco 2 estão representados na Figura 58, onde é possível avaliar as regiões espectrais ( $950\text{-}980\text{ cm}^{-1}$ ;  $1240\text{-}1280\text{ cm}^{-1}$ ;  $1630\text{-}1670\text{ cm}^{-1}$ ;  $1720\text{-}1760\text{ cm}^{-1}$ ) que possibilitam a diferenciação das amostras que contém apenas biodiesel das amostras que contém diesel, na forma de puros ou em misturas.

Figura 57 - Espectros de Raman (RAM) das amostras do Banco 2 após as etapas de correção de linha de base e de normalização.



Fonte: O autor, 2019.

Figura 58 - Espectros de Raman para cada um dos grupos de amostras que compõem o Banco 2: Biodiesel (B), diesel (D), misturas diesel-biodiesel (M) e misturas diesel-biodiesel-etanol (T).



Fonte: O autor, 2019.

### 6.2.3 Análise de componentes principais (PCA)

O estudo envolvendo a análise de componentes principais foi realizado para o banco de dados 2 contendo os espectros NIR), mid-IR e Raman com o objetivo de realizar uma análise exploratória dos dados espectrais.

Neste estudo, diferentes pré-processamentos foram testados na matriz espectral durante a construção dos modelos, como centralização na média, derivada de 1ª ordem, padronização normal de sinal, além da transformação do sinal espacial.

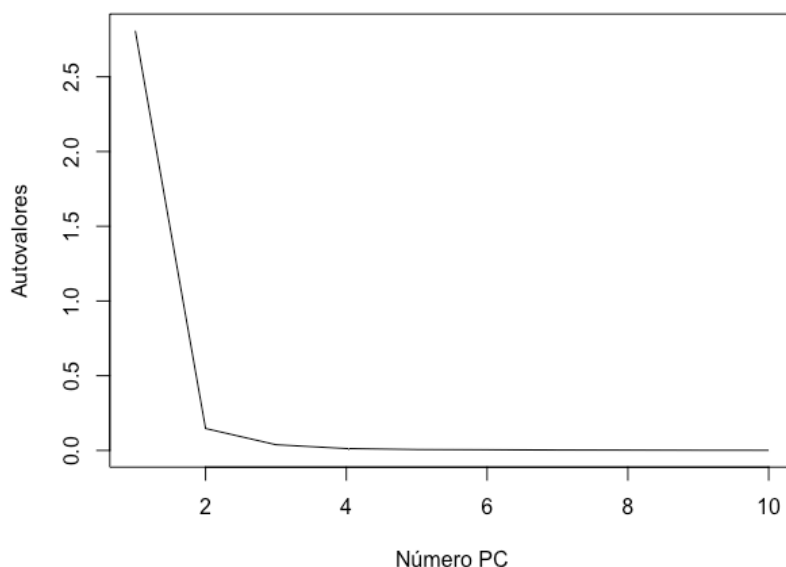
Para o banco de dados de espectros de infravermelho próximo, médio e Raman, o estudo foi realizado utilizando combinações de pré-processamentos, além do já aplicado correção de linha de base. Os pré-processamentos testados para a matriz dos espectros de NIR foram centralização dos dados na média (mc), 1ª derivada (drv) e padronização normal de sinal (snv). Na etapa de validação do modelo, utilizou-se o procedimento da validação cruzada randômica com a seleção de blocos com 10 divisões na métrica da ferramenta PCA clássico e PCA robusto (ROBPCA), o qual se mostrou adequado para o tamanho dos bancos.

O modelo PCA para os dados espectrais de NIR foi construído usando a correção da linha de base (base) combinada com a centralização média (mc). A Tabela 20 foi construída para comparar os autovalores e os valores de variância capturada (%) obtidos no modelo PCA para facilitar a escolha do número de componentes principais mais adequado para o modelo. Na Tabela 20 e Figura 59, pode-se observar que dois PCs são suficientes na construção do modelo PCA.

Tabela 20 - Comparação autovalores e variância explicada do modelo PCA a partir dos espectros de NIR (Banco 2) pré-processados com centralização na média

PCs	Autovalores	Variância capturada (%)	Variância acumulada (%)
1	3,43E-02	95,06	95,06
<b>2</b>	<b>9,02E-04</b>	<b>2,50</b>	<b>97,55</b>
3	4,16E-04	1,15	98,70
4	1,34E-04	0,37	99,08
5	8,91E-05	0,25	99,32

Figura 59 - Gráfico de autovalores *versus* número de componentes principais obtido pela PCA para os espectros NIR (Banco 2)



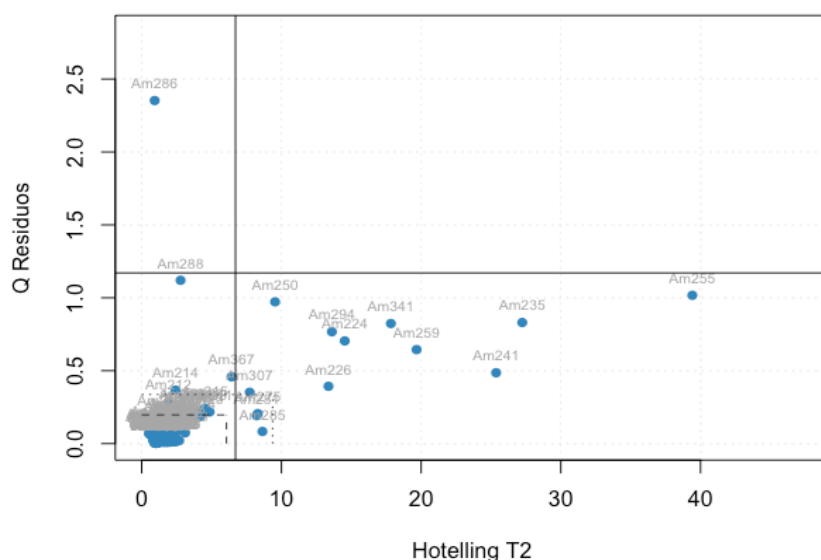
Fonte: O autor, 2019.

Na Figura 60 é observado a ausência de amostras anômalas, visto que a amostra que apresentou alto valor de resíduos possui baixa importância no modelo, e as amostras que apresentaram elevada importância no modelo apresentaram baixos valores de resíduos. Desta forma, nenhuma das amostras apresentaram comportamento anômalo segundo o PCA clássico.

O PCA robusto foi aplicado para confirmar o resultado obtido pelo PCA clássico, conforme Figura 61. No entanto, pode ser observado que três amostras apresentaram uma combinação de altos valores de distância ortogonal e distância entre os escores. Sendo assim, para os espectros de NIR do Banco 2 foi necessário realizar a normalização por transformação de sinal espacial (tss) para deslocar essas amostras para dentro do conjunto de dados sem a necessidade de removê-las.

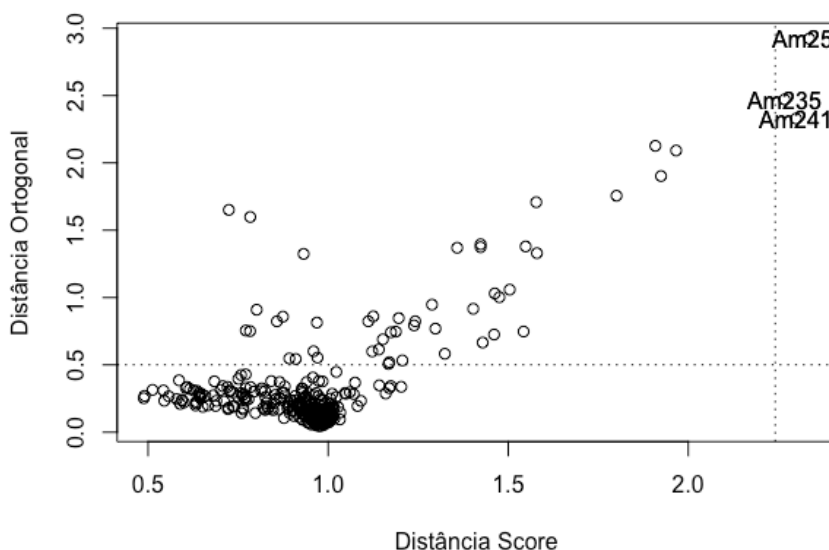


Figura 60 - Gráfico dos resíduos *versus* escores (Q resíduos *versus* Hotelling T2) obtido pela PCA para os espectros NIR do Banco 2



Fonte: O autor, 2019.

Figura 61 - Gráfico da combinação de altos valores distância ortogonal e distância entre os escores obtido pelo ROBPCA para os dados de NIR do Banco 2

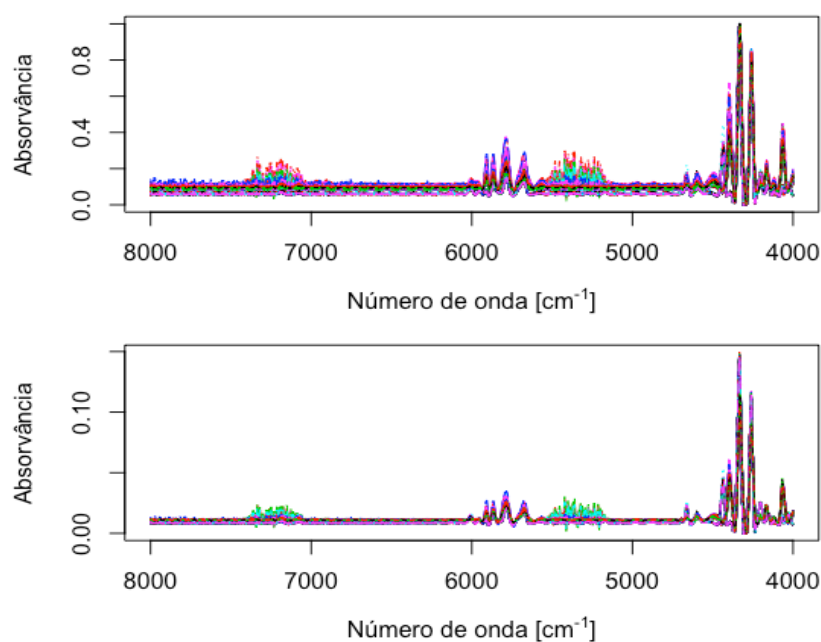


Fonte: O autor, 2019.

Os espectros de NIR antes e após a transformação de sinal espacial são apresentadas na Figura 62. Os espectros de NIR após a transformação do sinal espacial serão considerados na construção dos modelos de regressão de previsão do ponto de entupimento de filtro a frio (PEFF) e viscosidade cinemática contruídos a partir do conjunto espectral de NIR. Um novo modelo de PCA Robusto foi gerado de

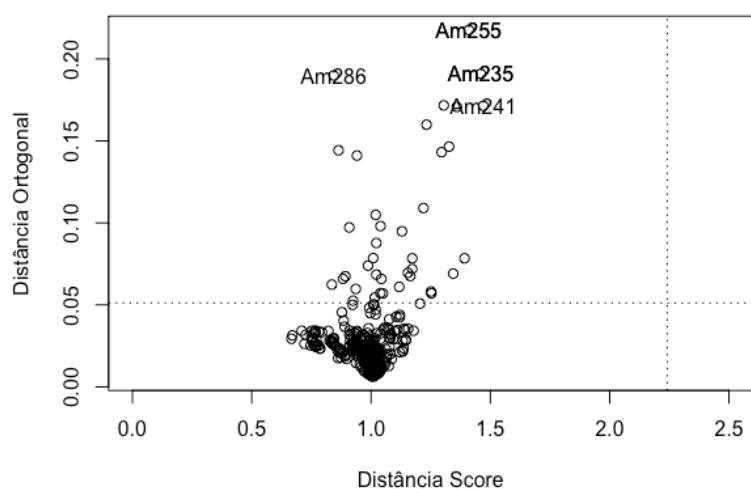
modo a verificar a inclusão das 3 amostras ao conjunto de dados, comprovando que estas amostras não devem ser tratadas como *outliers* e devem ser mantidas no conjunto de dados, conforme Figura 63.

Figura 62 - Espectros de NIR do Banco 2 antes e após a transformação de sinal espacial (tss)



Fonte: O autor, 2019.

Figura 63 - Gráfico da combinação de altos valores distância ortogonal e distância entre os escores obtido pelo ROBPCA para os dados de NIR do Banco 2 após a transformação de sinal espacial



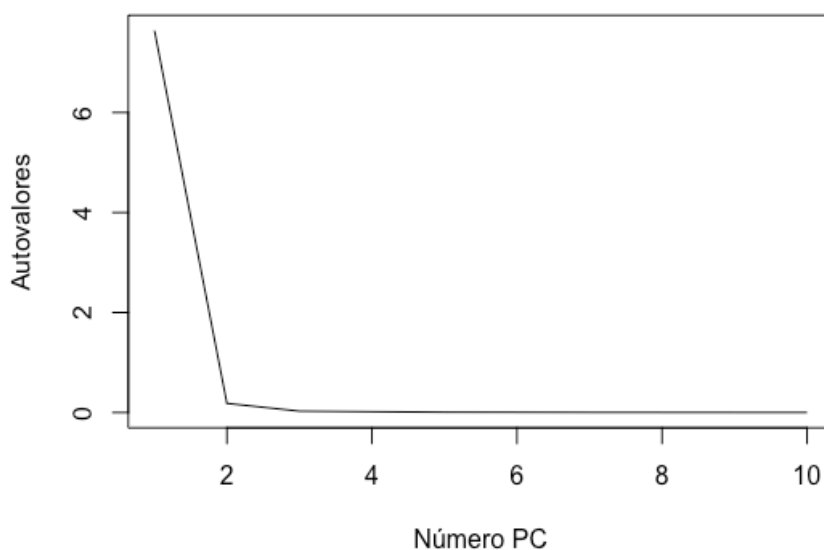
Fonte: O autor, 2019.

O modelo PCA para os dados espectrais de mid-IR foi obtido usando a correção da linha de base (base) combinada com a centralização média (mc). A Tabela 21 foi construída para comparar os autovalores e os valores de variância capturada (%) obtidos no modelo PCA para facilitar a escolha do número de componentes principais mais adequado para o modelo. Para a escolha do número de componentes principais foi adotado como critério os autovalores. Na Tabela 21 e Figura 64, pode-se observar que dois PCs são suficientes na construção do modelo PCA, o que foi confirmado com o gráfico *screeplot*.

Tabela 21 - Comparação autovalores e variância explicada do modelo PCA a partir dos espectros de mid-IR pré-processados com centralização na média (Banco 2)

PCs	Autovalores	Variância capturada (%)	Variância acumulada (%)
1	7,64E+00	96,88	96,88
<b>2</b>	<b>1,83E-01</b>	<b>2,32</b>	<b>99,20</b>
3	2,76E-02	0,35	99,55
4	1,61E-02	0,20	99,75
5	6,47E-03	0,08	99,83

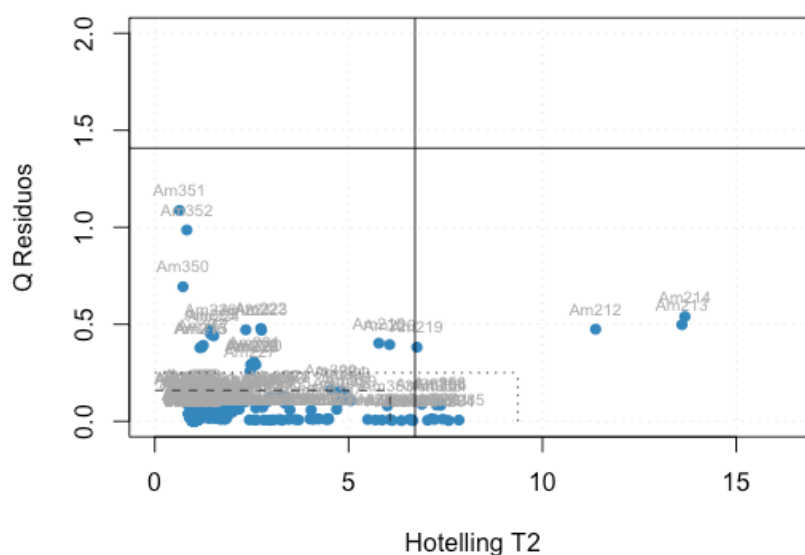
Figura 64 - Gráfico de autovalores *versus* número de componentes principais obtido pela PCA para os espectros mid-IR do Banco 2



Fonte: O autor, 2019.

A Figura 65 demonstra a ausência de amostras anômalas, na qual a maioria das amostras apresentaram baixos valores de resíduos, e as amostras que indicaram elevada importância no modelo apresentaram baixos valores de resíduos. Desta forma, todas as amostras podem ser mantidas no modelo.

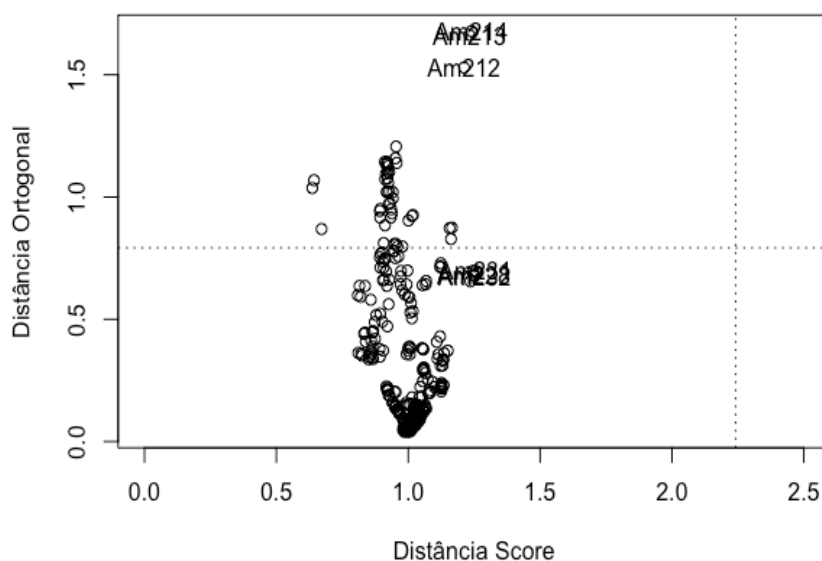
Figura 65 - Gráfico dos resíduos *versus* escores (Q resíduos *versus* Hotelling T2) obtido pela PCA para os espectros mid-IR do Banco 2



Fonte: O autor, 2019.

De modo a confirmar o resultado obtido pelo PCA clássico, foi aplicado o algoritmo ROBPCA e o resultado é apresentado na Figura 66, onde as amostras apresentaram uma combinação de baixos valores distância ortogonal e distância entre os escores, confirmando o resultado obtido pelo PCA clássico.

Figura 66 - Gráfico da combinação de altos valores distância ortogonal e distância entre os escores obtido pelo ROBPCA para os dados de mid-IR do Banco 2



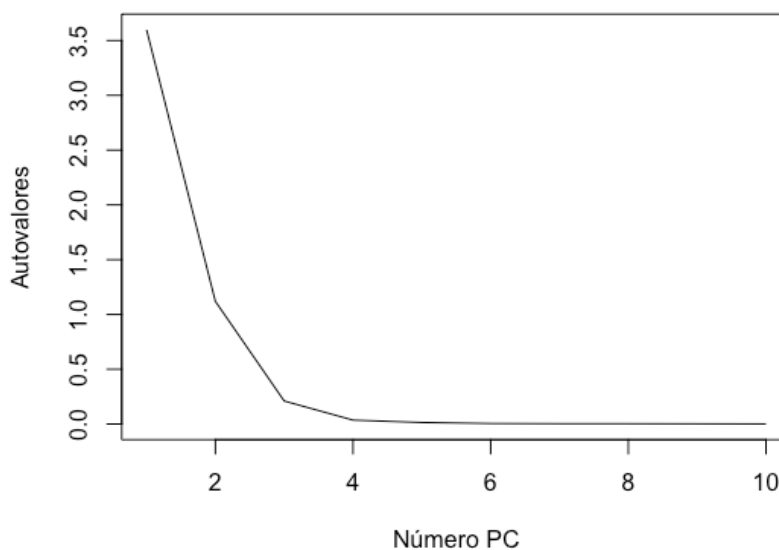
Fonte: O autor, 2019.

Para os espectros de Raman (RAM), o modelo PCA também foi construído usando a correção da linha de base (base) combinada com a centralização média (mc). A Tabela 22 foi construída para comparar os autovalores e os valores de variância capturada (%) obtidos no modelo PCA para facilitar a escolha do número de componentes principais mais adequado para o modelo. Para a escolha do número de componentes principais foi adotado como critério os autovalores. No caso deste estudo são necessários 3 componentes principais conforme mostrado na Tabela 22 e no gráfico de autovalores *versus* número de componentes principais, Figura 67.

Tabela 22 - Comparação autovalores e variância explicada do modelo PCA a partir dos espectros de Raman pré-processados com centralização na média (Banco 2)

PC	Autovalores	Variância Capturada (%)	Variância Acumulada (%)
1	3,60E+00	71,91	71,91
2	1,12E+00	22,37	94,29
<b>3</b>	<b>2,10E-01</b>	<b>4,20</b>	<b>98,49</b>
4	3,56E-02	0,71	99,20
5	1,40E-02	0,28	99,48

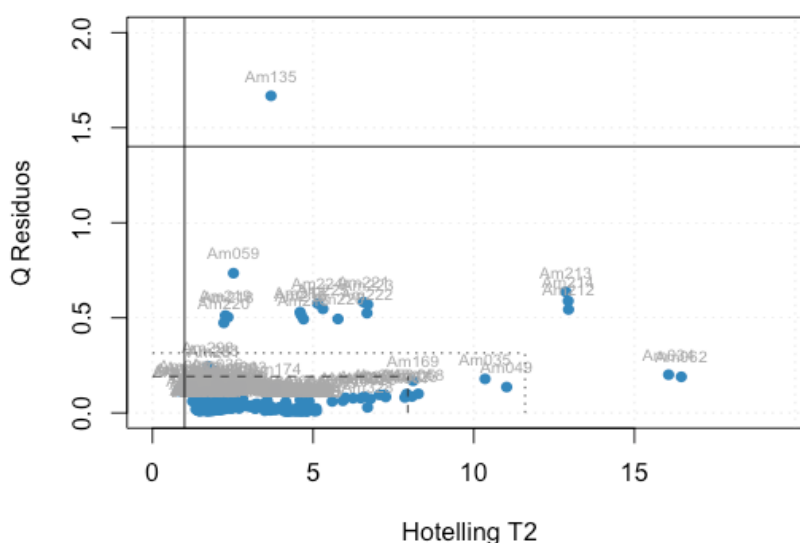
Figura 67 - Gráfico de autovalores *versus* número de componentes principais obtido pela PCA para os espectros Raman do Banco 2



Fonte: O autor, 2019.

Na Figura 68 é observado a ausência de amostras anômalas, visto que a amostra que apresentou alto valor de resíduos possui baixa importância no modelo, e as amostras que apresentaram alta importância no modelo apresentaram baixos valores de resíduos. Desta forma, nenhuma das amostras apresentaram comportamento anômalo segundo o PCA clássico.

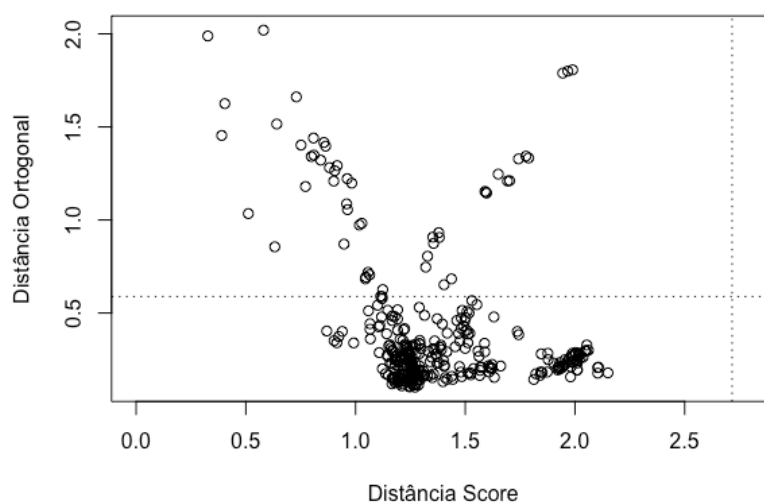
Figura 68 - Gráfico dos resíduos *versus* escores (Q resíduos *versus* Hotelling T2) obtido pela PCA para os espectros Raman do Banco 2



Fonte: O autor, 2019.

A PCA robusta foi aplicada para garantir o resultado obtido pelo PCA clássico, conforme Figura 69. Na Figura 69, pode ser observado que todas as amostras apresentaram uma combinação de baixos valores distância ortogonal e distância entre os escores, confirmando o resultado obtido pelo PCA clássico.

Figura 69 - Gráfico da combinação de altos valores distância ortogonal e distância entre os escores obtido pelo ROBPCA para os dados de Raman do Banco 2



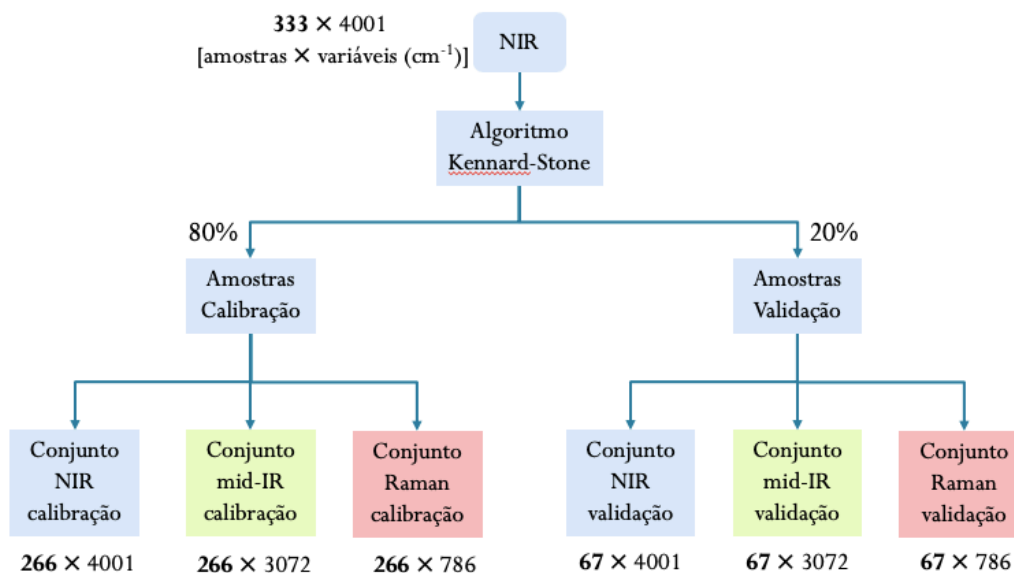
Fonte: O autor, 2019.

#### 6.2.4 Divisão do conjunto de amostras da calibração e validação

Após a PCA, as 333 amostras que compõe o Banco 2 foram divididas em dois conjuntos, a calibração (80% das amostras) e a validação (20% das amostras), utilizando o algoritmo de seleção Kennard-Stone (KS).

A partir do conjunto de dados espectrais de NIR, o algoritmo dividiu as amostras em conjunto de calibração, responsável pela construção do modelo de regressão, e em conjunto de validação, que visa avaliar a capacidade preditiva do modelo. As amostras foram divididas nos dois conjuntos de dados de acordo com esse algoritmo, 266 amostras para o conjunto de calibração (80%) e as 67 amostras restantes foram usadas para o conjunto de validação externa/teste (20%). Essa divisão das amostras foi aplicada aos dados espectrais de mid-IR e Raman, de modo a manter a equidade das amostras nos conjuntos de calibração e validação, visando facilitar a comparação dos modelos, conforme apresentado na Figura 70.

Figura 70 - Fluxograma da divisão das amostras em conjunto de calibração e validação aplicando o algoritmo Kennard-Stone nos espectros de NIR, mid-IR e Raman



Fonte: O autor, 2019.

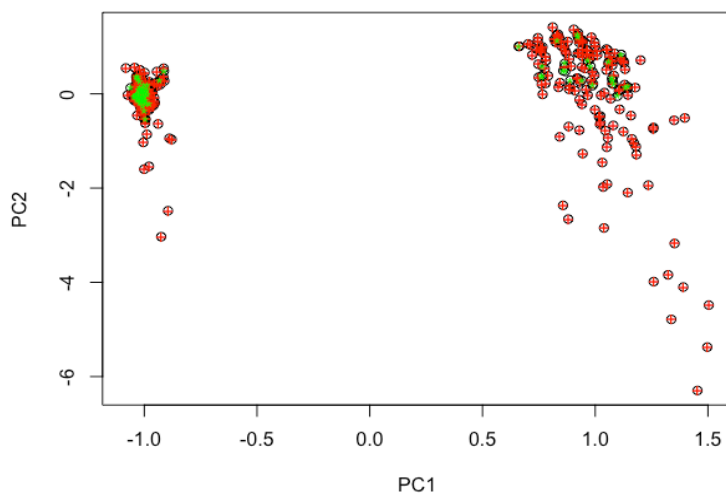
Após a divisão das amostras em calibração e validação, os conjuntos de dados de calibração de NIR, mid-IR e Raman, passaram a ser matrizes de tamanho  $266 \times 4001$ ,  $266 \times 3072$  e  $266 \times 786$  (amostras  $\times$  variáveis), respectivamente. Os conjuntos de dados de validação de NIR, mid-IR e Raman passaram a ser matrizes de  $67 \times 4001$ ,  $67 \times 3072$  e  $67 \times 786$  (amostras  $\times$  variáveis), respectivamente.

Na Figura 71 pode ser observado como ocorreu a distribuição das amostras de calibração (vermelho) e validação (verde) no conjunto total de amostras do Banco 2, considerando os espectros NIR e o algoritmo de Kennard-Stone.

O conjunto de calibração e validação dos espectros (NIR, mid-IR e Raman) formados apresentaram 266 amostras e 67 amostras, respectivamente. Na Tabela 23 são mostradas quais amostras foram selecionadas para o conjunto de calibração e validação.



Figura 71 - Distribuição das amostras de calibração (+) e validação (•) no conjunto total de amostras do Banco 2, considerando os espectros NIR.



Fonte: O autor, 2019.

Tabela 23 - Amostras selecionadas pelo algoritmo Kennard-Stone para compor os conjuntos de calibração e validação para os espectros de NIR, mid-IR e Raman do Banco 2.

Calibração	Validação
Am021 a Am047; Am049; Am057 a Am060; Am062;	Am061; Am067; Am069; Am071;
Am064; Am066; Am068; Am070; Am073 a Am075;	Am072; Am076; Am078; Am083;
Am077; Am079 a Am082; Am084 a Am086; Am088;	Am087; Am089 a Am092; Am097; Am099;
Am093 a Am096; Am098; Am100; Am102 a Am104;	Am101; Am105; Am108 a Am110;
Am106; Am107; Am111 a Am114; Am117; Am118; Am120; Am123;	Am115; Am116; Am119; Am121; Am122; Am124;
Am125 a Am127; Am130; Am133 a Am135; Am138; Am139;	Am128; Am129; Am131; Am132; Am136; Am137;
Am143; Am145 a Am147; Am151 a Am154; Am157;	Am140 a Am142; Am144; Am148 a Am150; Am155;
Am159; Am160; Am162 a Am170; Am172; Am174 a Am226;	Am156; Am161; Am171; Am173; Am227;
Am228 a Am230; Am232; Am233; Am235; Am237; Am239 a Am246;	Am231; Am234; Am236; Am238; Am247;
Am248 a Am250; Am252; Am254; Am255; Am257 a Am267;	Am251; Am253; Am256; Am268; Am270 a Am272;
Am269; Am273 a Am277; Am279 a Am286; Am288 a Am290;	Am278; Am287; Am291; Am297; Am313;
Am292 a Am296; Am298 a Am312; Am314 a Am336; Am338 a	Am337; Am346; Am360; Am 366; Am374
Am345; Am347 a Am359; Am361 a Am365;	
Am367 a Am373; Am375 a Am394	

### 6.2.5 Modelos de regressão

Um total de 432 modelos foram construídos, sendo 216 modelos para a propriedade PEFF e 216 modelos para a propriedade viscosidade. Todos os modelos

foram avaliados através da validação cruzada randômica com a seleção de blocos com 10 divisões.

Neste estudo, a qualidade dos modelos foram avaliados pelos parâmetros estatísticos: raiz quadrada do erro médio quadrático de calibração (RMSEC), de validação cruzada (RMSECV) e validação (RMSEP); erro-padrão relativo de validação (RSEP(%)); os coeficientes de determinação de calibração ( $R^2_{cal}$ ), de validação cruzada ( $R^2_{cv}$ ) e de validação ( $R^2_{val}$ ); o viés de calibração ( $bias_{cal}$ ), de validação cruzada ( $bias_{cv}$ ) e de validação ( $bias_{val}$ ). De acordo com Pomerantsev (2014), os modelos devem apresentar baixos valores de RMSEC e RMSEP, e com valores de bias muito próximos a zero.

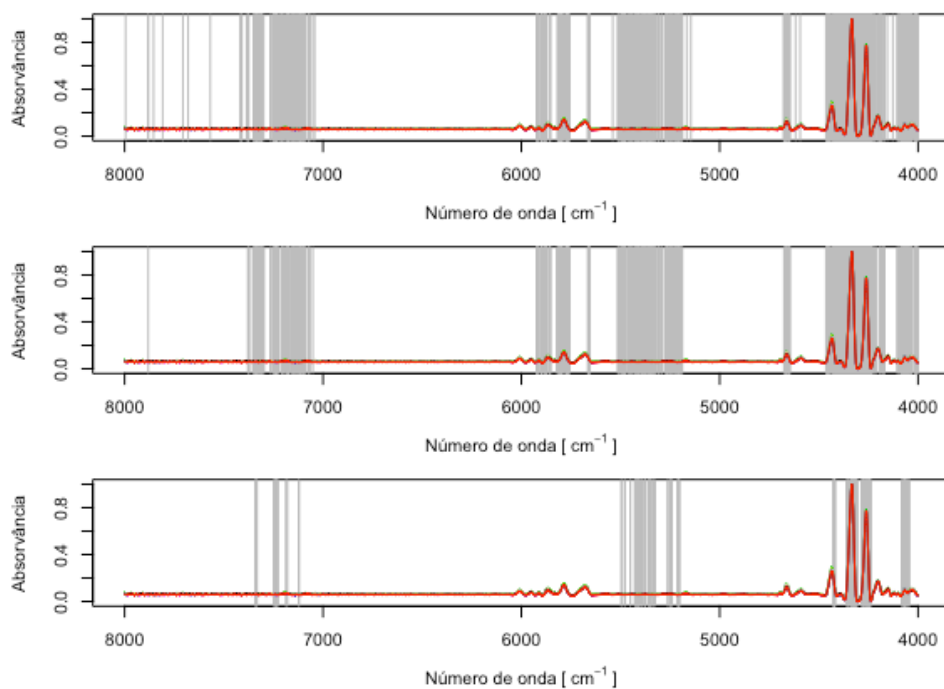
Além disso, a comparação entre os modelos foi realizada com a aplicação da ferramenta SWTP proposta no tópico 3.7.

Antes de iniciar a construção dos modelos de calibração multivariada, as amostras foram separadas em conjuntos de calibração e validação. O algoritmo de Kennard-Stone foi utilizado para realizar a divisão dessas amostras, conforme descrito no tópico 6.2.4.

Após da separação das amostras em calibração e validação, para os conjuntos de dados espectrais de NIR, mid-IR e Raman, foram adotados algumas estratégias para a construção dos modelos de previsão das propriedades, como a construção de 12 modelos de calibração multivariada de primeira ordem (PLS): i) utilizando todo o perfil espectral - PLS; ii) com seleção de variáveis por LW-PLS; iii) com seleção de variáveis por RC-PLS; iv) com seleção de variáveis por VIP-PLS; v) com seleção de variáveis por iPLS; vi) com seleção de variáveis por GA-PLS; vii) com seleção de variáveis por UVE-PLS; viii) com seleção de variáveis por BVE-PLS; ix) com seleção de variáveis por SPA-PLS; x) com seleção de variáveis por REP-PLS; xi) com seleção de variáveis por IPW-PLS; e xii) com seleção de variáveis pelo enPLS seguido pelo PLS.

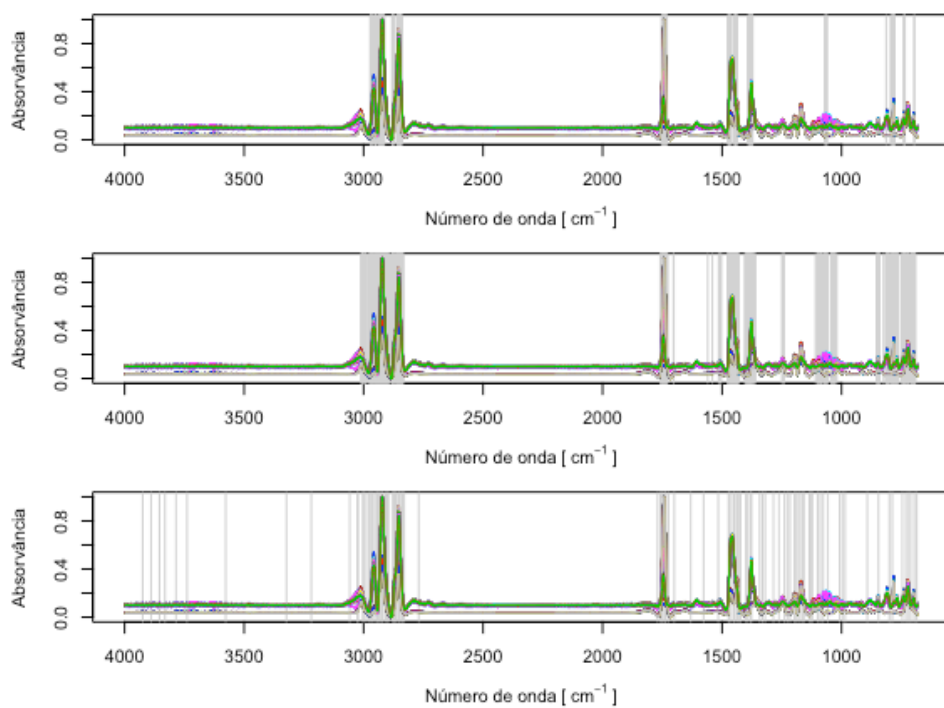
Nas Figuras 72, 73 e 74 estão representados alguns dos intervalos selecionados pelos algoritmos estudados para os espectros NIR, mid-IR e Raman, respectivamente. Como pode ser observado nas Figuras 72, 73 e 74, os intervalos selecionados pelos algoritmos VIP, BVE, IPW, GA e REP estão distribuídos pelas regiões onde há bandas de absorção.

Figura 72 - Intervalos espectrais do NIR (Banco 2) selecionados pelos algoritmos REP, BVE e VIP



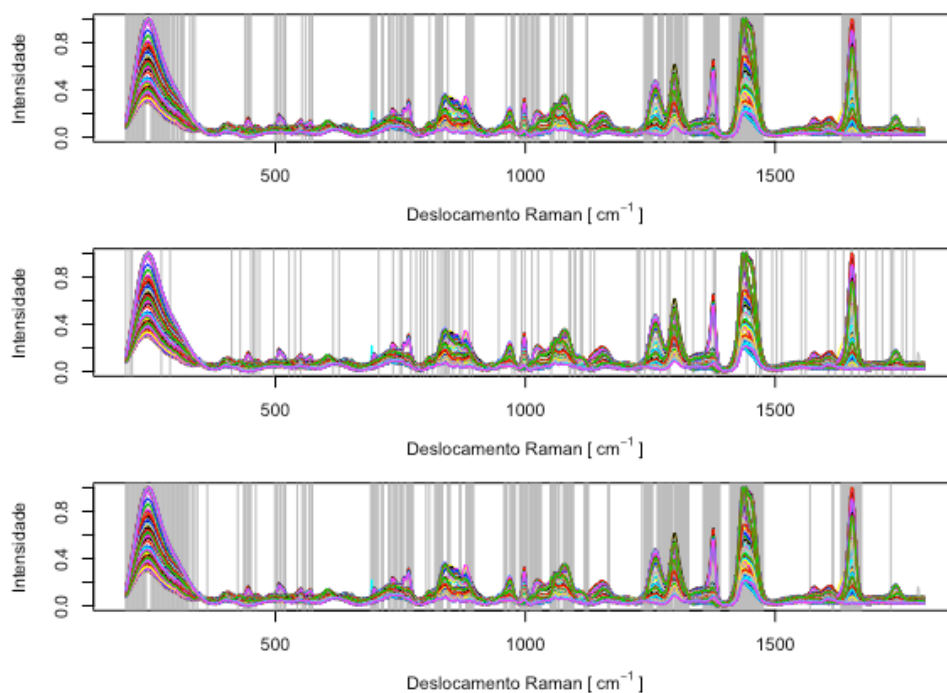
Fonte: O autor, 2019.

Figura 73 - Intervalos espectrais do mid-IR (Banco 2) selecionados pelos algoritmos VIP, BVE e IPW



Fonte: O autor, 2019.

Figura 74 - Intervalos espectrais do Raman (Banco 2) selecionados pelos algoritmos BVE, GA e REP



Fonte: O autor, 2019.

A Tabela 24 apresenta o número de variáveis selecionadas em cada método, para o conjunto de dados de NIR, mid-IR e Raman. Alguns dos métodos empregados conseguiram reduzir significativamente o número de variáveis dos três conjuntos de dados NIR, mid-IR e Raman das propriedades ponto de entupimento de filtro a frio (PEFF) e viscosidade cinemática a 40°C (VISCO). No entanto, o método SPA, RC e IPW selecionaram cerca de metade das variáveis para o conjunto de dados de Raman (RAM) para o PEFF, MIR para o PEFF e RAM para a viscosidade, respectivamente.

Tabela 24 - Quantidade de variáveis selecionadas por cada método de seleção de variáveis para cada um dos sensores (NIR, mid-IR e RAM), para o ponto de entupimento de filtro a frio (PEFF) e viscosidade cinemática do Banco 2

Propriedade	Sensor	Todo	LW	RC	VIP	iPLS	GA	UVE	BVE	SPA	REP	IPW
PEFF	NIR	4001	881	868	197	400	382	320	788	234	839	263
		100%	22,0%	21,7%	4,9%	10,0%	9,5%	8,0%	19,7%	5,8%	21,0%	6,6%
VISCO	NIR	4001	756	102	189	500	389	1523	520	222	607	89
		100%	18,9%	2,5%	4,7%	12,5%	9,7%	38,1%	13,0%	5,5%	15,2%	2,2%
PEFF	MIR	3072	640	1390	162	501	299	348	435	52	449	96
		100%	20,8%	45,2%	5,3%	16,3%	9,7%	11,3%	14,2%	1,7%	14,6%	3,1%
VISCO	MIR	3072	645	10	145	302	304	673	384	50	388	149
		100%	21,0%	0,3%	4,7%	9,8%	9,9%	21,9%	12,5%	1,6%	12,6%	4,9%
PEFF	RAM	786	234	55	217	147	92	153	252	407	312	275
		100%	29,8%	7,0%	27,6%	18,7%	11,7%	19,5%	32,1%	51,8%	39,7%	35,0%
VISCO	RAM	786	173	96	227	146	81	243	163	232	184	354
		100%	22,0%	12,2%	28,9%	18,6%	10,3%	30,9%	20,7%	29,5%	23,4%	45,0%

### 6.2.6 Modelos de previsão para o ponto de entupimento de filtro a frio

Os modelos PLS e SVM foram construídos considerando espectros completos e os intervalos selecionados pela seleção de variáveis com base em algoritmos e informações químicas. O critério R de Wold foi aplicado na escolha do número apropriado de variáveis latentes a serem incluídas em cada modelo de previsão por PLS (STONE, 1974; WOLD, 1978; KRZANOWSKI, 1987).

Na Tabela 25 são apresentados os parâmetros de mérito para os 10 melhores modelos de previsão do PEFF construídos a partir dos conjuntos espectrais de NIR, mid-IR e Raman. No ANEXO E estão disponíveis os parâmetros de mérito de todos os 216 modelos de previsão do PEFF obtidos neste estudo.

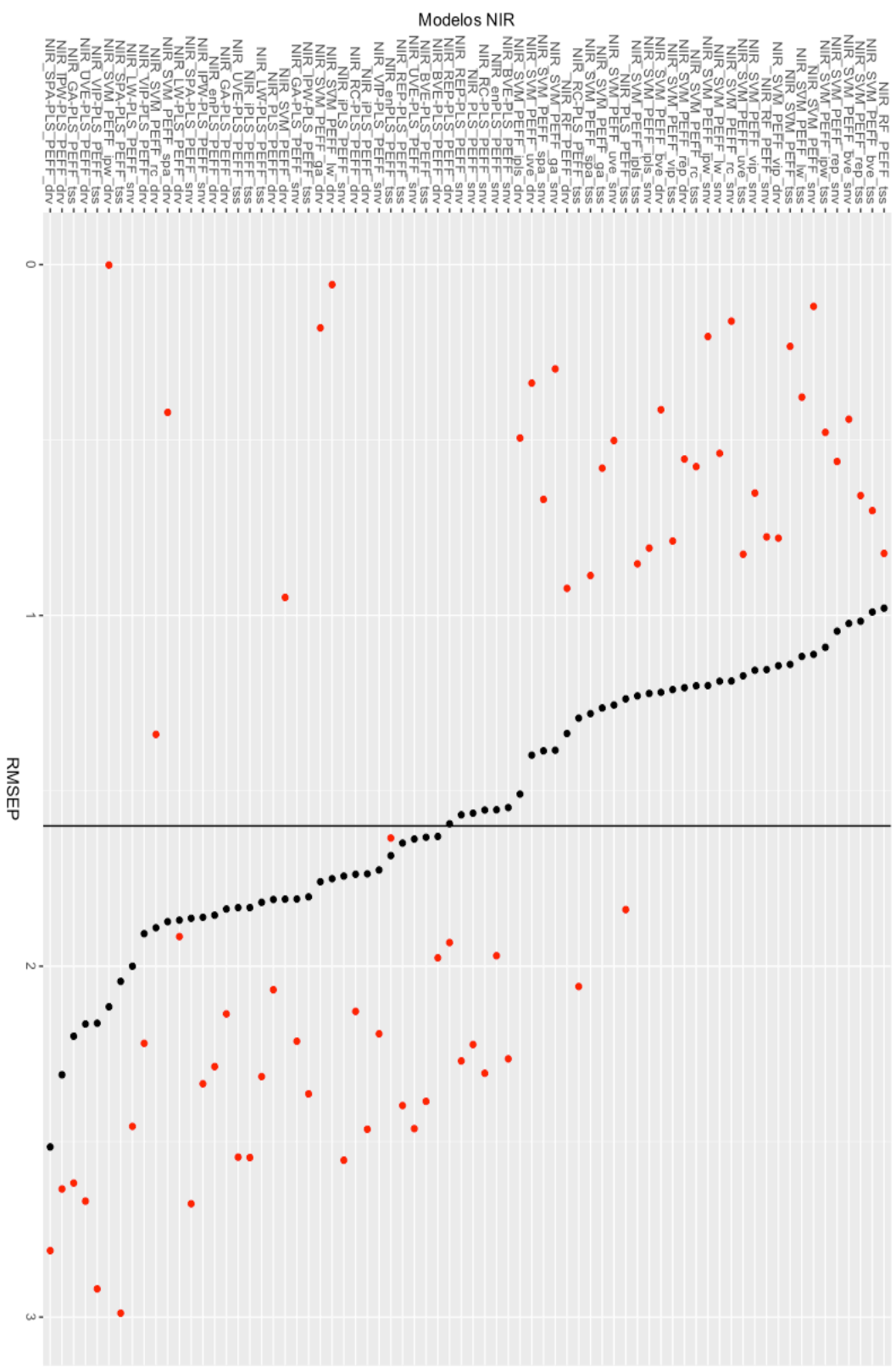
Em uma avaliação inicial, os modelos que apresentaram RMSEP menor que o erro experimental das propriedades estudadas poderiam ser considerados adequados para a previsão dessas propriedades. No entanto, é necessário comparar os modelos de modo a identificar o modelo que possui menor valor de resíduo dentre os modelos avaliados. Nas Figuras 75, 76 e 77, a linha vertical indica o erro experimental do PEFF ( $^{\circ}\text{C}$ ) dos modelos de NIR, mid-IR e Raman, respectivamente. Assim, os modelos que estão à esquerda da linha vertical apresentam RMSEP menor que o erro experimental,  $1,6^{\circ}\text{C}$  para o PEFF.

O teste de Wilcoxon pareado foi utilizado na comparação entre os modelos e uma matriz ( $216 \times 216$ ) com os valores de probabilidade resultantes do teste de Wilcoxon foi construída. Foi obtido um vetor contendo a soma da probabilidade do teste de soma de Wilcoxon (SWTP) e, o valor mínimo desse vetor revelou o melhor modelo para o PEFF.

Tabela 25 - Parâmetros de mérito dos 10 modelos com menor valor de SWTP obtidos a partir dos espectros de NIR, mid-IR e Raman para a previsão do Ponto de Entupimento de Filtro a Frio (PEFF)

Modelos	Espectros	Intervalos	PPx	nsV	kernel	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
MIR_RC-SVM_PEFF_mc	MIR	RC	mc	136	radial	0,8	0,98	0,7	0,94	6,6E-02	1,14E-02	0,4	0,58
MIR_BVE-SVM_PEFF_drv	MIR	BVE	drv	149	radial	0,6	0,99	0,8	0,93	4,4E-01	2,63E-01	1,7	2,99
MIR_IPW-SVM_PEFF_mc	MIR	IPW	mc	145	radial	0,8	0,98	0,8	0,93	5,7E-03	3,61E-02	0,6	3,69
MIR_RC-SVM_PEFF_snv	MIR	RC	snv	264	radial	0,3	1,00	0,8	0,93	2,7E-02	8,59E-02	1,0	3,99
MIR_LW-SVM_PEFF_snv	MIR	LW	snv	266	radial	0,2	1,00	0,9	0,91	1,2E-02	4,88E-04	0,1	4,74
MIR_LW-SVM_PEFF_mc	MIR	LW	mc	129	radial	0,8	0,98	0,8	0,93	1,3E-03	1,32E-02	0,4	6,37
MIR_RF_PEFF_mc	MIR	Todo	mc	-	-	0,6	0,99	0,9	0,91	1,3E-03	2,36E-01	1,6	6,77
MIR_iPLS-SVM_PEFF_mc	MIR	iPLS	mc	135	radial	0,9	0,97	0,9	0,92	2,9E-01	2,20E-01	1,5	9,77
MIR_REP-SVM_PEFF_drv	MIR	REP	drv	263	radial	0,3	1,00	0,9	0,92	7,4E-03	3,54E-02	0,6	9,82
MIR_GA-SVM_PEFF_mc	MIR	GA	mc	125	radial	0,9	0,97	0,8	0,93	3,5E-01	1,39E-01	1,2	9,87

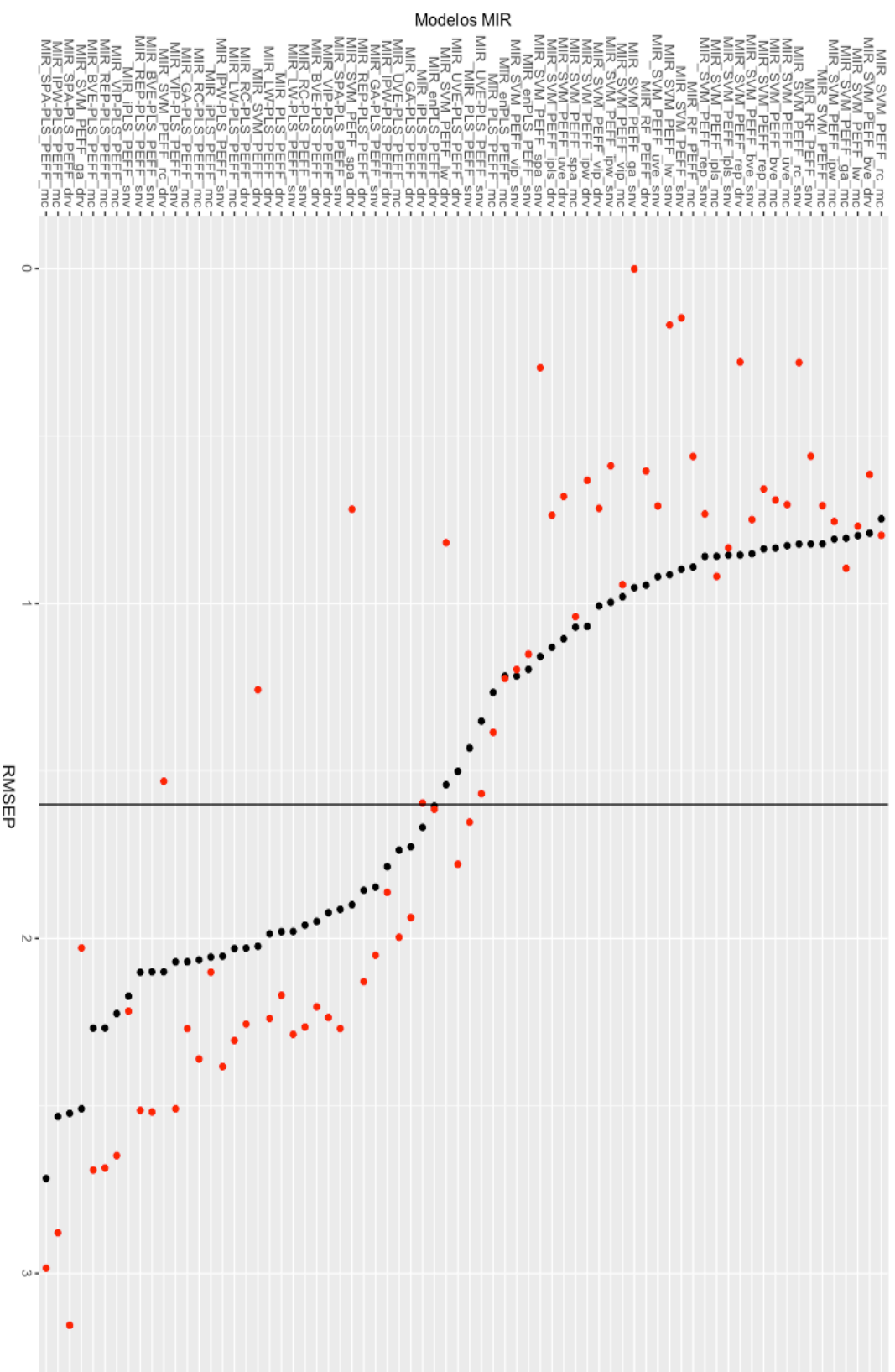
Figura 75 - Modelos de previsão do ponto de entupimento de filtro a frio (PEFF) versus RMSEP (•) e RMSEC (◦), em °C, obtidos a partir do conjunto espectros de NIR, a linha vertical indica o erro experimental



Fonte: O autor, 2019.

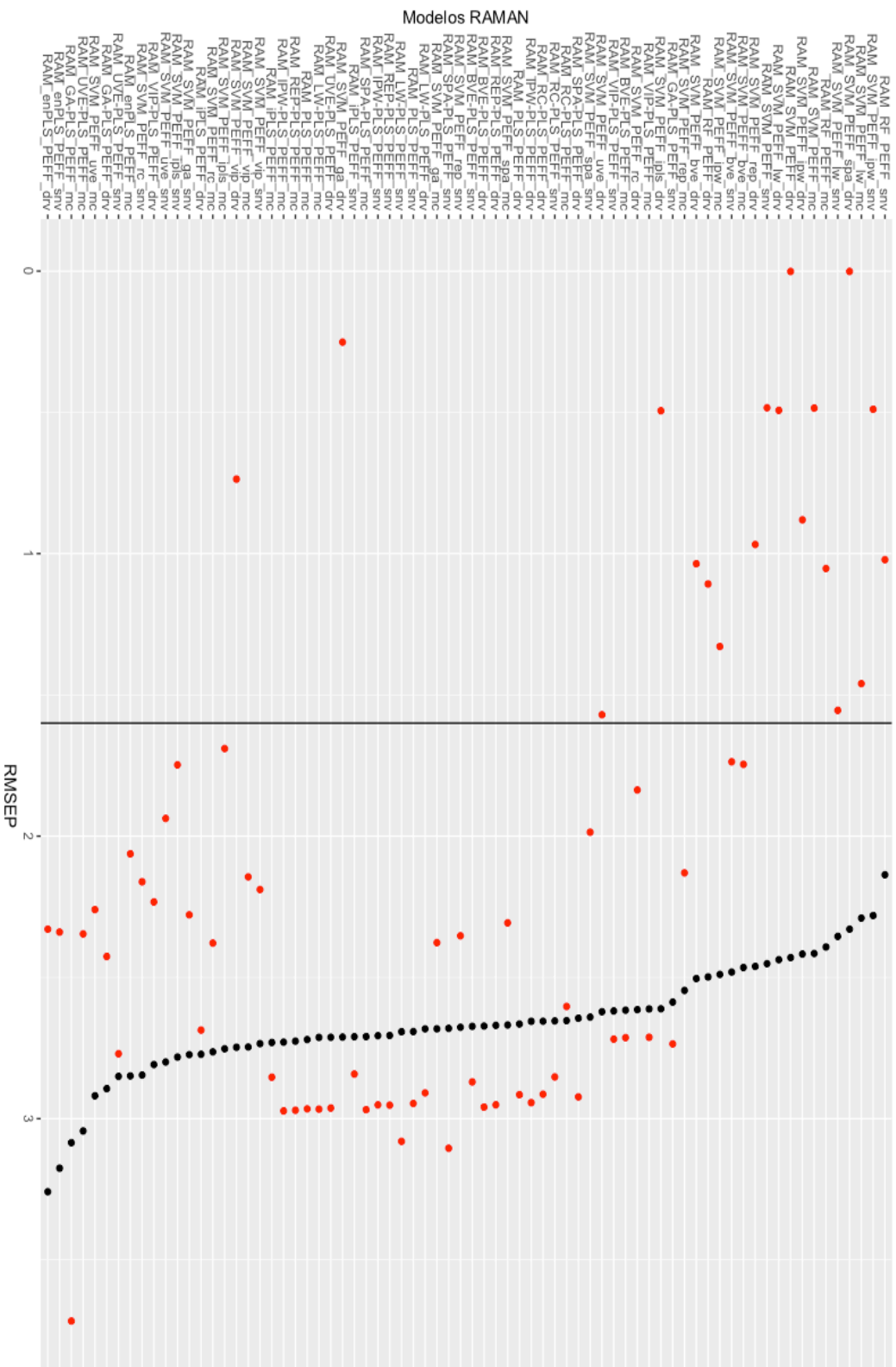


Figura 76 - Modelos de previsão do ponto de entupimento de filtro a frio (PEFF) versus RMSEP (•) e RMSEC (◊), em °C, obtidos a partir do conjunto espectros de mid-IR, a linha vertical indica o erro experimental



Fonte: O autor, 2019.

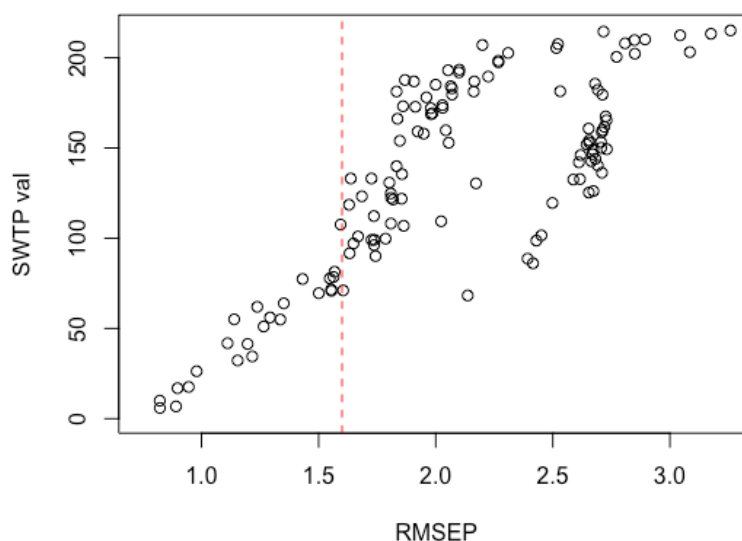
Figura 77 - Modelos de previsão do ponto de entupimento de filtro a frio (PEFF) versus RMSEP ( $\bullet$ ) e RMSEC ( $\bullet$ ), em °C, obtidos a partir do conjunto espectros de Raman, a linha vertical indica o erro experimental



Fonte: O autor, 2019.

Na Figura 78 é apresentado o comportamento dos modelos ao comparar os valores  $SWTP_{val}$  com os valores RMSEP para o PEFF. Os valores de  $SWTP_{val}$  apresentam uma grande variabilidade em função do RMSEP, isso ocorre porque há uma quantidade significativa de modelos lineares e estes por sua vez apresentam baixa capacidade preditiva para propriedades não lineares, como o PEFF.

Figura 78 - Soma da probabilidade do teste de Wilcoxon ( $SWTP_{val}$ ) versus RMSEP ( $^{\circ}C$ ) para todos os modelos de previsão do PEFF, a linha vertical indica o erro experimental



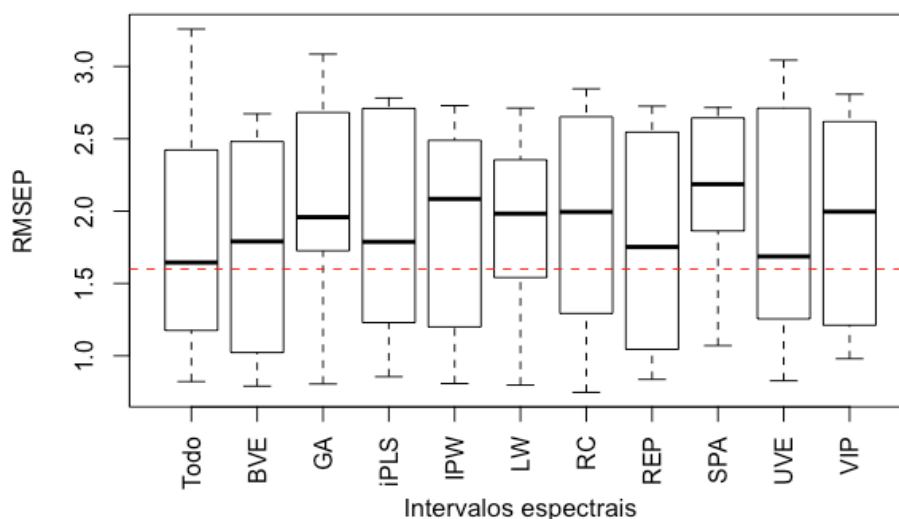
Fonte: O autor, 2019.

Os valores  $SWTP_{val}$  da validação para os modelos foram apresentados na Tabela 25, e o valor mínimo indica o melhor modelo entre os modelos estudados. No caso da previsão do PEFF, o modelo SVM utilizando os espectros de mid-IR e os intervalos obtidos pelo algoritmo RC e pré-processamento centralização da média (MIR\_RC-SVM\_PEFF\_mc), utilizando 136 vetores suportes e *kernel* radial, apresentou maior valor de  $R^2_{pred}$  0,94, menor valor de  $SWTP_{val}$  (0,58) e RMSEP (0,7 $^{\circ}C$ ) dentre os modelos avaliados.

Os próximos três modelos com baixos valores de  $SWTP_{val}$  são: MIR\_BVE-SVM\_PEFF\_drv ( $SWTP_{val}$  = 2,99) com 149 vetores suportes, RMSEP igual a 0,8 $^{\circ}C$  e  $R^2_{pred}$  0,93; MIR\_IPW-SVM\_PEFF\_mc ( $SWTP_{val}$  = 3,69) com 145 vetores suportes, RMSEP igual a 0,8 $^{\circ}C$  e  $R^2_{pred}$  igual a 0,93; e MIR\_RC-SVM\_PEFF\_snv ( $SWTP_{val}$  = 3,99) com 264 vetores suportes, RMSEP igual a 0,8 $^{\circ}C$  e  $R^2_{pred}$  igual a 0,93. Os melhores modelos de previsão do PEFF utilizando espectros de mid-IR foram obtidos

com a aplicação de seleção de variáveis (RC, BVE e IPW) combinada com máquina de vetores suporte (SVM). Ao avaliar as Figuras 79 e 80 em conjunto com a Tabela 25, observa-se que a aplicação de técnicas de seleção de variáveis associadas a uma ferramenta linear (PLS) não garantem a obtenção de modelos com baixos valores de RMSEP e  $SWTP_{val}$  na previsão do PEFF. Isto ocorre, principalmente, devido a dificuldade em prever o PEFF, que é uma propriedade física que apresenta comportamento não-linear. Os melhores modelos foram obtidos com seleção de variáveis combinados com máquina de vetores suporte (método não-linear), apresentando menores valores de RMSEP ( $^{\circ}\text{C}$ ). Conforme apresentado na Figura 79, os modelos apresentaram baixo desempenho e elevada variabilidade.

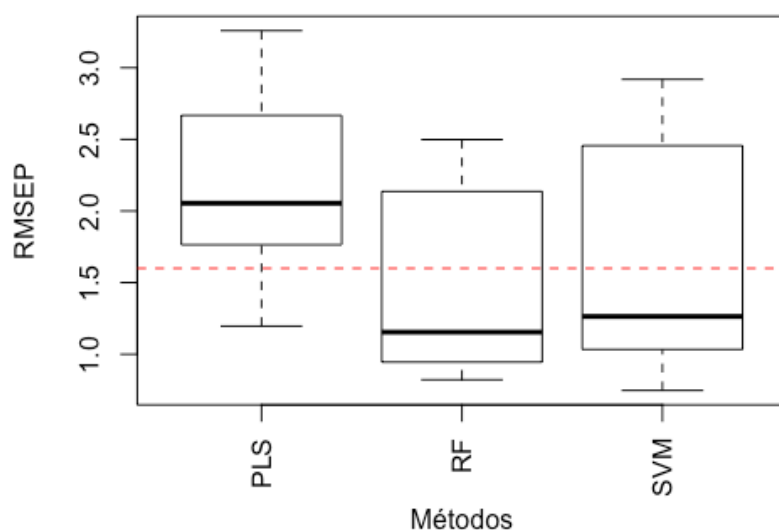
Figura 79 - RMSEP ( $^{\circ}\text{C}$ ) versus intervalos espectrais para todos os modelos de previsão do PEFF, a linha horizontal tracejada indica o erro experimental



Fonte: O autor, 2019.

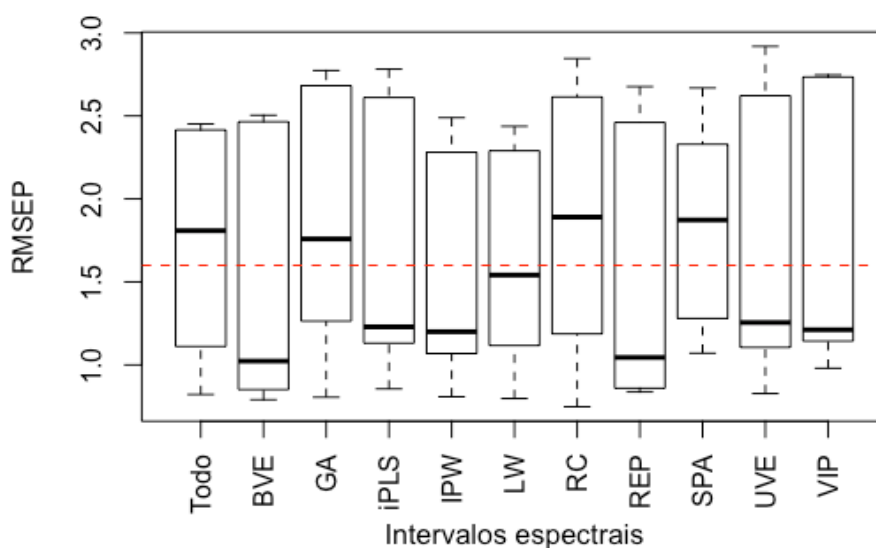
A Figura 80 foi obtida considerando todos os modelos PLS, RF e SVM com aplicação de seleção de variáveis. Sabendo disso, a Figura 81 foi construída para comprovar que a seleção de variáveis com o método não-linear, SVM, foi mais adequada na previsão do PEFF, obtendo valores de RMSEP próximo ou abaixo do erro experimental.

Figura 80 - RMSEP ( $^{\circ}\text{C}$ ) versus métodos PLS, RF e SVM para todos os modelos de previsão do PEFF, a linha horizontal tracejada indica o erro experimental



Fonte: O autor, 2019.

Figura 81 - RMSEP ( $^{\circ}\text{C}$ ) versus intervalos utilizados, considerando apenas os modelos SVM de previsão do PEFF, a linha horizontal tracejada indica o erro experimental

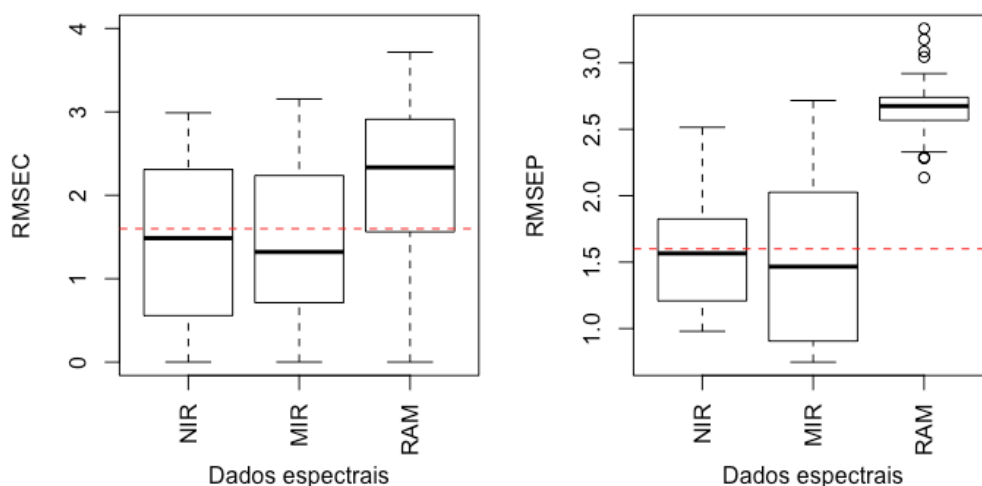


Fonte: O autor, 2019.

Os modelos de previsão do PEFF usando espectros NIR e Raman mais bem posicionados no ranking de menor valor de  $\text{SWTP}_{\text{val}}$  aparecem na 26<sup>a</sup> e 65<sup>a</sup> posição, respectivamente. Sabendo disso, concluímos que a utilização dos espectros NIR e Raman na previsão do PEFF não trouxeram melhorias aos resultados obtidos com

espectros mid-IR. De fato, este fenômeno pode ser confirmado com a Figura 82, onde pode ser visto que os modelos utilizando espectros mid-IR apresentam menores valores de RMSEC e RMSEP, embora os modelos com mid-IR tenham apresentado maior variabilidade. No caso do espectro Raman, observa-se que apesar da menor variabilidade dos resultados, os modelos Raman apresentaram o pior desempenho dentre os modelos estudados.

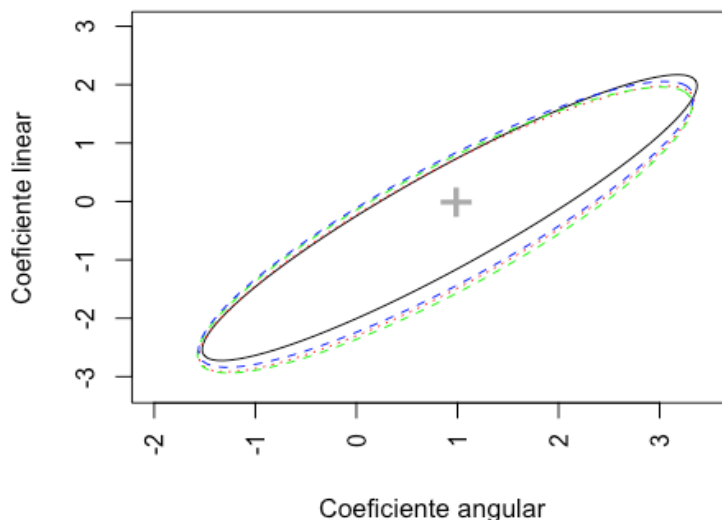
Figura 82 - RMSEC ( $^{\circ}\text{C}$ ) e RMSEP ( $^{\circ}\text{C}$ ) *versus* dados espectrais (NIR, mid-IR e Raman) para todos os modelos de previsão do PEFF, a linha horizontal tracejada indica o erro experimental



Fonte: O autor, 2019.

A elipse da região conjunta de confiança (EJCR) foi empregada nos 4 melhores modelos de previsão do PEFF, para avaliar a existência de erros sistemáticos e a precisão do método. Conforme mostrado na Figura 83, é possível observar que os modelos de previsão selecionados não apresentaram erro sistemático e o modelo MIR\_RC-SVM\_PEFF\_mc apresentou maior precisão quando comparado aos modelos apresentados no gráfico da EJCR da Figura 83.

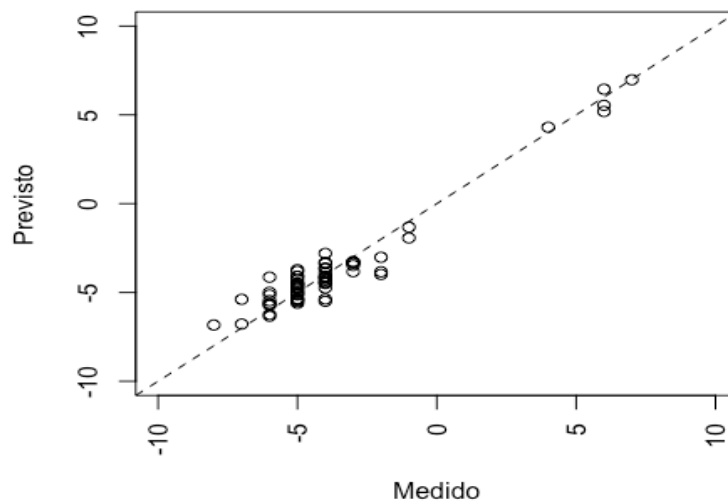
Figura 83 - Elipse da região conjunta de confiança (EJCR) do conjunto de validação para os modelos de previsão do PEFF: (—) MIR\_RC-SVM\_PEFF\_mc; (- -) MIR\_BVE-SVM\_PEFF\_drv; (- -) MIR\_IPW-SVM\_PEFF\_mc; (- -) MIR\_RC-SVM\_PEFF\_snv.



Fonte: O autor, 2019.

Ao comparar os modelos por SWTP obtidos e por não apresentar erro sistemático na EJCR, o modelo MIR\_RC-SVM\_PEFF\_mc foi selecionado como o melhor modelo, apresentando RMSEP igual a  $0,7^{\circ}\text{C}$  e  $R^2_{\text{pred}}$  0,94 (Tabela 25). O gráfico dos valores previstos ( $^{\circ}\text{C}$ ) versus medidos ( $^{\circ}\text{C}$ ) para o melhor modelo do PEFF é mostrado na Figura 84.

Figura 84 - Gráfico dos valores previstos ( $^{\circ}\text{C}$ ) versus medidos ( $^{\circ}\text{C}$ ) para o modelo MIR\_RC-SVM\_PEFF



Fonte: O autor, 2019.

Esse resultado pode ser comparado com alguns estudos da literatura, tais como Cunha *et al.* (2017) e Cunha, Torres e Luna (2020). Cunha *et al.* (2017) estudaram uma metodologia para determinar o ponto de entupimento de amostras de biodiesel usando espectroscopia no infravermelho médio (4000-680  $\text{cm}^{-1}$ ) e modelo PLS com três variáveis latentes e RMSEP igual a 1,1°C. Cunha, Torres e Luna (2020) desenvolveram uma metodologia para previsão do ponto de entupimento de amostras de biodiesel usando espectros de infravermelho próximo e obtiveram um modelo com RMSEP igual a 0,7°C.

#### 6.2.7 Modelos de previsão para a viscosidade cinemática

Os modelos PLS e SVM foram construídos considerando espectros completos e os intervalos selecionados pela seleção de variáveis com base em algoritmos. O critério R de Wold foi aplicado na escolha do número apropriado de variáveis latentes a serem incluídas nos modelos de previsão por PLS (STONE, 1974; WOLD, 1978; KRZANOWSKI, 1987).

Na Tabela 26 são apresentados os parâmetros de mérito para os 10 modelos de previsão da viscosidade com menor valor de SWTP, construídos a partir dos conjuntos espectrais de NIR, mid-IR (MIR) e Raman (RAM), respectivamente.

Nas Figuras 85, 86 e 87, a linha vertical indica o erro experimental da viscosidade cinemática ( $\text{mm}^2\cdot\text{s}^{-1}$ ) dos modelos de NIR, mid-IR e Raman, respectivamente. Assim, os modelos que estão à esquerda da linha vertical apresentam RMSEP menor que o erro experimental, 0,0267  $\text{mm}^2\cdot\text{s}^{-1}$  para a viscosidade.

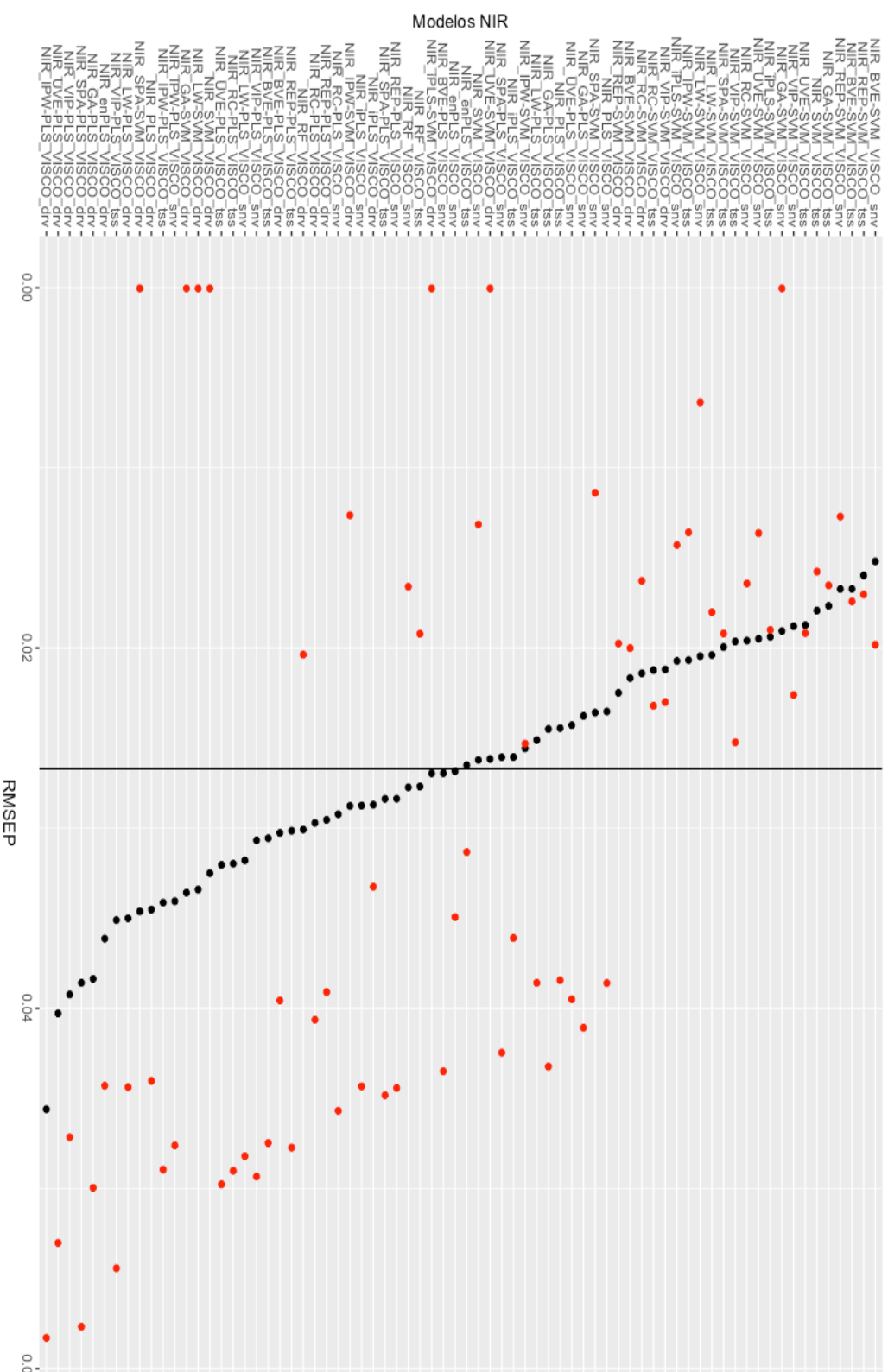
O teste de Wilcoxon pareado foi utilizado na comparação entre os modelos e uma matriz (216 × 216) com os valores de probabilidade resultantes do teste de Wilcoxon foi construída. Foi obtido um vetor contendo os valores de SWTP e, o valor mínimo desse vetor indicou o melhor modelo de previsão da viscosidade cinemática.



Tabela 26 - Parâmetros de mérito dos 10 modelos com menor valor de SWTP, SVM com seleção de variáveis e Florestas Aleatórias (RF) obtidos a partir dos espectros de NIR para a previsão da Viscosidade Cinemática (VISCO)

Modelos	Espectros	Intervalos	PPx	nsV	kernel	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
MIR_VIP-SVM_VISCO_snv	MIR	VIP	snv	266	radial	0,0212	0,98	0,0137	0,99	1,29E-03	5,40E-04	0,15	0,51
MIR_REP-SVM_VISCO_snv	MIR	REP	snv	266	radial	0,0202	0,98	0,0155	0,98	8,07E-04	9,62E-04	0,27	1,83
MIR_BVE-SVM_VISCO_snv	MIR	BVE	snv	266	radial	0,0206	0,98	0,0160	0,98	1,21E-03	1,06E-03	0,29	3,78
MIR_IPW-SVM_VISCO_mc	MIR	IPW	mc	266	radial	0,0207	0,98	0,0168	0,98	4,55E-04	4,79E-04	0,19	4,30
MIR_UVE-SVM_VISCO_mc	MIR	UVE	mc	266	radial	0,0138	0,99	0,0167	0,98	8,00E-04	5,30E-04	0,20	5,38
MIR_RF_VISCO_snv	MIR	Todo	snv	-	-	0,0119	0,99	0,0184	0,98	8,38E-06	1,34E-04	0,10	5,51
MIR_SVM_VISCO_mc	MIR	Todo	mc	266	radial	0,0126	0,99	0,0173	0,98	5,32E-04	7,39E-04	0,04	5,86
MIR_RF_VISCO_mc	MIR	Todo	mc	-	-	0,0115	0,99	0,0251	0,95	4,82E-05	2,85E-03	0,47	7,38
MIR_UVE-SVM_VISCO_snv	MIR	UVE	snv	266	radial	0,0156	0,99	0,0188	0,97	7,26E-04	1,45E-03	0,33	9,87
MIR_LW-SVM_VISCO_mc	MIR	LW	mc	266	radial	0,0133	0,99	0,0177	0,98	6,13E-04	7,11E-04	0,39	9,96

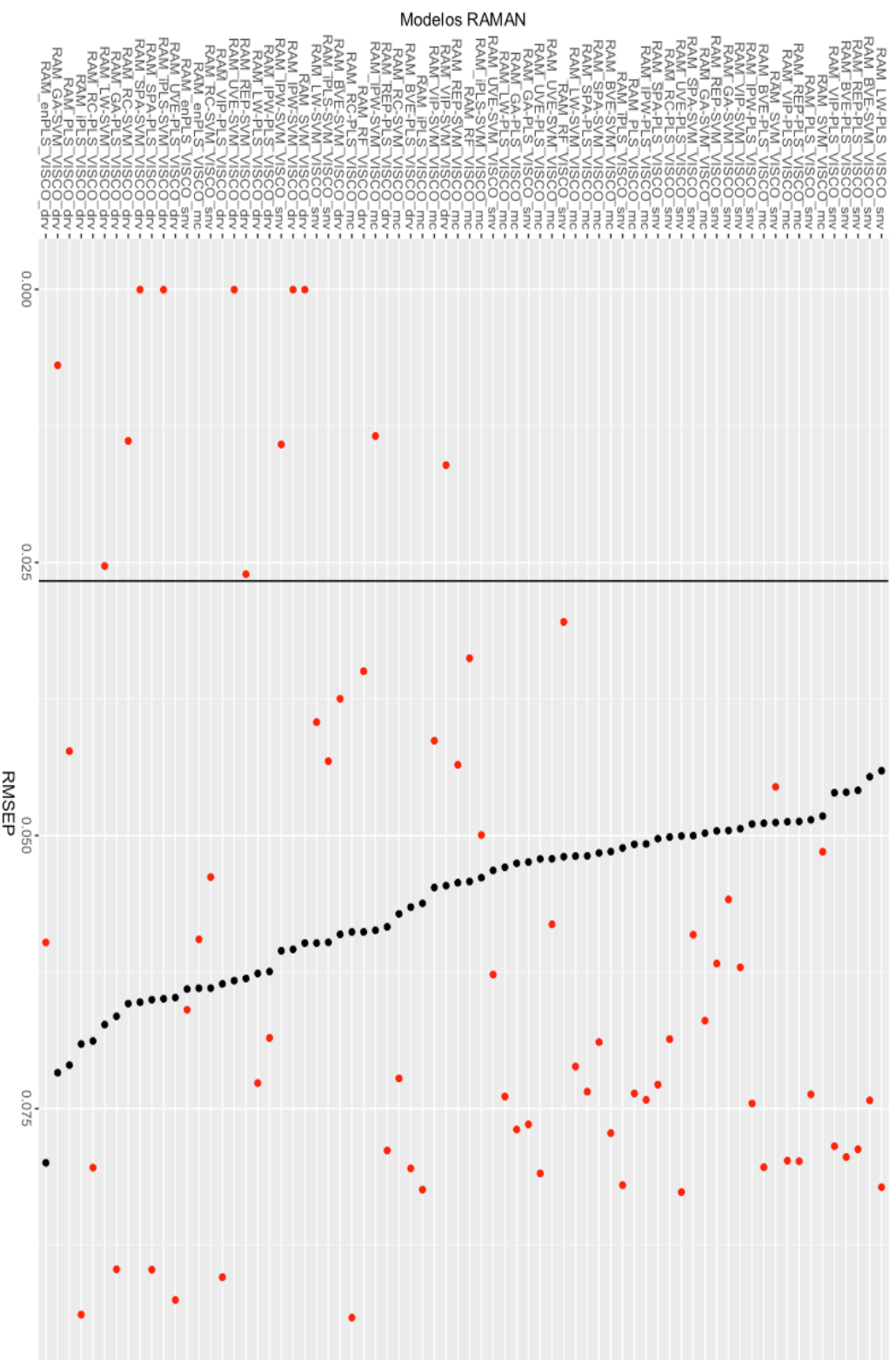
Figura 85 - Modelos de previsão da viscosidade (VISCO) versus RMSEP (•) e RMSEC (•), em  $\text{mm}^2\cdot\text{s}^{-1}$ , obtidos a partir do conjunto espectros de NIR, a linha vertical indica o erro experimental



Fonte: O autor, 2019.



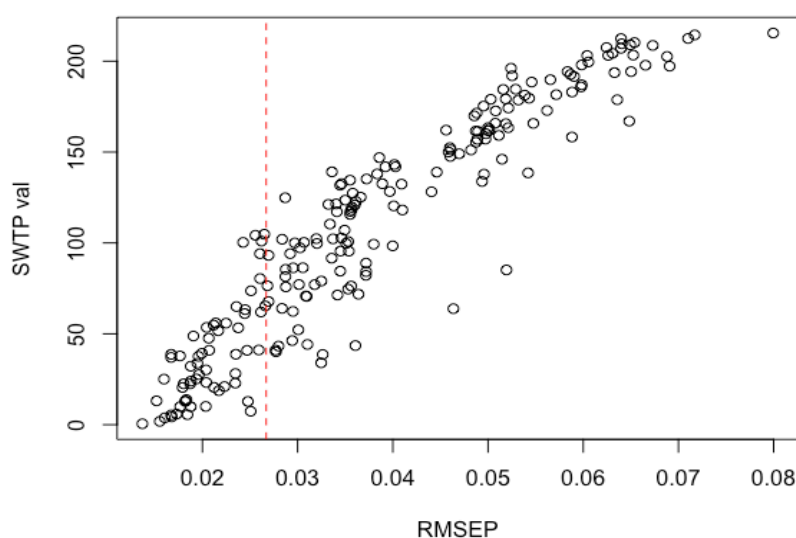
Figura 87 - Modelos de previsão da viscosidade (VISCO) versus RMSEP (•) e RMSEC (◦), em  $\text{mm}^2\cdot\text{s}^{-1}$ , obtidos a partir do conjunto espectros de Raman (RAM), a linha vertical indica o erro experimental



Fonte: O autor, 2019.

Na Figura 88 é apresentado o comportamento dos modelos ao comparar os valores  $SWTP_{val}$  com os valores RMSEP para a viscosidade. Os valores de  $SWTP_{val}$  apresentam uma grande variabilidade em função do RMSEP, isso ocorre porque há uma quantidade significativa de modelos lineares e estes por sua vez apresentam baixa capacidade preditiva para propriedades não lineares, como a viscosidade.

Figura 88 - Soma da probabilidade do teste de Wilcoxon ( $SWTP_{val}$ ) *versus* RMSEP ( $mm^2 \cdot s^{-1}$ ) para todos os modelos de previsão da Viscosidade, a linha vertical indica o erro experimental



Fonte: O autor, 2019.

Os parâmetros de mérito e os valores  $SWTP_{val}$  da validação para cada modelo foram disponibilizados na Tabela 26, e o valor mínimo indica o melhor modelo entre os modelos estudados. Para a previsão da viscosidade cinemática, o modelo SVM utilizando os intervalos espectrais de mid-IR obtidos pelo algoritmo VIP e pré-processamento snv (MIR\_VIP-SVM\_VISCO\_snv) apresentou o menor valor  $SWTP_{val}$  (0,51) entre os modelos avaliados.

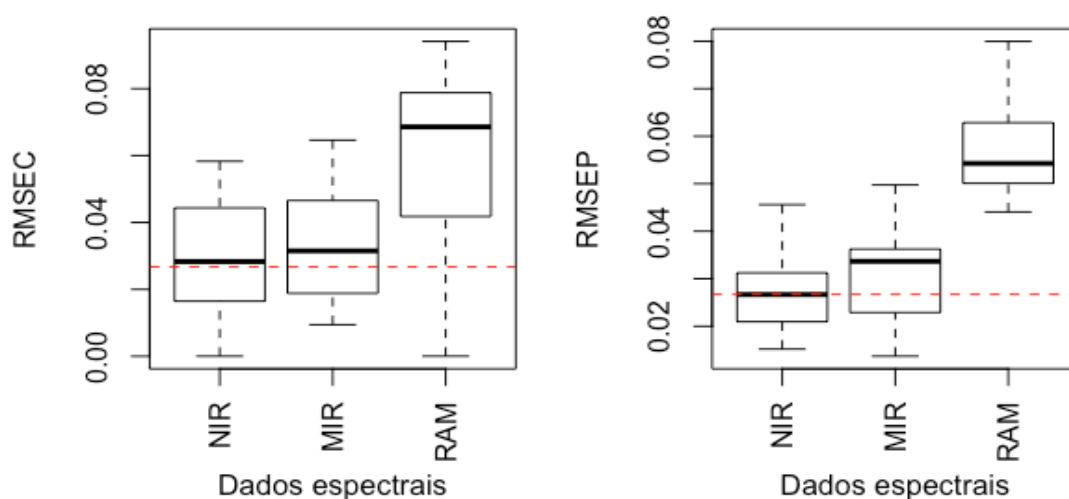
No entanto, o modelo de calibração MIR\_VIP-SVM\_VISCO\_snv, para viscosidade, foi construído com 266 vetores de suporte de um conjunto de dados contendo 266 amostras (calibração). O número de vetores de suporte (nSV) de um modelo deve ser menor que o número de amostras nele contidos. A igualdade entre o nSV e o número de amostras significa que, na construção do modelo, foi utilizado um vetor de suporte diferente para explicar cada amostra, o que indica um comportamento de memorização e não um comportamento de aprendizado. Neste

caso, não é recomendável utilizar o modelo MIR\_VIP-SVM\_snv na previsão da viscosidade. Em seguida, observou-se a sequência dos valores SWTP mais baixos, observa-se que o mesmo ocorre aos próximos 4 modelos com baixo SWTP: i) MIR\_REP-SVM\_VISCO\_snv com SWTP igual a 1,83 e 266 vetores de suporte; ii) MIR\_BVE-SVM\_VISCO\_snv com SWTP igual a 3,78 e 266 vetores de suporte; iii) MIR\_IPW-SVM\_VISCO\_mc com SWTP igual a 4,30 e 266 vetores de suporte; e iv) MIR\_UVE-SVM\_VISCO\_mc com SWTP igual a 5,38 e 266 vetores suporte.

O sexto modelo com menor valor de SWTP foi o modelo RF para o conjunto espectral de mid-IR com pré-processamento SNV (MIR\_RF\_VISCO\_snv), que pode ser considerado o mais adequado e apresentou baixo RMSEP ( $0,0184 \text{ mm}^2\cdot\text{s}^{-1}$ ) e SWTP (5,51).

Os modelos de previsão da viscosidade usando espectros NIR e Raman mais bem posicionados no *ranking* de menor valor de  $\text{SWTP}_{\text{val}}$  aparecem na 15ª e 85ª posição, respectivamente. Sabendo disso, concluímos que a utilização dos espectros de Raman na previsão do viscosidade não contribuiu com melhores resultados e não superou os modelos obtidos por mid-IR. De fato, este comportamento dos espectros de Raman pode ser confirmado com a Figura 89, onde pode ser observado que os modelos utilizando espectros Raman apresentam maiores valores de RMSEC e RMSEP.

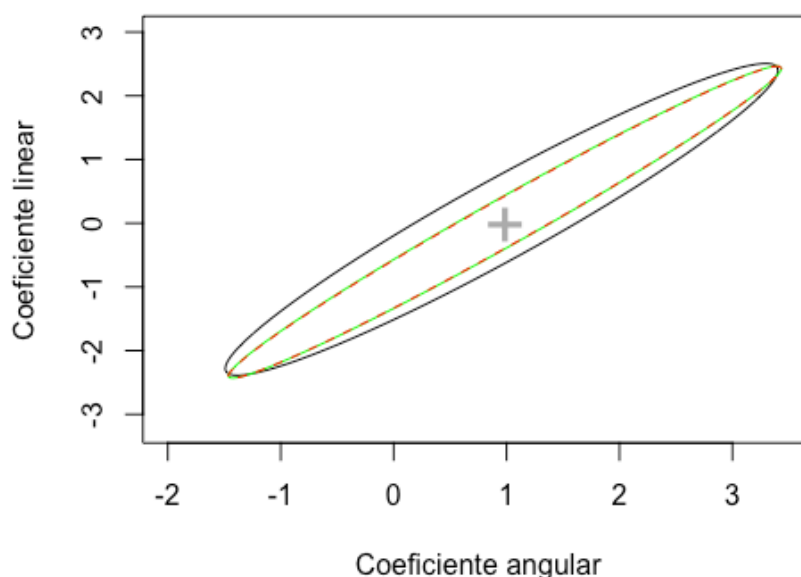
Figura 89 - RMSEC ( $\text{mm}^2\cdot\text{s}^{-1}$ ) e RMSEP ( $\text{mm}^2\cdot\text{s}^{-1}$ ) versus dados espectrais (NIR, mid-IR e Raman) para todos os modelos de previsão da viscosidade, a linha horizontal indica o erro experimental



Fonte: O autor, 2019.

A elipse da região conjunta de confiança (EJCR) foi aplicada no melhor modelo de previsão da viscosidade, para avaliar a existência de erros sistemáticos e a precisão do método. Conforme mostrado na Figura 90, é possível observar que o modelo de previsão selecionados são precisos e não apresentam erro sistemático. No caso, os modelos MIR\_VIP-SVM\_VISCO\_snv e MIR\_REP-SVM\_VISCO\_snv são mais precisos que o modelo MIR\_RF-VISCO\_snv.

Figura 90 - Elipse da região conjunta de confiança (EJCR) do conjunto de validação para os modelos de previsão da viscosidade: (–) MIR\_RF-VISCO\_snv, (–) MIR\_VIP-SVM\_VISCO\_snv e (-.-) MIR\_REP-SVM\_VISCO\_snv.

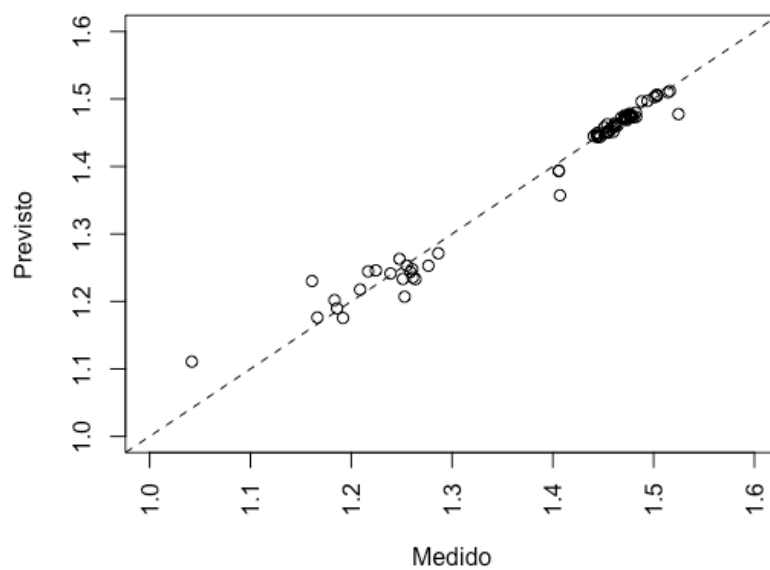


Fonte: O autor, 2019.

Ao comparar os modelos por SWTP obtidos e por não apresentar erro sistemático na EJCR, o modelo MIR\_RF\_VISCO\_snv foi confirmado como o melhor modelo, apresentando RMSEP igual a  $0,0184 \text{ mm}^2\cdot\text{s}^{-1}$  e  $R^2_{\text{pred}} 0,98$  (Tabela 26). O gráfico dos valores previstos ( $\text{mm}^2\cdot\text{s}^{-1}$ ) versus medidos ( $\text{mm}^2\cdot\text{s}^{-1}$ ) para o melhor modelo da viscosidade é mostrado na Figura 91.

Esse resultado pode ser comparado com o estudo desenvolvido por Zhang *et al.* (2012) que fizeram a previsão da viscosidade cinemática de misturas diesel-biodiesel utilizando espectros de NIR e mid-IR e obtiveram modelos com RMSEP de  $0,114$  e  $0,070 \text{ mm}^2\cdot\text{s}^{-1}$ , respectivamente.

Figura 91 - Gráfico dos valores previstos ( $\text{mm}^2\text{s}^{-1}$ ) versus medidos ( $\text{mm}^2\text{s}^{-1}$ ) para o modelo MIR\_RF\_VISCO\_snv



Fonte: O autor, 2019.



## CONCLUSÃO

Na literatura há muitos trabalhos que abordam a previsão de propriedades lineares de combustíveis ou biocombustíveis, e poucos trabalhos na previsão de propriedades físicas com comportamento não linear. Este trabalho foi desenvolvido de modo a contribuir e a enriquecer a literatura existente com novas informações e ferramentas utilizadas.

Neste estudo foram apresentados os resultados da aplicação de estatística multivariada para a análise exploratória e previsão do índice de refração, massa específica, ponto de entupimento de filtro a frio e viscosidade cinemática a 40°C, a partir de espectros de infravermelho, próximo e médio, e Raman de misturas de biodiesel, diesel-biodiesel e diesel-biodiesel-etanol.

A análise de componentes principais (PCA) foi aplicada como ferramenta de análise exploratória dos dados, a qual permitiu visualizar a distribuição das amostras, o comportamento e principalmente, fazer a identificação de amostras com comportamentos atípicos para os três conjuntos espectrais (NIR, mid-IR e Raman). O modelo apresentado pela PCA robusta (ROBPCA) mostrou-se consistente e permitiu a confirmação dos resultados obtidos pelo PCA.

Na etapa de análise exploratória, os espectros Raman (Banco 1) e os espectros NIR (Banco 2) apresentaram amostras atípicas que foram mantidas no conjunto de dados. A permanência dessas amostras só foi possível após uma etapa de transformação de sinal espacial dos conjuntos espectrais. Dessa forma, a transformação de sinal espacial se mostrou uma alternativa mais interessante que a remoção dessas amostras.

Como principal contribuição deste trabalho, a figura de mérito soma das probabilidades do teste de Wilcoxon (SWTP), baseada na comparação dos resíduos dos modelos, foi criada e aplicada neste estudo, e se mostrou robusta na comparação entre os modelos obtidos.

A elipse da região conjunta de confiança (*elliptical joint confidence region*, EJCR) foi uma ferramenta fundamental para verificar a precisão dos modelos e na detecção de ocorrência de erros sistemáticos presentes nos modelos de previsão.

Para a previsão da massa específica a regressão por florestas aleatórias utilizando espectros mid-IR pré-processados pela primeira derivada (drv) mostrou melhores resultados, apresentando RMSEP igual a  $1,16 \cdot 10^{-4} \text{ g} \cdot \text{cm}^{-3}$  e  $R^2_{\text{pred}} 0,98$ .

Para o índice de refração, o modelo PLS construído a partir de intervalos espectrais mid-IR selecionados pelo algoritmo de seleção de variáveis forneceu o modelo com RMSEP igual a  $1,0 \cdot 10^{-4}$  e  $R^2_{\text{pred}} 0,95$ .

Os melhores modelos de previsão do ponto de entupimento e viscosidade cinemática, por serem propriedades que não seguem um comportamento linear, foram obtidos com ferramentas quimiométricas não-lineares, como o SVM e florestas aleatórias.

Dentre os modelos de previsão do PEFF, o modelo SVM, utilizando os espectros de mid-IR e os intervalos obtidos pelo algoritmo RC e pré-processamento centralização da média (MIR\_RC-SVM\_PEFF\_mc), utilizando 136 vetores suportes e *kernel* radial, foi considerado o melhor modelo por apresentar menor valor de  $\text{SWTP}_{\text{val}}$  (0,58), com RMSEP igual a  $0,7^\circ\text{C}$  e  $R^2_{\text{pred}} 0,94$ .

Para a previsão da viscosidade cinemática, o modelo florestas aleatórias para o conjunto espectral de mid-IR com pré-processamento SNV (MIR\_RF\_VISCO\_snv) foi considerado o mais adequado e apresentou RMSEP igual a  $0,0184 \text{ mm}^2 \cdot \text{s}^{-1}$  e  $R^2_{\text{pred}} 0,98$ . Além disso, vale pontuar que para a previsão da viscosidade, os cinco modelos com menor valor de SWTP foram obtidos com SVM utilizando os intervalos espectrais de mid-IR. No entanto, esses modelos foram construídos com 266 vetores de suporte a partir de um conjunto de dados contendo 266 amostras (calibração), indicando um comportamento de memorização e não um comportamento de aprendizado. Logo, não é recomendável utilizar modelos SVM quando o número de vetores suporte for próximo ou igual a quantidade de amostras do conjunto de calibração.

De um modo geral, os resultados indicaram que a espectroscopia mid-IR foi considerada a mais adequada na previsão das propriedades estudadas. A utilização da espectroscopia NIR foi uma ferramenta válida e apresentou resultados medianos na previsão do índice de refração e massa específica. De fato, esse comportamento já era esperado, visto que as bandas de absorção mid-IR tendem a ser relativamente estreitas e mais bem definidas que as de NIR, que tendem a ser muito mais amplas e sobrepostas.

Havia uma expectativa sobre a possibilidade de se obter modelos de previsão com baixos valores de resíduos a partir de espectros de Raman. No entanto, essa expectativa não foi atendida, uma vez que os modelos obtidos a partir dos espectros de Raman estão entre os que apresentam maior valor de RMSEP e maior valor da soma da probabilidade do teste de Wilcoxon (SWTP). De fato, não era esperado que os modelos obtidos com os espectros de Raman fossem superiores aos modelos que utilizaram espectros mid-IR.

Analisando todos os resultados obtidos, verificou-se que a metodologia desenvolvida se mostra promissora para determinar as propriedades físico-químicas estudadas de forma rápida e não destrutível, sendo possível prever até mesmo propriedades não-lineares, como ponto de entupimento de filtro a frio e viscosidade cinemática, para controle da qualidade de biodiesel, diesel e suas misturas.

## REFERÊNCIAS

ABDEL-RAHMAN, E. M.; MUTANGA, O.; ODINDI, J.; ADAM, E.; ODINDO, A.; ISMAIL, R. Estimating Swiss chard foliar macro- and micronutrient concentrations under different irrigation water sources using ground-based hyperspectral data and four partial least squares (PLS)-based (PLS1, PLS2, SPLS1 and SPLS2) regression algorithms. *Computers and Electronics in Agriculture*, v. 132, p. 21–33, 2017. doi:10.1016/j.compag.2016.11.008

AGÊNCIA NACIONAL DO PETRÓLEO GÁS NATURAL E BIOCOMBUSTÍVEIS (ANP). Resolução ANP nº. 14 de maio de 2012. Disponível em: <<http://www.anp.gov.br/>>. Acesso em: 1 nov. 2019.

AGUERSSIF, N.; BENAMOR, M.; KACHBI, M.; DRAA, M. T. Simultaneous determination of Fe(III) and Al(III) by first-derivative spectrophotometry and partial least-squares (PLS-2) method – Application to post-haemodialysis fluids. *Journal of Trace Elements in Medicine and Biology*, v. 22, n. 3, p. 175–182, 2008. doi:10.1016/j.jtemb.2007.12.004

ALEME, H. G. *Utilização de curvas de destilação e métodos quimiométricos na previsão da origem e na determinação do teor alcoólico e massa específica da gasolina automotiva*. 2008. 81 f. Dissertação (Mestrado em Química) - Universidade Federal de Minas Gerais, Belo Horizonte, MG, 2008.

AMJAD, A.; ULLAH, R.; KHAN, S.; BILAL, M.; KHAN, A. Raman Spectroscopy based Analysis of Milk using random forests Classification. *Vibrational Spectroscopy*, v. 99, p. 124-129, 2018. doi:10.1016/j.vibspec.2018.09.003.

ARNOLD, S. A.; HARVEY, L. M.; MCNEIL, B.; HALL, J. W. Employing Near-Infrared Spectroscopic Methods of Analysis for Fermentation Monitoring and Control - Part 1, Method Development. *BioPharm International*, v. 16, p. 26-34, 2002.

ASTM International, Annual Book of ASTM standards. *Standards practices for infrared multivariate quantitative analysis*, E1655 ASTM International, West Conshohocken, 2005.

BALLABIO, D.; CONSONNI, V. Classification tools in chemistry. Part 1: linear models. *PLS\_DA. Anal. Methods*, v. 5, n. 2, p. 3790-3798, jun. 2013.

BAPTISTA P.; FELIZARDO, P.; MENEZES, J. C.; CORREIA, M. J. N. Multivariate near infrared spectroscopy models for predicting the iodine value, CFPP, kinematic viscosity at 40 °C and density at 15 °C of biodiesel. *Talanta*, v. 77, p. 144–151, jun. 2008.

BARBOSA, L. C. A. *Espectroscopia no infravermelho: na caracterização de compostos orgânicos*. Viçosa: UFV, 2013.

BISHOP, M. *Pattern Recognition and Machine Learning*. New York: Springer, 2006.

BORGES, U.; FREITAG, H.; HURTIENNE, T.; NITSCH, M. *Proálcool. Analyse und Evaluierung des brasilianischen Biotreibstoffprogramms*. Saarbrücken, Fort Lauderdale, 1984.

BREIMAN L. Random forests. *Machine Learning*, v. 45, p. 5-32, 2001. doi:10.1023/a:1010933404324.

BRETERON, R. G. *Chemometrics: data analysis for the laboratory and chemical plant*. England: John Wiley & Sons, 2003.

CANDEIA, R. A. *Biodiesel de Soja: síntese, degradação e misturas binárias*. 2008. 132 f. Tese (Doutorado em Química) – Universidade Federal da Paraíba, João Pessoa, PB, 2008.

CAO, D. S.; DENG, Z. K.; ZHU, M. F.; YAO, Z. J.; DONG, J.; ZHAO, R. G. Ensemble partial least squares regression for descriptor selection, outlier detection, applicability domain assessment, and ensemble modeling in QSAR/QSPR modeling. *J. Chemom.*, v. 31, p. 1-17, 2017. doi:10.1002/cem.2922.

CARTWRIGHT, H. M. *Applications of Artificial Intelligence in Chemistry*. New York: Oxford Science Publications, 1995.

CENTNER, V.; MASSART, D.; NOORD, O.; JONG, S.; VANDEGINSTE, B.; STERNA, C. Elimination of uninformative variables for multivariate calibration. *Analytical Chemistry*, v. 68, p. 3851–3858, 1996.

CHARNET, R.; FREIRE, C. A. L.; CHARNET, E. M. R.; BONVINO, H. *Análise de modelos de regressão linear com aplicações*. Campinas: Unicamp, 2008.

CHONG, G.; JUN, C. H. Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems*, v. 78, p. 103–112, 2005.

COSTA, P. A. *Aplicações da Espectroscopia de Radiação Infravermelha Próxima FT-NIR na Monitorização de Processos Farmacêuticos*. Lisboa, Portugal: Instituto Superior Técnico - Universidade Técnica de Lisboa, 2000.

COSTA FILHO, P. A.; POPPI, R. J. Aplicação de algoritmos genéticos na seleção de variáveis em espectroscopia no infravermelho médio. Determinação simultânea de glicose, maltose e frutose. *Química Nova*, v. 25, p. 46, 2002.

CUNHA, C. L. *Estudo da previsão de propriedades do biodiesel utilizando espectros de infravermelho e calibração multivariada*. 2014. 179 f. Dissertação (Mestrado em Engenharia Química) – Universidade do Estado do Rio de Janeiro, Rio de Janeiro, RJ, 2014.

CUNHA, C. L. ; LUNA, A. S.; OLIVEIRA, R. C. G.; XAVIER, G. M.; PAREDES, M. L. L.; TORRES, A. R. Predicting the properties of biodiesel and its blends using mid-FT-IR spectroscopy and first-order multivariate calibration. *Fuel*, v. 204, p. 185-194, 2017.

CUNHA, C. L.; TORRES, A. R.; LUNA, A. S. Multivariate regression models obtained from near-infrared spectroscopy data for prediction of the physical properties of biodiesel and its blend. *Fuel*, v. 261, 116344, 2020.

DEARING, T. I.; THOMPSON, W. J.; RECHSTEINER, C. E.; MARQUARDT, B. J. Characterization of Crude Oil Products Using Data Fusion of Process Raman, Infrared, and Nuclear Magnetic Resonance (NMR) Spectra. *Applied Spectroscopy*, v. 65, n. 2, p. 181-186, 2011.

DE SOUZA, A. M.; POPPI, R. J. Experimento didático de quimiometria para análise exploratória de óleos vegetais comestíveis por espectroscopia no infravermelho médio e análise de componentes principais: um tutorial, Parte I. *Química Nova*, v. 35, n. 1, p. 223-229, jul. 2012.

DE SOUZA, L. M.; MITSUTAKE, H.; GONTIJO, L. C.; BORGES NETO, W. Quantification of residual automotive lubricant oil as an adulterant in Brazilian S-10 diesel using MIR spectroscopy and PLS. *Fuel*, v. 130, p. 257-262, 2014. doi:10.1016/j.fuel.2014.03.051.

DECONINCK, E.; CAUWENBERGH, T.; BOTHY, J. L.; CUSTERS, D.; COURSELLE, P.; BEER, J. O. Analysis Detection of sibutramine in adulterated dietary supplements using attenuated total reflectance-infrared spectroscopy. *Journal of Pharmaceutical and Biomedical*, v. 100, p. 279-83, 2014. doi:10.1016/j.jpba.2014.08.009.

DRAGO, R. S. *Physical Methods in Chemistry*. 2nd edition. Saunders College Publishing, 1992.

EFRON, B.; TIBSHIRANI, R. Improvements on cross-validation: the 0.632 + bootstrap method, *J. Amer. Statist. Assoc.*, v. 92, n. 438, p. 548–560, jun. 1997.

\_\_\_\_\_. *An Introduction to the Bootstrap*. Chapman & Hall/CRC Press, 1993.

ERIKSSON, L.; JOHANSSON, E.; KETTANEH-WOLD, N.; WOLD, S. *Multi-and megavariable data analysis*. Umetrics: Umeå, 2001.

ESBENSEN, K. H.; GELADI, P. Principal Component Analysis: Concept, Geometrical Interpretation, Mathematical Background, Algorithms, History, Practice. In: BROWN, S. D.; TAULER, R.; WALCZAK, B. *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*, v. 2. Amsterdam: Elsevier Science B.V., 2009.

ESBENSEN, K. H. *Multivariate Data Analysis – in practice: an introduction to multivariate analysis and experimental design*. 5 ed. Aalborg University, Esbjerg: CAMO, 2002.

FANGRUI, M.; HANNA, M. A. Biodiesel production: a review. *Bioresource Technology*, v. 70, n. 1, p. 1-15, out. 1999.

FARRÉS, M.; PLATIKANOV, S.; TSAKOVSKI, S.; TAULER, R. Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for

variable selection and interpretation. *Journal of Chemometrics*, v. 29, p. 528-536, 2015.

FERRÃO, M. F.; VIERA, M. S.; PAZOS, R. E. P.; FACHINI, D.; GERBASE, A. E.; MARDER, L. Simultaneous determination of quality parameters of biodiesel/diesel blends using HATR-FTIR spectra and PLS, iPLS or siPLS regressions. *Fuel*, v. 90, n. 2, p. 701–706, 2011.

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

FERRER-RIQUELME, A. J. Statistical control of measures and process. Separata de: BROWN S. D.; TAULER, R.; WALCZAK, B. *Comprehensive chemometrics: chemical and biochemical data analysis*. Oxford: Elsevier Science, 2009, p. 97–126.

FILZMOSER, P.; GSCHWANDTNER, M. mvoutlier: Multivariate Outlier Detection Based on Robust Methods. R version 2.0.9, 2018.

FORINA, M.; CASOLINO, C.; MILLAN, C. P. Iterative predictor weighting (IPW) PLS: a technique for the elimination of useless predictors in regression problems. *Journal of Chemometrics*, v. 13, p. 165–184, 1999.

FRANK, I. Intermediate least squares regression method. *Chemometrics and Intelligent Laboratory Systems*, v. 1, p. 233–242, 1987.

FREEMAN, S. K. *Applications of Laser Raman Spectroscopy*. [S.l.: s.n.], 1974.

FRENICH, A.; JOUAN-RIMBAUD, D.; MASSART, D.; KUTTATHARMMAKUL, S.; GALERA, M.; VIDAL, J. Wavelength selection method for multicomponent spectrophotometric determinations using partial least squares. *Analyst*, v. 120, p. 2787–2792, 1995.

GAYDOU, V.; KISTER, J.; DUPUY, N. Evaluation of multiblock NIR/MIR PLS predictive models to detect adulteration of diesel/biodiesel blends by vegetal oil. *Chemometrics and Intelligent Laboratory Systems*, v.106, v.2, n. 15, p. 190 – 197, abril 2010.

GEHAN, E. A. A generalized two-sample Wilcoxon test for doubly censored data. *Biometrika*, v.52, p. 650–653, dez. 1965. doi:10.2307/2333721

GODINHO, M. S.; BLANCO, M. R.; GAMBARRA NETO, F. F.; LIÃO, L. M.; SENA, M. M.; TAULER, R.; OLIVEIRA, A. E. Evaluation of transformer insulating oil quality using NIR, fluorescence, and NMR spectroscopic data fusion. *Talanta*, v. 129, p. 143–149, nov. 2014.

HANSON, B. A. *ChemoSpec: Exploratory Chemometrics for Spectroscopy*. R package version 5.1.48, 2019.

HARRIS, D. C. *Análise química quantitativa*. 8. ed. Rio de Janeiro: LTC, 2012.

HASEGAWA, K.; MIYASHITA, Y.; FUNATSU, K. GA strategy for variable selection in QSAR studies: GA-based PLS analysis of calcium channel antagonists. *Journal of Chemical Information and Computer Sciences*, v. 37, p. 306-310, 1997.

HAYKIN, S. *Neural Networks: A Comprehensive Foundation*. 2. ed., New Jersey: Prentice-Hall, 1999.

HAWARE, R. V.; WRIGHT, P. R.; MORRIS, K. R.; HAMAD, M. L. Data fusion of Fourier transform infrared spectra and powder X-ray diffraction patterns for pharmaceutical mixtures. *Journal of Pharmaceutical and Biomedical Analysis*, v. 56, n. 5, p. 944– 949, dez. 2011.

HU, Y.; PENG, S.; PENG, J.; WEI, J. An improved ensemble partial least squares for analysis of near-infrared spectra. *Talanta*, v. 94, p. 301-307, 2012. doi:10.1016/j.talanta.2012.03.047.

HUANG, J. H.; YAN, J.; WU, Q. H.; FERRO, M. D.; YI, L. Z.; LU, H. M.; XU, Q. S.; LIANG, Y. Z. Selective of informative metabolites using random forests based on model population analysis. *Talanta*, v. 117, p. 549-555, 2013. doi:10.1016/j.talanta.2013.07.070.

KALOGERAS, K.; BEZERGIANNI, S.; KAZANTZI, V.; PILAVACHI, P. A. *On the prediction of properties for diesel / biodiesel mixtures featuring new environmental considerations*. 20th European Symposium on Computer Aided Process Engineering – ESCAPE20 S., p. 973-978, 2010.

KENNARD, R. W.; STONE, L. A.; Computer aided design of experiments. *Technometrics*, v. 11, p. 137-148, fev. 1969.

KOWALSKI, B. R. Chemometrics. *Chem. Ind.*, v. 22, p. 882, 1978.

KRZANOWSKI, W. J. Cross-validation in principal component analysis. *Biometrics*, v.43, p. 575–84, 1987.

KUCHERYAVSKIY, S. *mdatools: Multivariate Data Analysis for Chemometrics*. R package version 0.7.0, 2015.

LACERDA, E. G. M.; CARVALHO, A. C. P. L. Introdução aos algoritmos genéticos. Separata de: GALVÃO, C. O.; VALENÇA, M. J. S. (Ed.). *Sistemas inteligentes: aplicações a recursos hídricos e ciências ambientais*. Porto Alegre: Universidade/UFRGS: Associação Brasileira de Recursos Hídricos, (Coleção ABRH de Recursos Hídricos; 7), 1999. p. 99-150.

LI, H.; ZENG, M.; TAN, B.; LIANG, Y.; XU, Q.; CAO, D. Recipe for revealing informative metabolites based on model population analysis. *Metabolomics*, v. 6, p. 353–361, 2010.

LI, B.; WEI, Y.; DUAN, H.; XI, L.; WU, X. Discrimination of the geographical origin of *Codonopsis pilosula* using near infrared diffuse reflection spectroscopy coupled with



random forests and k-nearest neighbor methods. *Vib. Spectrosc.*, v. 62, p. 17-22, 2012. doi:10.1016/j.vibspec.2012.05.001.

LIAW, A.; WIENER, M. *Classification and Regression by randomForest*. R News v. 2, p. 18–22, 2002.

LILAND, K. H.; MEVIK, B. H. *baseline*: Baseline Correction of Spectra. R package version 1.2-1, 2015.

LIRA, L. F. B.; VASCONCELOS, F. V. C.; PEREIRA, C. F.; PAIM, A. P. S.; STRAGEVITCH, L.; PIMENTEL, M. F. Prediction of properties of diesel/ biodiesel blends by infrared spectroscopy and multivariate calibration. *Fuel*, v. 89, n. 2, p. 405-409, fev. 2010 (a).

LIU, W.; LIU, C.; YU, J.; ZHANG, Y.; LI, J.; CHEN, Y.; ZHENG, L. Discrimination of geographical origin of extra virgin olive oils using terahertz spectroscopy combined with chemometrics. *Food Chemistry*, v. 251, p. 86-92, 2018. doi:10.1016/j.foodchem.2018.01.081.

LÔBO, I. P.; FERREIRA, S. L. C.; CRUZ, R. S. Biodiesel: Parâmetros de qualidade e métodos analíticos. *Química Nova*, v. 32, n. 6, p. 1596, jul. 2009.

LOPES, F. N. *Espectroscopia Raman aplicada ao estudo de pigmentos em bens culturais: I – Pinturas rupestres*. Dissertação de mestrado, Instituto de Química, USP, São Paulo, 2005.

LU, W.; WANG, W. Potencial assessment of the “support vector machine” method in forecasting ambient air pollutant trends. *Chemosphere*, v. 59, n. 5, p. 693-701, dez. 2005.

LUCAS, D. C. *Algoritmos Genéticos: uma introdução* (apostila), p. 1-47 , 2002.

LUNA, A. S.; LIMA, I. C. A.; HENRIQUES, C. A.; DE ARAUJO, L. R. R.; ROCHA, W. F.; DA SILVA, J. V. Prediction of fatty methyl esters and physical properties of soybean oil/biodiesel blends from near and mid-infrared spectra using the data fusion strategy. *Analytical Methods*, v. 9, n. 33, p. 4808-4818, 2017.

MA, S.; AYBAT, N. S. Efficient Optimization Algorithms for Robust Principal Component Analysis and Its Variants. *Proceedings of the IEEE*, v. 106, n. 8, p. 1411–1426, 2018. doi:10.1109/jproc.2018.2846606

MADDEN, H. H. Comments on the Savitzky-Golay Convolution Method for Least-Squares Fit Smoothing and Differentiation of Digital Data. *Analytical Chemistry*, v. 50, n. 9, p. 1383-1386, ago. 1978.

MARINOVIC, S.; KRISTOVIC, M.; SPEHAR, B.; VINKO RUKAVINA, V.; JUKIC, A. Prediction of Diesel Fuel Properties by Vibrational Spectroscopy Using Multivariate Analysis. *Journal of Analytical Chemistry*, v. 67, n. 12, p. 939-949, 2012. doi:10.1134/S1061934812120039

MARTENS, M. Sensory and chemical quality criteria for white cabbage studied by multivariate data analysis. *Lebensmittel-Wissenschaft und Technologie*, v. 18, p. 100–104, 1985.

MEHMOOD, T.; MARTENS, H.; SÆBØ, S.; WARRINGER, J.; SNIPEN, L. A partial least squares based algorithm for parsimonious variable selection. *Algorithms for Molecular Biology*, v. 6, p. 1-12, 2011.

MEHMOOD, T.; LILAND, K. H.; SNIPEN, L.; SÆBØ, S. A review of variable selection methods in Partial Least Squares Regression. *Chemometrics and Intelligent Laboratory Systems*, v. 118, p. 62-69, 2012.

MENEZES, J. C.; FERREIRA, A. P.; RODRIGUES, L. O.; BRÁS, L. P.; ALVES, T. P. Chemometrics role within the PAT context: examples from primary pharmaceutical manufacturing. Separata de: BROWN, S. D.; TAULER, R.; WALCZAK, B. *Comprehensive chemometrics: chemical and biochemical data analysis*. Oxford: Elsevier Science, 2009. p. 313–355.

METROHM. NIR Spectroscopy. *A guide to near-infrared spectroscopic analysis of industrial manufacturing processes*. Herisau, Suíça: Metrohm, 2013.

MEVIK, B.; SEGTMAN, V. H.; NÆS, T. Ensemble methods and partial least squares regression. *Journal of Chemometrics*, v. 18, p. 498-507, 2005. doi:10.1002/cem.895.

MEVIK, B. H.; WEHRENS, R.; LILAND, K. H. *PLS: Partial Least Squares and Principal Component Regression*. R package version 2.5-0, 2015.

MEYER, D.; DIMITRIADOU, E.; HORNIK, K.; WEINGESSEL, A.; LEISCH, F. *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071)*, TU Wien. R Package Version 17-0, 2018.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos de aprendizado de máquina. Separata de: Rezende, S. O. (Ed.) *Sistemas Inteligentes - Fundamentos e Aplicações*, 1. ed., Barueri SP: Editora Manole, 2003. p. 89–114.

MONTEIRO, M. R.; AMBROZIN, A. R. P.; SANTOS, M. S.; BOFFO, E. F.; PEREIRA-FILHO, E. R.; LIÃO, L. M.; FERREIRA, A. G. Evaluation of biodiesel–diesel blends quality using <sup>1</sup>H NMR and chemometrics. *Talanta*, v. 78, n. 3, p. 660-664, maio 2009.

MONTGOMERY, D. C.; RUNGER, G. C. *Estatística aplicada e probabilidade para engenheiros*. 4. ed. Rio de Janeiro: LTC Editora, 2009.

MULLER, K. R.; SMOLA, A.; RATSCH, G.; SCHOLKOPF, B.; KOHLMORGEN, J.; VAPNIK, V. Predicting time series with support vector machines. Separata de: GERSTNER, W.; GERMOND, A.; HASLER, M.; MICORD, J. D. (Eds.). *Artificial Neural Networks*, Berlin, 1997. p. 999–1004.

MUNDO DA QUIMIOMETRIA (2010). Disponível em: < <http://mundodaquimiometria.blogspot.com/> >. Acesso em: 20 fev. 2016.

MURDOCH, D; CHOW, E. D. *ellipse: Functions for drawing ellipses and ellipse-like confidence regions*. R package version 0.3-8, 2013.

NADLER, B.; COIFMAN, R. R. Partial least squares, Beer's law and the net analyte signal: statistical modeling and analysis. *Journal of Chemometrics*, v. 19, p. 45–54, 2005.

NAES, T.; ISAKSSON, T.; FEARN, T.; DAVIES, T. *A user-friendly guide to Multivariate Calibration and Classification*. Chichester-UK: NIR Publications, 2002.

NAGUIB, I. A.; DARWISH, H. W. Support vector regression and artificial neural network models for stability indicating analysis of mebeverine hydrochloride and sulphuride mixtures in pharmaceutical preparation: a comparative study. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, v. 86, p. 515-526, 2012.

NI, D; JI, X; WU, M; WANG, W; DENG, X; HU, Z; WANG, T.; SHEN, D.; CHENG, J. Z.; WANG, H. Automatic cystocele severity grading in transperineal ultrasound by random forests regression. *Pattern Recognit*, v. 63, p. 551-60, 2017. doi:10.1016/j.patcog.2016.09.033.

NØRGAARD, L.; SAUDLAND, A.; WAGNER, J.; NIELSEN, J. P.; MUNCK, L. Interval Partial Least Square Regression (iPLS): A comparative Chemometric Study with an Example from near-infrared spectroscopy. *Applied Spectroscopy*, v. 54, n. 3, p. 413-419, mar. 2000.

OLIVEIRA, F. C.; BRANDÃO, C. R. R.; RAMALHO, H. F.; COSTA, L. A. F.; SUAREZ, P. A. Z.; RUBIM, J. C. Adulteration of diesel/biodiesel blends by vegetable oil as determined by Fourier transform (FT) near infrared spectrometry and FT–Raman spectroscopy. *Analytica Chimica Acta*, v. 587, p. 194, 2007.

PALERMO, G.; PIRAINO, P.; ZUCHT, H.D. Performance of PLS regression coefficients in selecting variables for each response of a multivariate PLS for omics-type data. *Advances and Applications in Bioinformatics and Chemistry*, n. 2, p. 57-70, 2009.

PARISOTTO, G.; MARQUES, M.; HERMES, A. L.; FLORES, E. M. M.; FERRÃO, M. F. Desenvolvimento de metodologias no infravermelho aplicadas para a quantificação de cloridrato de metformina em formulações farmacêuticas. *Orbital*, v. 1, n. 1, p. 38-48, 2009.

PASQUINI, C. Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. *Journal Brazilian Chemical Society*, v. 14, p. 198, 2003.

PAVIA, D. L.; LAMPMAN, G. M.; KRIZ, G. S.; VYVYAN, J. R. *Introdução à espectroscopia*. 4. ed. São Paulo: Cengage Learning, 2010.

PICARD, R.; COOK, D. Cross-validation of regression models. *J. Amer. Statist. Assoc.*, v. 79, n. 387, p. 575–583, set. 1984.

PIMENTEL, M. F.; RIBEIRO, G. M. G. S.; CRUZ, R. S.; STRAGEVITCH, L.; PACHECO FILHO, J. G. A.; TEIXEIRA, L. S. G. Determination of biodiesel content

when blended with mineral diesel fuel using infrared spectroscopy and multivariate calibration. *Microchemical Journal*, v. 82, p. 201, 2006.

PLATT, L. Fast training of SVM using sequential optimization. Separata de: SCHÜLKOPF, B.; BURGESS, B.; SMOLA, A. (Eds.). *Advances in Kernel Methods-support Vector Learning*. Cambridge: MIT Press, 1998. p. 185-208.

POMERANTSEV, A. L. *Chemometrics in Excel*. New Jersey: Wiley, 2014.

RAMAN, C. V.; KRISHNAN, K. S. A New Type of Secondary Radiation. *Nature*, v. 121, p. 501-502, 1928.

R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, 2019.

RESOLUÇÃO ANP N° 07/08. ANP – Agência Nacional do Petróleo, Gás Natural e Biocombustíveis, 19 de Mar. 2008. Disponível em: < [http://nxt.anp.gov.br/nxt/gateway.dll/leg/resolucao\\_anp/2008/mar%C3%A7o/ranp%207%20-%202008.xml](http://nxt.anp.gov.br/nxt/gateway.dll/leg/resolucao_anp/2008/mar%C3%A7o/ranp%207%20-%202008.xml)>. Acesso em: 20 fev. 2012.

SAEYS, Y.; INZA, I.; LARRAÑAGA, P. A review of feature selection techniques in bioinformatics. *Bioinformatics*, v. 23, p. 2507-2517, 2007. doi:10.1093/bioinformatics/btm344.

SALA, O. *Fundamentos da Espectroscopia Raman e no Infravermelho*. São Paulo: Unesp, 1996.

SANCHEZ, A. S.; GARCIA - NIETO, P. J.; FERNANDEZ, P. R.; DEL COZ DÍAZ, J. J.; IGLESIAS-RODRÍGUEZ, F. J. Application of an SVM-based regression model to the air quality study at local scale in the Avilés urban area (Spain). *Mathematical and Computer Modelling*, v. 54, n. 5-6, p. 1453-1466, abr. 2011.

SAVITZKY, A.; GOLAY, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, v. 36, n. 8, p. 1627–1639, jul. 1964.

SCHÖLKOPF, B.; SMOLA, A. J. *Learning with Kernels*. Cambridge: MIT Press, MA, 2002.

SERNEELS, S.; NOLF, E.; ESPEN, P. J. V. Spatial Sign Preprocessing: A simple way to impart moderate robustness to multivariate estimators. *Journal of Chemical Information and Modeling*, v. 46, n. 3, p. 1402-1409, 2006.

SHAHBAZI, B.; CHELGANI, S. C.; MATIN, S. S. Prediction of froth flotation responses based on various conditioning parameters by random forests method. *Colloids Surfaces A: Physicochem. Eng. Aspects*, v. 529, p. 936-41, 2017. doi:10.1016/j.colsurfa.2017.07.013.

SHAY, E.G. Diesel fuel from vegetable oils: Status and Opportunities. *Biomass Bioenergy*, v. 4, n. 4, p. 227-242, nov. 1993.

SILVA, F. E. B.; SABIN, G. P.; MÜLLER, E. I.; PARISOTTO, G.; FLORES, E. M. M.; FERRÃO, M. F. Otimização de métodos de regressão multivariada para quantificação de sulfametoxazol e Trimetoprima em medicamentos. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 27., 2007, Foz do Iguaçu. **Resumos...** Foz do Iguaçu: ABEPRO, 2007.

SILVA, P. R. F.; FREITAS, T. F. S. *Biodiesel: o ônus e o bônus de produzir combustível*. Ciência Rural, Santa Maria, v. 38, n. 3, p. 843-851, maio 2008.

SILVERSTEIN, R. M.; WEBSTER, F. X.; KIEMLE, D. J. Spectrometric identification of organic compounds 7 ed. 2005 – Silverstein, Webster & Kiemle.pdf. *Microchem. J.*, p. 21-496, 2005.

SKOOG, D. A.; HOLLER, F. J.; NIEMAN, T. A. *Princípios de análise Instrumental*. 6. ed. Porto Alegre: Bookman, 2009.

STEINWART, I.; CHRISTMANN, A. *Support Vector Machines*. New York: Springer, 2008(a).

\_\_\_\_\_. How SVMs can estimate quantiles and the median. In: \_\_\_\_\_. *Advances in Neural Information Processing Systems*, Cambridge, MA: MIT Press, 2008(b). v.20, p. 305-312.

STEVENS, A.; RAMIREZ-LOPEZ, L. *An introduction to the prospectr package*. R Package Vignette, 2013.

STONE, M. Cross-validatory choice and assessment of statistical predictions. *J. R. Stat. Soc.*, v. B36, p. 111–47, 1974.

SUAREZ, P. A. Z.; SANTOS, A. L. F.; RODRIGUES, J. P.; ALVES, M. B. Bicomcombustíveis a partir de óleos e gorduras: desafios tecnológicos para viabilizá-los. *Química Nova*, v. 32, n. 3, p. 768-775, mar. 2009.

TEÓFILO, R. F. 2013. *Métodos Quimiométricos: Uma Visão Geral - Conceitos básicos de quimiometria*. Universidade Federal de Viçosa, Viçosa, v. 1, 118 p.

TODOROV V., FILZMOSER P. An object-oriented framework for robust multivariate analysis. *Journal of Statistical Software*, v. 32, p. 1–47, 2009.

TRAN, T. N.; AFANADOR, N. L.; BUYDENS, L. M. C.; BLANCHET, L. Interpretation of variable importance in Partial Least Squares with Significance Multivariate Correlation (sMC). *Chemometrics and Intelligent Laboratory Systems*, v. 138, p. 153-160, 2014.

VALDERRAMA, P.; BRAGA, J. W. B.; POPPI, R.J. Variable selection, outlier detection, and figures de mert estimation in a partial least-squares regression multivariate calibration model. A case study for the determination of quality parameters in the alcohol industry by near-infrared spectroscopy. *Journal of Agricultural and Food Chemistry*, v. 55, p. 8331-8338, 2007. doi:10.1021/jf071538s.

VAPNIK, V. N. *The nature of Statistical learning theory*. New York: Springer-Verlag, 1995.

\_\_\_\_\_. *Statistical Learning Theory*. Huddersfield: John Wiley and Sons, 1998.

\_\_\_\_\_. *The Nature of Statistical Learning Theory*. New York: Springer, 1999.

\_\_\_\_\_.; CHERVONENKIS, A. Y. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, v. 16, n. 2, p. 264-280, maio 1971.

WANG, W.; MEN, C.; LU, W. Online prediction model based on support vector machine. *Neurocomputing*, v. 71, n. 4-6, p. 550-558, jan. 2008(a).

WANG, H.; TANG, H.; WILSON, J.; SALLEY, S. O.; NG, K. Y. S. Total Acid Number Determination of Biodiesel and Biodiesel Blends. *J. Am. Oil Chem. Soc.*, v. 85, n. 11, p. 1083-1086, ago. 2008(b).

WANG, L. B.; YU, H. Y.; HE, X. H.; LIU, R. Y. Influence of fatty acid composition of woody biodiesel plants on the fuel properties. *J. Fuel Chem. Technol.*, v. 40, n.4, p. 397-404, abr. 2012.

WANG, Y.; HUANG, H. Y.; ZUO, Z. T.; WANG, Y. Z. Comprehensive quality assessment of *Dendrobium officinale* using ATR-FTIR spectroscopy combined with random forests and support vector machine regression. *Spectrochim Acta - Part A Mol. Biomol. Spectrosc.*, v. 205, p. 637-48, 2018. doi:10.1016/j.saa.2018.07.086.

WISE, B. M.; GALLAGHER, N. B.; BRO, R.; SHAVER, J. M. *PLS\_Toolbox 3.01*, Eigenvector Research, Inc.: Manson, 2003.

WISE, B. M.; GALLAGHER, N. B.; BRO, R.; SHAVER, J. M.; WINDIG, W.; KOCH, R. S. *Chemometrics Tutorial for PLS\_Toolbox and Solo*. Eigenvector, 2006.

WOLD, H. Soft modeling: The basic design and some extensions. Separata de: JORESOG, K. G.; WOLD, H. (Ed.). *Systems under indirect observation*. Amsterdam: North-Holland, 1982. v.2, p. 1- 53.

WOLD, S.; SJÖSTROM, M.; ERIKSSON, L. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, v. 58, p. 109-130, 2001.

WOLD, S.; JOHANSSON, E., COCCHI, M. *PLS: Partial Least Squares Projections to Latent Structures*, 3D QSAR in drug design, v. 1, p. 523-550, 1993.

WOLD, S. Cross-validation estimation of the number of components in factor and principal component analysis. *Technometrics*, v. 24, p. 397-405, 1978.

WORKMAN, J.; WEYER, L. *Practical Guide to Interpretive Near-Infrared Spectroscopy*. CRC/Taylor & Francis, Boca Raton, 2008.

XIABO, Z., JIEWEN, Z., POVEY, M. J. W., HOLMES, M., HANPIN, M. Variables selection methods in near-infrared spectroscopy. *Analytica Chimica Acta*, v. 667, p. 14-32, 2010.

XIAO N, CAO D-S, LI M-Z, XU Q-S. *ENPLS: Ensemble Partial Least Squares Regression*. R Package Version 60, 2018.

YEGANEH, B.; MOTLAGH, M. S. P.; RASHIDI, Y.; KAMALAN, H. Prediction of CO concentrations based on a hybrid Partial Least Square and Support Vector Machine model. *Atmospheric Environment*, v. 55, p. 357-365, ago. 2012.

ZHANG, H.; WU, P.; YIN, A.; YANG, X.; ZHANG, M.; GAO, C. Prediction of soil organic carbon in an intensively managed reclamation zone of eastern China: A comparison of multiple linear regressions and the random forests model. *Sci. Total Environ.*, v. 592, p. 704-13, 2017. doi:10.1016/j.scitotenv.2017.02.146.

ZHANG, W.; YUAN, W.; ZHANG, X.; CORONADO, M. Predicting the dynamic and kinematic viscosities of biodiesel–diesel blends using mid- and near-infrared spectroscopy. *Applied Energy*, v. 98, p. 122–127, 2012.

ZHAO, C.; GAO, F.; WANG, F. Phase-based joint modeling and spectroscopy analysis for batch processes monitoring. *Ind. Eng. Chem. Res.*, v. 49, n. 2, p. 669–681, dez. 2010.

ZHOU, G. J. Ensemble partial least squares algorithm based on variable clustering for quantitative infrared spectrometric analysis. *Chinese J. Anal. Chem.*, v. 43, p. 1086-1091, 2015. doi:10.1016/S1872-2040(15)60842-8.

**ANEXO A - Informação da composição e nomeação das amostras de biodiesel nos conjuntos espectrais de NIR, Mid-IR e Raman para o BANCO 1**

Amostra	Biodiesel				Diesel			Massa Específica	Índice Refração	PEFF °C	Viscosidade mm <sup>2</sup> .s <sup>-1</sup>		
	Canola	Girassol	Milho	Soja	Biodiesel Master Rio	S10	S500					Padrão (lote: 02/17)	Etanol
Am021	0	0	0	1	0	0	0	0	0	0,8820	1,4575	-5,0	4,1568
Am022	0	0	0	1	0	0	0	0	0	0,8813	1,4568	-5,0	4,1705
Am023	0	0	0	1	0	0	0	0	0	0,8813	1,4569	-5,0	4,1638
Am024	0	0	0	1	0	0	0	0	0	0,8814	1,4570	-5,0	4,1630
Am025	0	0	0	1	0	0	0	0	0	0,8814	1,4570	-5,0	4,5205
Am026	0	0	0	1	0	0	0	0	0	0,8813	1,4568	-5,0	4,5615
Am027	0	0	0	1	0	0	0	0	0	0,8813	1,4569	-5,0	4,1630
Am028	0	0	0	1	0	0	0	0	0	0,8816	1,4571	-5,0	4,1646
Am029	0	0	1	0	0	0	0	0	0	0,8799	1,4559	-3,0	4,3538
Am030	0	0	1	0	0	0	0	0	0	0,8797	1,4558	-2,0	4,3638
Am031	0	0	1	0	0	0	0	0	0	0,8798	1,4558	-3,0	4,3622
Am032	0	0	1	0	0	0	0	0	0	0,8799	1,4559	-2,0	4,3725
Am033	1	0	0	0	0	0	0	0	0	0,8793	1,4556	-6,0	4,5286
Am034	1	0	0	0	0	0	0	0	0	0,8792	1,4557	-7,0	4,5481
Am035	1	0	0	0	0	0	0	0	0	0,8792	1,4556	-6,0	4,5334
Am036	1	0	0	0	0	0	0	0	0	0,8792	1,4556	-7,0	4,8524
Am037	0	1	0	0	0	0	0	0	0	0,8809	1,4570	-2,0	4,2712
Am038	0	1	0	0	0	0	0	0	0	0,8815	1,4575	-1,0	4,4143
Am039	0	1	0	0	0	0	0	0	0	0,8815	1,4575	-2,0	4,2666
Am040	0	1	0	0	0	0	0	0	0	0,8816	1,4575	-1,0	4,6060
Am041	0	0	0	1	0	0	0	0	0	0,8815	1,4570	-5,0	4,1741



Amostra	Biodiesel					Diesel			Massa Específica	Índice Refração	PEFF °C	Viscosidade mm <sup>2</sup> .s <sup>-1</sup>				
	Canola	Girassol	Milho	Soja	Região Sul	Biodiesel Master Rio	S10	S500					Padrão (lote: 02/17)	Etanol		
Am042	0	0	1	0	0	0	0	0	0	0	0	0	0.8799	1.4559	-2.6	4.3853
Am043	1	0	0	0	0	0	0	0	0	0	0	0	0.8790	1.4553	-6.6	4.5440
Am044	0	1	0	0	0	0	0	0	0	0	0	0	0.8814	1.4572	-1.6	4.2846
Am045	0	0	0	1	0	0	0	0	0	0	0	0	0.8813	1.4569	-5.0	4.1657
Am046	0	0	0	1	0	0	0	0	0	0	0	0	0.8814	1.4569	-5.0	4.1895
Am047	0	0	0	1	0	0	0	0	0	0	0	0	0.8813	1.4569	-5.0	4.1777
Am049	0	0	0	1	0	0	0	0	0	0	0	0	0.8816	1.4574	-5.0	4.1696
Am057	0	0	0	1	0	0	0	0	0	0	0	0	0.8816	1.4573	-5.0	4.1720
Am058	0	0	0	1	0	0	0	0	0	0	0	0	0.8818	1.4573	-5.0	4.1840
Am059	0	0	0	1	0	0	0	0	0	0	0	0	0.8814	1.4570	-5.0	4.1787
Am060	0	0	0	1	0	0	0	0	0	0	0	0	0.8818	1.4574	-4.0	4.1797
Am061	0	0	0	0	1	0	0	0	0	0	0	0	0.8777	1.4535	6.0	4.4288
Am062	0	0	0	0.5	0.5	0	0	0	0	0	0	0	0.8797	1.4554	-4.7	4.3272
Am064	0	0	0	1	0	0	0	0	0	0	0	0	0.8810	1.4569	-5.0	4.1879
Am066	0	0	0	1	0	0	0	0	0	0	0	0	0.8813	1.4571	-5.0	4.3463
Am067	0	0	0.9	0.1	0	0	0	0	0	0	0	0	0.8801	1.4560	-3.0	4.4045
Am068	0	0	0.9	0.1	0	0	0	0	0	0	0	0	0.8802	1.4560	-3.0	4.3662
Am069	0	0	0.9	0.1	0	0	0	0	0	0	0	0	0.8803	1.4560	-3.0	4.3922
Am070	0	0	0.7	0.3	0	0	0	0	0	0	0	0	0.8806	1.4563	-4.0	4.3335
Am071	0	0	0.7	0.3	0	0	0	0	0	0	0	0	0.8805	1.4562	-3.0	4.3596
Am072	0	0	0.7	0.3	0	0	0	0	0	0	0	0	0.8805	1.4562	-3.0	4.3648
Am073	0	0	0.5	0.5	0	0	0	0	0	0	0	0	0.8809	1.4564	-4.0	4.3183
Am074	0	0	0.5	0.5	0	0	0	0	0	0	0	0	0.8808	1.4564	-3.0	4.3147
Am075	0	0	0.5	0.5	0	0	0	0	0	0	0	0	0.8809	1.4565	-3.0	4.3240

Amostra	Biodiesel					Diesel			Massa Específica	Índice Refração	PEFF °C	Viscosidade mm <sup>2</sup> .s <sup>-1</sup>		
	Canola	Girassol	Milho	Soja	Região Sul	Biodiesel Master Rio	S10	S500					Padrão (lote: 02/17)	Etanol
Am076	0	0	0.3	0.7	0	0	0	0	0	0	0.8812	1.4567	-4.0	4.2678
Am077	0	0	0.3	0.7	0	0	0	0	0	0	0.8812	1.4567	-4.0	4.2791
Am078	0	0	0.3	0.7	0	0	0	0	0	0	0.8812	1.4567	-4.0	4.3023
Am079	0	0	0.1	0.9	0	0	0	0	0	0	0.8814	1.4569	-4.0	4.2391
Am080	0	0	0.1	0.9	0	0	0	0	0	0	0.8815	1.4569	-5.0	4.4037
Am081	0	0	0.1	0.9	0	0	0	0	0	0	0.8815	1.4569	-4.0	4.2278
Am082	0	0.9	0	0.1	0	0	0	0	0	0	0.8821	1.4575	-2.0	4.2531
Am083	0	0.9	0	0.1	0	0	0	0	0	0	0.8815	1.4572	-2.0	4.2503
Am084	0	0.9	0	0.1	0	0	0	0	0	0	0.8814	1.4572	-4.0	4.2370
Am085	0	0.7	0	0.3	0	0	0	0	0	0	0.8816	1.4572	-3.0	4.2546
Am086	0	0.7	0	0.3	0	0	0	0	0	0	0.8816	1.4573	-3.0	4.2502
Am087	0	0.7	0	0.3	0	0	0	0	0	0	0.8815	1.4572	-4.0	4.2739
Am088	0	0.5	0	0.5	0	0	0	0	0	0	0.8818	1.4573	-4.0	4.2371
Am089	0	0.5	0	0.5	0	0	0	0	0	0	0.8817	1.4572	-4.0	4.2394
Am090	0	0.5	0	0.5	0	0	0	0	0	0	0.8817	1.4572	-6.0	4.2234
Am091	0	0.3	0	0.7	0	0	0	0	0	0	0.8817	1.4571	-5.0	4.2398
Am092	0	0.3	0	0.7	0	0	0	0	0	0	0.8814	1.4571	-5.0	4.2402
Am093	0	0.3	0	0.7	0	0	0	0	0	0	0.8814	1.4571	-5.0	4.2835
Am094	0	0.1	0	0.9	0	0	0	0	0	0	0.8814	1.4570	-5.0	4.1944
Am095	0	0.1	0	0.9	0	0	0	0	0	0	0.8816	1.4571	-5.0	4.2290
Am096	0	0.1	0	0.9	0	0	0	0	0	0	0.8816	1.4571	-5.0	4.2321
Am097	0.9	0	0	0.1	0	0	0	0	0	0	0.8795	1.4557	-7.0	4.4968
Am098	0.9	0	0	0.1	0	0	0	0	0	0	0.8793	1.4556	-8.0	4.4923
Am099	0.9	0	0	0.1	0	0	0	0	0	0	0.8795	1.4557	-8.0	4.4973
Am100	0.7	0	0	0.3	0	0	0	0	0	0	0.8800	1.4560	-6.0	4.4586
Am101	0.7	0	0	0.3	0	0	0	0	0	0	0.8800	1.4560	-6.0	4.4545
Am102	0.7	0	0	0.3	0	0	0	0	0	0	0.8799	1.4560	-7.0	4.4331

Amostra	Biodiesel					Diesel			Massa Específica	Índice Refração	PEFF °C	Viscosidade mm <sup>2</sup> .s <sup>-1</sup>		
	Canola	Girassol	Milho	Soja	Região Sul	Biodiesel Master Rio	S10	S500					Padrão (lote: 02/17)	Etanol
Am103	0.5	0	0	0.5	0	0	0	0	0	0	0.8804	1.4563	-5.0	4.3761
Am104	0.5	0	0	0.5	0	0	0	0	0	0	0.8803	1.4563	-5.0	4.3685
Am105	0.5	0	0	0.5	0	0	0	0	0	0	0.8805	1.4564	-6.0	4.4037
Am106	0.3	0	0	0.7	0	0	0	0	0	0	0.8808	1.4566	-6.0	4.2696
Am107	0.3	0	0	0.7	0	0	0	0	0	0	0.8808	1.4566	-5.0	4.5188
Am108	0.3	0	0	0.7	0	0	0	0	0	0	0.8810	1.4567	-6.0	4.3558
Am109	0.1	0	0	0.9	0	0	0	0	0	0	0.8813	1.4569	-5.0	4.2358
Am110	0.1	0	0	0.9	0	0	0	0	0	0	0.8814	1.4569	-5.0	4.2408
Am111	0.1	0	0	0.9	0	0	0	0	0	0	0.8815	1.4570	-5.0	4.2488
Am112	0.1	0	0.4	0.5	0	0	0	0	0	0	0.8806	1.4564	-6.0	4.3665
Am113	0.1	0	0.4	0.5	0	0	0	0	0	0	0.8808	1.4564	-5.0	4.3883
Am114	0.1	0	0.4	0.5	0	0	0	0	0	0	0.8806	1.4564	-5.0	4.3389
Am115	0.3	0	0.3	0.4	0	0	0	0	0	0	0.8806	1.4562	-7.0	4.3865
Am116	0.3	0	0.3	0.4	0	0	0	0	0	0	0.8806	1.4563	-6.0	4.3796
Am117	0.3	0	0.3	0.4	0	0	0	0	0	0	0.8805	1.4562	-6.0	4.3798
Am118	0.4	0	0.1	0.5	0	0	0	0	0	0	0.8805	1.4563	-5.0	4.3497
Am119	0.4	0	0.1	0.5	0	0	0	0	0	0	0.8805	1.4563	-4.0	4.3773
Am120	0.4	0	0.1	0.5	0	0	0	0	0	0	0.8805	1.4563	-6.0	4.3796
Am121	0	0.1	0.4	0.5	0	0	0	0	0	0	0.8811	1.4566	-5.0	4.3236
Am122	0	0.1	0.4	0.5	0	0	0	0	0	0	0.8811	1.4566	-5.0	4.3113
Am123	0	0.1	0.4	0.5	0	0	0	0	0	0	0.8811	1.4566	-4.0	4.2330
Am124	0	0.3	0.3	0.4	0	0	0	0	0	0	0.8813	1.4568	-5.0	4.2917
Am125	0	0.3	0.3	0.4	0	0	0	0	0	0	0.8811	1.4568	-5.0	4.2868
Am126	0	0.3	0.3	0.4	0	0	0	0	0	0	0.8811	1.4568	-5.0	4.3150
Am127	0	0.4	0.1	0.5	0	0	0	0	0	0	0.8814	1.4570	-5.0	4.2397
Am128	0	0.4	0.1	0.5	0	0	0	0	0	0	0.8814	1.4570	-4.0	4.2854
Am129	0	0.4	0.1	0.5	0	0	0	0	0	0	0.8816	1.4570	-4.0	4.3059

Amostra	Biodiesel					Diesel			Massa Específica	Índice Refração	PEFF °C	Viscosidade mm <sup>2</sup> .s <sup>-1</sup>		
	Canola	Girassol	Milho	Soja	Região Sul	Biodiesel Master Rio	S10	S500					Padrão (lote: 02/17)	Etanol
Am130	0.4	0.1	0	0.5	0	0	0	0	0	0	0.8808	1.4565	-5.0	4.3537
Am131	0.4	0.1	0	0.5	0	0	0	0	0	0	0.8805	1.4564	-5.0	4.3576
Am132	0.4	0.1	0	0.5	0	0	0	0	0	0	0.8807	1.4565	-5.0	4.3798
Am133	0.3	0.3	0	0.4	0	0	0	0	0	0	0.8809	1.4567	-4.0	4.3416
Am134	0.3	0.3	0	0.4	0	0	0	0	0	0	0.8811	1.4568	-5.0	4.4051
Am135	0.3	0.3	0	0.4	0	0	0	0	0	0	0.8808	1.4566	-4.0	4.3181
Am136	0.1	0.4	0	0.5	0	0	0	0	0	0	0.8813	1.4569	-4.0	4.2854
Am137	0.1	0.4	0	0.5	0	0	0	0	0	0	0.8812	1.4569	-4.0	4.2852
Am138	0.1	0.4	0	0.5	0	0	0	0	0	0	0.8813	1.4569	-4.0	4.2449
Am139	0.25	0.25	0.25	0.25	0	0	0	0	0	0	0.8806	1.4564	-6.0	4.3919
Am140	0.25	0.25	0.25	0.25	0	0	0	0	0	0	0.8807	1.4565	-6.0	4.3775
Am141	0.25	0.25	0.25	0.25	0	0	0	0	0	0	0.8806	1.4564	-6.0	4.5936
Am142	0.1	0.1	0.4	0.4	0	0	0	0	0	0	0.8809	1.4565	-5.0	4.3403
Am143	0.1	0.1	0.4	0.4	0	0	0	0	0	0	0.8808	1.4564	-6.0	4.3493
Am144	0.1	0.1	0.4	0.4	0	0	0	0	0	0	0.8809	1.4566	-5.0	4.3606
Am145	0.4	0.1	0.1	0.4	0	0	0	0	0	0	0.8805	1.4563	-6.0	4.3757
Am146	0.4	0.1	0.1	0.4	0	0	0	0	0	0	0.8805	1.4563	-6.0	4.3877
Am147	0.4	0.1	0.1	0.4	0	0	0	0	0	0	0.8805	1.4563	-6.0	4.3969
Am148	0.1	0.4	0.1	0.4	0	0	0	0	0	0	0.8812	1.4568	-4.0	4.3163
Am149	0.1	0.4	0.1	0.4	0	0	0	0	0	0	0.8812	1.4568	-5.0	4.2803
Am150	0.1	0.4	0.1	0.4	0	0	0	0	0	0	0.8812	1.4569	-5.0	4.3130
Am151	0.4	0.1	0.4	0.1	0	0	0	0	0	0	0.8801	1.4560	-7.0	4.4478
Am152	0.4	0.1	0.4	0.1	0	0	0	0	0	0	0.8801	1.4560	-7.0	4.4350
Am153	0.4	0.1	0.4	0.1	0	0	0	0	0	0	0.8801	1.4560	-7.0	4.4477

Amostra	Biodiesel					Diesel			Massa Específica	Índice Refração	PEFF °C	Viscosidade mm <sup>2</sup> .s <sup>-1</sup>		
	Canola	Girassol	Milho	Soja	Região Sul	Biodiesel Master Rio	S10	S500					Padrão (lote: 02/17)	Etanol
Am154	0.1	0.4	0.4	0.1	0	0	0	0	0	0	0.8809	1.4566	-6.0	4.4020
Am155	0.1	0.4	0.4	0.1	0	0	0	0	0	0	0.8809	1.4565	-6.0	4.3546
Am156	0.1	0.4	0.4	0.1	0	0	0	0	0	0	0.8808	1.4565	-6.0	4.3718
Am157	0.4	0.4	0.1	0.1	0	0	0	0	0	0	0.8805	1.4564	-7.0	4.3926
Am158	0.4	0.4	0.1	0.1	0	0	0	0	0	0	0.8806	1.4564	-5.0	4.4155
Am159	0.4	0.4	0.1	0.1	0	0	0	0	0	0	0.8808	1.4565	-6.0	4.4186
Am160	0.1	0	0	0	0.9	0	0	0	0	0	0.8778	1.4537	4.0	4.4843
Am161	0.1	0	0	0	0.9	0	0	0	0	0	0.8778	1.4537	4.0	4.4848
Am162	0.1	0	0	0	0.9	0	0	0	0	0	0.8778	1.4537	4.0	4.4978
Am163	0.3	0	0	0	0.7	0	0	0	0	0	0.8782	1.4541	2.0	4.5002
Am164	0.3	0	0	0	0.7	0	0	0	0	0	0.8782	1.4541	2.0	4.5129
Am165	0.3	0	0	0	0.7	0	0	0	0	0	0.8782	1.4541	2.0	4.5012
Am166	0.5	0	0	0	0.5	0	0	0	0	0	0.8783	1.4544	-1.0	4.5312
Am167	0.5	0	0	0	0.5	0	0	0	0	0	0.8783	1.4545	-1.0	4.5245
Am168	0.5	0	0	0	0.5	0	0	0	0	0	0.8783	1.4544	0.0	4.5422
Am169	0.7	0	0	0	0.3	0	0	0	0	0	0.8786	1.4549	-3.0	4.5311
Am170	0.7	0	0	0	0.3	0	0	0	0	0	0.8786	1.4548	-4.0	4.5412
Am171	0.7	0	0	0	0.3	0	0	0	0	0	0.8786	1.4548	-4.0	4.5457
Am172	0.9	0	0	0	0.1	0	0	0	0	0	0.8789	1.4553	-6.0	4.4806
Am173	0.9	0	0	0	0.1	0	0	0	0	0	0.8789	1.4553	-5.0	4.5543
Am174	0.9	0	0	0	0.1	0	0	0	0	0	0.8789	1.4552	-5.0	4.5406
Am175	0	0	0	0	1	0	0	0	0	0	0.8776	1.4534	6.0	4.4764
Am176	0	0	0	1	0	0	0	0	0	0	0.8816	1.4570	-5.0	4.2645
Am177	1	0	0	0	0	0	0	0	0	0	0.8788	1.4554	-7.0	4.5379
Am180	1	0	0	0	0	0	0	0	0	0	0.8790	1.4554	-6.6	4.5962
Am181	0	0	0	0	1	0	0	0	0	0	0.8777	1.4535	6.0	4.4654

**ANEXO B - Informação da composição e nomeação das amostras de biodiesel nos conjuntos espectrais de NIR, Mid-IR e Raman para o BANCO 2**

Amostra	Biodiesel					Diesel			PEFF		Viscosidade mm <sup>2</sup> .s <sup>-1</sup>
	Canola	Girassol	Milho	Soja	Região Sul	Biodiesel Master Rio	S10	S500 (lote: 02/17)	Etanol	°C	
Am210	0	0	0	0.50	0.50	0	0	0	0	1.0	4.3207
Am211	0	0	0	0.50	0.50	0	0	0	0	1.0	4.3260
Am212	0	0	0	0	0	0	0	0	1	-26.0	2.6629
Am213	0	0	0	0	0	0	0	0	1	-26.0	2.6704
Am214	0	0	0	0	0	0	0	0	1	-26.0	2.7028
Am215	0	0	0	0	0	0.20	0.80	0	0	-15.0	2.9151
Am216	0	0	0	0	0	0.20	0.80	0	0	-15.0	2.8949
Am217	0	0	0	0	0	0.20	0.80	0	0	-15.0	2.9240
Am218	0	0	0	0	0	0.30	0.70	0	0	-12.0	3.0269
Am219	0	0	0	0	0	0.30	0.70	0	0	-12.0	3.0606
Am220	0	0	0	0	0	0.30	0.70	0	0	-12.0	2.9609
Am221	0	0	0	0.10	0	0	0	0	0.90	-18.0	2.9037
Am222	0	0	0	0.10	0	0	0	0	0.90	-18.0	2.8098
Am223	0	0	0	0.10	0	0	0	0	0.90	-18.0	2.8178
Am224	0	0	0	0	0	0.15	0	0	0.85	-16.0	2.8622
Am225	0	0	0	0	0	0.15	0	0	0.85	-16.0	2.8876
Am226	0	0	0	0	0	0.15	0	0	0.85	-16.0	2.8487
Am227	0	0	0	0	0	0	1	0	0	-1.0	3.4946
Am228	0	0	0	0	0	0	1	0	0	-2.0	3.4859
Am229	0	0	0	0	0	0	1	0	0	-1.0	3.5106
Am230	0	0	0	0	0	0	1	0	0	-1.0	3.4846
Am231	0	0	0	0	0	0	1	0	0	-1.0	3.5291
Am232	0	0	0	0	0	0	1	0	0	-1.0	3.4853
Am233	0	0	0	0.10	0	0	0.90	0	0	-3.0	3.5711
Am234	0	0	0	0.10	0	0	0.90	0	0	-2.0	3.5206
Am235	0	0	0	0.10	0	0	0.90	0	0	-3.0	3.5416

Amostra	Biodiesel					Diesel			PEFF		Viscosidade mm <sup>2</sup> .s <sup>-1</sup>	
	Canola	Girassol	Milho	Soja	Região Sul	Biodiesel Master Rio	S10	S500 (lote: 02/17)	Padrão (lote: 02/17)	Etanol		°C
Am236	0	0	0	0.10	0	0	0.88	0	0	0.02	-3.0	3.2736
Am237	0	0	0	0.10	0	0	0.88	0	0	0.02	-4.0	3.2996
Am238	0	0	0	0.10	0	0	0.88	0	0	0.02	-4.0	3.2931
Am239	0	0	0	0.10	0	0	0.85	0	0	0.05	-4.0	3.0647
Am240	0	0	0	0.10	0	0	0.85	0	0	0.05	-4.0	3.0355
Am241	0	0	0	0.10	0	0	0.85	0	0	0.05	-4.0	3.0266
Am242	0	0	0	0.10	0	0	0.83	0	0	0.07	-4.0	2.9954
Am243	0	0	0	0.10	0	0	0.83	0	0	0.07	-3.0	2.9739
Am244	0	0	0	0.10	0	0	0.83	0	0	0.07	-4.0	2.9964
Am245	0	0	0	0.10	0	0	0.80	0	0	0.10	-4.0	2.8354
Am246	0	0	0	0.10	0	0	0.80	0	0	0.10	-5.0	2.8586
Am247	0	0	0	0.10	0	0	0.80	0	0	0.10	-5.0	2.8344
Am248	0	0	0	0.15	0	0	0.85	0	0	0	-4.0	3.5362
Am249	0	0	0	0.15	0	0	0.85	0	0	0	-3.0	3.5759
Am250	0	0	0	0.15	0	0	0.85	0	0	0	-4.0	3.5196
Am251	0	0	0	0.15	0	0	0.83	0	0	0.02	-5.0	3.4024
Am252	0	0	0	0.15	0	0	0.83	0	0	0.02	-5.0	3.3459
Am253	0	0	0	0.15	0	0	0.83	0	0	0.02	-5.0	3.5006
Am254	0	0	0	0.15	0	0	0.80	0	0	0.05	-5.0	3.1480
Am255	0	0	0	0.15	0	0	0.80	0	0	0.05	-4.0	3.1411
Am256	0	0	0	0.15	0	0	0.80	0	0	0.05	-5.0	3.2106
Am257	0	0	0	0.15	0	0	0.78	0	0	0.07	-5.0	3.4651
Am258	0	0	0	0.15	0	0	0.78	0	0	0.07	-4.0	3.0847
Am259	0	0	0	0.15	0	0	0.78	0	0	0.07	-4.0	3.0947
Am260	0	0	0	0.15	0	0	0.75	0	0	0.10	-4.0	2.9327
Am261	0	0	0	0.15	0	0	0.75	0	0	0.10	-4.0	3.1532
Am262	0	0	0	0.15	0	0	0.75	0	0	0.10	-4.0	3.0114

Amostra	Biodiesel					Diesel			PEFF		Viscosidade mm <sup>2</sup> .s <sup>-1</sup>	
	Canola	Girassol	Milho	Soja	Região Sul	Biodiesel Master Rio	S10	S500 (lote: 02/17)	Padrão	Etanol		°C
Am263	0	0	0	0.20	0	0	0.80	0	0	0	-4.0	3.6541
Am264	0	0	0	0.20	0	0	0.80	0	0	0	-4.0	3.6476
Am265	0	0	0	0.20	0	0	0.80	0	0	0	-4.0	3.6182
Am266	0	0	0	0.20	0	0	0.78	0	0	0.02	-5.0	3.6051
Am267	0	0	0	0.20	0	0	0.78	0	0	0.02	-5.0	3.4390
Am268	0	0	0	0.20	0	0	0.78	0	0	0.02	-5.0	3.4521
Am269	0	0	0	0.20	0	0	0.75	0	0	0.05	-5.0	3.4668
Am270	0	0	0	0.20	0	0	0.75	0	0	0.05	-5.0	3.4830
Am271	0	0	0	0.20	0	0	0.75	0	0	0.05	-4.0	3.5076
Am272	0	0	0	0.20	0	0	0.73	0	0	0.07	-3.0	3.5843
Am273	0	0	0	0.20	0	0	0.73	0	0	0.07	-4.0	3.4233
Am274	0	0	0	0.20	0	0	0.73	0	0	0.07	-4.0	3.4288
Am275	0	0	0	0.20	0	0	0.70	0	0	0.10	-4.0	3.1726
Am276	0	0	0	0.20	0	0	0.70	0	0	0.10	-5.0	3.1526
Am277	0	0	0	0.20	0	0	0.70	0	0	0.10	-4.0	3.0053
Am278	0	0	0	0.25	0	0	0.75	0	0	0	-2.0	3.6198
Am279	0	0	0	0.25	0	0	0.75	0	0	0	-3.0	3.6023
Am280	0	0	0	0.25	0	0	0.75	0	0	0	-3.0	3.6107
Am281	0	0	0	0.25	0	0	0.73	0	0	0.02	-5.0	3.4603
Am282	0	0	0	0.25	0	0	0.73	0	0	0.02	-5.0	3.4715
Am283	0	0	0	0.25	0	0	0.73	0	0	0.02	-5.0	3.4696
Am284	0	0	0	0.25	0	0	0.70	0	0	0.05	-4.0	3.2821
Am285	0	0	0	0.25	0	0	0.70	0	0	0.05	-4.0	3.3514
Am286	0	0	0	0.25	0	0	0.70	0	0	0.05	-4.0	3.3499
Am287	0	0	0	0.25	0	0	0.68	0	0	0.07	-4.0	3.3763
Am288	0	0	0	0.25	0	0	0.68	0	0	0.07	-4.0	3.3856
Am289	0	0	0	0.25	0	0	0.68	0	0	0.07	-4.0	3.3594



Amostra	Biodiesel					Diesel			PEFF		Viscosidade mm <sup>2</sup> .s <sup>-1</sup>	
	Canola	Girassol	Milho	Soja	Região Sul	Biodiesel Master Rio	S10	S500	Padrão (lote: 02/17)	Etanol		°C
Am290	0	0	0	0.25	0	0	0.65	0	0	0.10	-4.0	3.2855
Am291	0	0	0	0.25	0	0	0.65	0	0	0.10	-5.0	3.1934
Am292	0	0	0	0.25	0	0	0.65	0	0	0.10	-4.0	3.3344
Am293	1	0	0	0	0	0	0	0	0	0	-7.0	4.5518
Am294	1	0	0	0	0	0	0	0	0	0	-7.0	4.6050
Am295	1	0	0	0	0	0	0	0	0	0	-6.0	4.5482
Am296	0.10	0	0	0	0	0	0.90	0	0	0	-3.0	3.6074
Am297	0.10	0	0	0	0	0	0.90	0	0	0	-4.0	3.5272
Am298	0.10	0	0	0	0	0	0.90	0	0	0	-3.0	3.5713
Am299	0.15	0	0	0	0	0	0.85	0	0	0	-2.0	3.6043
Am300	0.15	0	0	0	0	0	0.85	0	0	0	-3.0	3.8746
Am301	0.15	0	0	0	0	0	0.85	0	0	0	-3.0	3.5589
Am302	0.20	0	0	0	0	0	0.80	0	0	0	-3.0	3.9466
Am303	0.20	0	0	0	0	0	0.80	0	0	0	-3.0	3.5993
Am304	0.20	0	0	0	0	0	0.80	0	0	0	-3.0	3.6813
Am305	0.25	0	0	0	0	0	0.75	0	0	0	-4.0	3.6599
Am306	0.25	0	0	0	0	0	0.75	0	0	0	-5.0	4.0177
Am307	0.25	0	0	0	0	0	0.75	0	0	0	-4.0	3.6462
Am308	0	0	1	0	0	0	0	0	0	0	-2.0	4.3339
Am309	0	0	1	0	0	0	0	0	0	0	-1.0	4.4309
Am310	0	0	1	0	0	0	0	0	0	0	-1.0	4.3813
Am311	0	0	0.10	0	0	0	0.90	0	0	0	-4.0	3.5389
Am312	0	0	0.10	0	0	0	0.90	0	0	0	-4.0	3.5849
Am313	0	0	0.10	0	0	0	0.90	0	0	0	-4.0	3.5386
Am314	0	0	0.15	0	0	0	0.85	0	0	0	-4.0	3.6009
Am315	0	0	0.15	0	0	0	0.85	0	0	0	-5.0	3.6095
Am316	0	0	0.15	0	0	0	0.85	0	0	0	-4.0	3.5802

Amostra	Biodiesel					Diesel			PEFF		Viscosidade mm <sup>2</sup> .s <sup>-1</sup>	
	Canola	Girassol	Milho	Soja	Região Sul	Biodiesel Master Rio	S10	S500	Padrão (lote: 02/17)	Etanol		°C
Am317	0	0	0.20	0	0	0	0.80	0	0	0	-5.0	3.6121
Am318	0	0	0.20	0	0	0	0.80	0	0	0	-4.0	3.6168
Am319	0	0	0.20	0	0	0	0.80	0	0	0	-4.0	3.5453
Am320	0	0	0.25	0	0	0	0.75	0	0	0	-4.0	3.7598
Am321	0	0	0.25	0	0	0	0.75	0	0	0	-5.0	3.6712
Am322	0	0	0.25	0	0	0	0.75	0	0	0	-4.0	3.7300
Am323	0	1	0	0	0	0	0	0	0	0	-6.0	4.3220
Am324	0	1	0	0	0	0	0	0	0	0	-5.0	4.3323
Am325	0	1	0	0	0	0	0	0	0	0	-6.0	4.3467
Am326	0	0.10	0	0	0	0	0.90	0	0	0	-4.0	3.2485
Am327	0	0.10	0	0	0	0	0.90	0	0	0	-4.0	3.1820
Am328	0	0.10	0	0	0	0	0.90	0	0	0	-5.0	3.1893
Am329	0	0.15	0	0	0	0	0.85	0	0	0	-5.0	3.2187
Am330	0	0.15	0	0	0	0	0.85	0	0	0	-5.0	3.2622
Am331	0	0.15	0	0	0	0	0.85	0	0	0	-5.0	3.2557
Am332	0	0.20	0	0	0	0	0.80	0	0	0	-5.0	3.2967
Am333	0	0.20	0	0	0	0	0.80	0	0	0	-5.0	3.2942
Am334	0	0.20	0	0	0	0	0.80	0	0	0	-5.0	3.3042
Am335	0	0.25	0	0	0	0	0.75	0	0	0	-4.0	3.3809
Am336	0	0.25	0	0	0	0	0.75	0	0	0	-4.0	3.3799
Am337	0	0.25	0	0	0	0	0.75	0	0	0	-4.0	3.3495
Am338	0	0	0	0	1	0	0	0	0	0	6.0	4.5084
Am339	0	0	0	0	1	0	0	0	0	0	6.0	4.4753
Am340	0	0	0	0	1	0	0	0	0	0	5.0	4.4909
Am341	0	0	0	0	0.10	0	0.90	0	0	0	-4.0	3.2121
Am342	0	0	0	0	0.10	0	0.90	0	0	0	-4.0	3.2536
Am343	0	0	0	0	0.10	0	0.90	0	0	0	-4.0	3.2111

Amostra	Biodiesel					Diesel			PEFF		Viscosidade mm <sup>2</sup> .s <sup>-1</sup>	
	Canola	Girassol	Milho	Soja	Região Sul	Biodiesel Master Rio	S10	S500	Padrão (lote: 02/17)	Etanol		°C
Am344	0	0	0	0	0.15	0	0.85	0	0	0	-3.0	3.2675
Am345	0	0	0	0	0.15	0	0.85	0	0	0	-2.0	3.3006
Am346	0	0	0	0	0.15	0	0.85	0	0	0	-3.0	3.2655
Am347	0	0	0	0	0.20	0	0.80	0	0	0	-3.0	3.3483
Am348	0	0	0	0	0.20	0	0.80	0	0	0	-2.0	3.3014
Am349	0	0	0	0	0.20	0	0.80	0	0	0	-2.0	3.3347
Am350	0	0	0	0	0.25	0	0.75	0	0	0	-2.0	3.3733
Am351	0	0	0	0	0.25	0	0.75	0	0	0	-2.0	3.3849
Am352	0	0	0	0	0.25	0	0.75	0	0	0	-3.0	3.3786
Am353	0	0	0	0	0	0	0	1	0	0	7.0	4.0597
Am354	0	0	0	0	0	0	0	1	0	0	6.0	4.0220
Am355	0	0	0	0	0	0	0	1	0	0	6.0	4.0693
Am356	0	0	0	0	0	0	0	1	0	0	6.0	4.0570
Am357	0	0	0	0	0	0	0	1	0	0	6.0	4.0230
Am358	0	0	0	0	0	0	0	1	0	0	6.0	4.0301
Am359	0.10	0	0	0	0	0	0.90	0	0	0	7.0	4.0713
Am360	0.10	0	0	0	0	0	0.90	0	0	0	6.0	4.0791
Am361	0.10	0	0	0	0	0	0.90	0	0	0	6.0	4.0730
Am362	0.15	0	0	0	0	0	0.85	0	0	0	6.0	4.1024
Am363	0.15	0	0	0	0	0	0.85	0	0	0	6.0	4.0983
Am364	0.15	0	0	0	0	0	0.85	0	0	0	6.0	4.0935
Am365	0	0	0.10	0	0.10	0	0.90	0	0	0	7.0	4.1035
Am366	0	0	0.10	0	0.10	0	0.90	0	0	0	7.0	4.0840
Am367	0	0	0.10	0	0.10	0	0.90	0	0	0	7.0	4.1042
Am368	0	0	0.15	0	0.15	0	0.85	0	0	0	3.0	4.0919
Am369	0	0	0.15	0	0.15	0	0.85	0	0	0	2.0	3.7071
Am370	0	0	0.15	0	0.15	0	0.85	0	0	0	3.0	3.9861

Amostra	Biodiesel					Diesel			PEFF		Viscosidade mm <sup>2</sup> .s <sup>-1</sup>	
	Canola	Girassol	Milho	Soja	Região Sul	Biodiesel Master Rio	S10	S500	Padrão (lote: 02/17)	Etanol		°C
Am371	0	0.10	0	0	0	0	0	0.90	0	0	6.0	4.0542
Am372	0	0.10	0	0	0	0	0	0.90	0	0	6.0	4.0645
Am373	0	0.10	0	0	0	0	0	0.90	0	0	6.0	4.0573
Am374	0	0.15	0	0	0	0	0	0.85	0	0	6.0	4.0793
Am375	0	0.15	0	0	0	0	0	0.85	0	0	6.0	4.0760
Am376	0	0.15	0	0	0	0	0	0.85	0	0	6.0	4.0753
Am377	0	0	0	0	0	0	0	0.90	0	0	6.0	4.0784
Am378	0	0	0	0	0	0	0	0.90	0	0	6.0	4.0823
Am379	0	0	0	0	0	0	0	0.90	0	0	6.0	4.0947
Am380	0	0	0	0	0	0	0	0.85	0	0	6.0	4.0830
Am381	0	0	0	0	0	0	0	0.85	0	0	6.0	4.0848
Am382	0	0	0	0	0.15	0	0	0.85	0	0	6.0	4.0880
Am383	0	0	0	0	0.10	0	0	0.90	0	0	6.0	4.0850
Am384	0	0	0	0	0.10	0	0	0.90	0	0	6.0	4.0616
Am385	0	0	0	0	0.10	0	0	0.90	0	0	6.0	4.0790
Am386	0	0	0	0	0.15	0	0	0.85	0	0	6.0	4.0614
Am387	0	0	0	0	0.15	0	0	0.85	0	0	6.0	4.0908
Am388	0	0	0	0	0.15	0	0	0.85	0	0	6.0	4.0547
Am389	0	0	1	0	0	0	0	0	0	0	-3.0	4.3835
Am390	0	0	1	0	0	0	0	0	0	0	-3.0	4.3862
Am391	0	0	1	0	0	0	0	0	0	0	-3.0	4.4245
Am392	0	0	0	0	0	0	0	1	0	0	2.0	3.6376
Am393	0	0	0	0	0	0	0	1	0	0	3.0	3.6201
Am394	0	0	0	0	0	0	0	1	0	0	3.0	3.5312

## ANEXO C - Parâmetros de mérito dos modelos PLS com seleção de variáveis e Florestas Aleatórias (RF) obtidos a partir dos espectros de NIR, mid-IR e RAM para a previsão do índice de refração

Modelos	Espectros	Intervalos	PPx	VL	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>cal</sub>
NIR_PLUS_IndRefra_mc	NIR	Todo	mc	3	2.40E-04	0.94	2.55E-04	0.93	1.22E-04	0.96	1.27E-20	1.12E-09	7.79E-08	3.56E-03	58.05
NIR_PLUS_IndRefra_drv	NIR	Todo	drv	3	2.77E-04	0.92	2.90E-04	0.91	1.85E-04	0.91	-3.16E-20	1.64E-06	8.01E-08	3.61E-03	80.48
NIR_PLUS_IndRefra_snv	NIR	Todo	snv	4	1.54E-04	0.98	1.71E-04	0.97	7.18E-05	0.99	3.17E-20	2.49E-08	1.23E-10	1.41E-04	4.95
NIR_PLUS_IndRefra_mc_1	NIR	1	mc	4	3.28E-04	0.89	5.05E-04	0.73	2.82E-04	0.80	-1.37E-20	3.44E-05	8.10E-07	1.15E-02	112.71
NIR_PLUS_IndRefra_drv_1	NIR	1	drv	3	7.41E-04	0.43	9.99E-04	0.01	7.98E-04	0.31	-1.27E-20	1.52E-04	6.67E-06	3.29E-02	146.49
NIR_PLUS_IndRefra_snv_1	NIR	1	snv	3	3.90E-04	0.84	6.39E-04	0.58	3.75E-04	0.64	2.06E-20	1.45E-04	1.27E-06	1.44E-02	122.31
NIR_PLUS_IndRefra_mc_2	NIR	2	mc	4	3.58E-04	0.87	4.76E-04	0.76	3.21E-04	0.74	-5.67E-21	2.26E-06	2.99E-07	6.98E-03	111.65
NIR_PLUS_IndRefra_drv_2	NIR	2	drv	2	7.18E-04	0.46	9.55E-04	0.05	7.45E-04	0.21	-1.54E-20	2.89E-06	5.33E-06	2.94E-02	145.50
NIR_PLUS_IndRefra_snv_2	NIR	2	snv	3	4.07E-04	0.83	5.26E-04	0.71	3.77E-04	0.64	-1.76E-20	9.69E-06	4.10E-07	8.16E-03	118.85
NIR_PLUS_IndRefra_mc_3	NIR	3	mc	2	3.32E-04	0.88	3.44E-04	0.88	2.22E-04	0.87	-2.44E-21	3.82E-07	5.84E-08	3.08E-03	92.73
NIR_PLUS_IndRefra_drv_3	NIR	3	drv	2	3.17E-04	0.90	3.39E-04	0.88	2.31E-04	0.86	-1.94E-20	5.84E-07	9.86E-08	4.00E-03	96.04
NIR_PLUS_IndRefra_snv_3	NIR	3	snv	2	3.12E-04	0.90	3.37E-04	0.88	2.45E-04	0.85	2.79E-21	5.39E-07	1.07E-07	4.18E-03	97.27
NIR_PLUS_IndRefra_mc_4	NIR	4	mc	5	3.05E-04	0.90	3.78E-04	0.85	3.13E-04	0.75	3.43E-21	1.45E-06	3.67E-07	9.94E-03	109.18
NIR_PLUS_IndRefra_drv_4	NIR	4	drv	4	2.88E-04	0.91	6.18E-04	0.60	5.68E-04	0.18	-6.70E-21	2.77E-06	7.87E-07	1.13E-02	137.20
NIR_PLUS_IndRefra_snv_4	NIR	4	snv	4	3.01E-04	0.91	3.72E-04	0.86	3.26E-04	0.73	2.06E-20	5.42E-07	6.36E-07	1.02E-02	104.70
NIR_PLUS_IndRefra_mc_5	NIR	5	mc	3	2.77E-04	0.92	2.89E-04	0.91	2.20E-04	0.88	1.68E-20	7.24E-12	3.55E-09	7.60E-04	89.55
NIR_PLUS_IndRefra_drv_5	NIR	5	drv	2	3.38E-04	0.88	3.53E-04	0.87	2.58E-04	0.83	-2.85E-21	7.45E-08	5.32E-08	2.94E-03	97.00
NIR_PLUS_IndRefra_snv_5	NIR	5	snv	4	1.96E-04	0.96	2.13E-04	0.95	1.63E-04	0.93	5.86E-21	1.06E-08	4.59E-08	2.73E-03	77.43
NIR_LW-PLUS_IndRefra_mc	NIR	LW	mc	3	2.42E-04	0.94	2.54E-04	0.93	1.22E-04	0.96	1.67E-20	1.01E-08	9.04E-09	1.21E-03	60.25
NIR_LW-PLUS_IndRefra_drv	NIR	LW	drv	4	2.16E-04	0.95	2.66E-04	0.93	1.78E-04	0.92	2.12E-20	6.96E-06	2.39E-07	6.23E-03	79.47
NIR_LW-PLUS_IndRefra_snv	NIR	LW	snv	4	1.68E-04	0.97	1.83E-04	0.97	9.30E-05	0.98	-2.34E-20	2.15E-07	6.87E-09	1.06E-03	34.63
NIR_RC-PLUS_IndRefra_mc	NIR	RC	mc	2	5.15E-04	0.72	5.33E-04	0.70	3.53E-04	0.68	3.43E-22	2.08E-07	1.61E-06	1.62E-02	127.24
NIR_RC-PLUS_IndRefra_drv	NIR	RC	drv	3	7.85E-04	0.36	8.19E-04	0.30	5.34E-04	0.27	-1.47E-21	3.22E-10	8.99E-07	1.21E-02	134.84
NIR_RC-PLUS_IndRefra_snv	NIR	RC	snv	2	2.55E-04	0.93	2.64E-04	0.93	1.80E-04	0.92	-5.71E-21	1.42E-07	1.06E-07	4.15E-03	84.48
NIR_VIP-PLUS_IndRefra_mc	NIR	VIP	mc	4	1.63E-04	0.97	1.72E-04	0.97	8.30E-05	0.98	-1.49E-20	6.96E-08	2.74E-10	2.11E-04	9.53
NIR_VIP-PLUS_IndRefra_drv	NIR	VIP	drv	2	2.76E-04	0.92	2.89E-04	0.91	1.85E-04	0.91	-1.38E-20	1.69E-07	1.93E-08	1.77E-03	77.26
NIR_VIP-PLUS_IndRefra_snv	NIR	VIP	snv	3	1.63E-04	0.97	1.70E-04	0.97	8.01E-05	0.98	-3.26E-20	2.47E-08	1.52E-09	4.96E-04	4.80

Modelos	Espectros	Intervalos	PPx	VL	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
NIR_iPLS_IndRefra_mc	NIR	IPLS	mc	4	3.42E-04	0.88	3.91E-04	0.84	2.71E-04	0.81	1.51E-20	1.18E-07	3.83E-08	2.49E-03	110.28
NIR_iPLS_IndRefra_drv	NIR	iPLS	drv	3	3.64E-04	0.86	4.43E-04	0.80	2.63E-04	0.82	8.94E-21	3.57E-07	1.18E-08	1.39E-03	116.49
NIR_iPLS_IndRefra_snv	NIR	iPLS	snv	3	2.92E-04	0.91	3.16E-04	0.90	1.50E-04	0.94	-1.67E-20	3.03E-07	1.71E-09	5.27E-04	58.12
NIR_GA-PLS_IndRefra_mc	NIR	GA	mc	3	1.77E-04	0.97	1.88E-04	0.96	8.51E-05	0.98	-2.64E-21	7.45E-09	1.44E-08	1.53E-03	13.17
NIR_GA-PLS_IndRefra_drv	NIR	GA	drv	4	2.81E-04	0.92	3.64E-04	0.86	2.27E-04	0.87	-4.58E-21	5.84E-06	6.90E-08	3.35E-03	92.53
NIR_GA-PLS_IndRefra_snv	NIR	GA	snv	3	1.62E-04	0.97	1.73E-04	0.97	6.97E-05	0.99	4.34E-20	6.27E-08	5.85E-09	9.75E-04	13.22
NIR_UVE-PLS_IndRefra_mc	NIR	UVE	mc	3	2.48E-04	0.94	2.63E-04	0.93	1.50E-04	0.94	1.62E-20	5.36E-08	4.05E-08	2.57E-03	53.49
NIR_UVE-PLS_IndRefra_drv	NIR	UVE	drv	2	3.04E-04	0.90	3.20E-04	0.89	2.30E-04	0.86	-8.31E-22	1.07E-08	7.69E-09	1.12E-03	84.61
NIR_UVE-PLS_IndRefra_snv	NIR	UVE	snv	2	2.35E-04	0.94	2.45E-04	0.94	1.09E-04	0.97	-8.33E-21	9.62E-08	9.94E-11	1.27E-04	48.75
NIR_BVE-PLS_IndRefra_mc	NIR	BVE	mc	3	2.39E-04	0.94	2.54E-04	0.93	1.08E-04	0.97	-8.13E-22	4.48E-08	1.02E-09	4.08E-04	51.01
NIR_BVE-PLS_IndRefra_drv	NIR	BVE	drv	2	2.75E-04	0.92	2.89E-04	0.91	1.85E-04	0.91	2.06E-21	1.76E-07	1.60E-08	1.61E-03	78.80
NIR_BVE-PLS_IndRefra_snv	NIR	BVE	snv	3	1.63E-04	0.97	1.71E-04	0.97	8.04E-05	0.98	8.29E-21	6.27E-10	1.30E-09	4.60E-04	5.85
NIR_SPA-PLS_IndRefra_mc	NIR	SPA	mc	3	2.25E-04	0.95	2.39E-04	0.94	1.22E-04	0.96	-3.30E-20	2.77E-08	6.25E-08	3.19E-03	54.89
NIR_SPA-PLS_IndRefra_drv	NIR	SPA	drv	3	3.33E-04	0.88	3.65E-04	0.86	2.35E-04	0.86	-1.66E-20	1.53E-06	5.18E-07	9.18E-03	105.75
NIR_SPA-PLS_IndRefra_snv	NIR	SPA	snv	3	1.92E-04	0.96	2.05E-04	0.96	9.45E-05	0.98	1.28E-20	7.08E-08	9.41E-09	1.24E-03	38.19
NIR_REP-PLS_IndRefra_mc	NIR	REP	mc	4	1.62E-04	0.97	1.72E-04	0.97	8.29E-05	0.98	-3.91E-21	8.41E-10	6.38E-10	3.22E-04	11.06
NIR_REP-PLS_IndRefra_drv	NIR	REP	drv	2	2.75E-04	0.92	2.89E-04	0.91	1.84E-04	0.91	-3.06E-20	6.33E-07	8.56E-09	1.18E-03	75.87
NIR_REP-PLS_IndRefra_snv	NIR	REP	snv	3	1.63E-04	0.97	1.71E-04	0.97	8.02E-05	0.98	1.66E-20	8.85E-10	1.40E-09	4.77E-04	4.83
NIR_IPW-PLS_IndRefra_mc	NIR	IPW	mc	4	2.99E-04	0.91	3.40E-04	0.88	2.17E-04	0.88	1.23E-20	5.50E-07	7.08E-08	3.39E-03	97.26
NIR_IPW-PLS_IndRefra_drv	NIR	IPW	drv	5	5.35E-04	0.70	6.02E-04	0.62	3.14E-04	0.75	-8.30E-21	9.49E-07	1.15E-07	4.32E-03	122.95
NIR_IPW-PLS_IndRefra_snv	NIR	IPW	snv	3	3.45E-04	0.88	3.76E-04	0.85	1.93E-04	0.90	-1.54E-20	2.79E-07	6.15E-08	3.16E-03	89.06
NIR_enPLS_IndRefra_mc	NIR	Todo	mc	10	1.74E-04	0.97	2.11E-04	0.95	1.12E-04	0.97	8.72E-08	1.10E-08	2.56E-10	2.04E-04	50.53
NIR_enPLS_IndRefra_drv	NIR	Todo	drv	10	1.68E-04	0.97	2.18E-04	0.95	1.11E-04	0.97	3.50E-11	4.05E-09	3.42E-08	2.36E-03	45.80
NIR_enPLS_IndRefra_snv	NIR	Todo	snv	10	1.69E-04	0.97	2.00E-04	0.96	9.76E-05	0.98	7.60E-10	6.73E-10	9.92E-09	1.27E-03	39.83
NIR_RF_IndRefra_mc	NIR	Todo	mc	-	1.01E-04	0.99	2.73E-04	0.92	1.50E-04	0.94	5.16E-10	1.72E-09	8.30E-08	3.67E-03	69.02
NIR_RF_IndRefra_drv	NIR	Todo	drv	-	9.65E-05	0.99	2.74E-04	0.92	1.21E-04	0.96	8.57E-09	3.01E-08	4.30E-08	2.64E-03	57.69
NIR_RF_IndRefra_snv	NIR	Todo	snv	-	9.28E-05	0.99	2.33E-04	0.94	1.26E-04	0.96	4.28E-11	4.48E-10	5.99E-08	3.12E-03	62.19

Modelos	Espectros	Intervalos	PPx	VL	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
MIR_PLS_IndRetra_mc	MIR	Todo	mc	4	1.55E-04	0.97	1.69E-04	0.97	7.90E-05	0.98	-2.46E-20	5.90E-09	1.21E-10	1.40E-04	31.28
MIR_PLS_IndRetra_drv	MIR	Todo	drv	3	1.64E-04	0.97	1.76E-04	0.97	7.91E-05	0.98	6.72E-21	1.06E-08	3.66E-09	7.71E-04	31.04
MIR_PLS_IndRetra_snv	MIR	Todo	snv	4	1.52E-04	0.98	1.61E-04	0.97	7.03E-05	0.99	-2.61E-20	7.01E-09	3.86E-09	7.92E-04	17.92
MIR_PLS_IndRetra_mc_1	MIR	1	mc	3	2.06E-04	0.96	2.17E-04	0.95	1.44E-04	0.95	-2.56E-20	3.23E-07	6.29E-09	1.01E-03	71.19
MIR_PLS_IndRetra_drv_1	MIR	1	drv	4	1.44E-04	0.98	1.57E-04	0.97	6.71E-05	0.99	1.62E-21	4.96E-08	5.51E-10	2.99E-04	9.69
MIR_PLS_IndRetra_snv_1	MIR	1	snv	3	2.17E-04	0.95	2.29E-04	0.95	1.49E-04	0.94	9.07E-21	2.03E-07	8.33E-09	1.16E-03	74.84
MIR_PLS_IndRetra_mc_2	MIR	2	mc	4	8.27E-04	0.29	8.50E-04	0.25	6.10E-04	0.05	3.11E-21	2.04E-07	8.30E-07	1.16E-02	139.77
MIR_PLS_IndRetra_drv_2	MIR	2	drv	4	8.48E-04	0.25	8.73E-04	0.21	6.44E-04	0.16	7.93E-21	9.48E-06	1.07E-06	1.32E-02	142.97
MIR_PLS_IndRetra_snv_2	MIR	2	snv	2	8.76E-04	0.20	9.21E-04	0.12	6.17E-04	0.03	7.07E-21	1.47E-06	9.32E-07	1.23E-02	141.17
MIR_PLS_IndRetra_mc_3	MIR	3	mc	4	5.35E-04	0.70	5.99E-04	0.63	4.59E-04	0.46	-1.53E-20	1.68E-06	1.01E-07	4.06E-03	131.31
MIR_PLS_IndRetra_drv_3	MIR	3	drv	5	4.85E-04	0.76	7.27E-04	0.45	4.53E-04	0.48	1.87E-20	5.13E-06	1.93E-07	5.61E-03	133.29
MIR_PLS_IndRetra_snv_3	MIR	3	snv	4	4.30E-04	0.81	5.06E-04	0.73	3.86E-04	0.62	-1.90E-20	7.97E-07	1.01E-07	4.05E-03	125.11
MIR_PLS_IndRetra_mc_4	MIR	4	mc	4	1.54E-04	0.98	1.61E-04	0.97	9.22E-05	0.98	-1.85E-20	8.89E-08	2.00E-08	2.32E-03	41.59
MIR_PLS_IndRetra_drv_4	MIR	4	drv	4	1.47E-04	0.98	1.59E-04	0.97	7.27E-05	0.99	1.10E-20	5.06E-09	1.31E-11	4.61E-05	15.04
MIR_PLS_IndRetra_snv_4	MIR	4	snv	3	1.53E-04	0.98	1.60E-04	0.97	8.47E-05	0.98	3.60E-21	5.85E-08	1.67E-08	1.65E-03	26.15
MIR_PLS_IndRetra_mc_5	MIR	5	mc	4	7.58E-04	0.40	8.43E-04	0.26	6.04E-04	0.07	-5.17E-21	2.02E-06	2.17E-06	1.88E-02	142.04
MIR_PLS_IndRetra_drv_5	MIR	5	drv	2	8.22E-04	0.30	8.77E-04	0.20	6.33E-04	0.12	8.84E-21	5.10E-06	2.79E-06	2.13E-02	144.26
MIR_PLS_IndRetra_snv_5	MIR	5	snv	4	6.83E-04	0.51	7.95E-04	0.34	5.25E-04	0.30	2.30E-21	3.73E-06	1.04E-06	1.30E-02	137.90
MIR_LW-PLS_IndRetra_mc	MIR	LW	mc	4	1.71E-04	0.97	1.83E-04	0.97	1.21E-04	0.96	3.85E-21	5.55E-09	2.81E-08	2.14E-03	58.77
MIR_LW-PLS_IndRetra_drv	MIR	LW	drv	4	1.50E-04	0.98	1.64E-04	0.97	8.39E-05	0.98	5.47E-21	1.31E-09	8.37E-09	1.17E-03	31.37
MIR_LW-PLS_IndRetra_snv	MIR	LW	snv	4	1.61E-04	0.97	1.70E-04	0.97	9.88E-05	0.98	-1.00E-20	3.30E-08	1.04E-08	1.30E-03	41.66
MIR_RC-PLS_IndRetra_mc	MIR	RC	mc	2	2.88E-04	0.91	2.97E-04	0.91	3.29E-04	0.72	1.66E-20	4.43E-08	3.35E-07	7.37E-03	127.10
MIR_RC-PLS_IndRetra_drv	MIR	RC	drv	6	2.61E-04	0.93	2.88E-04	0.91	1.96E-04	0.90	7.38E-21	3.91E-08	2.33E-07	6.15E-03	85.91
MIR_RC-PLS_IndRetra_snv	MIR	RC	snv	2	3.16E-04	0.90	3.27E-04	0.89	3.72E-04	0.65	1.89E-20	6.99E-08	6.51E-07	1.03E-02	125.96
MIR_VIP-PLS_IndRetra_mc	MIR	VIP	mc	5	1.58E-04	0.97	1.70E-04	0.97	8.08E-05	0.98	-3.72E-20	4.92E-09	4.42E-09	8.48E-04	22.13
MIR_VIP-PLS_IndRetra_drv	MIR	VIP	drv	3	1.61E-04	0.97	1.68E-04	0.97	8.29E-05	0.98	2.35E-20	1.63E-08	1.87E-08	1.74E-03	14.12
MIR_VIP-PLS_IndRetra_snv	MIR	VIP	snv	4	1.56E-04	0.97	1.65E-04	0.97	7.05E-05	0.99	1.09E-20	1.37E-08	1.23E-09	4.47E-04	10.04

Modelos	Espectros	Intervalos	PPx	VL	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
MIR_iPLS_IndRefra_mc	MIR	iPLS	mc	6	3.73E-04	0.85	4.25E-04	0.81	4.43E-04	0.50	-2.13E-21	1.44E-06	7.50E-10	3.49E-04	133.22
MIR_iPLS_IndRefra_drv	MIR	iPLS	drv	5	2.89E-04	0.91	3.46E-04	0.88	3.18E-04	0.74	-1.12E-20	4.48E-08	3.28E-08	2.31E-03	136.74
MIR_iPLS_IndRefra_snv	MIR	iPLS	snv	5	3.25E-04	0.89	3.73E-04	0.86	4.20E-04	0.55	-6.85E-21	1.61E-07	2.39E-07	6.24E-03	132.59
MIR_GA-PLS_IndRefra_mc	MIR	GA	mc	4	1.46E-04	0.98	1.55E-04	0.98	6.38E-05	0.99	-9.97E-21	1.46E-10	8.66E-09	1.19E-03	13.30
MIR_GA-PLS_IndRefra_drv	MIR	GA	drv	3	1.56E-04	0.97	1.65E-04	0.97	8.92E-05	0.98	-7.98E-21	3.45E-10	6.91E-09	1.06E-03	46.43
MIR_GA-PLS_IndRefra_snv	MIR	GA	snv	3	1.49E-04	0.98	1.58E-04	0.97	6.58E-05	0.99	-1.60E-20	1.24E-09	2.72E-10	2.10E-04	15.58
MIR_UVE-PLS_IndRefra_mc	MIR	UVE	mc	6	2.85E-04	0.92	3.51E-04	0.87	1.97E-04	0.90	2.00E-21	8.72E-08	1.64E-07	5.16E-03	22.62
<b>MIR_UVE-PLS_IndRefra_drv</b>	<b>MIR</b>	<b>UVE</b>	<b>drv</b>	<b>5</b>	<b>1.69E-04</b>	<b>0.97</b>	<b>1.83E-04</b>	<b>0.97</b>	<b>1.44E-04</b>	<b>0.95</b>	<b>-3.33E-20</b>	<b>8.90E-08</b>	<b>6.46E-09</b>	<b>1.02E-03</b>	<b>1.00</b>
MIR_UVE-PLS_IndRefra_snv	MIR	UVE	snv	6	2.62E-04	0.93	3.11E-04	0.90	3.01E-04	0.77	-1.65E-20	8.98E-08	5.66E-09	9.59E-04	16.54
MIR_BVE-PLS_IndRefra_mc	MIR	BVE	mc	4	1.57E-04	0.97	1.73E-04	0.97	8.16E-05	0.98	-1.31E-20	1.28E-08	2.47E-11	6.34E-05	19.29
MIR_BVE-PLS_IndRefra_drv	MIR	BVE	drv	4	1.54E-04	0.98	1.64E-04	0.97	8.06E-05	0.98	8.65E-21	1.74E-08	2.82E-09	6.78E-04	23.66
MIR_BVE-PLS_IndRefra_snv	MIR	BVE	snv	3	1.59E-04	0.97	1.72E-04	0.97	8.09E-05	0.98	1.45E-20	6.04E-09	1.72E-09	5.29E-04	25.97
MIR_SPA-PLS_IndRefra_mc	MIR	SPA	mc	4	1.54E-04	0.98	1.68E-04	0.97	7.80E-05	0.98	5.60E-22	6.04E-09	3.82E-10	2.49E-04	35.90
MIR_SPA-PLS_IndRefra_drv	MIR	SPA	drv	4	1.53E-04	0.98	1.65E-04	0.97	7.51E-05	0.99	2.02E-20	1.15E-08	7.78E-10	3.56E-04	5.90
MIR_SPA-PLS_IndRefra_snv	MIR	SPA	snv	4	1.51E-04	0.98	1.61E-04	0.97	7.01E-05	0.99	-1.37E-20	6.57E-09	3.92E-09	7.99E-04	21.59
MIR_REP-PLS_IndRefra_mc	MIR	REP	mc	4	1.57E-04	0.97	1.73E-04	0.97	8.07E-05	0.98	1.47E-21	1.37E-08	5.24E-11	9.23E-05	24.59
MIR_REP-PLS_IndRefra_drv	MIR	REP	drv	4	1.54E-04	0.98	1.64E-04	0.97	8.16E-05	0.98	3.52E-21	1.84E-08	3.55E-09	7.59E-04	22.80
MIR_REP-PLS_IndRefra_snv	MIR	REP	snv	3	1.59E-04	0.97	1.73E-04	0.97	8.09E-05	0.98	-9.50E-21	1.25E-08	1.90E-09	5.56E-04	23.94
MIR_IPW-PLS_IndRefra_mc	MIR	IPW	mc	4	1.55E-04	0.97	1.64E-04	0.97	8.25E-05	0.98	-1.67E-21	3.16E-08	2.07E-09	5.81E-04	43.21
MIR_IPW-PLS_IndRefra_drv	MIR	IPW	drv	3	1.55E-04	0.97	1.64E-04	0.97	8.39E-05	0.98	-3.42E-20	3.80E-08	6.19E-09	1.00E-03	42.85
MIR_IPW-PLS_IndRefra_snv	MIR	IPW	snv	4	1.49E-04	0.98	1.57E-04	0.97	6.47E-05	0.99	4.00E-21	2.13E-08	3.71E-09	7.77E-04	26.80
MIR_enPLS_IndRefra_mc	MIR	Todo	mc	8	1.52E-04	0.98	1.70E-04	0.97	8.34E-05	0.98	2.24E-08	1.26E-09	1.54E-08	1.56E-03	16.95
MIR_enPLS_IndRefra_drv	MIR	Todo	drv	10	9.48E-05	0.99	1.78E-04	0.97	1.14E-04	0.97	1.40E-09	2.10E-09	4.27E-08	2.64E-03	55.92
MIR_enPLS_IndRefra_snv	MIR	Todo	snv	7	1.50E-04	0.98	1.67E-04	0.97	8.28E-05	0.98	1.61E-08	1.44E-09	2.09E-08	1.84E-03	32.47
MIR_RF_IndRefra_mc	MIR	Todo	mc	-	9.96E-05	0.99	2.56E-04	0.93	9.68E-05	0.98	2.63E-09	3.37E-09	1.81E-09	5.42E-04	35.33
MIR_RF_IndRefra_drv	MIR	Todo	drv	-	7.05E-05	0.99	1.92E-04	0.96	8.78E-05	0.98	3.63E-11	2.25E-11	6.44E-09	1.02E-03	35.01
MIR_RF_IndRefra_snv	MIR	Todo	snv	-	8.56E-05	0.99	2.36E-04	0.94	9.02E-05	0.98	1.85E-10	1.47E-11	1.14E-08	1.36E-03	40.82



Modelos	Espectros	Intervalos	PPx	VL	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
RAM_PLS_IndRefra_tss	RAM	Todo	tss	4	2.72E-04	0.92	7.28E-04	0.45	2.17E-04	0.92	5.60E-17	6.10E-05	1.27E-08	1.43E-03	68.18
RAM_PLS_IndRefra_drv	RAM	Todo	drv	3	2.16E-04	0.95	5.40E-04	0.70	1.37E-04	0.95	5.60E-17	2.31E-05	1.25E-07	4.52E-03	98.78
RAM_PLS_IndRefra_snv	RAM	Todo	snv	4	2.74E-04	0.92	7.41E-04	0.43	2.17E-04	0.92	5.60E-17	6.16E-05	1.29E-08	1.45E-03	66.70
RAM_LW-PLS_IndRefra_tss	RAM	LW	tss	3	2.80E-04	0.92	6.41E-04	0.57	1.62E-04	0.93	5.598E-17	3.441E-05	2.686E-09	6.61E-04	68.07
RAM_LW-PLS_IndRefra_drv	RAM	LW	drv	2	3.65E-04	0.86	5.14E-04	0.73	2.35E-04	0.86	5.599E-17	2.007E-05	1.705E-08	1.66E-03	103.59
RAM_LW-PLS_IndRefra_snv	RAM	LW	snv	3	2.83E-04	0.92	6.05E-04	0.62	1.69E-04	0.93	5.601E-17	3.458E-05	3.327E-09	7.35E-04	64.75
RAM_RC-PLS_IndRefra_tss	RAM	RC	tss	2	3.91E-04	0.84	4.16E-04	0.82	3.37E-04	0.71	5.599E-17	1.595E-05	2.802E-08	2.13E-03	113.82
RAM_RC-PLS_IndRefra_drv	RAM	RC	drv	2	6.98E-04	0.50	7.59E-04	0.40	5.82E-04	0.14	5.602E-17	1.113E-08	1.646E-06	1.64E-02	141.14
RAM_RC-PLS_IndRefra_snv	RAM	RC	snv	2	4.00E-04	0.83	4.26E-04	0.81	3.60E-04	0.67	5.598E-17	3.586E-05	4.446E-08	2.69E-03	118.84
RAM_VIP-PLS_IndRefra_tss	RAM	VIP	tss	5	3.17E-04	0.90	7.82E-04	0.37	2.58E-04	0.83	5.600E-17	3.446E-05	3.423E-08	2.36E-03	105.22
RAM_VIP-PLS_IndRefra_drv	RAM	VIP	drv	2	4.42E-04	0.80	5.56E-04	0.68	3.06E-04	0.76	5.600E-17	1.475E-05	2.868E-13	6.83E-06	123.72
RAM_VIP-PLS_IndRefra_snv	RAM	VIP	snv	3	4.58E-04	0.78	6.76E-04	0.53	3.25E-04	0.73	5.601E-17	3.666E-05	7.529E-09	1.11E-03	125.68
RAM_iPLS_IndRefra_tss	RAM	iPLS	tss	3	2.83E-04	0.92	3.63E-04	0.86	2.42E-04	0.85	5.599E-17	1.450E-05	5.972E-08	3.12E-03	103.62
RAM_iPLS_IndRefra_drv	RAM	iPLS	drv	2	3.08E-04	0.90	3.73E-04	0.86	2.39E-04	0.85	5.602E-17	8.651E-06	8.383E-08	3.69E-03	86.17
RAM_iPLS_IndRefra_snv	RAM	iPLS	snv	4	2.31E-04	0.94	3.83E-04	0.85	1.75E-04	0.92	5.598E-17	1.464E-05	6.721E-08	3.30E-03	77.23
RAM_GA-PLS_IndRefra_tss	RAM	GA	tss	3	2.72E-04	0.92	4.19E-04	0.82	2.10E-04	0.89	5.601E-17	8.476E-06	2.852E-08	2.15E-03	80.06
RAM_GA-PLS_IndRefra_drv	RAM	GA	drv	2	4.24E-04	0.81	5.53E-04	0.68	3.59E-04	0.67	5.597E-17	1.245E-05	3.467E-08	2.37E-03	126.43
RAM_GA-PLS_IndRefra_snv	RAM	GA	snv	5	1.59E-04	0.97	3.61E-04	0.87	1.52E-04	0.94	5.600E-17	6.507E-06	8.494E-09	1.17E-03	72.50
RAM_UVE-PLS_IndRefra_tss	RAM	UVE	tss	2	3.58E-04	0.87	5.47E-04	0.69	2.79E-04	0.80	5.599E-17	2.676E-05	1.659E-07	5.19E-03	97.46
RAM_UVE-PLS_IndRefra_drv	RAM	UVE	drv	3	3.35E-04	0.88	5.36E-04	0.70	4.60E-04	0.46	5.601E-17	3.976E-06	1.205E-09	4.43E-04	133.40
RAM_UVE-PLS_IndRefra_snv	RAM	UVE	snv	2	3.46E-04	0.88	3.91E-04	0.84	2.43E-04	0.85	5.602E-17	1.256E-05	6.889E-08	3.35E-03	91.61

Modelos	Espectros	Intervalos	PPx	VL	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
RAM_BVE-PLUS_IndRefra_tss	RAM	BVE	tss	3	4.19E-04	0.82	6.72E-04	0.53	2.78E-04	0.80	5.600E-17	6.343E-05	3.885E-09	7.95E-04	117.42
RAM_BVE-PLUS_IndRefra_drv	RAM	BVE	drv	3	3.31E-04	0.89	5.80E-04	0.65	2.80E-04	0.80	5.598E-17	2.774E-05	7.922E-08	3.59E-03	108.70
RAM_BVE-PLUS_IndRefra_snv	RAM	BVE	snv	4	3.09E-04	0.90	7.68E-04	0.39	2.03E-04	0.90	5.600E-17	6.837E-05	2.300E-08	1.93E-03	82.70
RAM_SPA-PLUS_IndRefra_tss	RAM	SPA	tss	4	2.72E-04	0.92	7.18E-04	0.47	1.83E-04	0.91	5.60E-17	4.71E-05	1.70E-08	1.66E-03	69.78
RAM_SPA-PLUS_IndRefra_drv	RAM	SPA	drv	2	3.98E-04	0.84	5.19E-04	0.72	2.60E-04	0.83	5.60E-17	1.12E-05	2.82E-09	6.76E-04	109.53
RAM_SPA-PLUS_IndRefra_snv	RAM	SPA	snv	4	2.72E-04	0.92	7.25E-04	0.45	1.82E-04	0.92	5.60E-17	4.80E-05	1.80E-08	1.71E-03	67.66
RAM_REP-PLUS_IndRefra_tss	RAM	REP	tss	4	3.02E-04	0.91	7.49E-04	0.42	1.84E-04	0.91	5.60E-17	5.99E-05	2.69E-08	2.09E-03	77.93
RAM_REP-PLUS_IndRefra_drv	RAM	REP	drv	3	3.17E-04	0.90	5.68E-04	0.67	2.63E-04	0.82	5.60E-17	2.47E-05	1.02E-07	4.08E-03	101.12
RAM_REP-PLUS_IndRefra_snv	RAM	REP	snv	3	4.22E-04	0.82	6.92E-04	0.50	2.81E-04	0.80	5.60E-17	6.37E-05	4.17E-09	8.23E-04	115.72
RAM_IPW-PLUS_IndRefra_tss	RAM	IPW	tss	4	2.76E-04	0.92	7.32E-04	0.44	1.75E-04	0.92	5.60E-17	5.68E-05	2.57E-08	2.04E-03	60.34
RAM_IPW-PLUS_IndRefra_drv	RAM	IPW	drv	2	3.88E-04	0.84	5.08E-04	0.73	2.48E-04	0.84	5.60E-17	1.04E-05	8.35E-08	3.68E-03	100.12
RAM_IPW-PLUS_IndRefra_snv	RAM	IPW	snv	4	2.78E-04	0.92	7.44E-04	0.43	1.79E-04	0.92	5.60E-17	5.97E-05	2.59E-08	2.05E-03	62.05
RAM_enPLUS_IndRefra_tss	RAM	Todo	tss	10	3.10E-04	0.90	6.02E-04	0.62	1.83E-04	0.91	5.03E-07	2.47E-07	1.07E-07	4.17E-03	92.35
RAM_enPLUS_IndRefra_drv	RAM	Todo	drv	10	2.00E-04	0.96	4.06E-04	0.83	2.85E-04	0.79	5.37E-08	9.25E-08	2.37E-09	6.20E-04	109.25
RAM_enPLUS_IndRefra_snv	RAM	Todo	snv	8	2.52E-04	0.93	5.07E-04	0.73	1.20E-04	0.96	1.51E-08	1.73E-07	8.13E-09	1.15E-03	49.25
RAM_RF_IndRefra_tss	RAM	Todo	tss	-	9.97E-05	0.99	2.74E-04	0.92	1.27E-04	0.96	5.71E-09	4.18E-09	9.45E-10	3.92E-04	54.26
RAM_RF_IndRefra_drv	RAM	Todo	drv	-	1.46E-04	0.98	3.77E-04	0.85	2.95E-04	0.78	3.82E-09	1.03E-08	4.32E-09	8.38E-04	116.81
RAM_RF_IndRefra_snv	RAM	Todo	snv	-	1.05E-04	0.99	2.99E-04	0.91	1.21E-04	0.96	5.93E-09	3.52E-09	2.58E-09	6.48E-04	52.03

## ANEXO D - Parâmetros de mérito dos modelos PLS com seleção de variáveis e Florestas Aleatórias (RF) obtidos a partir dos espectros de NIR, mid-IR e RAM para a previsão da massa específica

Modelos	Espectros	Intervalos	PPx	VL	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>cal</sub>
NIR_PLS_MasEsp_mc	NIR	Todo	mc	3	2.0E-04	0.97	2.1E-04	0.97	1.6E-04	0.96	5.016E-17	1.84E-07	2.97E-08	3.6E-03	28.61
NIR_PLS_MasEsp_drv	NIR	Todo	drv	3	3.0E-04	0.93	3.2E-04	0.92	2.7E-04	0.88	5.018E-17	1.85E-07	3.47E-08	3.9E-03	77.72
NIR_PLS_MasEsp_snv	NIR	Todo	snv	3	1.6E-04	0.98	1.7E-04	0.98	1.5E-04	0.96	5.022E-17	8.52E-08	9.79E-09	2.1E-03	33.21
NIR_PLS_MasEsp_mc_1	NIR	1	mc	3	4.3E-04	0.86	5.5E-04	0.77	3.4E-04	0.82	5.017E-17	3.89E-05	1.07E-06	2.2E-02	113.87
NIR_PLS_MasEsp_drv_1	NIR	1	drv	3	8.6E-04	0.44	1.2E-03	0.04	9.5E-04	0.12	5.019E-17	1.16E-04	8.03E-06	6.0E-02	146.43
NIR_PLS_MasEsp_snv_1	NIR	1	snv	2	5.8E-04	0.74	6.8E-04	0.66	4.5E-04	0.68	5.020E-17	1.55E-04	2.15E-06	3.1E-02	127.09
NIR_PLS_MasEsp_mc_2	NIR	2	mc	4	3.8E-04	0.89	5.0E-04	0.81	4.1E-04	0.74	5.019E-17	1.26E-06	2.19E-07	9.9E-03	119.45
NIR_PLS_MasEsp_drv_2	NIR	2	drv	2	8.5E-04	0.46	1.1E-03	0.03	9.1E-04	0.15	5.016E-17	2.94E-06	6.18E-06	5.2E-02	144.59
NIR_PLS_MasEsp_snv_2	NIR	2	snv	3	4.6E-04	0.84	5.9E-04	0.74	4.6E-04	0.66	5.018E-17	4.74E-06	3.42E-07	1.2E-02	120.31
NIR_PLS_MasEsp_mc_3	NIR	3	mc	2	4.7E-04	0.83	4.9E-04	0.82	3.5E-04	0.80	5.020E-17	4.26E-07	3.96E-08	4.2E-03	107.51
NIR_PLS_MasEsp_drv_3	NIR	3	drv	2	4.2E-04	0.87	4.5E-04	0.85	3.5E-04	0.81	5.022E-17	1.19E-07	1.07E-07	6.9E-03	106.18
NIR_PLS_MasEsp_snv_3	NIR	3	snv	2	4.3E-04	0.86	4.5E-04	0.85	3.8E-04	0.78	5.022E-17	9.01E-08	8.55E-08	6.2E-03	113.77
NIR_PLS_MasEsp_mc_4	NIR	4	mc	5	3.6E-04	0.90	4.5E-04	0.85	3.4E-04	0.82	5.018E-17	5.20E-06	3.89E-07	1.7E-02	98.93
NIR_PLS_MasEsp_drv_4	NIR	4	drv	4	3.3E-04	0.92	7.2E-04	0.61	6.1E-04	0.41	5.023E-17	9.73E-06	8.31E-07	1.9E-02	134.05
NIR_PLS_MasEsp_snv_4	NIR	4	snv	4	3.9E-04	0.89	4.7E-04	0.83	3.6E-04	0.80	5.019E-17	2.30E-06	7.64E-07	1.8E-02	96.14
NIR_PLS_MasEsp_mc_5	NIR	5	mc	3	3.4E-04	0.91	3.7E-04	0.90	3.1E-04	0.85	5.021E-17	1.22E-06	3.64E-08	4.0E-03	102.48
NIR_PLS_MasEsp_drv_5	NIR	5	drv	2	4.2E-04	0.87	4.4E-04	0.86	3.6E-04	0.79	5.025E-17	4.64E-06	2.43E-08	3.3E-03	100.80
NIR_PLS_MasEsp_snv_5	NIR	5	snv	4	2.2E-04	0.96	2.5E-04	0.95	2.3E-04	0.92	5.024E-17	1.17E-06	3.58E-08	4.0E-03	78.80
NIR_LW-PLS_MasEsp_mc	NIR	LW	base	3	2.4E-04	0.96	2.5E-04	0.95	2.1E-04	0.93	5.022E-17	8.39E-07	9.06E-09	2.0E-03	77.72
NIR_LW-PLS_MasEsp_drv	NIR	LW	drv	3	3.7E-04	0.89	4.2E-04	0.86	3.6E-04	0.80	5.019E-17	2.61E-06	9.95E-08	6.7E-03	113.10
NIR_LW-PLS_MasEsp_snv	NIR	LW	snv	3	2.2E-04	0.96	2.4E-04	0.96	2.1E-04	0.93	5.024E-17	9.89E-07	2.35E-08	3.2E-03	70.93
NIR_RC-PLS_MasEsp_mc	NIR	RC	base	2	5.0E-04	0.81	5.2E-04	0.79	4.0E-04	0.75	5.020E-17	2.95E-07	1.72E-06	2.8E-02	125.25
NIR_RC-PLS_MasEsp_drv	NIR	RC	drv	5	9.8E-04	0.28	1.1E-03	0.15	7.6E-04	0.10	5.021E-17	1.29E-08	3.20E-06	3.8E-02	139.73
NIR_RC-PLS_MasEsp_snv	NIR	RC	snv	2	2.8E-04	0.94	2.9E-04	0.93	2.5E-04	0.90	5.020E-17	2.28E-06	3.36E-07	1.2E-02	94.82
NIR_VIP-PLS_MasEsp_mc	NIR	VIP	base	3	2.1E-04	0.97	2.2E-04	0.96	1.6E-04	0.96	5.022E-17	4.40E-07	2.41E-08	3.3E-03	39.54
NIR_VIP-PLS_MasEsp_drv	NIR	VIP	drv	2	3.0E-04	0.93	3.1E-04	0.93	2.7E-04	0.89	5.020E-17	3.34E-06	1.15E-08	2.3E-03	81.21
NIR_VIP-PLS_MasEsp_snv	NIR	VIP	snv	2	2.0E-04	0.97	2.1E-04	0.97	1.7E-04	0.96	5.019E-17	3.30E-07	3.85E-08	4.1E-03	49.80

Modelos	Espectros	Intervalos	PPx	VL	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
NIR_IPLS_MaseEsp_mc	NIR	iPLS	base	4	7.4E-04	0.59	1.0E-03	0.20	5.5E-04	0.52	5.020E-17	2.39E-06	9.70E-08	6.6E-03	128.07
NIR_IPLS_MaseEsp_drv	NIR	iPLS	drv	3	9.9E-04	0.27	1.4E-03	0.24	9.5E-04	0.23	5.021E-17	3.28E-06	7.11E-06	5.6E-02	145.41
NIR_IPLS_MaseEsp_snv	NIR	iPLS	snv	2	9.3E-04	0.35	1.1E-03	0.12	5.5E-04	0.51	5.022E-17	1.30E-04	4.85E-07	1.5E-02	131.86
NIR_GA-PLS_MaseEsp_mc	NIR	GA	base	3	1.7E-04	0.98	1.9E-04	0.97	1.3E-04	0.97	5.023E-17	5.98E-08	1.67E-08	2.7E-03	6.23
NIR_GA-PLS_MaseEsp_drv	NIR	GA	drv	3	4.1E-04	0.87	4.6E-04	0.84	3.6E-04	0.79	5.019E-17	2.18E-07	8.45E-07	1.9E-02	118.45
NIR_GA-PLS_MaseEsp_snv	NIR	GA	snv	2	2.5E-04	0.95	2.5E-04	0.95	2.0E-04	0.94	5.017E-17	9.72E-08	2.87E-07	1.1E-02	67.75
NIR_UVE-PLS_MaseEsp_mc	NIR	UVE	base	2	2.1E-04	0.97	2.2E-04	0.96	1.7E-04	0.95	5.018E-17	3.10E-08	1.20E-08	2.3E-03	30.17
NIR_UVE-PLS_MaseEsp_drv	NIR	UVE	drv	3	2.7E-04	0.94	3.5E-04	0.91	3.0E-04	0.86	5.017E-17	6.60E-07	3.01E-07	1.2E-02	94.13
NIR_UVE-PLS_MaseEsp_snv	NIR	UVE	snv	2	2.1E-04	0.97	2.2E-04	0.96	1.8E-04	0.95	5.023E-17	1.81E-07	2.16E-09	9.8E-04	63.16
NIR_BVE-PLS_MaseEsp_mc	NIR	BVE	base	3	2.1E-04	0.97	2.2E-04	0.96	1.6E-04	0.96	5.019E-17	3.48E-07	2.59E-08	3.4E-03	44.07
NIR_BVE-PLS_MaseEsp_drv	NIR	BVE	drv	2	3.0E-04	0.93	3.1E-04	0.93	2.7E-04	0.89	5.019E-17	3.09E-06	1.23E-08	2.3E-03	80.75
NIR_BVE-PLS_MaseEsp_snv	NIR	BVE	snv	2	2.0E-04	0.97	2.1E-04	0.97	1.7E-04	0.96	5.023E-17	3.84E-07	4.15E-08	4.3E-03	53.45
NIR_SPA-PLS_MaseEsp_mc	NIR	SPA	base	3	2.0E-04	0.97	2.2E-04	0.96	1.5E-04	0.96	5.018E-17	8.80E-08	2.47E-08	3.3E-03	25.82
NIR_SPA-PLS_MaseEsp_drv	NIR	SPA	drv	3	3.2E-04	0.92	3.5E-04	0.91	3.0E-04	0.86	5.019E-17	7.62E-07	1.85E-08	2.9E-03	94.57
NIR_SPA-PLS_MaseEsp_snv	NIR	SPA	snv	2	2.7E-04	0.95	2.7E-04	0.94	2.2E-04	0.92	5.024E-17	4.79E-08	4.15E-07	1.4E-02	80.34
NIR_REP-PLS_MaseEsp_mc	NIR	REP	base	3	2.1E-04	0.97	2.2E-04	0.96	1.6E-04	0.96	5.022E-17	4.42E-07	2.44E-08	3.3E-03	40.87
NIR_REP-PLS_MaseEsp_drv	NIR	REP	drv	3	1.8E-04	0.97	2.6E-04	0.95	1.9E-04	0.94	5.017E-17	3.33E-06	4.84E-08	4.6E-03	48.04
NIR_REP-PLS_MaseEsp_snv	NIR	REP	snv	2	2.0E-04	0.97	2.1E-04	0.97	1.7E-04	0.96	5.024E-17	3.35E-07	3.90E-08	4.2E-03	51.00
NIR_IPW-PLS_MaseEsp_mc	NIR	IPW	base	3	1.9E-04	0.97	2.0E-04	0.97	1.5E-04	0.96	5.021E-17	4.61E-07	1.69E-08	2.7E-03	26.09
NIR_IPW-PLS_MaseEsp_drv	NIR	IPW	drv	3	4.0E-04	0.88	5.1E-04	0.80	4.4E-04	0.69	5.018E-17	5.78E-08	3.40E-08	3.9E-03	125.35
NIR_IPW-PLS_MaseEsp_snv	NIR	IPW	snv	2	2.2E-04	0.96	2.3E-04	0.96	1.8E-04	0.95	5.021E-17	7.34E-07	1.76E-07	8.8E-03	57.30
NIR_enPLS_MaseEsp_mc	NIR	Todo	mc	10	1.4E-04	0.98	2.3E-04	0.96	1.8E-04	0.95	1.172E-08	2.71E-08	4.64E-09	1.4E-03	48.10
NIR_enPLS_MaseEsp_drv	NIR	Todo	drv	10	1.1E-04	0.99	2.0E-04	0.97	1.7E-04	0.95	5.884E-12	1.40E-08	2.16E-08	3.1E-03	25.97
NIR_enPLS_MaseEsp_snv	NIR	Todo	snv	10	1.5E-04	0.98	2.0E-04	0.97	1.6E-04	0.96	5.504E-08	1.30E-09	5.10E-09	1.5E-03	18.15
NIR_RF_MaseEsp_mc	NIR	Todo	mc	-	9.7E-05	0.99	2.7E-04	0.95	1.6E-04	0.96	8.708E-10	4.70E-11	1.84E-08	2.9E-03	19.50
NIR_RF_MaseEsp_drv	NIR	Todo	drv	-	8.1E-05	0.99	2.4E-04	0.96	1.4E-04	0.97	2.911E-10	8.29E-09	8.31E-09	1.9E-03	9.62
NIR_RF_MaseEsp_snv	NIR	Todo	snv	-	8.9E-05	0.99	2.3E-04	0.96	1.3E-04	0.97	3.945E-09	4.45E-08	2.16E-09	9.8E-04	2.57

Modelos	Espectros	Intervalos	PPx	VL	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
MIR_PLS_MaseEsp_mc	MIR	Todo	mc	3	1.6E-04	0.98	1.7E-04	0.98	1.6E-04	0.96	5.021E-17	6.54E-08	1.28E-09	7.5E-04	47.61
MIR_PLS_MaseEsp_drv	MIR	Todo	drv	3	1.6E-04	0.98	1.7E-04	0.98	1.6E-04	0.96	5.018E-17	2.14E-07	5.87E-09	1.6E-03	30.38
MIR_PLS_MaseEsp_snv	MIR	Todo	snv	3	1.6E-04	0.98	1.7E-04	0.98	1.6E-04	0.96	5.021E-17	6.36E-08	3.64E-09	1.3E-03	53.28
MIR_PLS_MaseEsp_mc_1	MIR	1	mc	3	2.5E-04	0.95	2.7E-04	0.95	2.4E-04	0.91	5.021E-17	8.86E-07	1.17E-08	2.3E-03	90.12
MIR_PLS_MaseEsp_drv_1	MIR	1	drv	4	1.6E-04	0.98	1.8E-04	0.98	1.8E-04	0.95	5.021E-17	2.81E-07	4.69E-08	4.6E-03	54.45
MIR_PLS_MaseEsp_snv_1	MIR	1	snv	3	2.6E-04	0.95	2.7E-04	0.94	2.5E-04	0.90	5.020E-17	9.05E-07	1.15E-08	2.3E-03	92.39
MIR_PLS_MaseEsp_mc_2	MIR	2	mc	4	9.7E-04	0.30	9.9E-04	0.26	7.2E-04	0.18	5.021E-17	3.44E-09	9.84E-07	2.1E-02	138.04
MIR_PLS_MaseEsp_drv_2	MIR	2	drv	4	9.9E-04	0.26	1.0E-03	0.22	7.5E-04	0.10	5.019E-17	6.99E-06	1.23E-06	2.3E-02	140.88
MIR_PLS_MaseEsp_snv_2	MIR	2	snv	2	1.0E-03	0.23	1.1E-03	0.15	7.7E-04	0.05	5.019E-17	3.22E-07	1.26E-06	2.4E-02	140.42
MIR_PLS_MaseEsp_mc_3	MIR	3	mc	4	5.9E-04	0.74	6.5E-04	0.68	5.6E-04	0.51	5.017E-17	4.19E-06	7.39E-08	5.7E-03	128.60
MIR_PLS_MaseEsp_drv_3	MIR	3	drv	4	5.9E-04	0.74	8.1E-04	0.50	5.9E-04	0.45	5.024E-17	1.20E-05	9.44E-07	2.0E-02	131.63
MIR_PLS_MaseEsp_snv_3	MIR	3	snv	4	4.9E-04	0.82	5.6E-04	0.76	5.0E-04	0.60	5.022E-17	2.13E-07	1.01E-07	6.7E-03	127.32
MIR_PLS_MaseEsp_mc_4	MIR	4	mc	3	1.9E-04	0.97	2.0E-04	0.97	2.1E-04	0.93	5.018E-17	4.58E-07	5.34E-08	6.3E-03	59.04
MIR_PLS_MaseEsp_drv_4	MIR	4	drv	4	1.6E-04	0.98	1.7E-04	0.98	1.9E-04	0.94	5.016E-17	4.53E-07	6.01E-08	5.2E-03	31.80
MIR_PLS_MaseEsp_snv_4	MIR	4	snv	2	1.6E-04	0.98	1.7E-04	0.98	1.5E-04	0.97	5.023E-17	6.64E-07	2.36E-09	1.0E-03	22.18
MIR_PLS_MaseEsp_mc_5	MIR	5	mc	4	8.6E-04	0.44	9.5E-04	0.32	7.6E-04	0.09	5.020E-17	1.57E-06	3.11E-06	3.7E-02	142.39
MIR_PLS_MaseEsp_drv_5	MIR	5	drv	2	9.6E-04	0.31	1.0E-03	0.21	7.9E-04	0.01	5.020E-17	6.20E-06	4.08E-06	4.3E-02	143.42
MIR_PLS_MaseEsp_snv_5	MIR	5	snv	4	7.9E-04	0.53	8.8E-04	0.41	6.6E-04	0.31	5.022E-17	2.64E-06	1.33E-06	2.4E-02	136.97
MIR_LW-PLS_MaseEsp_mc	MIR	LW	mc	4	1.7E-04	0.98	2.1E-04	0.97	1.9E-04	0.94	5.020E-17	2.85E-07	1.63E-10	2.7E-04	33.95
MIR_LW-PLS_MaseEsp_drv	MIR	LW	drv	3	2.1E-04	0.97	2.3E-04	0.96	1.8E-04	0.95	5.019E-17	2.48E-07	2.51E-09	1.1E-03	57.04
MIR_LW-PLS_MaseEsp_snv	MIR	LW	snv	4	1.4E-04	0.98	1.6E-04	0.98	1.6E-04	0.96	5.019E-17	9.51E-08	7.98E-09	1.9E-03	19.74
MIR_RC-PLS_MaseEsp_mc	MIR	RC	mc	3	5.0E-04	0.81	5.2E-04	0.79	3.5E-04	0.81	5.022E-17	4.73E-09	1.97E-07	9.4E-03	116.93
MIR_RC-PLS_MaseEsp_drv	MIR	RC	drv	6	8.0E-04	0.52	9.2E-04	0.36	6.4E-04	0.35	5.020E-17	8.52E-07	1.01E-06	2.1E-02	130.30
MIR_RC-PLS_MaseEsp_snv	MIR	RC	snv	2	9.2E-04	0.37	9.4E-04	0.33	6.5E-04	0.34	5.022E-17	1.17E-07	3.24E-07	1.2E-02	136.39
MIR_VIP-PLS_MaseEsp_mc	MIR	VIP	mc	3	1.7E-04	0.98	1.7E-04	0.98	1.5E-04	0.97	5.021E-17	9.76E-08	1.24E-10	2.3E-04	14.95
MIR_VIP-PLS_MaseEsp_drv	MIR	VIP	drv	2	1.6E-04	0.98	1.6E-04	0.98	1.6E-04	0.96	5.017E-17	2.00E-08	9.91E-09	2.1E-03	43.42
MIR_VIP-PLS_MaseEsp_snv	MIR	VIP	snv	2	1.7E-04	0.98	1.7E-04	0.98	1.5E-04	0.97	5.017E-17	7.76E-08	1.07E-09	6.9E-04	9.82

Modelos	Espectros	Intervalos	PPx	VL	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
MIR_IPLS_MasEsp_mc	MIR	iPLS	mc	4	2.7E-04	0.94	2.9E-04	0.94	2.6E-04	0.90	5.021E-17	3.47E-07	4.94E-08	4.7E-03	54.31
MIR_IPLS_MasEsp_drv	MIR	iPLS	drv	4	1.9E-04	0.97	2.1E-04	0.97	2.1E-04	0.93	5.019E-17	3.01E-07	4.92E-08	4.7E-03	22.42
MIR_IPLS_MasEsp_snv	MIR	iPLS	snv	3	2.4E-04	0.96	2.6E-04	0.95	2.5E-04	0.90	5.018E-17	2.57E-09	1.32E-09	7.7E-04	90.20
MIR_GA-PLS_MasEsp_mc	MIR	GA	mc	3	1.6E-04	0.98	1.6E-04	0.98	1.4E-04	0.97	5.024E-17	5.29E-08	1.34E-08	2.4E-03	32.44
MIR_GA-PLS_MasEsp_drv	MIR	GA	drv	3	1.6E-04	0.98	1.7E-04	0.98	1.6E-04	0.96	5.017E-17	6.20E-08	8.04E-09	1.9E-03	44.90
MIR_GA-PLS_MasEsp_snv	MIR	GA	snv	3	1.4E-04	0.99	1.5E-04	0.98	1.5E-04	0.97	5.019E-17	6.95E-08	2.40E-08	3.3E-03	37.69
MIR_UVE-PLS_MasEsp_mc	MIR	UVE	mc	2	1.7E-04	0.98	1.8E-04	0.98	1.5E-04	0.96	5.018E-17	1.12E-07	1.35E-09	7.7E-04	44.40
MIR_UVE-PLS_MasEsp_drv	MIR	UVE	drv	5	1.3E-04	0.99	1.5E-04	0.98	1.5E-04	0.96	5.018E-17	1.51E-07	5.83E-08	5.1E-03	14.88
MIR_UVE-PLS_MasEsp_snv	MIR	UVE	snv	2	1.7E-04	0.98	1.7E-04	0.98	1.5E-04	0.97	5.020E-17	9.42E-08	1.16E-09	7.2E-04	35.46
MIR_BVE-PLS_MasEsp_mc	MIR	BVE	mc	2	2.0E-04	0.97	2.0E-04	0.97	1.7E-04	0.95	5.024E-17	1.37E-07	2.34E-08	3.2E-03	14.87
MIR_BVE-PLS_MasEsp_drv	MIR	BVE	drv	3	1.6E-04	0.98	1.7E-04	0.98	1.6E-04	0.96	5.017E-17	1.24E-07	8.12E-09	1.9E-03	31.75
MIR_BVE-PLS_MasEsp_snv	MIR	BVE	snv	3	1.5E-04	0.98	1.6E-04	0.98	1.5E-04	0.96	5.018E-17	1.72E-07	4.82E-09	1.5E-03	32.16
MIR_SPA-PLS_MasEsp_mc	MIR	SPA	mc	3	1.6E-04	0.98	1.8E-04	0.98	1.6E-04	0.96	5.019E-17	3.97E-08	6.97E-10	5.6E-04	49.93
MIR_SPA-PLS_MasEsp_drv	MIR	SPA	drv	2	1.7E-04	0.98	1.7E-04	0.98	1.7E-04	0.96	5.018E-17	1.48E-07	1.64E-10	2.7E-04	28.34
MIR_SPA-PLS_MasEsp_snv	MIR	SPA	snv	3	1.6E-04	0.98	1.7E-04	0.98	1.6E-04	0.96	5.023E-17	4.71E-08	3.16E-09	1.2E-03	55.68
MIR_REP-PLS_MasEsp_mc	MIR	REP	mc	2	2.0E-04	0.97	2.1E-04	0.97	1.7E-04	0.95	5.017E-17	1.33E-07	3.09E-08	3.7E-03	15.56
MIR_REP-PLS_MasEsp_drv	MIR	REP	drv	3	1.6E-04	0.98	1.7E-04	0.98	1.5E-04	0.96	5.019E-17	1.85E-07	7.49E-09	1.8E-03	33.96
MIR_REP-PLS_MasEsp_snv	MIR	REP	snv	3	1.5E-04	0.98	1.6E-04	0.98	1.5E-04	0.97	5.017E-17	1.70E-07	3.62E-09	1.3E-03	34.02
MIR_IPW-PLS_MasEsp_mc	MIR	IPW	mc	5	1.3E-04	0.99	1.4E-04	0.98	1.5E-04	0.97	5.021E-17	8.30E-08	5.72E-08	5.0E-03	12.68
MIR_IPW-PLS_MasEsp_drv	MIR	IPW	drv	2	1.7E-04	0.98	1.8E-04	0.98	1.7E-04	0.96	5.022E-17	5.90E-08	5.48E-10	4.9E-04	67.23
MIR_IPW-PLS_MasEsp_snv	MIR	IPW	snv	4	1.3E-04	0.99	1.4E-04	0.98	1.4E-04	0.97	5.020E-17	1.21E-07	5.85E-08	5.1E-03	10.96
MIR_enPLS_MasEsp_mc	MIR	Todo	mc	7	1.2E-04	0.99	1.4E-04	0.98	1.4E-04	0.97	1.455E-08	1.48E-09	3.36E-08	3.9E-03	20.89
MIR_enPLS_MasEsp_drv	MIR	Todo	drv	10	8.9E-05	0.99	1.9E-04	0.97	1.7E-04	0.95	2.629E-09	1.97E-08	4.76E-08	4.6E-03	32.63
MIR_enPLS_MasEsp_snv	MIR	Todo	snv	9	1.1E-04	0.99	1.2E-04	0.99	1.3E-04	0.97	2.136E-09	3.04E-09	5.15E-08	4.8E-03	5.57
MIR_RF_MasEsp_mc	MIR	Todo	mc	-	8.4E-05	0.99	2.1E-04	0.97	1.2E-04	0.98	5.581E-09	7.43E-09	3.28E-08	3.8E-03	2.44
MIR_RF_MasEsp_drv	MIR	Todo	drv	-	7.3E-05	0.99	1.8E-04	0.98	1.2E-04	0.98	8.228E-10	8.24E-09	3.52E-08	4.0E-03	2.06
MIR_RF_MasEsp_snv	MIR	Todo	snv	-	7.2E-05	0.99	1.8E-04	0.98	1.3E-04	0.97	3.237E-09	1.11E-08	7.30E-08	5.7E-03	5.29

Modelos	Espectros	Intervalos	PPx	VL	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>cal</sub>
RAM_PLS_MaseEsp_tss	RAM	Todo	tss	4	3.2E-04	0.92	9.0E-04	0.39	2.5E-04	0.92	2.705E-17	8.23E-05	8.47E-09	1.9E-03	76.14
RAM_PLS_MaseEsp_drv	RAM	Todo	drv	2	4.5E-04	0.85	6.4E-04	0.70	2.8E-04	0.85	2.704E-17	3.29E-05	5.89E-09	1.6E-03	87.66
RAM_PLS_MaseEsp_snv	RAM	Todo	snv	3	3.6E-04	0.90	8.6E-04	0.44	3.2E-04	0.90	2.701E-17	8.39E-05	1.30E-11	7.6E-05	54.29
RAM_LW-PLS_MaseEsp_tss	RAM	LW	tss	3	3.0E-04	0.93	7.3E-04	0.60	2.5E-04	0.90	2.703E-17	5.05E-05	1.72E-08	2.8E-03	74.46
RAM_LW-PLS_MaseEsp_drv	RAM	LW	drv	2	4.5E-04	0.85	6.4E-04	0.70	2.9E-04	0.87	2.704E-17	2.86E-05	5.89E-09	1.6E-03	87.66
RAM_LW-PLS_MaseEsp_snv	RAM	LW	snv	3	3.1E-04	0.93	7.4E-04	0.59	2.5E-04	0.90	2.709E-17	5.28E-05	3.01E-08	3.7E-03	78.32
RAM_RC-PLS_MaseEsp_tss	RAM	RC	tss	2	4.9E-04	0.82	5.6E-04	0.77	4.0E-04	0.75	2.708E-17	3.47E-05	2.20E-08	3.1E-03	108.60
RAM_RC-PLS_MaseEsp_drv	RAM	RC	drv	2	4.9E-04	0.82	5.9E-04	0.74	4.8E-04	0.64	2.702E-17	8.08E-06	1.65E-07	8.6E-03	131.08
RAM_RC-PLS_MaseEsp_snv	RAM	RC	snv	2	5.0E-04	0.81	5.5E-04	0.77	4.1E-04	0.73	2.701E-17	6.52E-05	2.02E-08	3.0E-03	114.83
RAM_VIP-PLS_MaseEsp_tss	RAM	VIP	tss	5	4.0E-04	0.88	9.7E-04	0.29	2.9E-04	0.87	2.704E-17	4.81E-05	2.30E-07	1.0E-02	81.09
RAM_VIP-PLS_MaseEsp_drv	RAM	VIP	drv	3	5.1E-04	0.80	7.2E-04	0.61	4.0E-04	0.75	2.703E-17	2.19E-05	1.22E-07	7.4E-03	120.35
RAM_VIP-PLS_MaseEsp_snv	RAM	VIP	snv	4	5.0E-04	0.81	1.1E-03	0.16	4.0E-04	0.75	2.703E-17	4.81E-05	3.53E-08	4.0E-03	111.24
RAM_iPLS_MaseEsp_tss	RAM	iPLS	tss	4	4.7E-04	0.83	9.9E-04	0.26	4.4E-04	0.69	2.704E-17	9.83E-05	7.85E-08	5.9E-03	125.39
RAM_iPLS_MaseEsp_drv	RAM	iPLS	drv	5	3.6E-04	0.90	8.9E-04	0.40	6.9E-04	0.25	2.705E-17	1.92E-05	4.46E-07	1.4E-02	137.01
RAM_iPLS_MaseEsp_snv	RAM	iPLS	snv	4	3.9E-04	0.88	9.4E-04	0.33	3.6E-04	0.80	2.703E-17	9.40E-05	5.29E-10	4.9E-04	111.35
RAM_GA-PLS_MaseEsp_tss	RAM	GA	tss	3	2.5E-04	0.95	6.6E-04	0.67	2.8E-04	0.87	2.702E-17	2.67E-05	4.11E-08	4.3E-03	89.01
RAM_GA-PLS_MaseEsp_drv	RAM	GA	drv	2	4.7E-04	0.84	5.9E-04	0.73	4.7E-04	0.64	2.707E-17	7.52E-08	2.91E-07	1.1E-02	122.62
RAM_GA-PLS_MaseEsp_snv	RAM	GA	snv	5	2.0E-04	0.97	5.1E-04	0.80	2.3E-04	0.92	2.704E-17	2.91E-05	3.05E-08	3.7E-03	68.45
RAM_UVE-PLS_MaseEsp_tss	RAM	UVE	tss	3	3.9E-04	0.89	8.9E-04	0.40	2.6E-04	0.89	2.704E-17	8.02E-05	1.10E-11	7.0E-05	63.10
RAM_UVE-PLS_MaseEsp_drv	RAM	UVE	drv	2	4.8E-04	0.82	6.7E-04	0.66	3.0E-04	0.86	2.703E-17	3.32E-05	4.26E-09	1.4E-03	95.30
RAM_UVE-PLS_MaseEsp_snv	RAM	UVE	snv	4	3.2E-04	0.92	9.3E-04	0.34	2.6E-04	0.89	2.704E-17	8.11E-05	1.05E-08	2.2E-03	87.45

Modelos	Espectros	Intervalos	PPx	VL	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
RAM_BVE-PLS_MasEsp_tss	RAM	BVE	tss	4	3.5E-04	0.91	9.4E-04	0.33	2.7E-04	0.88	2.702E-17	8.37E-05	4.41E-08	4.4E-03	101.89
RAM_BVE-PLS_MasEsp_drv	RAM	BVE	drv	2	5.0E-04	0.81	6.8E-04	0.66	3.0E-04	0.85	2.699E-17	4.67E-05	3.33E-11	1.2E-04	97.78
RAM_BVE-PLS_MasEsp_snv	RAM	BVE	snv	4	3.5E-04	0.91	9.4E-04	0.33	2.7E-04	0.88	2.703E-17	7.75E-05	3.49E-08	3.9E-03	93.69
RAM_SPA-PLS_MasEsp_tss	RAM	SPA	tss	3	3.9E-04	0.89	9.1E-04	0.37	2.5E-04	0.90	2.702E-17	8.31E-05	9.48E-10	6.5E-04	65.34
RAM_SPA-PLS_MasEsp_drv	RAM	SPA	drv	4	2.4E-04	0.95	6.0E-04	0.73	3.6E-04	0.79	2.701E-17	2.18E-05	2.07E-07	9.6E-03	96.34
RAM_SPA-PLS_MasEsp_snv	RAM	SPA	snv	4	3.4E-04	0.91	9.2E-04	0.36	2.6E-04	0.90	2.703E-17	8.61E-05	1.43E-08	2.5E-03	83.67
RAM_REP-PLS_MasEsp_tss	RAM	REP	tss	4	3.6E-04	0.90	9.5E-04	0.32	2.8E-04	0.88	2.702E-17	8.59E-05	3.52E-08	4.0E-03	100.67
RAM_REP-PLS_MasEsp_drv	RAM	REP	drv	3	3.9E-04	0.88	6.9E-04	0.64	3.3E-04	0.83	2.699E-17	5.00E-05	9.22E-08	6.4E-03	100.27
RAM_REP-PLS_MasEsp_snv	RAM	REP	snv	6	2.3E-04	0.96	8.1E-04	0.50	1.9E-04	0.94	2.706E-17	8.25E-05	2.21E-08	3.1E-03	29.31
RAM_IPW-PLS_MasEsp_tss	RAM	IPW	tss	4	3.2E-04	0.92	8.9E-04	0.41	2.5E-04	0.90	2.704E-17	6.24E-05	3.58E-08	4.0E-03	86.45
RAM_IPW-PLS_MasEsp_drv	RAM	IPW	drv	2	4.8E-04	0.82	6.6E-04	0.67	3.0E-04	0.86	2.704E-17	1.84E-05	9.11E-08	6.4E-03	91.81
RAM_IPW-PLS_MasEsp_snv	RAM	IPW	snv	4	3.2E-04	0.92	9.0E-04	0.39	2.4E-04	0.91	2.702E-17	6.42E-05	3.26E-08	3.8E-03	82.39
RAM_enPLS_MasEsp_tss	RAM	Todo	tss	10	3.3E-04	0.92	7.2E-04	0.61	2.2E-04	0.93	3.622E-07	2.98E-07	2.64E-08	3.4E-03	71.38
RAM_enPLS_MasEsp_drv	RAM	Todo	drv	6	3.7E-04	0.90	4.8E-04	0.83	2.9E-04	0.87	1.282E-07	6.48E-09	3.50E-10	3.9E-04	77.46
RAM_enPLS_MasEsp_snv	RAM	Todo	snv	9	1.4E-04	0.98	6.7E-04	0.66	2.0E-04	0.94	1.208E-08	4.15E-07	3.28E-08	3.8E-03	61.85
RAM_RF_MasEsp_tss	RAM	Todo	tss	-	1.1E-04	0.99	3.0E-04	0.93	1.8E-04	0.95	9.516E-09	2.67E-08	2.20E-08	3.1E-03	32.28
RAM_RF_MasEsp_drv	RAM	Todo	drv	-	1.6E-04	0.98	4.7E-04	0.83	4.0E-04	0.75	7.806E-10	1.02E-07	2.09E-08	3.0E-03	115.73
RAM_RF_MasEsp_snv	RAM	Todo	snv	-	1.1E-04	0.99	3.1E-04	0.93	1.7E-04	0.96	5.777E-10	1.21E-08	1.67E-08	2.7E-03	8.79



## ANEXO E - Parâmetros de mérito dos modelos PLS com seleção de variáveis e Florestas Aleatórias (RF) obtidos a partir dos espectros de NIR, mid-IR e RAM para a previsão do PEFF

Modelos	Espectros	Intervalos	PPx	VL	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
NIR_PLS_PEFF_tss	NIR	Todo	tss	8	1,8	0,88	2,2	0,83	1,2	0,83	9,2E-30	1,86E-01	1,28E+00	3,6	61,97
NIR_PLS_PEFF_drv	NIR	Todo	drv	6	2,1	0,85	2,6	0,77	1,8	0,64	4,4E-29	1,64E+00	3,97E+00	7,0	122,39
NIR_PLS_PEFF_snv	NIR	Todo	snv	6	2,2	0,82	2,6	0,75	1,6	0,73	1,2E-28	2,48E-02	8,01E+00	10,3	78,51
NIR_LW-PLS_PEFF_tss	NIR	LW	tss	7	2,3	0,81	2,8	0,72	1,8	0,64	4,9E-30	1,22E+01	4,83E-01	2,2	121,43
NIR_LW-PLS_PEFF_drv	NIR	LW	drv	8	1,9	0,87	2,6	0,76	1,9	0,61	8,8E-29	1,85E+02	2,05E-03	0,1	187,68
NIR_LW-PLS_PEFF_snv	NIR	LW	snv	6	2,5	0,78	2,9	0,69	2,0	0,56	1,7E-29	1,78E+01	1,87E+00	4,7	185,07
NIR_RC-PLS_PEFF_tss	NIR	RC	tss	7	2,1	0,85	2,4	0,80	1,3	0,82	7,4E-31	5,85E+00	2,77E+00	5,2	56,08
NIR_RC-PLS_PEFF_drv	NIR	RC	drv	6	2,1	0,84	2,5	0,78	1,7	0,67	8,0E-29	6,03E+01	2,18E+00	5,1	99,11
NIR_RC-PLS_PEFF_snv	NIR	RC	snv	6	2,3	0,81	2,6	0,76	1,6	0,73	1,9E-29	1,06E+01	6,24E-02	0,8	70,85
NIR_VIP-PLS_PEFF_tss	NIR	VIP	tss	5	2,9	0,70	3,1	0,65	2,2	0,48	5,4E-32	2,56E-03	1,05E+01	11,9	181,28
NIR_VIP-PLS_PEFF_drv	NIR	VIP	drv	6	2,2	0,82	2,5	0,77	1,9	0,60	3,9E-29	1,10E+01	1,24E+01	13,1	186,89
NIR_VIP-PLS_PEFF_snv	NIR	VIP	snv	8	2,2	0,83	2,5	0,78	1,7	0,67	1,3E-30	1,27E+00	4,28E+00	7,3	133,10
NIR_iPLS_PEFF_tss	NIR	iPLS	tss	5	2,5	0,77	2,8	0,72	1,8	0,63	2,0E-29	5,85E+00	5,98E-03	0,3	139,91
NIR_iPLS_PEFF_drv	NIR	iPLS	drv	5	2,5	0,78	2,8	0,72	1,7	0,67	1,0E-28	2,47E+00	6,11E+00	7,5	96,18
NIR_iPLS_PEFF_snv	NIR	iPLS	snv	5	2,6	0,77	2,8	0,72	1,7	0,67	5,8E-30	5,73E-01	4,83E-01	2,2	90,09
NIR_GA-PLS_PEFF_tss	NIR	GA	tss	6	2,6	0,76	3,1	0,66	2,2	0,47	6,5E-29	4,40E+00	2,58E+00	5,6	207,07
NIR_GA-PLS_PEFF_drv	NIR	GA	drv	6	2,1	0,84	2,6	0,76	1,8	0,63	4,0E-29	5,78E-01	1,04E-02	0,3	166,19
NIR_GA-PLS_PEFF_snv	NIR	GA	snv	7	2,2	0,83	2,6	0,75	1,8	0,64	2,2E-28	4,74E+00	6,00E-03	0,3	124,80
NIR_UVE-PLS_PEFF_tss	NIR	UVE	tss	5	2,5	0,77	2,9	0,71	1,8	0,63	9,2E-29	6,71E+00	5,12E-01	2,3	181,28
NIR_UVE-PLS_PEFF_drv	NIR	UVE	drv	5	2,7	0,75	3,0	0,67	2,2	0,48	9,7E-29	1,30E+00	1,19E-01	1,1	186,89
NIR_UVE-PLS_PEFF_snv	NIR	UVE	snv	6	2,5	0,78	2,9	0,69	1,6	0,70	1,6E-29	8,76E+00	1,72E+00	4,1	133,10

Modelos	Espectros	Intervalos	Ppx	VL	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
NIR_BVE-PLS_PEFF_tss	NIR	BVE	tss	6	2,4	0,80	2,7	0,75	1,6	0,71	5,2E-29	5,96E+00	2,30E+01	18,7	91,65
NIR_BVE-PLS_PEFF_drv	NIR	BVE	drv	7	2,0	0,86	2,3	0,80	1,6	0,71	8,6E-29	1,87E+01	4,36E+00	7,4	118,56
NIR_BVE-PLS_PEFF_snv	NIR	BVE	snv	6	2,3	0,82	2,6	0,75	1,5	0,74	3,1E-29	1,01E+01	1,32E+01	13,6	77,72
NIR_SPA-PLS_PEFF_tss	NIR	SPA	tss	6	3,0	0,68	3,3	0,62	2,0	0,54	6,2E-30	4,49E+00	1,96E+00	4,8	159,81
NIR_SPA-PLS_PEFF_drv	NIR	SPA	drv	5	2,8	0,72	3,3	0,61	2,5	0,30	4,4E-28	1,63E-02	8,18E-01	3,1	205,38
NIR_SPA-PLS_PEFF_snv	NIR	SPA	snv	6	2,7	0,74	3,1	0,67	1,9	0,62	1,4E-28	3,74E-01	2,76E-01	1,8	106,88
NIR_REP-PLS_PEFF_tss	NIR	REP	tss	6	2,4	0,79	2,7	0,74	1,6	0,70	2,2E-29	4,52E+00	2,15E+01	18,0	97,08
NIR_REP-PLS_PEFF_drv	NIR	REP	drv	7	1,9	0,87	2,3	0,81	1,6	0,72	7,6E-29	2,58E+01	3,61E+00	6,7	107,58
NIR_REP-PLS_PEFF_snv	NIR	REP	snv	6	2,3	0,82	2,6	0,75	1,6	0,73	5,3E-29	9,94E+00	1,25E+01	13,1	81,42
NIR_IPW-PLS_PEFF_tss	NIR	IPW	tss	6	2,4	0,80	2,6	0,75	1,8	0,64	1,1E-30	5,38E-01	1,31E+01	13,5	130,79
NIR_IPW-PLS_PEFF_drv	NIR	IPW	drv	6	2,6	0,75	3,0	0,67	2,3	0,41	3,1E-30	2,78E+01	5,18E-01	2,4	202,65
NIR_IPW-PLS_PEFF_snv	NIR	IPW	snv	6	2,3	0,81	2,6	0,75	1,9	0,62	1,1E-30	5,20E+00	4,17E+00	7,2	173,14
NIR_enPLS_PEFF_tss	NIR	Todo	tss	10	1,6	0,90	2,3	0,82	1,7	0,69	2,7E+00	1,30E-01	1,50E+01	11,3	123,20
NIR_enPLS_PEFF_drv	NIR	Todo	drv	9	2,3	0,81	3,5	0,55	1,9	0,62	2,3E+01	5,58E+00	1,54E-02	0,4	121,89
NIR_enPLS_PEFF_snv	NIR	Todo	snv	10	2,0	0,86	3,0	0,69	1,6	0,73	1,5E+00	9,37E-01	1,00E+01	9,4	71,49
NIR_RF_PEFF_tss	NIR	Todo	tss	-	0,8	0,98	2,1	0,84	1,0	0,89	2,4E-03	4,79E-02	4,77E-01	2,2	26,30
NIR_RF_PEFF_drv	NIR	Todo	drv	-	0,9	0,97	2,7	0,74	1,3	0,80	3,6E-02	4,62E-02	5,98E-02	0,8	54,93
NIR_RF_PEFF_snv	NIR	Todo	snv	-	0,8	0,98	2,0	0,86	1,2	0,85	1,1E-02	1,78E-03	1,51E+00	3,9	32,25

Modelos	Espectros	Intervalos	PRx	nsV	kernel	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
NIR_SVM_PEFF_tss	NIR	Todo	tss	266	radial	0,2	1,00	1,1	0,86	4,7E-03	6,01E+00	7,5	55,05
NIR_SVM_PEFF_drv	NIR	Todo	drv	183	linear	0,9	0,97	1,8	0,64	2,6E+00	4,15E-01	2,1	108,12
NIR_SVM_PEFF_snv	NIR	Todo	snv	266	radial	0,1	1,00	1,1	0,86	2,5E-03	4,32E+00	6,4	41,79
NIR_LW-SVM_PEFF_tss	NIR	LW	tss	266	radial	0,4	0,99	1,1	0,86	8,3E-04	8,42E+00	8,7	39,67
NIR_LW-SVM_PEFF_drv	NIR	LW	drv	266	radial	0,1	1,00	1,8	0,66	3,1E-03	6,39E+00	7,7	96,39
NIR_LW-SVM_PEFF_snv	NIR	LW	snv	266	radial	0,5	0,99	1,2	0,84	3,3E-01	3,67E+00	5,9	45,68
NIR_RC-SVM_PEFF_tss	NIR	RC	tss	266	radial	0,6	0,99	1,2	0,84	5,4E-02	3,30E+00	5,6	50,94
NIR_RC-SVM_PEFF_drv	NIR	RC	drv	155	linear	1,3	0,94	1,9	0,61	1,8E+01	3,40E+00	6,5	114,17
NIR_RC-SVM_PEFF_snv	NIR	RC	snv	266	radial	0,2	1,00	1,2	0,84	2,6E-02	1,58E+00	4,0	49,72
NIR_VIP-SVM_PEFF_tss	NIR	VIP	tss	189	radial	0,8	0,98	1,2	0,84	4,2E-01	1,95E+00	4,4	32,32
NIR_VIP-SVM_PEFF_drv	NIR	VIP	drv	200	radial	0,8	0,98	1,1	0,86	1,9E-01	6,42E-01	2,7	38,06
NIR_VIP-SVM_PEFF_snv	NIR	VIP	snv	266	radial	0,7	0,98	1,2	0,85	2,0E-02	2,24E+00	4,7	34,89
NIR_iPLS-SVM_PEFF_tss	NIR	iPLS	tss	200	radial	0,9	0,97	1,2	0,83	7,3E-02	2,38E+00	4,8	53,45
NIR_iPLS-SVM_PEFF_drv	NIR	iPLS	drv	212	radial	0,5	0,99	1,5	0,75	9,1E-03	6,62E-01	2,6	76,09
NIR_iPLS-SVM_PEFF_snv	NIR	iPLS	snv	266	radial	0,8	0,98	1,2	0,84	1,4E-01	2,50E+00	4,9	49,62
NIR_GA-SVM_PEFF_tss	NIR	GA	tss	197	radial	0,6	0,99	1,3	0,82	1,1E-01	1,12E+01	9,9	60,70
NIR_GA-SVM_PEFF_drv	NIR	GA	drv	266	radial	0,2	1,00	1,8	0,66	3,5E-03	6,68E+00	7,8	93,59
NIR_GA-SVM_PEFF_snv	NIR	GA	snv	266	radial	0,3	1,00	1,4	0,79	1,1E-01	1,14E+01	10,0	71,24

Modelos	Espectros	Intervalos	PPx	nsV	kernel	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
NIR_UVE-SVM_PEFF_tss	NIR	UVE	tss	266	radial	0,8	0,98	1,2	0,85	3,7E-01	4,77E+00	6,7	55,73
NIR_UVE-SVM_PEFF_dlv	NIR	UVE	dlv	266	radial	0,3	1,00	1,4	0,78	6,8E-03	4,77E-02	0,7	64,48
NIR_UVE-SVM_PEFF_snv	NIR	UVE	snv	266	radial	0,5	0,99	1,3	0,83	2,1E-02	6,31E+00	7,6	62,21
NIR_BVE-SVM_PEFF_tss	NIR	BVE	tss	176	radial	0,7	0,98	1,0	0,89	3,4E-03	2,27E+00	4,7	31,39
NIR_BVE-SVM_PEFF_dlv	NIR	BVE	dlv	266	radial	0,4	0,99	1,2	0,84	6,8E-03	3,77E+00	6,0	62,83
NIR_BVE-SVM_PEFF_snv	NIR	BVE	snv	266	radial	0,4	0,99	1,0	0,88	6,9E-03	1,46E+00	3,8	30,32
NIR_SPA-SVM_PEFF_tss	NIR	SPA	tss	179	radial	0,9	0,97	1,3	0,82	1,3E-02	5,38E+00	7,1	61,36
NIR_SPA-SVM_PEFF_dlv	NIR	SPA	dlv	266	radial	0,4	0,99	1,9	0,61	1,7E-03	2,17E+00	4,6	138,16
NIR_SPA-SVM_PEFF_snv	NIR	SPA	snv	214	radial	0,7	0,98	1,4	0,79	1,8E-02	5,45E+00	7,1	56,52
NIR_REP-SVM_PEFF_tss	NIR	REP	tss	185	radial	0,7	0,98	1,0	0,89	6,1E-02	3,79E+00	6,0	36,04
NIR_REP-SVM_PEFF_dlv	NIR	REP	dlv	265	radial	0,6	0,99	1,2	0,84	1,7E-03	3,60E+00	5,9	52,61
NIR_REP-SVM_PEFF_snv	NIR	REP	snv	266	radial	0,6	0,99	1,0	0,88	1,4E-04	1,33E+00	3,7	31,77
NIR_IPW-SVM_PEFF_tss	NIR	IPW	tss	192	radial	0,5	0,99	1,1	0,87	1,4E-02	5,76E+00	7,3	35,08
NIR_IPW-SVM_PEFF_dlv	NIR	IPW	dlv	266	radial	0,0	1,00	2,1	0,51	1,9E-07	1,73E-02	0,4	170,09
NIR_IPW-SVM_PEFF_snv	NIR	IPW	snv	266	radial	0,2	1,00	1,2	0,84	3,6E-02	2,26E+00	4,7	55,63

Modelos	Espectros	Intervalos	PPx	VL	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
MIR_PLS_PEFF_mc	MIR	Todo	mc	8	1,4	0,93	1,5	0,91	1,3	0,82	6,5E-29	6,27E-02	1,50E-01	1,3	51,06
MIR_PLS_PEFF_dhv	MIR	Todo	dhv	4	2,2	0,83	2,3	0,81	2,0	0,57	2,2E-28	4,35E-02	5,34E-01	2,5	168,71
MIR_PLS_PEFF_snv	MIR	Todo	snv	6	1,7	0,90	1,8	0,88	1,4	0,77	2,4E-29	2,68E-01	2,67E-01	1,7	77,39
MIR_LW-PLS_PEFF_mc	MIR	LW	mc	4	2,3	0,81	2,5	0,78	2,0	0,55	4,2E-29	9,79E+00	1,85E-03	0,1	172,21
MIR_LW-PLS_PEFF_dhv	MIR	LW	dhv	4	2,2	0,82	2,4	0,80	2,0	0,57	2,8E-28	5,04E-01	1,19E-01	1,1	169,23
MIR_LW-PLS_PEFF_snv	MIR	LW	snv	4	2,3	0,81	2,4	0,79	2,0	0,57	3,5E-29	1,71E+00	3,32E-03	0,2	172,06
MIR_RC-PLS_PEFF_mc	MIR	RC	mc	4	2,4	0,80	2,5	0,78	2,1	0,53	2,6E-29	4,73E-01	3,69E+00	6,7	184,24
MIR_RC-PLS_PEFF_dhv	MIR	RC	dhv	4	2,3	0,82	2,4	0,80	2,0	0,55	3,5E-30	1,11E-01	4,19E-01	2,2	173,76
MIR_RC-PLS_PEFF_snv	MIR	RC	snv	4	2,3	0,82	2,4	0,80	2,0	0,58	5,7E-30	6,65E+00	1,09E+00	3,6	178,05
MIR_VIP-PLS_PEFF_mc	MIR	VIP	mc	3	2,6	0,75	2,7	0,74	2,2	0,45	6,9E-29	1,38E+00	9,73E+00	11,4	189,58
MIR_VIP-PLS_PEFF_dhv	MIR	VIP	dhv	4	2,2	0,82	2,3	0,81	1,9	0,59	1,3E-28	7,14E+00	1,09E-01	1,1	159,13
MIR_VIP-PLS_PEFF_snv	MIR	VIP	snv	3	2,5	0,78	2,6	0,75	2,1	0,53	6,7E-33	2,28E+00	3,86E+00	6,9	179,59
MIR_iPLS_PEFF_mc	MIR	iPLS	mc	5	2,1	0,84	2,2	0,83	2,1	0,53	2,1E-29	2,43E+00	9,85E-01	3,4	152,88
MIR_iPLS_PEFF_dhv	MIR	iPLS	dhv	6	1,6	0,91	1,8	0,89	1,7	0,69	3,2E-29	3,50E-01	6,60E+00	7,8	100,98
MIR_iPLS_PEFF_snv	MIR	iPLS	snv	4	2,2	0,82	2,3	0,80	2,2	0,48	1,4E-29	1,99E+00	1,28E+00	3,9	130,39
MIR_GA-PLS_PEFF_mc	MIR	GA	mc	4	2,3	0,82	2,4	0,80	2,1	0,53	3,6E-28	8,22E-02	4,09E+00	7,1	183,02
MIR_GA-PLS_PEFF_dhv	MIR	GA	dhv	4	1,9	0,87	2,0	0,85	1,7	0,67	9,0E-28	4,46E-02	9,09E-02	1,0	99,09
MIR_GA-PLS_PEFF_snv	MIR	GA	snv	4	2,1	0,85	2,1	0,84	1,8	0,62	7,1E-29	2,68E+00	1,61E+00	4,3	154,00
MIR_UVE-PLS_PEFF_mc	MIR	UVE	mc	5	2,0	0,86	2,1	0,84	1,7	0,67	1,5E-28	8,13E+00	8,03E-01	3,0	112,26
MIR_UVE-PLS_PEFF_dhv	MIR	UVE	dhv	5	1,8	0,89	1,9	0,87	1,5	0,75	2,2E-28	4,36E-01	2,15E-01	1,5	69,53
MIR_UVE-PLS_PEFF_snv	MIR	UVE	snv	6	1,6	0,91	1,7	0,90	1,4	0,80	4,8E-28	1,26E+00	1,72E-01	1,4	63,91

Modelos	Espectros	Intervalos	PPx	VL	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
MIR_BVE-PLS_PEFF_mc	MIR	BVE	mc	3	2,7	0,74	2,8	0,73	2,3	0,43	3,8E-28	3,86E+00	7,85E+00	10,1	197,48
MIR_BVE-PLS_PEFF_dtv	MIR	BVE	dtv	4	2,2	0,83	2,3	0,81	1,9	0,58	3,1E-28	2,30E+00	1,79E-02	0,4	157,94
MIR_BVE-PLS_PEFF_snv	MIR	BVE	snv	3	2,5	0,77	2,6	0,76	2,1	0,51	2,3E-29	2,89E+00	3,02E+00	6,1	191,89
MIR_SPA-PLS_PEFF_mc	MIR	SPA	mc	4	3,0	0,68	3,1	0,66	2,7	0,19	3,6E-28	1,21E+00	3,02E+01	22,1	214,51
MIR_SPA-PLS_PEFF_dtv	MIR	SPA	dtv	4	3,2	0,64	3,4	0,58	2,5	0,30	1,4E-28	1,85E+01	2,40E-04	0,1	207,57
MIR_SPA-PLS_PEFF_snv	MIR	SPA	snv	6	2,3	0,82	2,5	0,77	1,9	0,60	1,3E-28	4,86E+00	2,25E+00	5,2	172,83
MIR_REP-PLS_PEFF_mc	MIR	REP	mc	3	2,7	0,74	2,8	0,73	2,3	0,43	1,7E-28	4,38E+00	7,95E+00	10,2	198,33
MIR_REP-PLS_PEFF_dtv	MIR	REP	dtv	4	2,1	0,84	2,3	0,81	1,9	0,62	1,5E-31	2,98E+00	1,01E-02	0,3	135,49
MIR_REP-PLS_PEFF_snv	MIR	REP	snv	3	2,5	0,77	2,6	0,76	2,1	0,51	1,2E-28	3,75E+00	2,99E+00	6,0	193,37
MIR_IPW-PLS_PEFF_mc	MIR	IPW	mc	3	2,9	0,70	3,0	0,68	2,5	0,29	4,7E-29	1,18E-02	5,96E-01	2,5	181,43
MIR_IPW-PLS_PEFF_dtv	MIR	IPW	dtv	6	1,9	0,88	2,1	0,85	1,8	0,65	1,4E-31	2,59E+00	2,65E+00	5,7	99,74
MIR_IPW-PLS_PEFF_snv	MIR	IPW	snv	4	2,4	0,80	2,5	0,78	2,1	0,54	2,6E-28	1,19E+00	9,71E-01	3,3	193,08
MIR_enPLS_PEFF_mc	MIR	Todo	mc	10	1,2	0,95	1,4	0,93	1,2	0,84	1,8E-02	3,19E-03	2,05E-01	1,5	34,49
MIR_enPLS_PEFF_dtv	MIR	Todo	dtv	10	1,6	0,91	3,0	0,68	1,6	0,72	1,7E+00	4,05E+00	5,28E-02	0,8	71,08
MIR_enPLS_PEFF_snv	MIR	Todo	snv	10	1,2	0,95	1,5	0,92	1,2	0,84	5,6E+00	5,18E-01	1,06E+00	3,3	41,40
MIR_RF_PEFF_mc	MIR	Todo	mc	-	0,6	0,99	1,4	0,93	0,9	0,91	1,3E-03	1,32E-02	2,36E-01	1,6	6,77
MIR_RF_PEFF_dtv	MIR	Todo	dtv	-	0,6	0,99	1,6	0,91	0,9	0,90	2,6E-03	7,91E-02	7,76E-02	0,9	17,59
MIR_RF_PEFF_snv	MIR	Todo	snv	-	0,6	0,99	1,6	0,91	0,8	0,93	1,8E-02	1,14E+00	4,40E-03	0,2	5,91

Modelos	Espectros	Intervalos	PPx	nSV	kernel	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
MIR_SVM_PEFF_mc	MIR	Todo	mc	265	radial	0,7	0,98	0,8	0,93	1,2E-02	1,79E-01	1,4	9,98
MIR_SVM_PEFF_dhv	MIR	Todo	dhv	123	linear	1,3	0,94	2,0	0,55	1,5E-01	5,23E-01	2,3	109,38
MIR_SVM_PEFF_snv	MIR	Todo	snv	266	radial	0,1	1,00	0,9	0,91	9,7E-03	5,46E-02	0,8	16,88
MIR_LW-SVM_PEFF_mc	MIR	LW	mc	129	radial	0,8	0,98	0,8	0,93	1,3E-03	1,32E-02	0,4	6,37
MIR_LW-SVM_PEFF_dhv	MIR	LW	dhv	266	polinomial	0,8	0,98	1,5	0,74	7,1E-01	2,11E+00	4,6	47,35
MIR_LW-SVM_PEFF_snv	MIR	LW	snv	266	radial	0,2	1,00	0,9	0,91	1,2E-02	4,88E-04	0,1	4,74
MIR_RC-SVM_PEFF_mc	MIR	RC	mc	136	radial	0,8	0,98	0,7	0,94	6,6E-02	1,14E-02	0,4	0,58
MIR_RC-SVM_PEFF_dhv	MIR	RC	dhv	83	linear	1,5	0,92	2,1	0,51	7,8E+00	3,68E-01	2,0	97,10
MIR_RC-SVM_PEFF_snv	MIR	RC	snv	264	radial	0,3	1,00	0,8	0,93	2,7E-02	8,59E-02	1,0	3,99
MIR_VIP-SVM_PEFF_mc	MIR	VIP	mc	266	radial	0,9	0,97	1,0	0,89	1,2E-02	3,39E-02	0,6	32,29
MIR_VIP-SVM_PEFF_dhv	MIR	VIP	dhv	266	radial	0,7	0,98	1,0	0,89	8,9E-02	3,18E-01	1,9	23,94
MIR_VIP-SVM_PEFF_snv	MIR	VIP	snv	96	radial	1,2	0,95	1,2	0,84	7,1E-02	2,81E-01	1,8	50,28
MIR_iPLS-SVM_PEFF_mc	MIR	iPLS	mc	135	radial	0,9	0,97	0,9	0,92	2,9E-01	2,20E-01	1,5	9,77
MIR_iPLS-SVM_PEFF_dhv	MIR	iPLS	dhv	266	polinomial	0,7	0,98	1,1	0,86	1,9E+00	2,42E-01	1,6	27,93
MIR_iPLS-SVM_PEFF_snv	MIR	iPLS	snv	142	radial	0,8	0,98	0,9	0,92	7,6E-03	1,91E-01	1,4	14,92
MIR_GA-SVM_PEFF_mc	MIR	GA	mc	125	radial	0,9	0,97	0,8	0,93	3,5E-01	1,39E-01	1,2	9,87
MIR_GA-SVM_PEFF_dhv	MIR	GA	dhv	57	linear	2,0	0,85	2,5	0,31	7,1E+00	8,97E+00	10,9	203,18
MIR_GA-SVM_PEFF_snv	MIR	GA	snv	266	radial	0,0	1,00	1,0	0,90	1,6E-06	2,85E-02	0,6	21,71

Modelos	Espectros	Intervalos	PPx	nSV	kernel	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
MIR_UVE-SVM_PEFF_mc	MIR	UVE	mc	139	radial	0,7	0,98	0,8	0,92	1,5E-01	1,52E-01	1,3	16,96
MIR_UVE-SVM_PEFF_drv	MIR	UVE	drv	266	radial	0,7	0,98	1,1	0,87	9,1E-01	6,19E-01	2,5	27,18
MIR_UVE-SVM_PEFF_snv	MIR	UVE	snv	266	radial	0,7	0,98	0,9	0,91	5,8E-02	6,00E-01	2,6	15,97
MIR_BVE-SVM_PEFF_mc	MIR	BVE	mc	266	radial	0,7	0,98	0,8	0,92	8,0E-02	1,21E+00	3,7	18,73
MIR_BVE-SVM_PEFF_drv	MIR	BVE	drv	149	radial	0,6	0,99	0,8	0,93	4,4E-01	2,63E-01	1,7	2,99
MIR_BVE-SVM_PEFF_snv	MIR	BVE	snv	266	radial	0,7	0,98	0,9	0,92	8,2E-02	3,52E-01	2,0	14,45
MIR_SPA-SVM_PEFF_mc	MIR	SPA	mc	88	radial	1,0	0,96	1,1	0,87	9,6E-01	1,24E-01	1,2	42,29
MIR_SPA-SVM_PEFF_drv	MIR	SPA	drv	266	radial	0,7	0,98	1,9	0,60	1,6E+00	9,54E+00	9,2	69,55
MIR_SPA-SVM_PEFF_snv	MIR	SPA	snv	266	radial	0,3	1,00	1,2	0,85	3,2E-03	1,76E-01	1,4	45,08
MIR_REP-SVM_PEFF_mc	MIR	REP	mc	266	radial	0,7	0,98	0,8	0,92	4,0E-02	8,75E-01	3,2	17,42
MIR_REP-SVM_PEFF_drv	MIR	REP	drv	263	radial	0,3	1,00	0,9	0,92	7,4E-03	3,54E-02	0,6	9,82
MIR_REP-SVM_PEFF_snv	MIR	REP	snv	266	radial	0,7	0,98	0,9	0,92	8,0E-02	5,08E-01	2,4	17,85
MIR_IPW-SVM_PEFF_mc	MIR	IPW	mc	145	radial	0,8	0,98	0,8	0,93	5,7E-03	3,61E-02	0,6	3,69
MIR_IPW-SVM_PEFF_drv	MIR	IPW	drv	266	radial	0,6	0,99	1,1	0,87	3,2E-01	8,29E-02	0,9	18,90
MIR_IPW-SVM_PEFF_snv	MIR	IPW	snv	266	radial	0,6	0,99	1,0	0,89	3,2E-02	5,88E-01	2,6	23,68



Modelos	Espectros	Intervalos	PPx	VL	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
RAM_PLS_PEFF_mc	RAM	Todo	mc	4	3,0	0,69	3,1	0,66	2,7	0,18	6,7E-31	9,36E-04	6,28E+00	9,0	161,74
RAM_PLS_PEFF_drv	RAM	Todo	drv	3	2,9	0,70	3,1	0,66	2,7	0,22	3,1E-30	9,22E-05	6,68E+00	9,3	142,65
RAM_PLS_PEFF_snv	RAM	Todo	snv	4	2,9	0,69	3,1	0,66	2,7	0,20	9,8E-28	7,57E-07	8,81E+00	10,8	140,31
RAM_LW-PLS_PEFF_mc	RAM	LW	mc	4	3,0	0,69	3,1	0,66	2,7	0,19	2,9E-28	1,88E+01	6,89E+00	9,4	159,87
RAM_LW-PLS_PEFF_drv	RAM	LW	drv	3	2,9	0,70	3,0	0,67	2,7	0,21	6,6E-29	1,47E+02	9,81E+00	11,5	143,97
RAM_LW-PLS_PEFF_snv	RAM	LW	snv	3	3,1	0,66	3,2	0,64	2,7	0,20	4,8E-28	1,03E+01	4,18E-01	2,2	182,21
RAM_RC-PLS_PEFF_mc	RAM	RC	mc	6	2,6	0,76	2,9	0,69	2,7	0,22	1,0E-29	1,86E+01	2,67E+01	20,5	160,75
RAM_RC-PLS_PEFF_drv	RAM	RC	drv	3	2,9	0,70	3,0	0,67	2,7	0,22	1,8E-28	5,00E+01	5,86E+00	8,6	154,18
RAM_RC-PLS_PEFF_snv	RAM	RC	snv	5	2,9	0,71	3,0	0,68	2,7	0,22	5,1E-28	2,56E+00	1,86E+01	16,5	125,29
RAM_VIP-PLS_PEFF_mc	RAM	VIP	mc	6	2,7	0,74	2,9	0,70	2,6	0,25	2,6E-28	8,30E+00	1,63E+01	15,3	142,06
RAM_VIP-PLS_PEFF_drv	RAM	VIP	drv	5	2,2	0,82	2,8	0,72	2,8	0,13	1,5E-28	9,46E+01	1,03E+01	11,8	208,00
RAM_VIP-PLS_PEFF_snv	RAM	VIP	snv	6	2,7	0,74	2,9	0,70	2,6	0,24	3,6E-29	4,75E+00	1,60E+01	15,1	146,04
RAM_iPLS_PEFF_mc	RAM	iPLS	mc	4	2,9	0,71	3,0	0,68	2,7	0,18	5,7E-29	2,66E-01	9,32E+00	11,2	149,33
RAM_iPLS_PEFF_drv	RAM	iPLS	drv	2	2,7	0,74	3,1	0,65	2,8	0,15	1,5E-32	3,36E+01	3,68E+01	24,9	200,50
RAM_iPLS_PEFF_snv	RAM	iPLS	snv	4	2,8	0,71	3,0	0,68	2,7	0,19	1,8E-28	2,07E+01	9,85E+00	11,5	136,29
RAM_GA-PLS_PEFF_mc	RAM	GA	mc	2	3,7	0,51	3,8	0,48	3,1	-0,05	1,3E-28	1,29E+01	4,40E+00	7,4	203,16
RAM_GA-PLS_PEFF_drv	RAM	GA	drv	4	2,4	0,79	2,8	0,71	2,9	0,08	1,4E-30	4,94E+00	2,36E+01	19,0	210,14
RAM_GA-PLS_PEFF_snv	RAM	GA	snv	5	2,7	0,73	2,9	0,70	2,6	0,26	2,6E-29	1,04E+01	5,91E+00	8,7	132,51
RAM_UVE-PLS_PEFF_mc	RAM	UVE	mc	7	2,3	0,80	2,9	0,69	3,0	-0,02	3,1E-29	1,34E+01	3,54E-01	1,9	212,45
RAM_UVE-PLS_PEFF_drv	RAM	UVE	drv	4	3,0	0,69	3,3	0,62	2,7	0,19	7,3E-29	5,87E+00	1,43E+01	14,2	179,57
RAM_UVE-PLS_PEFF_snv	RAM	UVE	snv	5	2,8	0,73	3,1	0,67	2,9	0,10	2,7E-31	9,85E+00	1,69E+00	4,5	202,24

Modelos	Espectros	Intervalos	PPx	VL	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
RAM_BVE-PLS_PEFF_mc	RAM	BVE	mc	6	2,7	0,74	2,9	0,70	2,6	0,25	1,7E-29	6,65E+00	1,37E+01	13,8	132,76
RAM_BVE-PLS_PEFF_dhv	RAM	BVE	dhv	3	3,0	0,69	3,1	0,66	2,7	0,21	8,7E-30	2,93E+02	6,50E+00	9,2	149,01
RAM_BVE-PLS_PEFF_snv	RAM	BVE	snv	5	2,9	0,71	3,0	0,68	2,7	0,21	3,1E-30	4,75E+00	1,58E+01	15,0	126,11
RAM_SPA-PLS_PEFF_mc	RAM	SPA	mc	4	3,0	0,69	3,1	0,66	2,7	0,19	4,1E-28	9,14E+00	6,96E+00	9,5	158,71
RAM_SPA-PLS_PEFF_dhv	RAM	SPA	dhv	3	2,9	0,69	3,1	0,66	2,6	0,23	9,2E-28	1,19E+02	4,61E+00	7,6	152,01
RAM_SPA-PLS_PEFF_snv	RAM	SPA	snv	3	3,1	0,66	3,2	0,64	2,7	0,21	5,4E-31	1,11E+01	1,41E-01	1,3	185,61
RAM_REP-PLS_PEFF_mc	RAM	REP	mc	4	3,0	0,68	3,1	0,66	2,7	0,18	1,7E-29	7,48E+00	6,06E+00	8,8	167,48
RAM_REP-PLS_PEFF_dhv	RAM	REP	dhv	3	3,0	0,69	3,1	0,66	2,7	0,21	6,2E-30	2,30E+02	6,80E+00	9,4	146,55
RAM_REP-PLS_PEFF_snv	RAM	REP	snv	4	3,0	0,69	3,1	0,66	2,7	0,19	1,2E-28	5,37E+00	7,83E+00	10,1	153,29
RAM_IPW-PLS_PEFF_mc	RAM	IPW	mc	4	3,0	0,68	3,1	0,66	2,7	0,18	2,5E-28	4,55E+00	6,31E+00	9,0	165,33
RAM_IPW-PLS_PEFF_dhv	RAM	IPW	dhv	3	2,9	0,69	3,1	0,66	2,7	0,22	1,9E-28	1,15E+01	5,73E+00	8,5	152,98
RAM_IPW-PLS_PEFF_snv	RAM	IPW	snv	4	3,0	0,69	3,1	0,66	2,7	0,19	2,1E-29	5,75E+00	8,19E+00	10,4	150,07
RAM_enPLS_PEFF_mc	RAM	Todo	mc	10	2,1	0,85	3,1	0,65	2,8	0,11	3,0E-02	2,94E+00	2,23E+01	18,4	209,79
RAM_enPLS_PEFF_dhv	RAM	Todo	dhv	9	2,3	0,81	3,4	0,58	3,3	0,17	4,6E+00	4,07E+00	3,51E+01	24,2	215,13
RAM_enPLS_PEFF_snv	RAM	Todo	snv	10	2,3	0,80	3,2	0,63	3,2	0,11	2,0E-01	4,64E+00	1,73E+01	15,9	213,28
RAM_RF_PEFF_mc	RAM	Todo	mc	-	1,1	0,96	2,7	0,75	2,4	0,37	2,5E-04	1,05E+00	1,25E+01	13,2	88,66
RAM_RF_PEFF_dhv	RAM	Todo	dhv	-	1,1	0,96	2,9	0,70	2,5	0,31	1,8E-01	3,01E-01	1,75E+01	16,0	119,53
RAM_RF_PEFF_snv	RAM	Todo	snv	-	1,0	0,96	2,6	0,75	2,1	0,50	7,9E-02	3,76E-01	1,23E+01	13,1	68,26

Modelos	Espectros	Intervalos	PPx	nSV	kernel	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
RAM_SVM_PEFF_mc	RAM	Todo	mc	201	radial	0,5	0,99	2,4	0,36	1,7E-01	1,07E+01	12,0	86,11
RAM_SVM_PEFF_drv	RAM	Todo	drv	265	radial	0,0	1,00	2,4	0,35	3,4E-09	1,42E+01	14,1	98,67
RAM_SVM_PEFF_snv	RAM	Todo	snv	200	radial	0,5	0,99	2,5	0,34	5,6E-02	1,71E+01	15,7	101,57
RAM_LW-SVM_PEFF_mc	RAM	LW	mc	155	radial	1,5	0,92	2,3	0,42	6,1E+00	2,08E+01	17,6	62,95
RAM_LW-SVM_PEFF_drv	RAM	LW	drv	207	radial	0,5	0,99	2,4	0,34	1,3E-03	1,72E+01	15,8	106,13
RAM_LW-SVM_PEFF_snv	RAM	LW	snv	142	radial	1,6	0,91	2,4	0,39	5,6E+00	2,24E+01	18,4	74,12
RAM_RC-SVM_PEFF_mc	RAM	RC	mc	130	radial	2,4	0,80	2,8	0,16	2,3E+01	1,99E+01	17,2	96,53
RAM_RC-SVM_PEFF_drv	RAM	RC	drv	112	radial	1,8	0,88	2,6	0,25	6,5E+00	2,95E+01	21,7	178,46
RAM_RC-SVM_PEFF_snv	RAM	RC	snv	191	radial	2,2	0,83	2,8	0,11	4,7E+00	1,69E+01	15,6	110,16
RAM_VIP-SVM_PEFF_mc	RAM	VIP	mc	82	radial	2,1	0,84	2,7	0,17	2,8E+01	4,65E+01	28,9	146,01
RAM_VIP-SVM_PEFF_drv	RAM	VIP	drv	203	radial	0,7	0,98	2,7	0,17	1,7E+00	3,35E+01	23,5	177,43
RAM_VIP-SVM_PEFF_snv	RAM	VIP	snv	105	radial	2,2	0,83	2,7	0,18	3,7E+01	4,63E+01	28,8	146,67
RAM_iPLS-SVM_PEFF_mc	RAM	iPLS	mc	113	radial	1,7	0,90	2,8	0,16	1,1E+01	2,67E+01	20,5	138,40
RAM_iPLS-SVM_PEFF_drv	RAM	iPLS	drv	204	radial	0,5	0,99	2,6	0,25	2,1E+00	2,06E+01	17,6	174,63
RAM_iPLS-SVM_PEFF_snv	RAM	iPLS	snv	138	radial	1,7	0,89	2,8	0,15	1,3E+01	4,04E+01	26,4	93,90
RAM_GA-SVM_PEFF_mc	RAM	GA	mc	84	radial	2,4	0,80	2,7	0,21	1,9E+01	2,79E+01	21,0	109,54
RAM_GA-SVM_PEFF_drv	RAM	GA	drv	266	radial	0,3	1,00	2,7	0,19	1,6E-01	2,63E+01	20,3	108,32
RAM_GA-SVM_PEFF_snv	RAM	GA	snv	105	radial	2,3	0,81	2,8	0,15	4,2E+01	3,65E+01	24,8	124,92

Modelos	Espectros	Intervalos	PPx	nsV	kernel	RMSEC	$R^2_{cal}$	RMSEP	$R^2_{pred}$	BIAS <sub>cal</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
RAM_UVE-SVM_PEFF_mc	RAM	UVE	mc	107	radial	2,3	0,82	2,9	0,06	3,2E+01	2,81E+01	21,1	116,67
RAM_UVE-SVM_PEFF_drv	RAM	UVE	drv	104	radial	1,6	0,91	2,6	0,24	1,7E-01	9,48E+00	11,3	176,57
RAM_UVE-SVM_PEFF_snv	RAM	UVE	snv	158	radial	1,9	0,87	2,8	0,14	1,5E+01	2,85E+01	21,3	111,62
RAM_BVE-SVM_PEFF_mc	RAM	BVE	mc	109	radial	1,7	0,89	2,5	0,33	7,2E+00	2,78E+01	20,9	111,86
RAM_BVE-SVM_PEFF_drv	RAM	BVE	drv	168	radial	1,0	0,96	2,5	0,31	1,9E+00	2,11E+01	17,8	143,04
RAM_BVE-SVM_PEFF_snv	RAM	BVE	snv	110	radial	1,7	0,89	2,5	0,32	1,3E+01	2,72E+01	20,7	110,11
RAM_SPA-SVM_PEFF_mc	RAM	SPA	mc	91	radial	2,3	0,81	2,7	0,21	2,7E+01	3,46E+01	24,0	114,84
RAM_SPA-SVM_PEFF_drv	RAM	SPA	drv	266	radial	0,0	1,00	2,3	0,40	7,3E-07	1,75E+01	15,9	106,64
RAM_SPA-SVM_PEFF_snv	RAM	SPA	snv	104	radial	2,0	0,86	2,6	0,23	2,7E+01	4,07E+01	26,5	110,73
RAM_REP-SVM_PEFF_mc	RAM	REP	mc	90	radial	2,1	0,84	2,5	0,29	2,1E+01	3,81E+01	25,5	112,75
RAM_REP-SVM_PEFF_drv	RAM	REP	drv	155	radial	1,0	0,97	2,5	0,33	2,0E+00	2,21E+01	18,3	122,98
RAM_REP-SVM_PEFF_snv	RAM	REP	snv	85	radial	2,4	0,80	2,7	0,21	3,8E+01	5,14E+01	30,8	135,07
RAM_IPW-SVM_PEFF_mc	RAM	IPW	mc	149	radial	1,3	0,94	2,5	0,32	5,2E+00	2,25E+01	18,5	105,35
RAM_IPW-SVM_PEFF_drv	RAM	IPW	drv	155	radial	0,9	0,97	2,4	0,36	3,8E+00	1,15E+01	12,5	129,74
RAM_IPW-SVM_PEFF_snv	RAM	IPW	snv	203	radial	0,5	0,99	2,3	0,43	2,1E-01	6,28E+00	9,0	92,62

## ANEXO F - Parâmetros de mérito dos modelos PLS com seleção de variáveis e Florestas Aleatórias (RF) obtidos a partir dos espectros de NIR, mid-IR e RAM para a previsão da viscosidade

Modelos	Espectros	Intervalos	PPx	VL	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
NIR_PLS_VISCO_tss	NIR	Todo	tss	7	0,0384	0,93	0,0431	0,91	0,0244	0,96	1,45E-28	4,52E-04	1,28E-05	0,03	61,13
NIR_PLS_VISCO_drv	NIR	Todo	drv	6	0,0440	0,91	0,0544	0,86	0,0345	0,91	1,48E-28	9,12E-04	2,85E-05	0,05	95,53
NIR_PLS_VISCO_snv	NIR	Todo	snv	7	0,0386	0,93	0,0431	0,91	0,0235	0,96	1,66E-28	9,28E-04	4,89E-05	0,06	38,75
NIR_LW+PLS_VISCO_tss	NIR	LW	tss	7	0,0386	0,93	0,0432	0,91	0,0251	0,95	1,40E-28	5,60E-04	5,61E-07	0,01	73,67
NIR_LW+PLS_VISCO_drv	NIR	LW	drv	6	0,0444	0,91	0,0545	0,86	0,0350	0,91	1,40E-28	8,46E-02	1,25E-04	0,10	123,57
NIR_LW+PLS_VISCO_snv	NIR	LW	snv	5	0,0482	0,89	0,0522	0,87	0,0318	0,93	1,31E-28	1,74E-03	4,28E-05	0,06	77,05
NIR_RC+PLS_VISCO_tss	NIR	RC	tss	5	0,0490	0,88	0,0519	0,87	0,0320	0,93	1,66E-28	5,40E-03	3,39E-04	0,16	102,35
NIR_RC+PLS_VISCO_drv	NIR	RC	drv	7	0,0406	0,92	0,0497	0,88	0,0297	0,94	1,51E-28	9,10E-05	4,38E-03	0,58	100,01
NIR_RC+PLS_VISCO_snv	NIR	RC	snv	6	0,0457	0,90	0,0483	0,89	0,0292	0,94	1,46E-28	4,90E-03	6,33E-04	0,22	94,11
NIR_VIP+PLS_VISCO_tss	NIR	VIP	tss	6	0,0544	0,86	0,0629	0,81	0,0351	0,91	1,67E-28	3,59E-03	9,29E-06	0,03	99,70
NIR_VIP+PLS_VISCO_drv	NIR	VIP	drv	7	0,0472	0,89	0,0541	0,86	0,0392	0,89	1,42E-28	1,60E-03	2,52E-05	0,04	141,89
NIR_VIP+PLS_VISCO_snv	NIR	VIP	snv	7	0,0493	0,88	0,0536	0,86	0,0307	0,93	1,37E-28	4,65E-03	1,78E-05	0,04	100,29
NIR_iPLS_VISCO_tss	NIR	iPLS	tss	7	0,0361	0,94	0,0396	0,92	0,0260	0,95	1,42E-28	1,91E-06	7,93E-04	0,25	94,09
NIR_iPLS_VISCO_drv	NIR	iPLS	drv	8	0,0333	0,95	0,0439	0,91	0,0287	0,94	1,72E-28	1,28E-02	2,31E-03	0,42	124,86
NIR_iPLS_VISCO_snv	NIR	iPLS	snv	6	0,0443	0,91	0,0477	0,89	0,0287	0,94	1,44E-28	1,48E-06	3,18E-04	0,16	85,50
NIR_GA+PLS_VISCO_tss	NIR	GA	tss	6	0,0432	0,91	0,0473	0,89	0,0245	0,96	1,54E-28	6,68E-03	1,12E-03	0,29	63,31
NIR_GA+PLS_VISCO_drv	NIR	GA	drv	5	0,0500	0,88	0,0550	0,85	0,0384	0,89	1,77E-28	1,91E-02	2,31E-04	0,13	137,97
NIR_GA+PLS_VISCO_snv	NIR	GA	snv	6	0,0411	0,92	0,0455	0,90	0,0238	0,96	1,30E-28	8,27E-03	4,34E-06	0,02	53,29
NIR_UVE+PLS_VISCO_tss	NIR	UVE	tss	5	0,0498	0,88	0,0543	0,86	0,0320	0,93	1,39E-28	5,52E-03	1,81E-03	0,37	99,70
NIR_UVE+PLS_VISCO_drv	NIR	UVE	drv	5	0,0530	0,86	0,0581	0,84	0,0403	0,88	1,57E-28	1,78E-02	6,25E-04	0,22	141,89
NIR_UVE+PLS_VISCO_snv	NIR	UVE	snv	7	0,0395	0,92	0,0439	0,91	0,0243	0,96	1,33E-28	9,98E-03	6,85E-06	0,02	100,29

Modelos	Espectros	Intervalos	PPX	VL	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
NIR_BVE-PLS_VISCO_tss	NIR	BVE	tss	6	0,0475	0,89	0,0513	0,87	0,0306	0,93	1,43E-28	1,40E-03	2,64E-04	0,14	86,36
NIR_BVE-PLS_VISCO_drv	NIR	BVE	drv	7	0,0396	0,92	0,0483	0,89	0,0303	0,93	1,46E-28	2,84E-03	2,98E-05	0,05	97,29
NIR_BVE-PLS_VISCO_snv	NIR	BVE	snv	6	0,0435	0,91	0,0478	0,89	0,0269	0,95	1,65E-28	3,06E-03	5,26E-04	0,20	67,75
NIR_SPA-PLS_VISCO_tss	NIR	SPA	tss	7	0,0448	0,90	0,0503	0,88	0,0284	0,94	1,35E-28	3,30E-03	1,29E-03	0,32	102,01
NIR_SPA-PLS_VISCO_drv	NIR	SPA	drv	5	0,0577	0,84	0,0640	0,80	0,0386	0,89	1,75E-28	2,77E-04	2,80E-03	0,47	146,95
NIR_SPA-PLS_VISCO_snv	NIR	SPA	snv	7	0,0425	0,91	0,0480	0,89	0,0261	0,95	1,41E-28	4,87E-03	1,81E-05	0,04	80,41
NIR_REP-PLS_VISCO_tss	NIR	REP	tss	6	0,0477	0,89	0,0524	0,87	0,0301	0,93	1,46E-28	2,30E-03	2,90E-04	0,15	77,11
NIR_REP-PLS_VISCO_drv	NIR	REP	drv	7	0,0391	0,93	0,0478	0,89	0,0295	0,94	1,45E-28	1,82E-02	6,61E-06	0,02	86,29
NIR_REP-PLS_VISCO_snv	NIR	REP	snv	6	0,0444	0,90	0,0486	0,89	0,0284	0,94	1,77E-28	4,23E-03	9,09E-04	0,26	63,99
NIR_IPW-PLS_VISCO_tss	NIR	IPW	tss	5	0,0489	0,88	0,0520	0,87	0,0341	0,92	1,46E-28	5,38E-03	1,56E-03	0,35	117,15
NIR_IPW-PLS_VISCO_drv	NIR	IPW	drv	6	0,0583	0,84	0,0655	0,79	0,0456	0,85	1,59E-28	2,03E-02	4,40E-04	0,18	162,07
NIR_IPW-PLS_VISCO_snv	NIR	IPW	snv	5	0,0476	0,89	0,0504	0,88	0,0341	0,92	1,56E-28	6,32E-03	1,38E-04	0,10	121,48
NIR_enPLS_VISCO_tss	NIR	Todo	tss	10	0,0313	0,95	0,0411	0,92	0,0265	0,95	7,43E-05	3,70E-05	2,51E-03	0,44	104,79
NIR_enPLS_VISCO_drv	NIR	Todo	drv	10	0,0443	0,91	0,0673	0,78	0,0361	0,91	2,54E-05	1,12E-02	2,91E-03	0,47	122,83
NIR_enPLS_VISCO_snv	NIR	Todo	snv	10	0,0349	0,94	0,0448	0,90	0,0268	0,95	2,96E-04	7,94E-05	2,72E-03	0,46	76,43
NIR_RF_VISCO_tss	NIR	Todo	tss	-	0,0192	0,98	0,0514	0,87	0,0277	0,94	4,97E-06	8,50E-04	7,55E-06	0,02	41,05
NIR_RF_VISCO_drv	NIR	Todo	drv	-	0,0204	0,98	0,0570	0,84	0,0301	0,93	5,19E-04	9,15E-04	4,72E-06	0,02	52,22
NIR_RF_VISCO_snv	NIR	Todo	snv	-	0,0166	0,99	0,0441	0,91	0,0277	0,94	8,42E-05	3,02E-04	1,28E-04	0,10	40,12

Modelos	Espectros	Intervalos	PPx	nSV	kernel	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
NIR_SVM_VISCO_tss	NIR	Todo	tss	266	radial	0,0158	0,99	0,0179	0,98	4,50E-05	4,48E-04	0,19	20,48
NIR_SVM_VISCO_drv	NIR	Todo	drv	266	radial	0,0000	1,00	0,0325	0,92	4,52E-10	1,08E-03	0,29	79,03
NIR_SVM_VISCO_snv	NIR	Todo	snv	201	linear	0,0131	0,99	0,0262	0,95	1,77E-04	3,10E-03	0,49	100,99
NIR_LW-SVM_VISCO_tss	NIR	LW	tss	151	radial	0,0180	0,98	0,0204	0,97	5,87E-05	6,44E-04	0,22	30,13
NIR_LW-SVM_VISCO_drv	NIR	LW	drv	266	radial	0,0000	1,00	0,0334	0,92	1,14E-09	2,10E-03	0,40	110,37
NIR_LW-SVM_VISCO_snv	NIR	LW	snv	266	radial	0,0064	1,00	0,0204	0,97	2,85E-05	5,32E-04	0,20	53,54
NIR_RC-SVM_VISCO_tss	NIR	RC	tss	266	radial	0,0232	0,97	0,0212	0,97	4,66E-05	2,49E-04	0,14	20,55
NIR_RC-SVM_VISCO_drv	NIR	RC	drv	266	radial	0,0163	0,99	0,0214	0,97	5,18E-05	2,93E-04	0,15	55,99
NIR_RC-SVM_VISCO_snv	NIR	RC	snv	146	radial	0,0164	0,99	0,0196	0,97	2,01E-05	6,75E-04	0,23	37,59
NIR_VIP-SVM_VISCO_tss	NIR	VIP	tss	148	radial	0,0252	0,97	0,0196	0,97	1,30E-05	1,18E-03	0,30	27,56
NIR_VIP-SVM_VISCO_drv	NIR	VIP	drv	266	radial	0,0230	0,97	0,0212	0,97	1,77E-04	4,37E-04	0,18	54,79
NIR_VIP-SVM_VISCO_snv	NIR	VIP	snv	266	radial	0,0226	0,98	0,0188	0,97	1,74E-04	1,14E-03	0,30	23,97
NIR_iPLS-SVM_VISCO_tss	NIR	iPLS	tss	266	radial	0,0190	0,98	0,0194	0,97	1,78E-06	2,62E-04	0,14	25,17
NIR_iPLS-SVM_VISCO_drv	NIR	iPLS	drv	266	radial	0,0000	1,00	0,0270	0,95	7,62E-10	2,76E-03	0,46	93,16
NIR_iPLS-SVM_VISCO_snv	NIR	iPLS	snv	266	radial	0,0143	0,99	0,0207	0,97	8,17E-05	2,37E-03	0,43	40,91
NIR_GA-SVM_VISCO_tss	NIR	GA	tss	145	radial	0,0165	0,99	0,0177	0,98	5,31E-04	2,58E-04	0,14	37,82
NIR_GA-SVM_VISCO_drv	NIR	GA	drv	266	radial	0,0000	1,00	0,0336	0,92	8,08E-10	2,87E-05	0,05	91,60
NIR_GA-SVM_VISCO_snv	NIR	GA	snv	265	radial	0,0000	1,00	0,0191	0,97	9,45E-10	1,63E-03	0,36	48,79

Modelos	Espectros	Intervalos	PPx	NSV	kernel	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
NIR_UVE-SVM_VISCO_tss	NIR	UVE	tss	265	radial	0,0192	0,98	0,0187	0,97	6,26E-06	4,66E-04	0,19	22,49
NIR_UVE-SVM_VISCO_drv	NIR	UVE	drv	266	radial	0,0000	1,00	0,0262	0,95	4,42E-13	9,23E-05	0,08	61,98
NIR_UVE-SVM_VISCO_snv	NIR	UVE	snv	266	radial	0,0136	0,99	0,0195	0,97	5,16E-05	1,12E-03	0,29	33,80
NIR_BVE-SVM_VISCO_tss	NIR	BVE	tss	266	radial	0,0174	0,99	0,0167	0,98	2,22E-04	5,30E-04	0,20	37,01
NIR_BVE-SVM_VISCO_drv	NIR	BVE	drv	169	radial	0,0200	0,98	0,0217	0,97	3,11E-04	3,81E-04	0,17	51,69
NIR_BVE-SVM_VISCO_snv	NIR	BVE	snv	266	radial	0,0198	0,98	0,0152	0,98	3,21E-04	6,32E-04	0,22	13,12
NIR_SPA-SVM_VISCO_tss	NIR	SPA	tss	151	radial	0,0192	0,98	0,0199	0,97	5,57E-04	5,41E-07	0,01	39,25
NIR_SPA-SVM_VISCO_drv	NIR	SPA	drv	265	radial	0,0000	1,00	0,0346	0,91	5,20E-10	3,00E-04	0,15	132,63
NIR_SPA-SVM_VISCO_snv	NIR	SPA	snv	265	radial	0,0114	0,99	0,0236	0,96	9,75E-05	1,38E-03	0,33	64,98
NIR_REP-SVM_VISCO_tss	NIR	REP	tss	266	radial	0,0170	0,99	0,0160	0,98	2,68E-04	5,11E-04	0,20	25,07
NIR_REP-SVM_VISCO_drv	NIR	REP	drv	170	radial	0,0197	0,98	0,0225	0,96	4,34E-04	2,52E-04	0,14	55,88
NIR_REP-SVM_VISCO_snv	NIR	REP	snv	144	radial	0,0127	0,99	0,0167	0,98	7,21E-05	3,42E-04	0,16	38,70
NIR_IPW-SVM_VISCO_tss	NIR	IPW	tss	266	radial	0,0136	0,99	0,0207	0,97	7,42E-05	3,28E-04	0,16	47,56
NIR_IPW-SVM_VISCO_drv	NIR	IPW	drv	176	radial	0,0126	0,99	0,0288	0,94	1,91E-04	8,29E-04	0,25	75,81
NIR_IPW-SVM_VISCO_snv	NIR	IPW	snv	102	linear	0,0253	0,97	0,0256	0,95	1,39E-04	8,16E-06	0,03	104,21



Modelos	Espectros	Intervalos	PPx	VL	RMSEC	kernel	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
MIR_PLS_VISCO_mc	MIR	Todo	mc	6	0,0358	0,94	0,0380	0,93	0,0336	0,92	1,61E-28	2,75E-06	7,18E-05	0,07	139,11
MIR_PLS_VISCO_drv	MIR	Todo	drv	4	0,0438	0,91	0,0468	0,89	0,0358	0,91	1,79E-28	2,40E-05	1,46E-03	0,34	127,38
MIR_PLS_VISCO_snv	MIR	Todo	snv	4	0,0447	0,90	0,0460	0,90	0,0349	0,91	1,29E-28	9,47E-06	1,10E-03	0,29	107,02
MIR_LW-PLS_VISCO_mc	MIR	LW	mc	4	0,0496	0,88	0,0517	0,87	0,0353	0,91	1,64E-28	1,46E-05	4,83E-04	0,19	74,58
MIR_LW-PLS_VISCO_drv	MIR	LW	drv	4	0,0451	0,90	0,0477	0,89	0,0372	0,90	1,31E-28	2,87E-03	1,90E-03	0,38	135,20
MIR_LW-PLS_VISCO_snv	MIR	LW	snv	4	0,0464	0,90	0,0481	0,89	0,0355	0,91	1,62E-28	1,19E-03	2,46E-03	0,44	115,71
MIR_RC-PLS_VISCO_mc	MIR	RC	mc	4	0,0604	0,82	0,0614	0,82	0,0410	0,88	1,56E-28	4,87E-04	1,69E-03	0,36	118,10
MIR_RC-PLS_VISCO_drv	MIR	RC	drv	4	0,0539	0,86	0,0556	0,85	0,0360	0,91	1,68E-28	4,70E-04	7,43E-03	0,75	120,87
MIR_RC-PLS_VISCO_snv	MIR	RC	snv	3	0,0638	0,80	0,0647	0,80	0,0470	0,84	1,79E-28	7,35E-05	3,79E-04	0,17	149,07
MIR_VIP-PLS_VISCO_mc	MIR	VIP	mc	3	0,0534	0,86	0,0547	0,86	0,0380	0,90	1,58E-28	1,54E-03	1,94E-08	0,00	99,31
MIR_VIP-PLS_VISCO_drv	MIR	VIP	drv	5	0,0393	0,93	0,0419	0,92	0,0332	0,92	1,52E-28	2,91E-03	3,34E-04	0,16	121,21
MIR_VIP-PLS_VISCO_snv	MIR	VIP	snv	3	0,0494	0,88	0,0510	0,87	0,0364	0,90	1,76E-28	3,54E-03	2,78E-04	0,15	71,77
MIR_iPLS_VISCO_mc	MIR	iPLS	mc	4	0,0461	0,90	0,0477	0,89	0,0354	0,91	1,64E-28	1,01E-03	3,60E-03	0,53	95,50
MIR_iPLS_VISCO_drv	MIR	iPLS	drv	4	0,0485	0,89	0,0511	0,87	0,0366	0,90	1,73E-28	1,40E-03	5,82E-03	0,67	125,21
MIR_iPLS_VISCO_snv	MIR	iPLS	snv	4	0,0475	0,89	0,0489	0,88	0,0402	0,88	1,32E-28	1,80E-03	8,89E-03	0,83	143,15
MIR_GA-PLS_VISCO_mc	MIR	GA	mc	4	0,0464	0,90	0,0483	0,89	0,0342	0,92	1,79E-28	3,35E-04	9,97E-04	0,28	71,35
MIR_GA-PLS_VISCO_drv	MIR	GA	drv	4	0,0387	0,93	0,0410	0,92	0,0355	0,91	1,42E-28	1,12E-03	4,08E-05	0,06	134,52
MIR_GA-PLS_VISCO_snv	MIR	GA	snv	4	0,0442	0,91	0,0460	0,90	0,0346	0,91	1,62E-28	1,52E-04	7,05E-04	0,23	102,69
MIR_UVE-PLS_VISCO_mc	MIR	UVE	mc	4	0,0432	0,91	0,0458	0,90	0,0355	0,91	1,53E-28	2,09E-04	3,97E-03	0,56	117,33
MIR_UVE-PLS_VISCO_drv	MIR	UVE	drv	3	0,0494	0,88	0,0518	0,87	0,0389	0,89	1,52E-28	1,82E-04	8,95E-03	0,84	132,58
MIR_UVE-PLS_VISCO_snv	MIR	UVE	snv	3	0,0512	0,87	0,0530	0,86	0,0401	0,88	1,81E-28	2,23E-04	5,19E-03	0,64	120,38

Modelos	Espectros	Intervalos	PPx	VL	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
MIR_BVE-PLS_VISCO_mc	MIR	BVE	mc	3	0,0531	0,86	0,0543	0,86	0,0372	0,90	1,34E-28	1,38E-03	5,59E-08	0,001	82,19
MIR_BVE-PLS_VISCO_drv	MIR	BVE	drv	4	0,0454	0,90	0,0476	0,89	0,0357	0,91	1,50E-28	3,13E-03	3,78E-03	0,54	119,59
MIR_BVE-PLS_VISCO_snv	MIR	BVE	snv	3	0,0490	0,88	0,0503	0,88	0,0356	0,91	1,65E-28	2,33E-03	3,40E-04	0,16	76,29
MIR_SPA-PLS_VISCO_mc	MIR	SPA	mc	4	0,0530	0,86	0,0573	0,84	0,0397	0,89	1,57E-28	3,72E-03	5,23E-04	0,20	128,29
MIR_SPA-PLS_VISCO_drv	MIR	SPA	drv	3	0,0646	0,80	0,0709	0,76	0,0497	0,82	1,42E-28	4,48E-04	4,38E-04	0,18	157,12
MIR_SPA-PLS_VISCO_snv	MIR	SPA	snv	6	0,0453	0,90	0,0490	0,88	0,0345	0,91	1,53E-28	2,72E-03	1,36E-03	0,32	131,77
MIR_REP-PLS_VISCO_mc	MIR	REP	mc	3	0,0532	0,86	0,0544	0,86	0,0372	0,90	1,32E-28	1,35E-03	3,09E-08	0,001	84,23
MIR_REP-PLS_VISCO_drv	MIR	REP	drv	4	0,0455	0,90	0,0476	0,89	0,0357	0,91	1,56E-28	3,53E-03	3,72E-03	0,54	118,27
MIR_REP-PLS_VISCO_snv	MIR	REP	snv	2	0,0553	0,85	0,0587	0,83	0,0372	0,90	1,80E-28	2,42E-03	3,51E-04	0,16	88,91
MIR_IPW-PLS_VISCO_mc	MIR	IPW	mc	3	0,0545	0,86	0,0559	0,85	0,0400	0,89	1,57E-28	6,54E-04	1,89E-03	0,38	98,39
MIR_IPW-PLS_VISCO_drv	MIR	IPW	drv	4	0,0467	0,89	0,0488	0,89	0,0337	0,92	1,30E-28	5,40E-03	7,82E-04	0,25	102,25
MIR_IPW-PLS_VISCO_snv	MIR	IPW	snv	4	0,0446	0,90	0,0458	0,90	0,0353	0,91	1,50E-28	1,62E-03	1,23E-03	0,31	100,75
MIR_enPLS_VISCO_mc	MIR	Todo	mc	10	0,0310	0,95	0,0315	0,95	0,0287	0,94	1,08E-03	2,15E-07	7,89E-04	0,25	81,52
MIR_enPLS_VISCO_drv	MIR	Todo	drv	10	0,0338	0,95	0,0509	0,88	0,0325	0,92	3,15E-04	1,86E-05	7,47E-05	0,08	34,02
MIR_enPLS_VISCO_snv	MIR	Todo	snv	10	0,0275	0,96	0,0375	0,93	0,0310	0,93	3,78E-04	2,67E-05	4,10E-03	0,56	70,57
MIR_RF_VISCO_mc	MIR	Todo	mc	-	0,0115	0,99	0,0298	0,96	0,0251	0,95	4,82E-05	2,18E-05	2,85E-03	0,47	7,38
MIR_RF_VISCO_drv	MIR	Todo	drv	-	0,0129	0,99	0,0353	0,94	0,0204	0,97	6,85E-06	7,95E-04	2,61E-04	0,14	10,04
MIR_RF_VISCO_snv	MIR	Todo	snv	-	0,0119	0,99	0,0303	0,96	0,0184	0,98	8,38E-06	2,68E-04	1,34E-04	0,10	5,51

Modelos	Espectros	Intervalos	PPx	nSV	kernel	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>pred</sub>	0,24	SWTP <sub>val</sub>
MIR_SVM_VISCO_mc	MIR	Todo	mc	266	radial	0,0126	0,99	0,0173	0,98	5,32E-04	7,39E-04	0,04	5,86
MIR_SVM_VISCO_drv	MIR	Todo	drv	128	linear	0,0230	0,97	0,0309	0,93	9,15E-04	2,57E-05	0,23	70,83
MIR_SVM_VISCO_snv	MIR	Todo	snv	266	radial	0,0094	1,00	0,0217	0,97	2,73E-04	6,83E-04	0,23	18,73
MIR_LW-SVM_VISCO_mc	MIR	LW	mc	266	radial	0,0133	0,99	0,0177	0,98	6,13E-04	7,11E-04	0,39	9,96
MIR_LW-SVM_VISCO_drv	MIR	LW	drv	266	radial	0,0097	1,00	0,0310	0,93	3,41E-04	1,99E-03	0,23	44,14
MIR_LW-SVM_VISCO_snv	MIR	LW	snv	266	radial	0,0120	0,99	0,0235	0,96	1,27E-04	6,76E-04	0,11	22,87
MIR_RC-SVM_VISCO_mc	MIR	RC	mc	124	radial	0,0235	0,97	0,0188	0,97	1,11E-04	1,63E-04	0,26	32,16
MIR_RC-SVM_VISCO_drv	MIR	RC	drv	161	radial	0,0331	0,95	0,0266	0,95	1,29E-05	8,97E-04	0,22	65,34
MIR_RC-SVM_VISCO_snv	MIR	RC	snv	266	radial	0,0254	0,97	0,0294	0,94	4,45E-04	6,45E-04	0,16	46,27
MIR_VIP-SVM_VISCO_mc	MIR	VIP	mc	116	radial	0,0220	0,98	0,0180	0,98	4,92E-04	3,19E-04	0,33	22,49
MIR_VIP-SVM_VISCO_drv	MIR	VIP	drv	127	radial	0,0192	0,98	0,0259	0,95	1,62E-04	1,41E-03	0,20	41,10
MIR_VIP-SVM_VISCO_snv	MIR	VIP	snv	266	radial	0,0212	0,98	0,0137	0,99	1,29E-03	5,40E-04	0,15	0,51
MIR_iPLS-SVM_VISCO_mc	MIR	iPLS	mc	130	radial	0,0208	0,98	0,0183	0,98	4,31E-05	2,99E-04	0,07	12,97
MIR_iPLS-SVM_VISCO_drv	MIR	iPLS	drv	85	linear	0,0320	0,95	0,0460	0,85	3,13E-03	7,04E-05	0,34	152,47
MIR_iPLS-SVM_VISCO_snv	MIR	iPLS	snv	266	radial	0,0171	0,99	0,0223	0,96	5,55E-04	1,49E-03	0,26	20,98
MIR_GA-SVM_VISCO_mc	MIR	GA	mc	266	polinomial	0,0185	0,98	0,0235	0,96	4,90E-04	8,61E-04	0,06	28,15
MIR_GA-SVM_VISCO_drv	MIR	GA	drv	86	linear	0,0307	0,95	0,0409	0,88	7,94E-04	4,46E-05	0,24	132,37
MIR_GA-SVM_VISCO_snv	MIR	GA	snv	140	radial	0,0139	0,99	0,0280	0,94	2,78E-05	7,66E-04	0,24	43,24

Modelos	Espectros	Intervalos	PPx	nsV	Kernel	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
MIR_UVE-SVM_VISCO_mc	MIR	UVE	mc	266	radial	0,0138	0,99	0,0167	0,98	8,00E-04	5,30E-04	0,20	5,38
MIR_UVE-SVM_VISCO_drv	MIR	UVE	drv	266	radial	0,0100	1,00	0,0326	0,92	3,46E-04	2,92E-03	0,48	38,63
MIR_UVE-SVM_VISCO_snv	MIR	UVE	snv	266	radial	0,0156	0,99	0,0188	0,97	7,26E-04	1,45E-03	0,33	9,87
MIR_BVE-SVM_VISCO_mc	MIR	BVE	mc	120	radial	0,0200	0,98	0,0183	0,98	2,08E-04	7,42E-04	0,24	13,82
MIR_BVE-SVM_VISCO_drv	MIR	BVE	drv	266	radial	0,0159	0,99	0,0248	0,96	4,38E-04	1,87E-03	0,38	12,80
MIR_BVE-SVM_VISCO_snv	MIR	BVE	snv	266	radial	0,0206	0,98	0,0160	0,98	1,21E-03	1,06E-03	0,29	3,78
MIR_SPA-SVM_VISCO_mc	MIR	SPA	mc	103	radial	0,0279	0,96	0,0295	0,94	2,23E-03	1,63E-05	0,04	62,31
MIR_SPA-SVM_VISCO_drv	MIR	SPA	drv	176	radial	0,0159	0,99	0,0464	0,85	3,74E-04	7,67E-04	0,24	63,83
MIR_SPA-SVM_VISCO_snv	MIR	SPA	snv	159	radial	0,0212	0,98	0,0345	0,91	2,88E-04	5,75E-03	0,67	84,48
MIR_REP-SVM_VISCO_mc	MIR	REP	mc	119	radial	0,0200	0,98	0,0182	0,98	2,15E-04	7,29E-04	0,24	12,85
MIR_REP-SVM_VISCO_drv	MIR	REP	drv	144	polynomial	0,0159	0,99	0,0247	0,96	3,35E-06	1,32E-04	0,10	40,91
MIR_REP-SVM_VISCO_snv	MIR	REP	snv	266	radial	0,0202	0,98	0,0155	0,98	8,07E-04	9,62E-04	0,27	1,83
MIR_IPW-SVM_VISCO_mc	MIR	IPW	mc	266	radial	0,0207	0,98	0,0168	0,98	4,55E-04	4,79E-04	0,19	4,30
MIR_IPW-SVM_VISCO_drv	MIR	IPW	drv	266	radial	0,0114	0,99	0,0361	0,91	5,31E-04	2,34E-03	0,43	43,54
MIR_IPW-SVM_VISCO_snv	MIR	IPW	snv	135	radial	0,0218	0,98	0,0204	0,97	1,41E-04	1,03E-04	0,09	23,21

Modelos	Espetros	Intervalos	PPx	VL	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
RAM_PLS_VISCO_mc	RAM	Todo	mc	7	0,0736	0,74	0,0844	0,66	0,0508	0,81	1,68E-28	7,67E-05	4,72E-03	0,60	172,64
RAM_PLS_VISCO_drv	RAM	Todo	drv	6	0,0423	0,91	0,0923	0,59	0,0710	0,64	1,51E-28	1,42E-03	3,42E-03	0,51	212,50
RAM_PLS_VISCO_snv	RAM	Todo	snv	7	0,0737	0,74	0,0846	0,65	0,0486	0,83	1,55E-28	4,56E-05	2,68E-03	0,45	169,89
RAM_LW-PLS_VISCO_mc	RAM	LW	mc	7	0,0739	0,74	0,0839	0,66	0,0529	0,80	1,59E-28	9,84E-03	4,39E-03	0,58	184,58
RAM_LW-PLS_VISCO_drv	RAM	LW	drv	4	0,0727	0,75	0,0917	0,60	0,0626	0,72	1,58E-28	1,54E-02	1,80E-04	0,12	203,18
RAM_LW-PLS_VISCO_snv	RAM	LW	snv	5	0,0822	0,67	0,0868	0,64	0,0441	0,86	1,54E-28	1,02E-02	7,08E-04	0,23	128,15
RAM_RC-PLS_VISCO_mc	RAM	RC	mc	4	0,0941	0,57	0,0965	0,55	0,0588	0,75	1,78E-28	5,91E-04	5,07E-04	0,20	183,00
RAM_RC-PLS_VISCO_drv	RAM	RC	drv	4	0,0804	0,69	0,0895	0,61	0,0688	0,66	1,43E-28	2,98E-03	1,19E-05	0,03	202,59
RAM_RC-PLS_VISCO_snv	RAM	RC	snv	6	0,0686	0,77	0,0763	0,72	0,0501	0,82	1,52E-28	2,48E-04	1,41E-03	0,33	162,71
RAM_VIP-PLS_VISCO_mc	RAM	VIP	mc	7	0,0798	0,69	0,0862	0,64	0,0487	0,83	1,37E-28	2,14E-03	2,20E-03	0,41	155,36
RAM_VIP-PLS_VISCO_drv	RAM	VIP	drv	3	0,0904	0,61	0,0934	0,58	0,0636	0,71	1,53E-28	1,07E-02	3,01E-03	0,48	178,79
RAM_VIP-PLS_VISCO_snv	RAM	VIP	snv	7	0,0784	0,70	0,0850	0,65	0,0461	0,85	1,41E-28	1,79E-03	2,62E-03	0,45	151,26
RAM_iPLS_VISCO_mc	RAM	iPLS	mc	5	0,0824	0,67	0,0905	0,61	0,0562	0,77	1,62E-28	1,43E-04	4,50E-03	0,59	172,86
RAM_iPLS_VISCO_drv	RAM	iPLS	drv	2	0,0938	0,58	0,1015	0,50	0,0691	0,66	1,53E-28	4,03E-02	3,28E-03	0,50	197,25
RAM_iPLS_VISCO_snv	RAM	iPLS	snv	5	0,0820	0,68	0,0902	0,61	0,0511	0,81	1,67E-28	1,01E-03	2,27E-03	0,42	159,14
RAM_GA-PLS_VISCO_mc	RAM	GA	mc	6	0,0769	0,72	0,0858	0,65	0,0525	0,80	1,53E-28	5,26E-04	7,81E-04	0,24	191,96
RAM_GA-PLS_VISCO_drv	RAM	GA	drv	3	0,0897	0,61	0,0957	0,56	0,0665	0,68	1,69E-28	1,56E-03	8,97E-04	0,26	197,76
RAM_GA-PLS_VISCO_snv	RAM	GA	snv	6	0,0764	0,72	0,0850	0,65	0,0524	0,80	1,42E-28	4,40E-04	5,73E-04	0,21	196,11
RAM_UVE-PLS_VISCO_mc	RAM	UVE	mc	5	0,0809	0,68	0,0891	0,62	0,0521	0,80	1,36E-28	4,22E-05	9,52E-06	0,03	174,24
RAM_UVE-PLS_VISCO_drv	RAM	UVE	drv	2	0,0925	0,59	0,0948	0,57	0,0648	0,70	1,43E-28	1,07E-02	4,44E-03	0,58	167,01
RAM_UVE-PLS_VISCO_snv	RAM	UVE	snv	5	0,0826	0,67	0,0891	0,62	0,0500	0,82	1,45E-28	1,52E-05	7,70E-04	0,24	161,67

Modelos	Espectros	Intervalos	PPx	VL	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSECV	R <sup>2</sup> <sub>cv</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>cv</sub>	BIAS <sub>pred</sub>	RSEP	SWTP <sub>val</sub>
RAM_BVE-PLS_VISCO_mc	RAM	BVE	mc	7	0,0804	0,69	0,0862	0,64	0,0489	0,83	1,33E-28	2,66E-03	2,24E-03	0,41	157,02
RAM_BVE-PLS_VISCO_div	RAM	BVE	div	4	0,0805	0,69	0,0925	0,59	0,0566	0,77	1,64E-28	1,68E-02	1,78E-04	0,12	189,81
RAM_BVE-PLS_VISCO_snv	RAM	BVE	snv	7	0,0794	0,70	0,0852	0,65	0,0460	0,85	1,47E-28	1,16E-03	2,67E-03	0,45	147,58
RAM_SPA-PLS_VISCO_mc	RAM	SPA	mc	7	0,0734	0,74	0,0847	0,65	0,0519	0,81	1,59E-28	2,00E-03	2,49E-03	0,44	179,23
RAM_SPA-PLS_VISCO_div	RAM	SPA	div	3	0,0897	0,61	0,0955	0,56	0,0650	0,70	1,46E-28	2,15E-02	2,44E-03	0,43	194,26
RAM_SPA-PLS_VISCO_snv	RAM	SPA	snv	7	0,0728	0,74	0,0845	0,66	0,0503	0,82	1,43E-28	1,44E-03	1,44E-03	0,33	179,03
RAM_REP-PLS_VISCO_mc	RAM	REP	mc	7	0,0798	0,69	0,0857	0,65	0,0487	0,83	1,36E-28	6,01E-04	2,24E-03	0,41	161,56
RAM_REP-PLS_VISCO_div	RAM	REP	div	4	0,0788	0,70	0,0924	0,59	0,0584	0,76	1,56E-28	3,09E-03	6,86E-06	0,02	194,29
RAM_REP-PLS_VISCO_snv	RAM	REP	snv	7	0,0787	0,70	0,0844	0,66	0,0459	0,85	1,46E-28	1,06E-04	2,67E-03	0,45	150,01
RAM_IPW-PLS_VISCO_mc	RAM	IPW	mc	7	0,0742	0,73	0,0834	0,66	0,0508	0,81	1,54E-28	9,53E-04	4,89E-03	0,61	165,89
RAM_IPW-PLS_VISCO_div	RAM	IPW	div	4	0,0685	0,77	0,0903	0,61	0,0624	0,72	1,53E-28	1,26E-02	1,12E-03	0,29	207,49
RAM_IPW-PLS_VISCO_snv	RAM	IPW	snv	7	0,0745	0,73	0,0834	0,66	0,0489	0,83	1,40E-28	1,22E-03	2,84E-03	0,46	161,39
RAM_enPLS_VISCO_mc	RAM	Todo	mc	10	0,0595	0,83	0,0894	0,61	0,0640	0,71	7,67E-05	1,17E-04	3,96E-03	0,55	212,43
RAM_enPLS_VISCO_div	RAM	Todo	div	10	0,0598	0,83	0,0940	0,57	0,0800	0,54	2,74E-03	2,69E-03	3,53E-04	0,17	215,49
RAM_enPLS_VISCO_snv	RAM	Todo	snv	10	0,0659	0,79	0,0901	0,61	0,0641	0,71	1,19E-02	7,95E-06	8,29E-04	0,25	209,59
RAM_RF_VISCO_mc	RAM	Todo	mc	-	0,0338	0,95	0,0867	0,64	0,0542	0,79	4,14E-06	9,03E-04	4,03E-03	0,55	138,53
RAM_RF_VISCO_div	RAM	Todo	div	-	0,0350	0,94	0,0903	0,61	0,0588	0,75	2,14E-05	9,19E-05	9,42E-04	0,27	158,28
RAM_RF_VISCO_snv	RAM	Todo	snv	-	0,0304	0,96	0,0828	0,67	0,0519	0,81	1,50E-04	6,58E-05	7,72E-03	0,76	85,15

Modelos	Espectros	Intervalos	PPx	nsV	kernel	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>pred</sub>	RSEP	SWTP	val
RAM_SVM_VISCO_mc	RAM	Todo	mc	165	radial	0,0515	0,87	0,0482	0,83	3,43E-04	1,59E-03	0,35	151,15	
RAM_SVM_VISCO_drv	RAM	Todo	drv	266	radial	0,0000	1,00	0,0598	0,74	9,61E-10	3,65E-04	0,17	187,01	
RAM_SVM_VISCO_snv	RAM	Todo	snv	175	polinomial	0,0455	0,90	0,0488	0,83	7,42E-03	8,56E-08	0,00	171,46	
RAM_LW-SVM_VISCO_mc	RAM	LW	mc	124	radial	0,0711	0,76	0,0519	0,81	7,29E-04	1,02E-03	0,28	165,62	
RAM_LW-SVM_VISCO_drv	RAM	LW	drv	177	radial	0,0253	0,97	0,0673	0,67	7,77E-04	1,35E-03	0,32	208,66	
RAM_LW-SVM_VISCO_snv	RAM	LW	snv	134	radial	0,0396	0,92	0,0598	0,74	1,29E-03	1,43E-05	0,03	197,94	
RAM_RC-SVM_VISCO_mc	RAM	RC	mc	131	radial	0,0722	0,75	0,0572	0,77	2,29E-04	2,31E-04	0,13	181,58	
RAM_RC-SVM_VISCO_drv	RAM	RC	drv	234	radial	0,0139	0,99	0,0654	0,69	1,34E-04	7,43E-05	0,08	210,23	
RAM_RC-SVM_VISCO_snv	RAM	RC	snv	116	linear	0,0538	0,86	0,0640	0,71	2,58E-05	1,01E-03	0,28	207,06	
RAM_VIP-SVM_VISCO_mc	RAM	VIP	mc	151	radial	0,0413	0,92	0,0548	0,78	5,00E-04	1,31E-02	0,99	165,77	
RAM_VIP-SVM_VISCO_drv	RAM	VIP	drv	219	radial	0,0161	0,99	0,0546	0,79	5,46E-05	1,87E-03	0,38	188,50	
RAM_VIP-SVM_VISCO_snv	RAM	VIP	snv	129	radial	0,0621	0,81	0,0494	0,82	6,78E-04	9,54E-03	0,85	133,81	
RAM_iPLS-SVM_VISCO_mc	RAM	iPLS	mc	161	radial	0,0499	0,88	0,0539	0,79	4,29E-03	4,43E-03	0,58	181,30	
RAM_iPLS-SVM_VISCO_drv	RAM	iPLS	drv	266	radial	0,0000	1,00	0,0649	0,70	1,24E-09	1,84E-03	0,37	208,99	
RAM_iPLS-SVM_VISCO_snv	RAM	iPLS	snv	203	radial	0,0432	0,91	0,0598	0,74	2,87E-03	4,65E-03	0,59	185,81	
RAM_GA-SVM_VISCO_mc	RAM	GA	mc	162	polinomial	0,0669	0,78	0,0498	0,82	6,31E-03	1,73E-04	0,12	160,11	
RAM_GA-SVM_VISCO_drv	RAM	GA	drv	266	radial	0,0070	1,00	0,0717	0,63	9,58E-05	7,70E-04	0,24	214,44	
RAM_GA-SVM_VISCO_snv	RAM	GA	snv	167	polinomial	0,0558	0,85	0,0495	0,82	2,33E-04	1,12E-04	0,09	175,28	

Modelos	Espectros	Intervalos	PPx	nSV	kernel	RMSEC	R <sup>2</sup> <sub>cal</sub>	RMSEP	R <sup>2</sup> <sub>pred</sub>	BIAS <sub>cal</sub>	BIAS <sub>pred</sub>	RSEP	SWTP val
RAM_UVE-SVM_VISCO_mc	RAM	UVE	mc	136	radial	0,0581	0,84	0,0521	0,80	6,53E-04	1,30E-03	0,32	163,40
RAM_UVE-SVM_VISCO_drv	RAM	UVE	drv	266	radial	0,0000	1,00	0,0633	0,71	2,75E-10	8,24E-04	0,25	193,70
RAM_UVE-SVM_VISCO_snv	RAM	UVE	snv	171	polinomial	0,0627	0,81	0,0532	0,80	2,63E-04	1,08E-03	0,29	178,61
RAM_BVE-SVM_VISCO_mc	RAM	BVE	mc	124	radial	0,0772	0,71	0,0515	0,81	1,61E-03	1,11E-03	0,29	146,06
RAM_BVE-SVM_VISCO_drv	RAM	BVE	drv	148	radial	0,0375	0,93	0,0590	0,75	1,34E-03	1,27E-05	0,03	191,56
RAM_BVE-SVM_VISCO_snv	RAM	BVE	snv	121	polinomial	0,0742	0,73	0,0446	0,86	1,40E-03	3,84E-03	0,54	138,94
RAM_SPA-SVM_VISCO_mc	RAM	SPA	mc	127	radial	0,0689	0,77	0,0516	0,81	1,01E-03	1,62E-03	0,35	184,44
RAM_SPA-SVM_VISCO_drv	RAM	SPA	drv	266	radial	0,0000	1,00	0,0653	0,69	2,75E-10	5,94E-04	0,21	203,33
RAM_SPA-SVM_VISCO_snv	RAM	SPA	snv	158	radial	0,0591	0,83	0,0500	0,82	2,07E-03	9,00E-04	0,26	163,44
RAM_REP-SVM_VISCO_mc	RAM	REP	mc	123	radial	0,0435	0,91	0,0543	0,79	2,49E-03	3,28E-03	0,50	179,60
RAM_REP-SVM_VISCO_drv	RAM	REP	drv	173	radial	0,0261	0,97	0,0631	0,71	1,73E-04	9,53E-05	0,09	204,33
RAM_REP-SVM_VISCO_snv	RAM	REP	snv	190	radial	0,0617	0,82	0,0496	0,82	2,25E-03	3,93E-03	0,55	137,85
RAM_IPW-SVM_VISCO_mc	RAM	IPW	mc	217	radial	0,0134	0,99	0,0587	0,75	7,70E-06	2,25E-03	0,41	192,66
RAM_IPW-SVM_VISCO_drv	RAM	IPW	drv	266	radial	0,0000	1,00	0,0604	0,74	8,85E-10	5,60E-04	0,21	203,12
RAM_IPW-SVM_VISCO_snv	RAM	IPW	snv	209	radial	0,0142	0,99	0,0605	0,74	1,17E-05	8,55E-04	0,26	199,57