

## Artigo

## Método do Enxame de Partículas para Otimização de Modelos de Regressão Multivariada Empregados na Determinação de Biodiesel em Blendas Biodiesel/Óleo Vegetal/Diesel

Rizzi, M.; Furtado, J. C.; da Costa, A. B.; Gerbase, A. E.; Ferrão, M. F.\*

Rev. Virtual Quim., 2016, 8 (6), 1877-1892. Data de publicação na Web: 7 de novembro de 2016

<http://rvq.sbq.org.br>

### Particle Swarm Method for Optimization of Multivariate Regression Models Employes for Biodiesel Determination in Biodiesel/Vegetable Oil/Diesel Blends

**Abstract:** This work compares different multivariate partial least squares regression models for the quantification of biodiesel in binary and ternary blends biodiesel/vegetable oil/diesel using the Fourier transform infrared by attenuated total reflectance (FTIR/ATR). Eighty ternary or binary blends were used for the calibration and validation of models. The particle swarm optimization (PSO) was implemented to select the spectral regions combined with the interval partial least squares regression (iPLS). Three models obtained by regression by partial least squares (PLS) using the completely spectral range ( $4000-650\text{ cm}^{-1}$ ), the spectral regions suggested by ASTM NBR 15568 ( $1750-1735\text{ cm}^{-1}$  and  $1200-1100\text{ cm}^{-1}$ ) and the regions selected by the hybrid model iPLS-PSO were builded. The best results of  $R^2_{cal}$ , RMSEC, RMSECV and RMSEP were obtained for iPLS-PSO, which employs a deterministic approach (iPLS) combined with a heuristic approach (PSO). These results using multivariate regression to quantify the biodiesel content in real samples of commercial diesel (biodiesel blends/diesel), even when these were intentionally adulterated completely or in part with vegetable oil.

**Keywords:** Particle swarm optimization; FTIR/ATR; multivariate calibration; adulteration; biodiesel; interval partial least squares regression (iPLS).

### Resumo

Foram comparados diferentes modelos de regressão multivariada por mínimos quadrados parciais para a quantificação de biodiesel nas blendas binárias e ternárias biodiesel/óleo vegetal/diesel empregando espectrofotometria no infravermelho médio com reflexão total atenuada (FTIR/ATR). Um total de oitenta blendas binárias ou ternárias foram empregadas para a construção e validação dos modelos. O método do enxame de partículas (PSO) foi implementado para a otimização das regiões espectrais juntamente com a regressão por mínimos quadrados parciais por intervalo (iPLS). Dentre os modelos por PLS empregados: toda a faixa espectral  $4000-650\text{ cm}^{-1}$ , as regiões espectrais sugeridas pela norma ASTM NBR 15568 ( $1750-1735\text{ cm}^{-1}$  e  $1200-1100\text{ cm}^{-1}$ ) e as regiões selecionadas pelo modelo híbrido iPLS-PSO, os melhores resultados de  $R^2_{cal}$ , RMSEC, RMSECV e RMSEP foram obtidos para o iPLS-PSO, que emprega uma abordagem determinística (iPLS) combinada a uma abordagem heurística (PSO). Assim é possível utilizar a regressão por PLS para quantificação dos teores de biodiesel em amostras reais de diesel comercial (blendas de biodiesel/diesel), mesmo quando estas foram intencionalmente adulteradas total ou parcialmente com óleo vegetal.

**Palavras-chave:** Enxame de partículas; infravermelho; calibração multivariada; biodiesel; regressão por mínimos quadrados parciais por intervalos (iPLS).

\* Universidade Federal do Rio Grande do Sul, Departamento de Química Inorgânica, Av. Bento Gonçalves, 9500, CEP 91501-970, Porto Alegre-RS, Brasil.

✉ [marco.ferrao@ufrgs.br](mailto:marco.ferrao@ufrgs.br)

DOI: [10.21577/1984-6835.20160127](https://doi.org/10.21577/1984-6835.20160127)

## Método do Enxame de Partículas para Otimização de Modelos de Regressão Multivariada Empregados na Determinação de Biodiesel em Blendas Biodiesel/Óleo Vegetal/Diesel

Marlize Rizzi,<sup>a</sup> João Carlos Furtado,<sup>a</sup> Adilson B. da Costa,<sup>a</sup> Annelise E. Gerbase,<sup>b</sup> Marco F. Ferrão<sup>b,c,\*</sup>

<sup>a</sup> Universidade de Santa Cruz do Sul, Mestrado em Sistemas e Processos Industriais, Av. Independência, 2293, Universitário, CEP 96815-900, Santa Cruz do Sul-RS, Brasil.

<sup>b</sup> Universidade Federal do Rio Grande do Sul, Departamento de Química Inorgânica, Av. Bento Gonçalves, 9500, CEP 91501-970, Porto Alegre-RS, Brasil.

<sup>c</sup> Instituto Nacional de Ciência e Tecnologia de Bioanalítica (INCT -Bioanalítica), CP 6154, CEP 13084-971, Campinas-SP, Brasil.

\* [marco.ferrao@ufrgs.br](mailto:marco.ferrao@ufrgs.br)

*Recebido em 6 de março de 2016. Aceito para publicação em 26 de outubro de 2016*

### 1. Introdução

1.1. O enxame de partículas na otimização de modelos de regressão multivariados

### 2. Materiais e Métodos

2.1. Amostragem

2.2. Obtenção dos espectros no infravermelho

2.3. Construção dos modelos multivariados

### 3. Resultados e Discussão

3.1. Resultados do modelo PLS global

3.2. Resultados do modelo PLS empregando as regiões descritas na ABNT

3.3. Resultados do modelo híbrido iPLS-PSO

### 4. Conclusões

## 1. Introdução

A maior parte da energia consumida mundialmente provém do petróleo, do carvão e do gás natural, fontes estas limitadas e com previsão de se esgotar no futuro; portando a busca por fontes

alternativas de produção de energia é de suma importância. Neste contexto surgem inúmeros estudos, que envolvem a produção de combustíveis alternativos ao óleo diesel, provenientes de óleos vegetais, denominados biocombustíveis.<sup>1</sup> Os primeiros estudos envolvendo o uso de óleos vegetais como combustíveis foram realizados na época do

desenvolvimento dos motores a diesel, porém neste período a questão ambiental não era considerada importante. Atualmente, quando as atenções se voltam aos problemas ambientais e à redução da emissão dos gases causadores do efeito estufa, a busca por fontes de produção de combustíveis “limpos” tem sido constante no mundo.

No Brasil, a lei nº 11.097, de 13 de janeiro de 2005, determinou percentuais de mistura do biodiesel ao óleo diesel, bem como o monitoramento da inserção deste novo combustível ao mercado pela Agência Nacional do Petróleo, Gás Natural e Biocombustíveis – ANP. A adição obrigatória iniciou com um percentual de 2% de biodiesel no diesel sendo denominada de mistura B2.<sup>2</sup> Em 2014 o percentual foi ampliado para 6% (B6) para os combustíveis comercializados a partir de 1 de julho e para 7% (B7) para os comercializados a partir de 1 de novembro, segundo a lei nº 13.033, de 24 de setembro de 2014.<sup>3</sup> Mais recentemente, a lei 13.263 de 23 de março de 2016 amplia gradativamente o aumento do percentual para 8% (em até 12 meses), 9% (em até 24 meses) e 10% (em até 36 meses) de biodiesel a ser adicionado no diesel.<sup>4</sup>

Segundo a lei nº 11.097, de 13 de janeiro de 2005, o biodiesel é um biocombustível derivado de biomassa renovável para uso em motores a combustão interna com ignição por compressão ou, conforme regulamento, para geração de outro tipo de energia, que possa substituir parcial ou totalmente combustíveis de origem fóssil.<sup>2</sup> O biodiesel pode ser obtido a partir de óleos vegetais ou gorduras animais, novos ou residuais, utilizando processos tecnológicos para conversão de óleos em combustíveis.

No controle de qualidade do biodiesel, diversas metodologia analíticas são adotadas tentando estabelecer suas características.<sup>5</sup> Segundo a resolução nº 45 da ANP de 25 de agosto de 2014,<sup>6</sup> dentre as principais metodologias que vem sendo utilizadas para a quantificação de derivados, subprodutos e contaminações, destacam-se as descritas nas normas brasileiras que empregam cromatografia gasosa (GC),<sup>7-9</sup> porém estas

metodologias são laboriosas e consomem insumos químicos.

Desde a regulamentação da adição do biodiesel ao diesel comercializado, o emprego da espectroscopia no infravermelho com calibração univariada já vem sendo utilizada com base na norma EN 14078 (British Standards Institution)<sup>10</sup> para a quantificação do teor do biodiesel na blenda biodiesel/diesel. Entretanto esta norma não contempla a mais comum adulteração das blends, que se trata da substituição total ou parcial do biodiesel por óleo vegetal ou gordura animal que não foi transesterificada, visto que tanto a mistura de metil ésteres (biodiesel) como as matérias primas empregadas na produção do biodiesel (óleo vegetal e gordura animal) apresentam a mesma banda da carbonila empregada na quantificação. Em virtude disso, metodologias multivariadas empregando o infravermelho podem ser empregadas para quantificar o biodiesel, mesmo na presença de adulterantes.<sup>11</sup> Além disso, no desenvolvimento de metodologias multivariadas empregando dados espectrais, a utilização de todo o espectro nem sempre produz os melhores modelos quantitativos, sendo comum a utilização de algoritmos de otimização para a seleção da(s) faixa(s) espectral(is) mais correlacionada(s) com a propriedade ou analito que se deseja determinar.<sup>12</sup>

No presente trabalho propõem-se a utilização do método do enxame de partículas para otimização de modelos de regressão multivariada por regressão por mínimos quadrados parciais (PLS), empregando dados de ATR-FTIR, visando a determinação de biodiesel em blends biodiesel/óleo vegetal/diesel.

### 1.1. O enxame de partículas na otimização de modelos de regressão multivariados

A técnica de otimização por enxame de partículas (PSO – do inglês *Particle Swarm*

*Optimization*), foi criada por Kennedy e Eberhart,<sup>13</sup> a partir da análise comportamental dos pássaros a procura de alimento ou de um local para construção do ninho. Na procura pelo ninho ou alimento, todo o indivíduo (partícula) pode lucrar com as experiências dos membros do grupo (enxame). Kennedy e Eberhart propuseram um algoritmo de otimização robusto, simples e de fácil implementação computacional.

A PSO é uma técnica de computação evolutiva, motivada pela simulação do “comportamento social”, que trabalha em uma população potencial de soluções em um espaço de busca, onde, através da cooperação e da competição entre estas soluções, é possível encontrar uma solução ótima, quando aplicadas em problemas de otimização.<sup>14</sup>

Dentre as aplicações recentes, diferentes métodos de seleção de amostras e algoritmos de seleção de variáveis foram comparados para o desenvolvimento de modelos de transferência de calibração empregando dados de infravermelho. Para tanto, métodos baseados em um índice de similaridade considerando distância euclidiana entre os coeficientes de Fourier foram introduzidos e testados contra abordagens mais comuns, bem como algoritmos genéticos e diferentes modelos baseados na otimização por enxame de partículas (PSO) foram implementados e comparados.<sup>15</sup>

Vários algoritmos de seleção de variáveis, tanto determinísticos quanto heurísticos, têm sido aplicados a fim de selecionar os comprimentos de onda mais informativos e, assim, otimizar modelos de regressão por mínimos quadrados parciais (PLS) a partir de dados de espectrometria molecular. Dentre estes métodos encontramos aqueles que se baseiam nos coeficientes de regressão, como o por projeções das variáveis importantes (VIP, do inglês *variable importance in projection*) e o por eliminação das variáveis não informativas (UVE, do inglês *uninformative variable elimination*). Destacam-se também aqueles que otimizam

uma função custo, geralmente o erro de validação, como a regressão por mínimos quadrados parciais por intervalo (iPLS, do inglês *interval partial least squares*), por janelas móveis e de tamanho variado (VS-MW, do inglês *variable-size moving-window*), empregando algoritmos genéticos (GA, do inglês *genetic algorithm*) e otimização por enxame de partículas (PSO).<sup>16</sup>

Xue *et al.* (2012) também apresentam o PSO para a otimização de modelos multivariados empregando dados de espectrometria no visível e infravermelho próximo para a implementação de métodos rápidos e não destrutivos para a detecção de resíduos de pesticidas em frutas. A seleção dos comprimentos de onda, através do algoritmo PSO, apresentou melhor capacidade preditiva na regressão por PLS comparados ao modelo que empregou toda a região espectral analisada.<sup>17</sup>

Em cenário semelhante, porém buscando o desenvolvimento de um método não-destrutivo para a determinação quantitativa e direta de Enterobacteriaceae em filés de frango através de imagens hiperespectrais, Feng *et al.* (2013) empregaram com sucesso o PSO para otimizar modelos de regressão PLS.<sup>18</sup>

O PSO também têm sido utilizado para a otimização das informações espectroscópicas necessárias para métodos de resolução multivariada de curvas com mínimos quadrados alternantes (MCR-ALS), tanto para análises por cromatografia gasosa com detecção por espectrometria de massas (GC-MS), como para cromatografia líquida de alto desempenho com detecção por arranjo de diodos (HPLC-DAD).<sup>19</sup>

Além das aplicações descritas anteriormente, o método de otimização por enxame de partículas (PSO) foi utilizado por Jin *et al.* (2013) na seleção de comprimentos de onda de sensores NIR no desenvolvimento de modelos de regressão multivariada por mínimos quadrados parciais (PLS) e por máquinas de vetor de suporte com mínimos quadrados (LS-SVM) para o monitoramento

de processos.<sup>20</sup> E mais recentemente por Lou et al. (2014) para selecionar os hiperparâmetros, como a regularização e os parâmetros do *kernel*, de modelos de calibração multivariados por máquina de vetor de suporte com mínimos quadrados (LS-SVM), também empregando dados no infravermelho próximo (NIR).<sup>21</sup>

#### O algoritmo básico do PSO

O algoritmo do PSO é baseado em um espaço de busca, onde é inicialmente gerada uma população aleatória de partículas e a estas associadas uma velocidade, ou seja, cada partícula possui uma posição ( $x_i$ ) e uma velocidade ( $v_i$ ). A cada iteração, cada partícula atualiza sua velocidade, pela

$$x_i^{it+1} = x_i^{it} + v_i^{it+1} \quad (\text{Eq.1})$$

$$v_i^{it+1} = v_i^{it} + (c_1^{it} * rand()^{it} * (pbest_i^{it} - x_i^{it})) + (c_2^{it} * rand()^{it} * Gbest^{it} - x_i^{it}) \quad (\text{Eq.2})$$

As partículas enquanto percorrem seu espaço de busca geram informações que são armazenadas em vetores, onde:

$X_i = (x_1, x_2, x_3, \dots, x_n)$  corresponde às posições das partículas;

$V_i = (v_1, v_2, v_3, \dots, v_n)$  indica as velocidades das partículas;

$$Pbest_i = (pbest_1, pbest_2, pbest_3, \dots, pbest_n)$$

corresponde a melhor posição obtida de cada partícula. E  $n$  indica o tamanho do enxame.

Armazena-se também, durante o processo de exploração, a melhor posição (*fitness*) já encontrada entre todas as partículas do enxame, que é representado pelo *Gbest*.

O algoritmo de PSO apresenta um conceito simples, de fácil implementação e de eficiência computacional.

Equação 2, e uma nova posição é gerada pela soma de sua posição atual e a nova velocidade, conforme a Equação 1.<sup>22</sup>

A Equação 2, que calcula a velocidade da partícula, é composta também pela função  $rand()^{it}$  que gera números aleatórios entre 0 e 1, pelos parâmetros de confiança  $c_1^{it}$  e  $c_2^{it}$  que indicam quanto a partícula confia em sua experiência ( $c_1^{it}$ ) e na do enxame ( $c_2^{it}$ ); e pelos termos de aceleração: ( $pbest_i^{it} - x_i^{it}$ ) que representa a distância entre a melhor posição encontrada para a partícula  $i$  e a posição atual dessa mesma partícula, e ( $Gbest^{it} - x_i^{it}$ ), que representa a distância entre a melhor posição de todas as partículas do enxame e a posição da partícula  $i$ .

O pseudocódigo do PSO pode ser escrito como:<sup>22</sup>

Criar uma população aleatória de partículas, denominada enxame;

Inicializar esta população com posições e velocidades aleatórias;

Repetir para cada partícula  $i$

Calcular a função objetivo (*fitness*);

Comparar o novo valor resultante da *fitness* da partícula  $i$ . Se o valor for melhor que o valor do *pbest* da partícula  $i$ , atualizar o *pbest* com o novo valor da *fitness*.

Comparar o novo valor resultante da *fitness* da partícula  $i$  com o valor do *Gbest*. Se o valor da *fitness* da partícula  $i$  for melhor que o valor do *Gbest*, atualizar o valor do *Gbest* com este novo valor;

Alterar a posição e a velocidade da partícula  $i$ ;

Até que a condição de parada seja satisfeita.

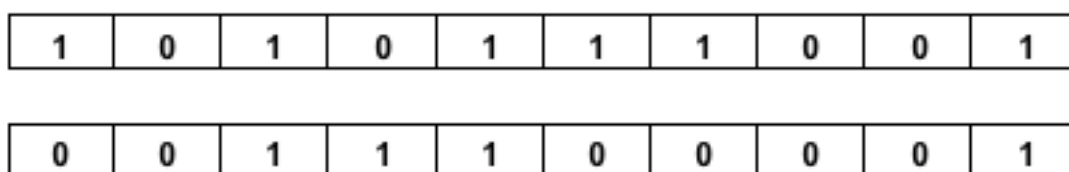
*Considerações gerais sobre as implementações da meta-heurística enxame de partículas para seleção variáveis de espectros no infravermelho*

O princípio básico do algoritmo enxame de partícula é que as partículas se movimentam no espaço de busca,

objetivando encontrar uma solução, orientadas pela sua experiência e pela experiência do enxame.

Como o propósito desta pesquisa é otimizar modelos de regressão multivariados, abaixo são explicados alguns conceitos do enxame de partículas neste contexto e que se mantêm para todas as implementações propostas.

**Partícula:** é uma possível solução para o problema. A posição da partícula é representada pelo conjunto de comprimentos de onda selecionados, ou seja, a configuração da mesma. A Figura 1 apresenta exemplo de duas partículas.



**Figura 1.** Representação gráfica de duas partículas

**Enxame:** um conjunto de soluções (partículas).

**Unidade de experiência:** corresponde a um intervalo do espectro, podendo este ser ou não utilizado na construção do modelo de regressão.

**Fitness:** é a qualidade da posição (configuração) da partícula, obtido pelo erro quadrático médio de validação cruzada (RMSECV) gerado no modelo de regressão conforme Equação 3:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (\text{Eq.3})$$

onde  $n$  é o número de espectros médios,  $y_i$  e  $\hat{y}_i$  são os valores de % utilizados na formulação das blendas e preditos pelo modelo do PLS, respectivamente, utilizando as amostras do conjunto de calibração, com as quais é realizada a validação cruzada.

**Tamanho do enxame:** corresponde ao número total de partículas utilizadas.

**Velocidade:** é a magnitude em que a solução corrente é alterada nas “direções” do  $G_{best}$  e do  $p_{best}_i$ .

**Fator de Inércia:** a intensidade em que a solução corrente é alterada (o percentual transferido do *Gbest* e do *pbest<sub>i</sub>*).

*Estrutura básica do enxame de partícula implementado para a otimização dos modelos de regressão*

O enxame inicia sua execução criando as partículas, onde cada partícula é composta de um número pré-definido de intervalos, codificadas com 0 e 1 aleatoriamente e, assim, cada partícula possui uma posição (solução inicial) e uma velocidade inicial. Para este problema, cada um dos intervalos da

partícula corresponde a uma região do espectro, e os modelos serão construídos somente com os intervalos codificados com 1, conforme Figura 2.

Quando o enxame é criado a posição inicial de cada partícula *i* é considerada como a sua melhor posição encontrada, ou seja o *pbest<sub>i</sub>*, e a melhor posição entre todas as partículas, passa a ser considerada o *Gbest* do enxame.

A cada iteração, cada partícula *i*, se desloca para uma outra posição, com base na soma da sua posição atual e nova velocidade gerada. Nestas implementações a velocidade será calculada através da Equação 4:

$$v_i^{it+1} = w * v_i^{it} + (c_1^{it} * rand()^{it} * (pbest_i^{it} - x_i^{it})) + (c_2^{it} * rand()^{it} * (Gbest^{it} - x_i^{it})) \quad (Eq.4)$$

Assim, com base no deslocamento da partícula *i* é gerado um vetor somente com os intervalos selecionados, e este é passado como parâmetro para a função *plsmodel* do pacote *iToolBox*,<sup>23</sup> juntamente com um modelo iPLS, com o mesmo número de intervalos da partícula *i* que está sendo avaliada, que gera um modelo de regressão empregando apenas estes intervalos. Este modelo gera um vetor de erros (RMSECV), onde cada valor deste vetor corresponde a utilização de um número diferente de variáveis latentes. O resultado do deslocamento da partícula (*fitness*) será o menor RMSECV, juntamente com o respectivo número de variáveis latentes.

Esta nova posição (valor resultante da *fitness*) da partícula *i* é comparada com a melhor posição que esta partícula *i* obteve até o presente momento (*pbest<sub>i</sub>*). Da mesma maneira compara-se a *fitness* da partícula *i* com o valor do *Gbest* do enxame. Neste trabalho em que se pretende minimizar o erro RMSECV, ambos, *pbest<sub>i</sub>* e *Gbest* somente serão atualizados se o valor obtido pela *fitness* da partícula *i* tiver sido menor que a última melhor solução armazenada por esta partícula e pelo enxame, respectivamente.

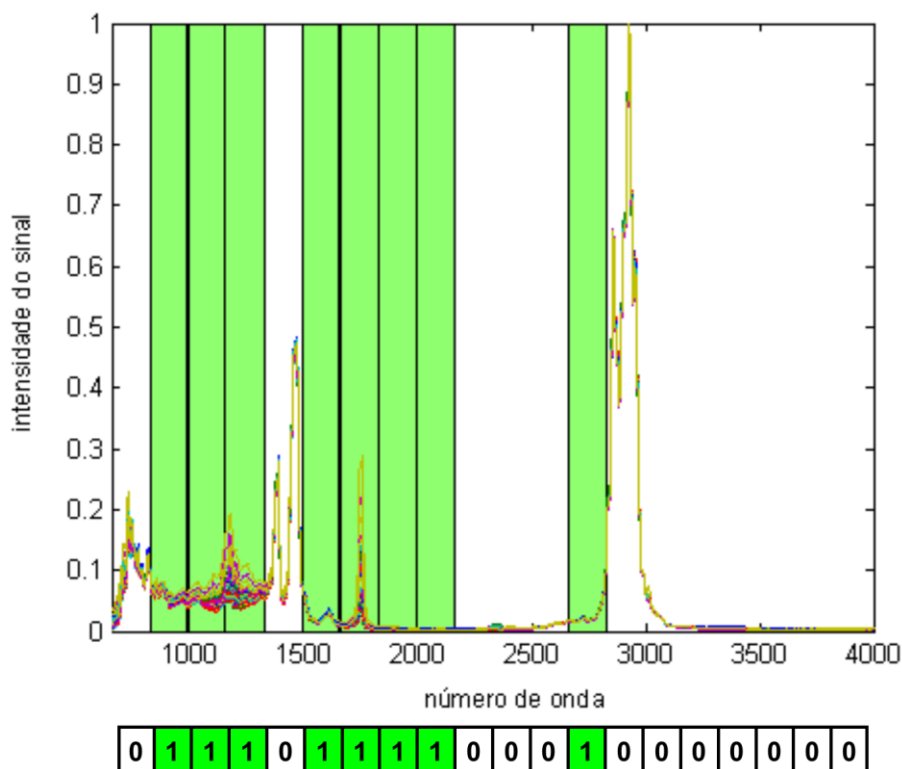


Figura 2. Exemplo de uma partícula e a representação dos seus intervalos selecionados

## 2. Materiais e Métodos

### 2. Materiais e Métodos

#### 2.1. Amostragem

Para a formulação das misturas foram combinados planejamentos experimentais que resultaram nas composições apresentadas na Figura 3. Um total de 80 blendas foi formulado em balão volumétrico de 25 mL, sendo que 40 misturas binárias e ternárias foram obtidas a partir de uma amostra de óleo diesel, de uma amostra de biodiesel de palma (doador pela Agropalma S/A), uma amostra de biodiesel de soja e uma amostra de óleo de soja (ambas fornecidas pela OleoPlan S.A). As 40 misturas binárias e ternárias restantes foram obtidas a partir de uma segunda amostra de óleo diesel, da

mesma amostra de biodiesel de soja e da mesma amostra de óleo de soja (ambas fornecida pela OleoPlan S.A). Cabe ressaltar que na apresentação do diagrama os dois diferentes óleos diesel, bem como os dois diferentes biodieseis utilizados, foram considerados como um único.

#### 2.2. Obtenção dos espectros no infravermelho

As amostras foram homogeneizadas antes da aquisição dos espectros, sendo os mesmos obtidos em duplicata utilizando um acessório de reflexão total atenuada (ATR) com cristal de ZnSe, acoplado a um espectrofotômetro Nicolet Magna 550, na faixa de 4000-600  $\text{cm}^{-1}$ , conforme parâmetros descritos na Tabela 1. Antes de ser obtido o espectro de cada amostra foi realizado o espectro do branco, sem nenhum composto sobre o cristal.



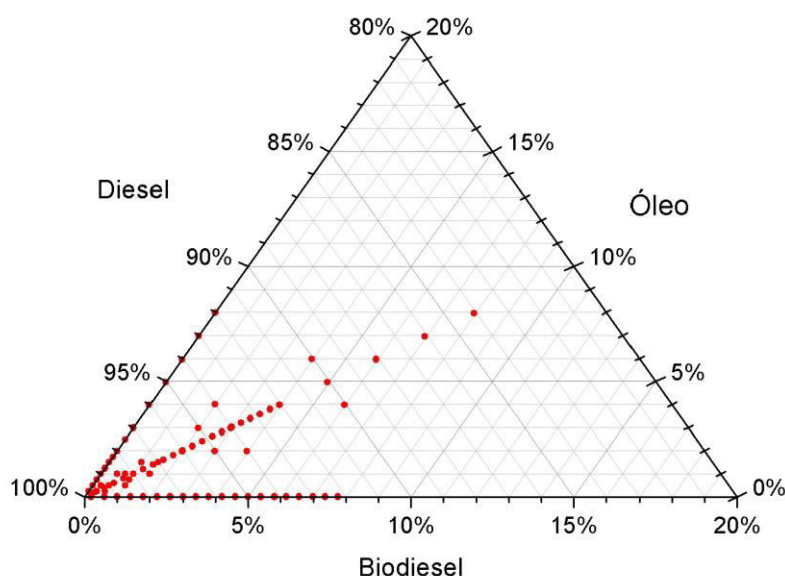


Figura 3. Diagrama ternário para as misturas biodiesel/óleo vegetal/diesel

Tabela 1. Parâmetros instrumentais para a aquisição dos espectros por FT-IR/ATR

Parâmetros	Condições Experimentais
Região espectral	4.000-650 $\text{cm}^{-1}$
Detector	DTGS
Divisor de feixe	KBr
Varreduras	32
Resolução	4 $\text{cm}^{-1}$
Acessório	Reflexão Total Atenuada - PIKE

### 2.3. Construção dos modelos multivariados

A seleção das amostras para a constituição dos conjuntos de calibração e previsão teve dois objetivos principais: construir conjuntos que contenham a informação da matriz de dados utilizados neste estudo e que contenha a maior variabilidade possível em relação à faixa de concentração. Os 80 espectros médios das amostras foram divididos em dois conjuntos, sendo 48 empregados na construção do

conjunto de calibração e 32 no conjunto de previsão.

Todos os modelos construídos neste trabalho, utilizaram o MATLAB<sup>®</sup> versão 6.5 (*The Math Works*, <http://www.mathworks.com>, EUA), juntamente com o pacote *iToolbox* versão 2 desenvolvido por Lars Norgaard (*iToolbox for MATLAB*, <http://www.models.kvl.dk>, Dinamarca).<sup>23</sup>

Nesta implementação, manteve-se a estrutura básica do exame de partículas implementado, mas a intensidade em que a partícula  $i$  é alterada (o percentual

transferido do  $G_{best}$  e do  $p_{best_i}$ ) foi definida e mantida constantemente durante todas as iterações. Os percentuais utilizados no presente trabalho foram 10% do  $G_{best}$  e 15 % para o  $p_{best_i}$ .

### 3. Resultados e Discussão

O algoritmo de enxame de partículas implementado busca combinar várias unidades de informação (regiões do espectro), para construção dos modelos de regressão. Neste trabalho foram utilizados enxames com tamanho de 25, 50 e 100 partículas, podendo cada partícula ser constituída por 20, 40 ou 60 intervalos totalizando 9 combinações, sendo que cada combinação foi executada em triplicata. Os resultados obtidos destas soluções foram comparados com os modelos obtidos

utilizando o espectro inteiro das misturas e as sub-regiões indicadas pela norma ABNT (NBR 15568).<sup>24</sup>

Visando quantificar o biodiesel nas misturas, foram desenvolvidos diversos modelos de regressão, e os resultados obtidos estão descritos a seguir. Primeiramente são apresentados os modelos usando o PLS e, em seguida, são descritos os modelos obtidos com a otimização através da meta-heurística enxame de partículas combinada ao iPLS.

#### 3.1. Resultados do modelo PLS global

O primeiro modelo, para a quantificação do biodiesel, foi construído utilizando toda região espectral ( $4000-650\text{ cm}^{-1}$ ), e o desempenho do modelo empregando 6 variáveis latentes é apresentado na Figura 4.

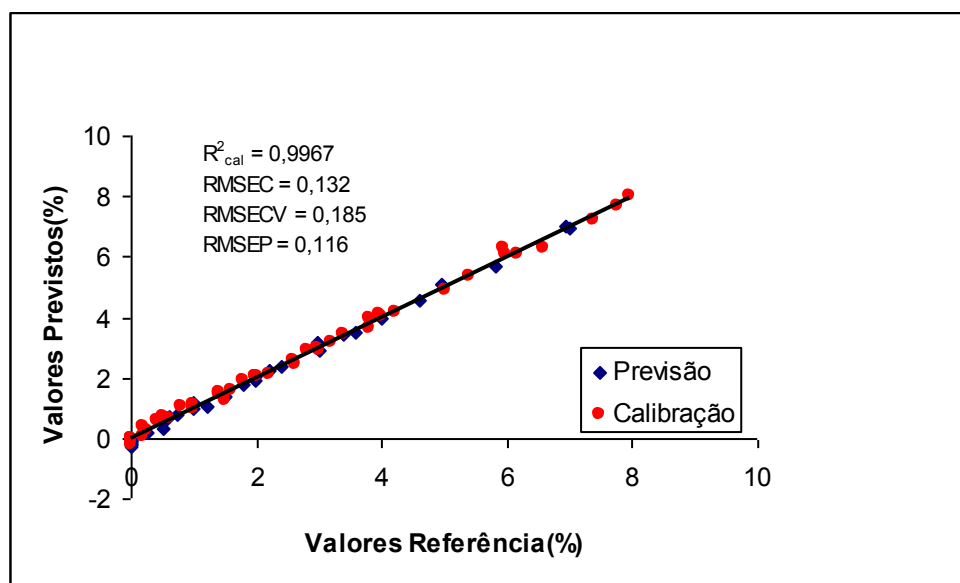


Figura 4. Valores de referência versus previstos para o modelo de previsão do biodiesel utilizando toda a região espectral das misturas

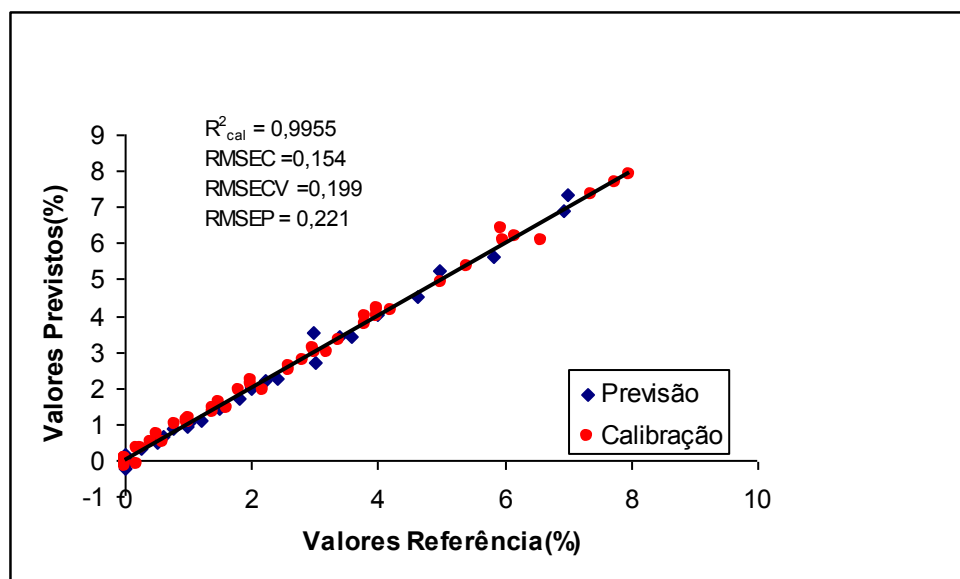
#### 3.2. Resultados do modelo PLS empregando as regiões descritas na ABNT

O segundo modelo construído para a quantificação do biodiesel, empregou somente as duas regiões espectrais (de  $1750$  a  $1735\text{ cm}^{-1}$  e de  $1200$  a  $1100\text{ cm}^{-1}$ ) sugeridas pela proposta de norma ABNT NBR 15568,

sendo o desempenho do modelo empregando 5 variáveis latentes apresentado na Figura 5.

Podemos observar que o modelo gerado utilizando o espectro inteiro apresentou uma melhor capacidade preditiva (RMSEP=0,116%), em relação ao modelo que levou em consideração somente às regiões

sugeridas pela proposta de norma ABNT (RMSEP=0,221%). Um aspecto que justifica este resultado pode ser a presença dos sinais interferentes do óleo nas mesmas regiões espectrais que o biodiesel, além de que outras regiões do espectro podem conter informações importantes para construção de um modelo mais robusto.



**Figura 5.** Valores de referência versus previstos para o modelo de previsão do biodiesel, utilizando somente as regiões espectrais sugeridas pela proposta de norma ABNT

### 3.3. Resultados do modelo híbrido iPLS-PSO

Na busca de gerar modelos com maior capacidade preditiva para a quantificação do biodiesel nas misturas, foi implementado um algoritmo híbrido para a seleção das variáveis (números de onda) para construção dos modelos de regressão multivariados de espectros no infravermelho, que consiste na utilização do algoritmo iPLS,<sup>25-27</sup> otimizado pela meta-heurística chamada enxame de partículas (PSO).

Na Figura 6 é apresentado um exemplo da evolução das partículas, mostrando as mesmas nas suas posições iniciais, bem como após 10 iterações, as posições das mesmas na 100ª iteração e, finalmente, quando atingem o número máximo de iterações

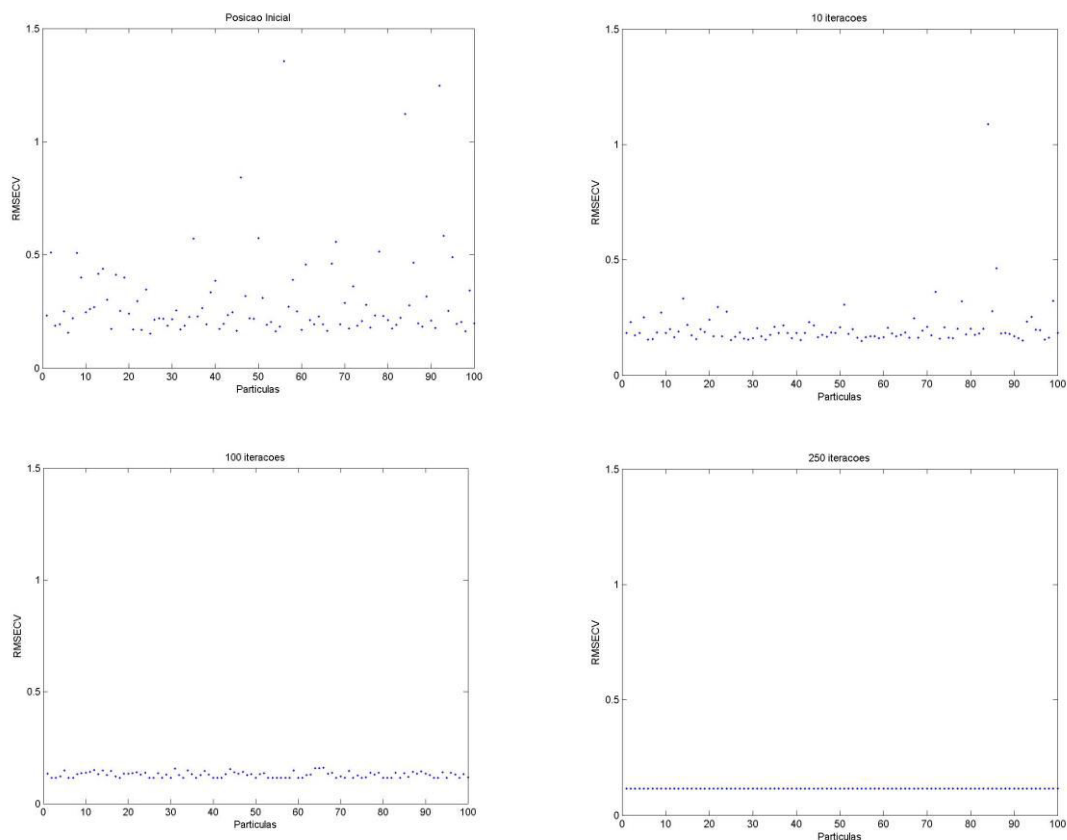
definido para o problema. Pode-se observar que com o passar das iterações as partículas convergem para uma mesma solução, isto é, mesmos valores de erro quadrático médio de validação cruzada (RMSECV).

Na Figura 7 é apresentado um comparativo das evoluções para as 3 repetições para a combinação do iPLS-PSO com 60 intervalos e 100 partículas. Observa-se nesta proposta que existe uma tendência de convergência de todos os modelos, sendo necessário em média 230 iterações. Apesar de não ser atingindo um mesmo mínimo para todas as repetições, em virtude do tempo computacional gasto (ou custo computacional – que representa o tempo necessário para serem realizados um número pré-determinado de ciclos), foi definido o valor de 250 iterações para a parada do algoritmo. Experimentos com valores

maiores que 250 foram conduzidos, porém não foi observado melhores desempenhos para estas condições que justificassem sua implementação.

Dentre as 9 combinações experimentadas com diferentes números de partículas e de intervalos, o melhor modelo iPLS-PSO para a quantificação do biodiesel foi aquele que

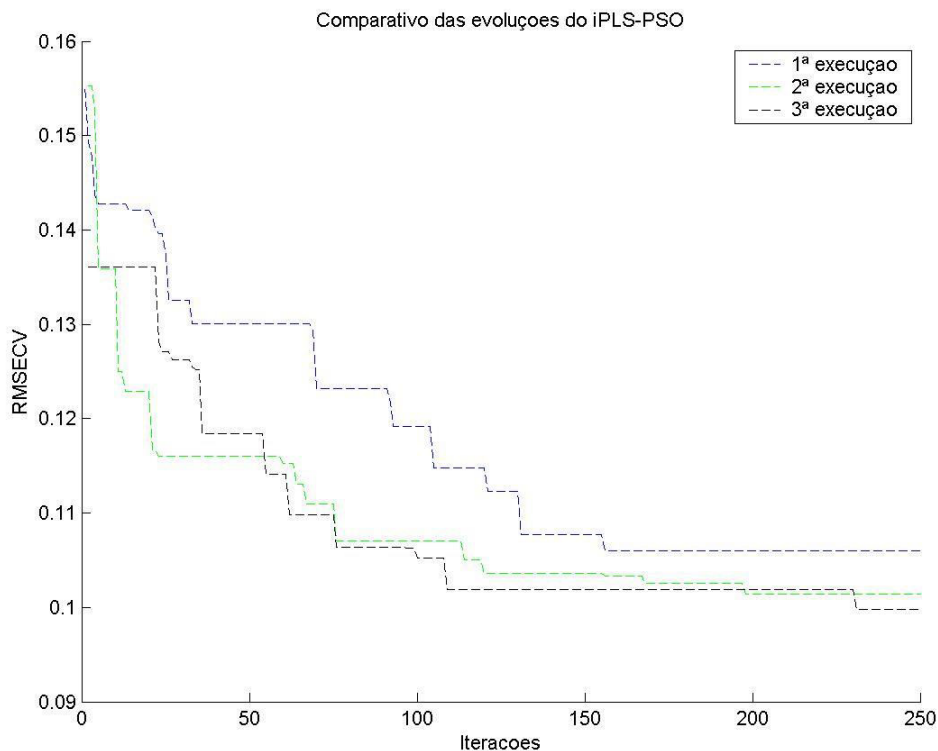
combinou as sub-regiões espectrais que podem visualizadas na Figura 8. Neste caso os espectros foram subdivididos em 60 intervalos e foram empregadas 100 partículas, sendo o desempenho deste modelo apresentado na Figura 9, quando são empregadas 6 variáveis latentes para quantificação do biodiesel.



**Figura 6.** Exemplo da evolução das partículas para a quantificação do biodiesel, usando o método iPLS-PSO com 40 intervalos e 100 partículas, considerando a posição inicial, 10, 100 e 250 iterações

Em todos os 3 modelos apresentados a faixa espectral 1100 - 1200  $\text{cm}^{-1}$  foi comum sendo característica do estiramento CO-OR tanto dos grupos ésteres do biodiesel, quanto dos ésteres dos triglicerídeos do óleo vegetal. Este resultado indica que o dado do estiramento CO-OR é mais informativo sobre

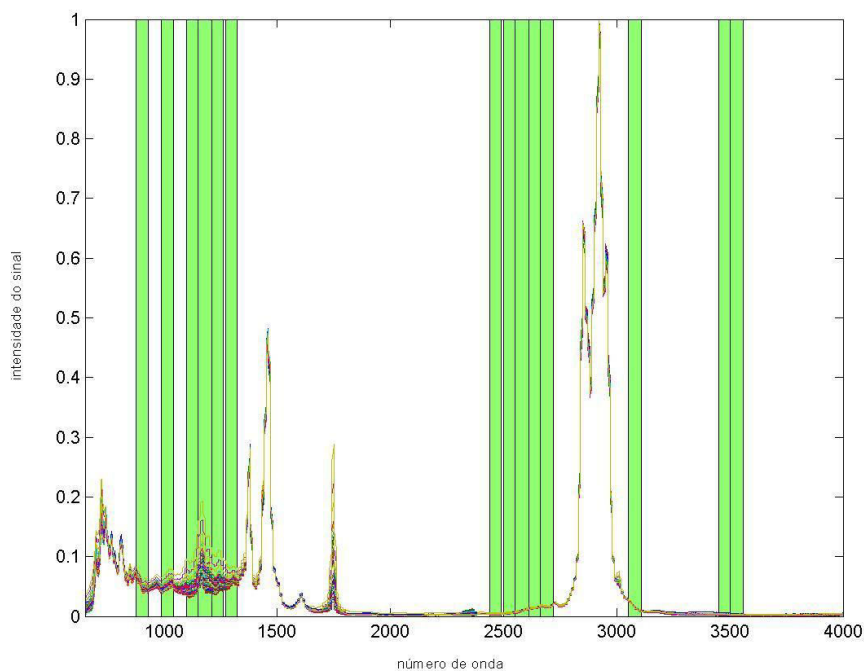
as diferenças estruturais do biodiesel e óleo vegetal do que a característica banda da carbonila (C=O) também presente nestas cadeias carbônicas e que é comumente empregada pela ANP para quantificar o teor de biodiesel presente no diesel comercializado no Brasil.



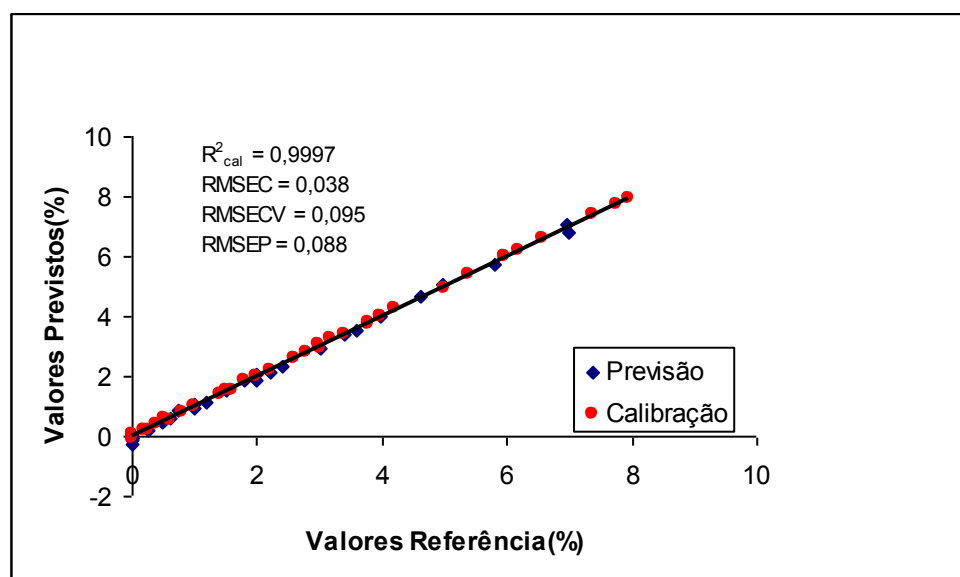
**Figura 7.** Evolução das três execuções do iPLS-PSO com 60 intervalos e 100 partículas, para quantificação do biodiesel

Comparando todos os parâmetros de mérito ( $R^2_{cal}$ , RMSEC, RMSECV e RMSEP), o modelo híbrido iPLS-PSO que emprega uma abordagem determinística (iPLS) combinada a uma abordagem heurística (PSO) foi o que resultou nos menores erros e melhor coeficiente de determinação quando comparado ao modelo desenvolvido com toda a faixa espectral ou com as regiões determinadas pela norma ABNT NBR 15568.

Outro aspecto importante que deve ser salientado é que, apesar de terem sido utilizados dois tipos distintos de biodiesel (biodiesel de palma e biodiesel de soja) na formulação das blends binárias e ternárias, os modelos para a quantificação do total de biodiesel presente em cada amostra apresentaram coeficientes de determinação maiores ou iguais a 0,9955.



**Figura 8.** Regiões espectrais utilizadas na modelagem para a quantificação do biodiesel usando o método iPLS-PSO com 60 intervalos e 100 partículas



**Figura 9.** Valores de referência versus previstos para o modelo de previsão do biodiesel, empregando o método iPLS-PSO com 60 intervalos e 100 partículas

#### 4. Conclusões

Neste trabalho são comparados diferentes

modelos de regressão multivariada por mínimos quadrados parciais para a quantificação de biodiesel nas blends binárias e ternárias biodiesel/óleo vegetal/diesel empregando os sinais de FT-

IR/ATR resultando em modelos com coeficientes de determinação iguais ou superiores a 0,9955.

A implementação de um processo de otimização denominado de modelo híbrido iPLS-PSO, que emprega uma abordagem determinística (iPLS) combinada a uma abordagem heurística (PSO) visando selecionar as melhores regiões espectrais, foi o que resultou nos menores erros e melhor coeficiente de determinação quando comparado ao modelo desenvolvido com toda a faixa espectral ou com o modelo que empregou as regiões determinadas pela norma ABNT NBR 15568.

Também foi possível observar que a faixa espectral 1100 - 1200  $\text{cm}^{-1}$  foi comum a todos os 3 modelos apresentados, sendo característica do estiramento CO-OR tanto dos grupos ésteres do biodiesel, quanto dos ésteres dos triglicerídeos do óleo vegetal. Dessa forma a mesma deve ser considerada na construção dos modelos de regressão multivariada garantindo maior robustez, principalmente quando podem estar sendo avaliadas amostras adulteradas com óleo vegetal.

## Agradecimentos

Os autores agradecem ao apoio financeiro da FAPERGS, do CNPq, da CAPES e ao INCT de Bioanalítica pelo fomento.

## Referências Bibliográficas

<sup>1</sup> Nascimento, M. G.; Costa Neto, P. R.; Mazzuco, L. M. Biotransformação de óleos e gorduras. *Bioteχνologia Ciência & Desenvolvimento* **2001**, *19*, 28. [Link]

<sup>2</sup> BRASIL. Lei nº 11.097, de 13 de janeiro de 2005. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/\\_Ato2004-2006/2005/Lei/L11097.htm#art2](http://www.planalto.gov.br/ccivil_03/_Ato2004-2006/2005/Lei/L11097.htm#art2)>. Acesso em: 29 janeiro 2016.

<sup>3</sup> BRASIL. Lei nº 13.033, de 24 de setembro de 2014. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/\\_Ato2011-2014/2014/Lei/L13033.htm](http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2014/Lei/L13033.htm)>. Acesso em: 12 julho 2016.

<sup>4</sup> BRASIL. Lei nº 13.263, de 23 de março de 2016. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/\\_Ato2015-2018/2016/Lei/L13263.htm#art1](http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2016/Lei/L13263.htm#art1)>. Acesso em: 12 julho 2016.

<sup>5</sup> Knothe, G. Analyzing biodiesel: standards and other methods. *Journal of the American Oil Chemists' Society* **2006**, *10*, 823. [Crossref]

<sup>6</sup> BRASIL, Agência Nacional Do Petróleo, Gás Natural E Biocombustíveis, Resolução ANP Nº 45 de 25 de agosto de 2014, Disponível em: <<http://nxt.anp.gov.br/NXT/gateway.dll?f=templates&fn=default.htm&vid=anp:10.1048/enu>>. Acesso em: 17 julho 2016.

<sup>7</sup> Associação Brasileira de Normas Técnicas - NBR 15342 Biodiesel - Determinação de monoglicerídeos e diglicerídeos em biodiesel de mamona por cromatografia gasosa, Rio de Janeiro, RJ 2016. [Link]

<sup>8</sup> Associação Brasileira de Normas Técnicas - NBR 15343 Biodiesel - Determinação da concentração de metanol e/ou etanol por cromatografia gasosa, Rio de Janeiro, RJ 2012. [Link]

<sup>9</sup> Associação Brasileira de Normas Técnicas - NBR 15764 Biodiesel - Determinação do teor total de ésteres por cromatografia gasosa, Rio de Janeiro, RJ 2015. [Link]

<sup>10</sup> British Standards Institution- EN 14078-Liquid petroleum products. Determination of fatty acid methyl ester (FAME) content in middle distillates. Infrared spectrometry method, 2014. [Link]

<sup>11</sup> Correa, C.; Ruschel, C. F. C. ; Bento, F. M. ; Ferrão, M. F. Aplicação de Cartas de Controle Multivariadas no Controle e Garantia da Qualidade do Biodiesel (B100). *Revista Virtual de Química* 2015, *7*, 2273. [CrossRef]

<sup>12</sup> 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* **2011**, *90*, 701. [CrossRef]

- <sup>13</sup> Kennedy, J.; Eberhart, R. C. *The 1995 IEEE International Conference on Neural Networks*, **1995**, p.1942-1948.
- <sup>14</sup> Pomeroy, P. *An introduction to particle swarm optimization*. 2003. Disponível em: <[http://web.archive.org/web/20060314230949/http://www.adaptiveview.com/articles/ip\\_soprnt.html](http://web.archive.org/web/20060314230949/http://www.adaptiveview.com/articles/ip_soprnt.html)>. Acesso em: 29 fevereiro 2015.
- <sup>15</sup> Igne, B.; Hurburgh Jr., C. R. Local chemometrics for samples and variables: optimizing calibration and standardization processes. *Journal of Chemometrics* **2010**, *24*, 75. [CrossRef]
- <sup>16</sup> Sorol, N.; Arancibia, E.; Bortolato, S. A.; Olivieri, A. C. Visible/near infrared-partial least-squares analysis of Brix in sugar cane juice: A test field for variable selection methods. *Chemometrics and Intelligent Laboratory Systems* **2010**, *102*, 100. [CrossRef]
- <sup>17</sup> Xue, L.; Cai, J.; Li, j.; Liu, M. Application of Particle Swarm Optimization (PSO) Algorithm to Determine Dichlorvos Residue on the Surface of Navel Orange with Vis-NIR Spectroscopy. *Procedia Engineering* **2012**, *29*, 4124. [CrossRef]
- <sup>18</sup> Feng, Y. Z.; ElMasry, G; Sun, D. W.; Scannell, A. G. M.; Walsh, D.; Morcy, N. Near-infrared hyperspectral imaging and partial least squares regression for rapid and reagentless determination of Enterobacteriaceae on chicken fillets. *Food Chemistry* **2013**, *138*, 1829. [CrossRef]
- <sup>19</sup> Parastar, H.; Ebrahimi-Najafabadi, H.; Jalali-Heravi, M. Multivariate curve resolution-particle swarm optimization: A high-throughput approach to exploit pure information from multi-component hyphenated chromatographic signals. *Analytica Chimica Acta* **2013**, *772*, 16. [CrossRef]
- <sup>20</sup> Jin, Y.; Wu, Z.; Liu, X.; Wu, Y. Near infrared spectroscopy in combination with chemometrics as a process analytical technology (PAT) tool for on-line quantitative monitoring of alcohol precipitation. *Journal of Pharmaceutical and Biomedical Analysis* **2013**, *77*, 32. [CrossRef]
- <sup>21</sup> Lou, W.; Yang, K.; Zhu, M.; Wu, Y.; Liu, X.; Jin, Y. Application of particle swarm optimization-based least square support vector machine in quantitative analysis of extraction solution of yangxinshi tablet using near infrared spectroscopy. *Journal of Innovative Optical Health Sciences* **2014**, *7*, 14500111. [CrossRef]
- <sup>22</sup> Engelbrecht, A. P.; *Fundamentals of computational swarm intelligence*. John Wiley & Sons, 2005.
- <sup>23</sup> Norgaard, L.; Saudland, A.; Wagner, J.; Nielsen J. P.; Munck, L.; Engelsen, S. B. Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. *Applied Spectroscopy*, **2000**, *54*, 413. [CrossRef]
- <sup>24</sup> Associação Brasileira de Normas Técnicas - NBR 15568 - Biodiesel - Determinação do teor de biodiesel em óleo diesel por espectroscopia na região do infravermelho médio, Rio de Janeiro, RJ 2008. [Link]
- <sup>25</sup> Balabin, R. M.; Smimov, S. V. Variable selection in near-infrared spectroscopy: benchmarking of feature selection methods on biodiesel data. *Analytica Chimica Acta* **2011**, *692*, 63. [CrossRef]
- <sup>26</sup> Muller, A. L. H.; Flores, É. M. M.; Müller, E. I.; Silva, F. E. B.; Ferrão, M. F. Attenuated total reflectance with Fourier transform infrared spectroscopy (ATR/FTIR) and different PLS Algorithms for simultaneous determination of clavulanic acid and amoxicillin in powder pharmaceutical formulation. *Journal of the Brazilian Chemical Society* **2011**, *22*, 1903. [CrossRef]
- <sup>27</sup> Silva, F. E. B.; Flores, É. M. M.; Parisotto, G.; Müller, E. I.; Ferrão, M. F. Green method by diffuse reflectance infrared spectroscopy and spectral region selection for the quantification of sulphamethoxazole and trimethoprim in pharmaceutical formulations. *Anais da Academia Brasileira de Ciências* **2016**, *88*, 1. [CrossRef]