

PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS E PROCESSOS INDUSTRIAIS –  
MESTRADO  
ÁREA DE CONCENTRAÇÃO EM  
CONTROLE E OTIMIZAÇÃO DE PROCESSOS INDUSTRIAIS

Márcia Werlang

**QUANTIFICAÇÃO DE BIODIESEL EM BLENDA BIODIESEL/DIESEL:  
COMPARAÇÃO DE MODELOS *i*PLS E LS-SVM COM APLICAÇÃO DO MÉTODO  
DA TRANSFORMADA DISCRETA *WAVELET***

Santa Cruz do Sul, fevereiro de 2011.

Márcia Werlang

**QUANTIFICAÇÃO DE BIODIESEL EM BLENDA BIODIESEL/DIESEL:  
COMPARAÇÃO DE MODELOS IPLS E LS-SVM COM APLICAÇÃO DO MÉTODO  
DA TRANSFORMADA DISCRETA *WAVELET***

Dissertação apresentada ao Programa de Pós-Graduação em Sistemas e Processos Industriais – Mestrado, Área de Concentração em Controle e Otimização de Processos Industriais, Universidade de Santa Cruz do Sul, para obtenção do título de Mestre em Sistemas e Processos Industriais.

Orientador: Prof. Dr. Rubén Edgardo Panta Pazos

Co-orientador: Prof. Dr. Marco Flôres Ferrão

Santa Cruz do Sul, fevereiro de 2011.

## **AGRADECIMENTOS**

A Deus por me guiar e proteger.

A minha família, responsável, em grande parte, pela pessoa na qual me tornei.

Aos amigos por sempre me apoiarem e pelas horas de descontração.

Aos colegas pelas boas horas de estudos.

As secretárias pela presteza em ajudar.

Aos professores pela atenção e ensinamentos.

Ao meu co-orientador, professor Dr. Marco Flores Ferrão, pela paciência, pelas muitas orientações e disponibilidade e também pelo senso de direção.

Ao meu orientador, professor Dr. Ruben Edgardo Panta Pazos, pela disponibilidade sempre prestada, incentivo e orientações.

A CAPES pela bolsa concedida.

E todos que de alguma forma contribuíram para a realização desse trabalho.

*Jamais considere seus estudos como uma obrigação, mas como uma oportunidade invejável para aprender a conhecer a influência libertadora da beleza do reino do espírito, para seu próprio prazer pessoal e para proveito da comunidade à qual seu futuro trabalho pertencer.*

Albert Einstein

## RESUMO

Desde o início do ano de 2008 vem sendo obrigatória a comercialização do biodiesel adicionado no diesel. O percentual comercializado vem crescendo com o passar o tempo, sendo atualmente 5% e tende a aumentar. A ANP<sup>1</sup> – responsável pela regulação e fiscalização do biodiesel, já concede autorização para utilização de blendas biodiesel/diesel com percentual de biodiesel acima do que estabelece a legislação vigente. Dessa forma, se faz necessário estabelecer metodologias capazes de atestar se o percentual comercializado está dentro da legislação, a fim de proteger o consumidor. Para isso esse trabalho utilizou espectroscopia no infravermelho médio associado a duas ferramentas: regressão por mínimos quadrados parciais por intervalo (*i*PLS) e máquina vetor de suporte com mínimos quadrados (LS-SVM). Além disso, utilizou-se o método da transformada discreta *wavelet* (TDW) para reduzir o número de variáveis e assim os custos computacionais. Esse trabalho estudou amostras de blendas biodiesel/diesel, com concentrações de biodiesel na faixa de 8% (v/v) – 30% (v/v). Os modelos estudados foram construídos utilizando a faixa espectral que contém o estiramento C-O dos ésteres presentes no biodiesel, que foi selecionada pelo algoritmo *i*PLS empregado. Os resultados obtidos mostraram uma grande capacidade do método TDW em reter informação em novas variáveis, criadas a partir da decomposição das variáveis originais, conferindo economia e rapidez na construção de modelos multivariados. Os modelos construídos revelaram que as ferramentas *i*PLS e LS-SVM são equivalentes para a quantificação de biodiesel em blendas biodiesel/diesel para concentrações estudadas, até mesmo quando modelam simultaneamente espectros oriundos de espectrofotômetros diferentes, no caso um Nicolet e um Shimadzu, apresentando erros de previsão, na faixa de 0,2% (v/v) – 0,4% (v/v), inferior ao limite estabelecido pela Norma NBR 15568 para essa faixa de concentração, 1,0% (v/v).

Palavras-chave: biodiesel, blendas biodiesel/diesel, transformada discreta *wavelet*, espectroscopia no infravermelho médio, regressão por mínimos quadrados parciais por intervalo, máquina de vetor de suporte com mínimos quadrados.

---

<sup>1</sup> ANP – Agência Nacional do Petróleo, Gás Natural e Biocombustíveis.

## ABSTRACT

Since the beginning of the year of 2008 it comes being obligatory the commercialization of biodiesel increase in diesel. The commercialized percentage comes growing with passing time, being currently 5% and tends to increase. The ANP (Brazilian National Agency for Petroleum, Natural Gas and Biofuels) - responsible for the regulation and fiscalization of biodiesel, already grants to authorization for biodiesel/diesel blend with percentage of biodiesel above of what it establishes the current legislation. In this way it makes necessary to establish methodologies capable to certify the commercialized percentage is inside of the legislation, in order to protect the consumer. For this work used mid-infrared spectroscopy associated the two tools: interval partial least squares (iPLS) and least-squares support vector machines (LS-SVM). And also, the discrete wavelet transform method (DWT) was used to reduce the number of variables and thus the computational costs. This work studied samples biodiesel/diesel blend, with concentrations of biodiesel in the range of 8% (v/v) - 30% (v/v). The models had been constructed using the spectral band that contains stretching C-O of esters present in biodiesel, that iPLS was selected by the algorithm used. The gotten results had shown a great capacity of method TDW in holding back information in new variables, created from the decomposition of the original variable, conferring economy and quickness in the construction of multivariate models. The constructed models had showing that the tools iPLS and LS-SVM are equivalents for the quantification of biodiesel in biodiesel/diesel blends for studied concentrations, even though when models simultaneous deriving spectrum of different spectrofotometer, in the case a Nicolet and Shimadzu, with error of prediction of 0.2% (v/v) - 0.4% (v/v), inferior to the limit established for Norm NBR 15568 for this band of concentration, 1.0% (v/v).

Keywords: biodiesel, biodiesel/diesel blends, discrete wavelet transform, mid-infrared spectroscopy, interval partial least squares regression, least-squares support vector machine.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Estado do Rio Grande do Sul com as divisões das regiões para coleta das amostras. ....	14
Figura 2 – Exemplo de um sinal de espectro de absorção de uma amostra de diesel no infravermelho médio.....	17
Figura 3 – Representação da deformação axial: simétrica (a) e assimétrica (b). ....	18
Figura 4 – Representação da deformação angular simétrica (a), deformação angular assimétrica (b), deformação angular simétrica fora do plano (c), deformação angular assimétrica fora do plano (d). Os sinais $x$ e $*$ representam os movimentos para dentro e para fora do plano do desenho.....	18
Figura 5 – Gráfico gerado pelo algoritmo <i>i</i> PLS, com os valores de RMSECV para o modelo construído com o espectro inteiro (linha pontilhada) e para os modelos construídos com os intervalos (barras). Os números dentro das barras correspondem ao número de variáveis latentes para cada intervalo. Os números abaixo das barras são os números de onda. ....	24
Figura 6 – Representação do comportamento das variáveis latentes de um modelo. ....	25
Figura 7 – Representação da LS-SVM como uma rede neural. ....	26
Figura 8 – Exemplo de superfície de otimização dos parâmetros $\gamma$ e $\sigma^2$ em função do erro (RMSECV). ....	29
Figura 9 – Esquema do processo de decomposição das transformadas discretas <i>wavelets</i> . ....	30
Figura 10 – Representação do sinal de uma função (a); representação do sinal da função com um nível de decomposição (b). ....	31
Figura 11 – Fluxograma das etapas executadas.....	33
Figura 12 – Sinal dos espectros das blendas biodiesel/diesel originais.....	38
Figura 13 – Sinal dos espectros de amostras de blendas biodiesel/diesel com aplicação de 4 coeficientes, obtendo 1 nível de decomposição, contendo 868 novas variáveis, e retendo 99,93% de informação.....	39
Figura 14 – Sinal dos espectros de amostras de blendas biodiesel/diesel com aplicação de 4 coeficientes obtendo o segundo nível de decomposição, contendo 434 novas variáveis, e retendo 99,61% de informação.....	39
Figura 15 – Sinal dos espectros de amostras de blendas biodiesel/diesel com aplicação de 6 coeficientes obtendo 1 nível de decomposição, contendo 868 novas variáveis, retendo 99,36% de informação. ....	40
Figura 16 – Sinal dos espectros de amostras de blendas biodiesel/diesel com aplicação de 6 coeficientes obtendo o segundo nível de decomposição, contendo 434 novas variáveis, retendo 97,42% de informação.....	40
Figura 17 – Sinal dos espectros de amostras de blendas biodiesel/diesel com aplicação de 8 coeficientes obtendo 1 nível de decomposição, contendo 868 novas variáveis, retendo 98,31% de informação. ....	41
Figura 18 – Sinal dos espectros de amostras de blendas biodiesel/diesel com aplicação de 8 coeficientes obtendo o segundo nível de decomposição, contendo 434 novas variáveis, retendo 94,04% de informação.....	41
Figura 19 – Sinal dos espectros das blendas biodiesel/diesel subdivididos em 10 intervalos. Para esse caso os intervalos 9 e 10 apresentam os menores erros (RMSECV) indicado pela altura da barra, a linha pontilhada corresponde ao erro (RMSECV) para o modelo que emprega o espectro inteiro.....	42
Figura 20 – Sinal dos espectros das blendas biodiesel/diesel subdivididos em 15 intervalos. Para esse caso os intervalos 13 e 15 apresentam os menores erros (RMSECV) indicados pela altura das barras, a linha pontilhada corresponde ao erro (RMSECV) para o modelo que emprega o espectro inteiro.....	43

Figura 21 – Sinal dos espectros das blendas biodiesel/diesel subdivididos em 20 intervalos. Para esse caso os intervalos 16, 17, 19 e 20 apresentam os menores erros (RMSECV) indicados pela altura das barras, a linha pontilhada corresponde ao erro (RMSECV) para o modelo que emprega o espectro inteiro.....	43
Figura 22 – Modelo construído com o espectro inteiro (faixa espectral de $4.000\text{ cm}^{-1}$ – $650\text{ cm}^{-1}$ ) empregando 5 variáveis latentes com os espectros obtidos nos dois espectrofotômetros. ....	45
Figura 23 – Modelo construído com o espectro inteiro (faixa espectral de $4.000\text{ cm}^{-1}$ – $650\text{ cm}^{-1}$ ) empregando 5 variáveis latentes com espectros obtidos no espectrofotômetro Shimadzu. ....	46
Figura 24 – Modelo construído com o espectro inteiro (faixa espectral de $4.000\text{ cm}^{-1}$ – $650\text{ cm}^{-1}$ ) empregando 5 variáveis latentes com espectros obtidos no espectrofotômetro Nicolet.	46
Figura 25 – Modelo <i>i</i> PLS subdividindo os espectros originais em 10 intervalos, empregando faixa espectral de $1.321\text{ cm}^{-1}$ – $985\text{ cm}^{-1}$ , 4 variáveis latentes e com dados lidos nos dois espectrofotômetros.....	47
Figura 26 – Modelo <i>i</i> PLS subdividindo os espectros originais em 10 intervalos, empregando faixa espectral de $1.321\text{ cm}^{-1}$ – $985\text{ cm}^{-1}$ , 4 variáveis latentes e com dados lidos no espectrofotômetro Shimadzu. ....	47
Figura 27 – Modelo <i>i</i> PLS subdividindo os espectros originais em 10 intervalos, empregando faixa espectral de $1.321\text{ cm}^{-1}$ – $985\text{ cm}^{-1}$ , 4 variáveis latentes e com dados lidos no espectrofotômetro Nicolet. ....	48
Figura 28 – Modelo <i>i</i> PLS com subdivisão dos espectros em 20 intervalos, empregando a faixa espectral de $1.165\text{ cm}^{-1}$ – $995\text{ cm}^{-1}$ , 2 variáveis latentes, com 2 níveis de decomposição e construído com dados lidos nos dois espectrofotômetros. ....	48
Figura 29 – Modelo <i>i</i> PLS com subdivisão dos espectros em 20 intervalos, empregando a faixa espectral de $1.165\text{ cm}^{-1}$ – $995\text{ cm}^{-1}$ , 2 variáveis latentes, com 2 níveis de decomposição e construído com dados lidos no espectrofotômetro Shimadzu. ....	49
Figura 30 – Modelo <i>i</i> PLS com subdivisão dos espectros em 20 intervalos, empregando a faixa espectral de $1.165\text{ cm}^{-1}$ – $995\text{ cm}^{-1}$ , 2 variáveis latentes, com 2 níveis de decomposição e construído com dados lidos no espectrofotômetro Nicolet. ....	49
Figura 31 – Superfície de otimização dos parâmetros $\gamma$ e $\sigma^2$ em função do erro (RMSECV) para o modelo LS-SVM com dados originais e subdivididos em 10 intervalos. ....	51
Figura 32 – Modelo LS-SVM subdividindo os espectros originais em 10 intervalos, empregando a faixa espectral de $1.321\text{ cm}^{-1}$ – $985\text{ cm}^{-1}$ construído com dados lidos em ambos espectrofotômetros.....	52
Figura 33 – Modelo LS-SVM subdividindo os espectros originais em 10 intervalos, empregando a faixa espectral de $1.321\text{ cm}^{-1}$ – $985\text{ cm}^{-1}$ construído com dados lidos no espectrofotômetro Shimadzu. ....	52
Figura 34 – Modelo LS-SVM subdividindo os espectros originais em 10 intervalos, empregando a faixa espectral de $1.321\text{ cm}^{-1}$ – $985\text{ cm}^{-1}$ construído com dados lidos no espectrofotômetro Nicolet. ....	53
Figura 35 – Superfície de otimização dos parâmetros $\gamma$ e $\sigma^2$ em função do erro (RMSECV) para o modelo LS-SVM com dados decompostos em 2 níveis, com 6 coeficientes e subdivididos em 15 intervalos. ....	54
Figura 36 – Modelo LS-SVM com dados decompostos em 2 níveis, com 6 coeficientes e subdividido em 15 intervalos, empregando a faixa espectral de $1.326\text{ cm}^{-1}$ – $1.103\text{ cm}^{-1}$ , construído com dados lidos em ambos espectrofotômetros. ....	54
Figura 37 – Modelo LS-SVM com dados decompostos em 2 níveis, com 6 coeficientes e subdividido em 15 intervalos, empregando a faixa espectral de $1.326\text{ cm}^{-1}$ – $1.103\text{ cm}^{-1}$ , construído com dados lidos no espectrofotômetro Shimadzu. ....	55



Figura 38 – Modelo LS-SVM com dados decompostos em 2 níveis, com 6 coeficientes e subdividido em 15 intervalos, empregando a faixa espectral de  $1.326 \text{ cm}^{-1} - 1.103 \text{ cm}^{-1}$ , construído com dados lidos no espectrofotômetro Nicolet. ....55

## LISTA DE TABELAS

Tabela 1 – Cristais empregados para a reflexão total atenuada.....	20
Tabela 2 – Concentrações das blendas biodiesel/diesel. ....	35
Tabela 3 – Resultados dos melhores modelos <i>i</i> PLS para a quantificação de biodiesel em blendas biodiesel/diesel com espectros originais e decompostos em até 2 níveis com 4, 6 e 8 coeficientes. ....	44
Tabela 4 – Resultados dos melhores modelos LS-SVM para a quantificação de biodiesel em blendas biodiesel/diesel com espectros originais e decompostos em até 2 níveis com 4, 6 e 8 coeficientes. ....	50
Tabela 5 – Resultados do teste F para os modelos <i>i</i> PLS. ....	56
Tabela 6 – Resultados do teste F para os modelos LS-SVM. ....	57
Tabela 7 – Resultados do teste F para os modelos <i>i</i> PLS e LS-SVM.....	58
Tabela 8 – Resultados dos modelos para a determinação de biodiesel em blendas biodiesel/diesel com espectros originais e decompostos em até 2 níveis com 4, 6 e 8 coeficientes, seguindo as regiões espectrais e o pré-processamento da Norma NBR 15.568..	59
Tabela 9 – Resultados do teste F para os erros de previsão dos modelos <i>i</i> PLS das Tabelas 3 e 8. ....	60
Tabela 10 – Resultados do tempo de execução da transformada discreta <i>wavelet</i> Daubechies com 4, 6 e 8 coeficientes. ....	60
Tabela 11 – Resultados do tempo de execução dos modelos <i>i</i> PLS com e sem aplicação da transformada discreta <i>wavelet</i> Daubechies com 4 coeficientes empregando 2 pré-processamentos e com subdivisão espectral de 10 intervalos. ....	61

## SUMÁRIO

1 INTRODUÇÃO.....	12
1.1 Justificativa.....	13
1.2 Objetivos.....	13
1.2.1 Objetivo geral .....	13
1.2.2 Objetivos específicos.....	14
1.3 Determinação do teor de biodiesel em blendas biodiesel/diesel no Brasil.....	14
1.4 Organização do texto .....	15
2 FUNDAMENTAÇÃO TEÓRICA .....	16
2.1 Espectroscopia no infravermelho .....	17
2.1.1 Principais técnicas de reflexão no infravermelho médio.....	19
2.1.1.1 Reflexão total atenuada .....	19
2.2 Análise multivariada.....	20
2.2.1 Mínimos quadrados parciais .....	21
2.2.2 Mínimos quadrados parciais por intervalo .....	23
2.2.3 Máquina de vetor de suporte com mínimos quadrados .....	25
2.3 Transformada discreta <i>wavelets</i> .....	29
3 MÉTODO .....	33
3.1 Planejamento experimental.....	34
3.2 Preparo de amostras.....	36
3.3 Obtenção dos espectros das amostras .....	36
3.4 Aplicação da transformada discreta <i>wavelet</i> .....	36
3.5 Modelagem dos dados .....	36
3.8 Comparação dos resultados .....	37
4 RESULTADOS .....	38
4.1 Comparação dos resultados obtidos com aplicação do método da transformada discreta <i>wavelet</i> .....	38
4.2 Resultados para determinação do biodiesel em blendas biodiesel/diesel.....	41
4.3 Descrição dos espectrofotômetros .....	58

4.4 Resultados da quantificação de biodiesel em blendas biodiesel/diesel seguindo a Norma NBR 15.568.....	58
4.5 Resultados comparativos dos tempos necessários para a aplicação do método da TDW ..	60
4.6 Resultados comparativos dos tempos necessários para a construção de modelos <i>i</i> PLS com e sem aplicação do método da TDW .....	61
4.7 Considerações sobre os resultados para a quantificação de biodiesel em blendas biodiesel/diesel .....	61
5 CONCLUSÕES .....	64
REFERÊNCIAS .....	65
ANEXOS.....	70

## 1 INTRODUÇÃO

A sociedade mundial tem discutido a substituição gradual do combustível fóssil pelo combustível produzido a partir de fontes renováveis tais como: a soja, o dendê, a canola, a gordura animal, o óleo de fritura entre outros. Este é chamado de biodiesel. No Brasil se tornou obrigatório no início de 2008 a comercialização de biodiesel adicionado ao diesel, ou seja, blendas biodiesel/diesel, com base na lei nº 11.097 de 13 de janeiro de 2005 (BRASIL, 2005). O percentual obrigatório foi de 2% (v/v) de biodiesel, sendo aumentado gradativamente até chegar atualmente em 5% (v/v) de biodiesel nas blendas biodiesel/diesel (Resolução nº 6 de 16.09.2009), percentual que só começaria a vigorar em 2013, segundo a lei 11.097. Diante disto é importante estabelecer métodos que sejam capazes de verificar de maneira adequada se o percentual de biodiesel adicionado ao diesel está dentro dos limites estabelecidos pela legislação vigente, a fim de proteger o consumidor.

Em outros países também vem se observando o aumento do percentual de biodiesel adicionado ao diesel mineral, como é o caso da Espanha. Neste país pretende-se de chegar a 7% de biodiesel até o final de 2011 e em 10% em 2013. Na Colômbia se pretende atingir o valor de 9% ou 9,5% até o final de 2011 e 15% de biodiesel em 2015. A Argentina espera adicionar 20% de biodiesel ao diesel em 2014. Em um ritmo menor está o Peru, que pretende chegar à marca de 5% de biodiesel até o final 2011. **(fontes)**

Uma das técnicas analíticas para determinar o percentual de biodiesel presente na mistura com diesel é a espectroscopia no infravermelho com o acessório de reflexão total atenuada (ATR-FTIR). Esta técnica tem a vantagem de não destruir a amostra analisada, não precisa de reagentes e por consequência os resíduos são mínimos. É uma técnica analítica rápida, não seletiva e que permite inferir sobre os grupos funcionais e a estrutura química das moléculas, bem como, de misturas complexas. Nestas variáveis é necessário a aplicação de ferramentas quimiométricas, pois elas têm a capacidade de modelar muitas variáveis simultaneamente. Quando são obtidos espectros no infravermelho médio a absorvância (ou  $\log(1/R)$ ) é medida para diversos números de onda (variáveis).

Como na região do espectro eletromagnético correspondente ao infravermelho médio, podem ser registrados os sinais de absorvância (ou  $\log(1/R)$ ) para uma elevada quantidade de números de onda (de 1.000 até 4.000) dependendo da resolução utilizada, nesse trabalho foi

empregado o método da transformada discreta *wavelets* (TDW), que além de reduzir o número de dados, permite diminuir os ruídos do sinal. Posteriormente é selecionado a faixa espectral mais significativa, e por fim, construir modelos multivariados preditivos, na região selecionada, que quantifiquem a presença de biodiesel nas blendas biodiesel/diesel de maneira satisfatória.

## **1.1 Justificativa**

Tendo em vista a comercialização do biodiesel em mistura com diesel em percentuais crescentes, se faz necessário o estudo de métodos que consigam determinar o percentual estabelecido pela legislação. Neste contexto, a espectroscopia no infravermelho associada às ferramentas quimiométricas é uma método ambientalmente correto, já utilizado em diferentes áreas. A espectroscopia no infravermelho conserva a integridade da amostra; o uso de reagentes para as análises das amostras é desnecessário, bem como a utilização de solventes. As análises das amostras podem ser realizadas à temperatura e pressão ambientes. Estas análises são fáceis e rápidas de serem realizadas. A espectroscopia no infravermelho associada às ferramentas quimiométricas é um método capaz de determinar o biodiesel com baixíssimas concentrações até valores elevados, na faixa de 30%, ou superiores (OLIVEIRA *et al.*, 2006).

A espectroscopia no infravermelho dá origem a conjuntos de dados com elevado número de variáveis. Com a intenção de obter modelos multivariados capazes de determinar a concentração de biodiesel de forma satisfatória nas blendas biodiesel/diesel, empregando para isso um conjunto menor de variáveis, optou-se pela utilização do método da transformada discreta *wavelet* (TDW) e dos algoritmos *i*PLS (regressão por mínimos quadrados parciais) e LS-SVM (máquina de vetor de suporte com mínimos quadrados).

## **1.2 Objetivos**

### **1.2.1 Objetivo geral**

O objetivo geral desse trabalho é estudar métodos ambientalmente corretos – que seguem os princípios da química verde – para a determinação do teor de biodiesel nas blendas

biodiesel/diesel, empregando espectroscopia no infravermelho médio associada às ferramentas quimiométricas.

### 1.2.2 Objetivos específicos

- Utilizar metodologia analítica por espectroscopia no infravermelho e TDW empregando *i*PLS visando à identificação da(s) faixa(s) espectral(is) para quantificação de biodiesel em blendas biodiesel/diesel;

- Desenvolver e comparar modelos multivariados de regressão LS-SVM empregando a(s) mesma(s) a(s) faixa(s) espectral(is) de infravermelho identificadas pelo *i*PLS, com e sem aplicação da TDW, para quantificação de biodiesel presente nas blendas biodiesel/diesel.

### 1.3 Determinação do teor de biodiesel em blendas biodiesel/diesel no Brasil

No Brasil a determinação do biodiesel em blendas biodiesel/diesel é realizada por uma rede de laboratórios. Todos os postos de gasolina são amostrados periodicamente e sem aviso prévio. Para facilitar o trabalho dos laboratórios, as unidades da federação foram divididas em regiões. Cada laboratório fica responsável por um grupo de regiões, que pode ser um Estado inteiro ou parte dele. No Rio Grande do Sul a divisão contém 10 regiões conforme Figura 1.

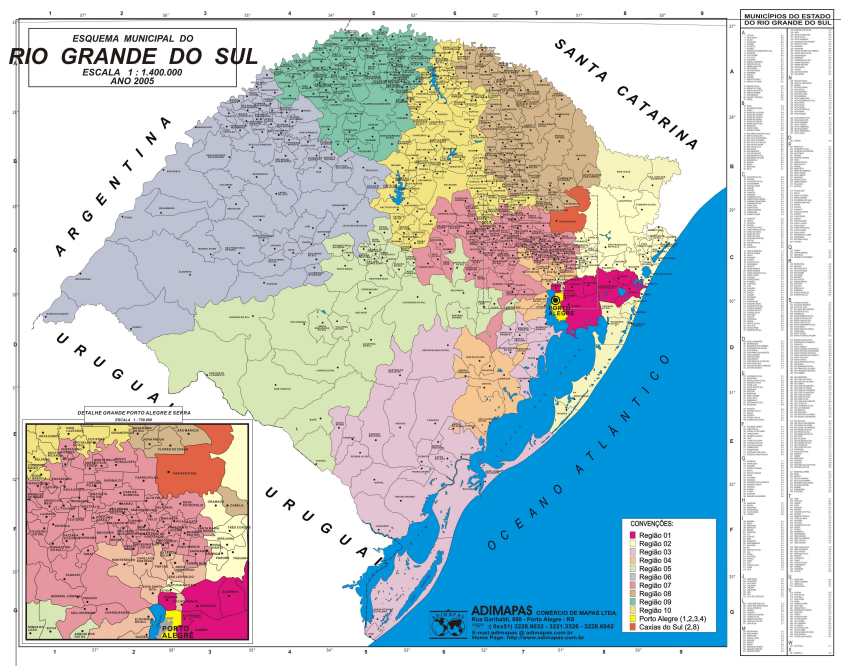


Figura 1 – Estado do Rio Grande do Sul com as divisões das regiões para coleta das amostras.  
Fonte: CECOM

Estas regiões e respectivos municípios foram definidos pelo Centro de Combustíveis da Universidade Federal do Rio Grande do Sul (CECOM) sob a supervisão da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP). As análises seguem a Norma BS EN 14.078: *Liquid petroleum products. Determination of fatty acid methyl esters (FAME) in middle distillates. Infrared spectroscopy method*, apesar de já existir a Norma NBR 15.568: Biodiesel - Determinação do teor de biodiesel em óleo diesel por espectroscopia na região do infravermelho médio. A Norma NBR prevê o emprego de espectroscopia no infravermelho médio e modelos de regressão com mínimos quadrados parciais para realizar a quantificação do biodiesel nas blendas. Ela estabelece dois conjuntos de concentrações. O primeiro conjunto tem concentração que varia de 0 – 8% (v/v) e o erro de previsão não deve ultrapassar 0,1% (v/v). O segundo conjunto abrange as concentrações de 8% – 30% (v/v) e o erro de previsão não deve ultrapassar de 1% (v/v). Para o primeiro conjunto são sugeridas no mínimo 160 amostras (blendas biodiesel/diesel), e para o segundo grupo são sugeridas no mínimo 45 amostras (blendas biodiesel/diesel) nas quais se empregam biodiesel e diesel de diferentes procedências. As faixas espectrais estabelecidas para realizar a modelagem vão de  $1.200\text{ cm}^{-1}$  a  $1.100\text{ cm}^{-1}$ , referente ao sinal C-O-C e de  $1.750\text{ cm}^{-1}$  a  $1.735\text{ cm}^{-1}$ , referente ao sinal C=O. A Norma NBR ainda estabelece que 2/3 dos espectros devem ser empregados na calibração do modelo e os outros 1/3 dos espectros devem ser empregados na validação externa do modelo (NBR, 2008).

#### **1.4 Organização do texto**

Esta dissertação está organizada da seguinte maneira. O segundo capítulo é dedicado a temas importantes para a compreensão do desenvolvimento desse trabalho. É apresentada uma explanação geral sobre biodiesel, informações a respeito de espectroscopia no infravermelho, análise multivariada e transformada discreta *wavelet*.

No terceiro capítulo é descrito o método empregado para a realização desse trabalho. No quarto capítulo são discutidos os resultados obtidos para a quantificação do biodiesel em blendas biodiesel/diesel. No capítulo seguinte são apresentadas as principais conclusões e sugestões de trabalhos futuros.



## 2 FUNDAMENTAÇÃO TEÓRICA

A maior parte da energia consumida no mundo é de origem mineral e não renovável. A emissão crescente de gases tóxicos, oriundos desse consumo, tem agravado os danos ambientais, como o aquecimento do planeta. Problemas desta natureza impulsionaram a busca por novas fontes de energia. Nesse cenário, uma das possibilidades que tem recebido apoio mundial é a utilização de combustíveis mais limpos, de fontes renováveis, em veículos automotores como uma substituição gradual ao combustível fóssil.

No Brasil tem se concedido grande importância ao biodiesel, produto derivado de biomassa renovável, em decorrência da enorme diversidade e abundância de fontes de matérias-primas, como por exemplo, os óleos vegetais e gordura animal, novos ou residuais. Para a obtenção do biodiesel a rota mais comum é a reação de transesterificação, onde um óleo reage com um álcool na presença de um catalisador dando origem ao biodiesel e ao glicerol (GERIS *et al.*, 2007).

Conforme 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 (BRASIL, 2005). Essa mesma lei determinou a adição de 2% (v/v) de biodiesel ao diesel comercializado no início do ano de 2008, bem como, a progressão gradual desse percentual (BRASIL, 2005).

Uma das maneiras de quantificar o percentual de biodiesel nas misturas com diesel é através da espectroscopia com infravermelho médio. Essa metodologia analítica permite a identificação da assinatura química da amostra estudada (CORREIA *et al.*, 2007). E a lei de Lambert-Beer, assegura que a intensidade do sinal analítico é proporcional à concentração da respectiva amostra (DENNEY *et al.*, 1993). Assim é possível realizar a quantificação com o auxílio das ferramentas quimiométricas, sejam elas lineares ou não. Existem na literatura vários estudos realizados com essa metodologia (COSTA FILHO & POPPI, 2002; BORGES NETO, 2005; CORDEIRO, 2006; BORIN, 2007; MARETTO 2007; DANTAS FILHO, 2007; RIZZI, 2008; VIERA *et al.*, 2009; GERBASE *et al.*, 2009).

A espectroscopia no infravermelho contempla alguns dos principais pontos da química verde, a saber (LENARDÃO *et al.*, 2003):

- a) A não utilização de solventes e reagentes para a realização das análises;
- b) Análises realizadas à temperatura e pressão ambientes;
- c) O risco de acidentes com sequelas graves, devido aos produtos químicos envolvidos, é praticamente inexistente;
- d) A mínima geração de resíduos;
- e) Análise rápida das amostras.

Para uma melhor compreensão deste trabalho se faz necessário uma explanação sobre espectroscopia no infravermelho, ferramentas quimiométricas, método da transformada discreta *wavelet*.

## 2.1 Espectroscopia no infravermelho

A espectroscopia no infravermelho é o estudo da interação da molécula com a radiação infravermelha, onde se procura conhecer os níveis de energia e as transições de espécies **atômicas** e moleculares (MCMURRY, 2005; BORGES NETO, 2005).

Quando uma amostra é exposta a radiação infravermelha, ela absorve energia em certos comprimentos de onda e transmite em outros. Devido a essa “troca” de energia pode-se definir o espectro de absorção da amostra através um gráfico de absorbância por número de onda, Figura 2.

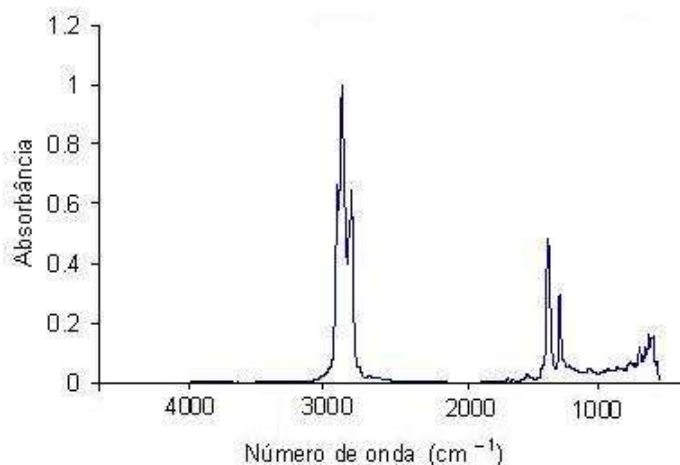


Figura 2 – Exemplo de um sinal de espectro de absorção de uma amostra de diesel no infravermelho médio.

Nesse gráfico são verificadas bandas indicando onde e em que grau houve a absorção (MCMURRY, 2005). “Essa absorção é originária da transição entre dois níveis de energia vibracional e rotacional de uma molécula, geralmente no estado eletrônico fundamental, arranjada na forma de bandas em função do número de onda ou do comprimento de onda da radiação” (BUENO 1989, p.172).

Além das transições eletrônicas, as moléculas exibem dois tipos adicionais de transições induzidas pela radiação infravermelha: **transições vibracionais** e **transições rotacionais** (Figuras 3 e 4). As transições vibracionais ocorrem porque a molécula apresenta um número muito grande de níveis energéticos quantizados (ou estados vibracionais) associados com as ligações que mantêm a molécula unida (SKOOG *et al.*, 2006).



Figura 3 – Representação da deformação axial: simétrica (a) e assimétrica (b).  
Fonte: Adaptado de SILVERTEIN *et al.*, 2007.

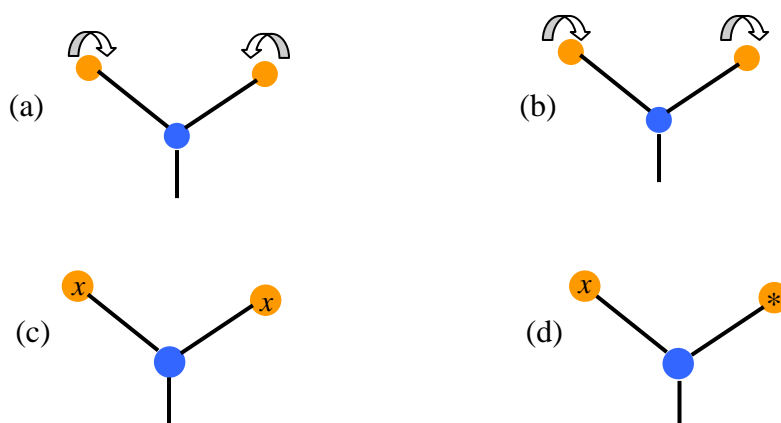


Figura 4 – Representação da deformação angular simétrica (a), deformação angular assimétrica (b), deformação angular simétrica fora do plano (c), deformação angular assimétrica fora do plano (d). Os sinais  $x$  e  $*$  representam os movimentos para dentro e para fora do plano do desenho.  
Fonte: Adaptado de SILVERTEIN *et al.*, 2007.

Há dois tipos de vibrações moleculares: deformações angulares, variação do ângulo entre duas ligações; e deformações axiais ou estiramento, variação rítmica da distância interatômica ao longo do eixo de ligação de dois átomos. Estes dois tipos de vibrações

moleculares podem acontecer da forma simétrica ou assimétrica (SILVERSTEIN *et al.*, 2007).

### **2.1.1 Principais técnicas de reflexão no infravermelho médio**

Para a obtenção dos espectros de infravermelho médio são empregadas técnicas de reflexão. Sendo as reflexões especular, difusa e total atenuada, algumas das principais técnicas empregadas. Nesse trabalho foram empregados os espectros de infravermelho médio<sup>2</sup>, obtidos pela reflexão total atenuada (ATR), para o desenvolvimento e comparação dos modelos de regressão multivariada. A ATR foi escolhida por ser uma técnica que permite obter espectros qualitativos de sólidos, e também de materiais muito absorventes, que são difíceis para serem analisados na espectroscopia por transmissão.

#### **2.1.1.1 Reflexão total atenuada**

A característica da reflexão total atenuada (ATR) são as várias reflexões, da radiação infravermelha que acontecem dentro do cristal, que interagem somente com o material em contato com o cristal. Quando essa radiação atravessa o cristal e atinge a amostra, com densidade menor que a do cristal, há uma reflexão dessa radiação criando uma onda evanescente. Assim, a amostra pode atenuar a intensidade da radiação que retorna ao cristal, ou seja, absorvendo parte da radiação original (FERRÃO, 2000).

A reflexão total atenuada permite obter espectros de amostras líquidas, espessas e de materiais muito absorventes, que são difíceis de serem analisados na espectroscopia por transmissão. Além de ser rápida, não necessita de muito preparo da amostra. Para essa reflexão o cristal precisa ter alto índice de refração (FERRÃO, 2001).

A profundidade de penetração da radiação varia proporcionalmente ao comprimento de onda, e o índice de refração da amostra também influencia esse grau de penetração. É possível controlar a intensidade de penetração através do ângulo de incidência e da escolha do material

---

<sup>2</sup> A região do infravermelho, situada no espectro eletromagnético entre as regiões do visível e das microondas, é dividida em três: próximo (14.290 a 4.000  $\text{cm}^{-1}$ ), médio (4.000 a 666  $\text{cm}^{-1}$ ) e distante (700 a 200  $\text{cm}^{-1}$ ). O surgimento de bandas de vibração/rotação são observadas em toda faixa do espectro no infravermelho. São essas bandas que permitem obter a assinatura química de uma amostra (SILVERSTEIN *et al.*, 2007).

de ATR (FERRÃO, 2001). Na Tabela 1 são apresentados alguns cristais que podem ser empregados para essa reflexão.

Tabela 1 – Cristais empregados para a reflexão total atenuada.

<b>Cristal</b>	<b>Índice de reflexão</b>	<b>Dureza (kg/mm<sup>2</sup>)</b>	<b>Ponto de fusão (C°)</b>	<b>Faixa espectral (cm<sup>-1</sup>)</b>
Diamante	2,4	7.000	3.500	4.500-2.500 1.667-33
Germânio	4,0	550	926	5.000-900
Tolureto de cádmio	2,7	56	1.040	10.000-500
Seleneto de zinco	2,4	135	1.520	20.000-700
Sulfeto de zinco	2,2	178	1.830	14.000-1.000

Fonte: Adaptado de HIND *et al.*, 2001.

O cristal a ser utilizado vai depender do tipo de amostra a ser analisada e da faixa espectral que se pretende empregar, uma vez que amostras pesadas quebram os cristais frágeis, como o seleneto de zinco, ou ainda, alguns cristais não abrangem algumas faixas espectrais que se pretende empregar. Esses são alguns dos fatores envolvidos para uma escolha adequada do cristal a ser empregado.

## 2.2 Análise multivariada

A análise multivariada é uma ferramenta caracterizada pela manipulação simultânea de muitos dados, podendo ser de origens distintas. Ela se divide em qualitativa e quantitativa. Sendo análise hierárquica de agrupamentos (HCA) (ARAÚJO *et al.*, 2008; SANTOS, 2005), análise de componentes principais (PCA) (GODINHO *et al.*, 2008; SANTOS, 2005) e mínimos quadrados parciais com análise discriminante (PLS-DA) (PATACA *et al.*, 2006) as principais técnicas qualitativas. Já as quantitativas utilizam regressão por componentes principais (PCR) e regressão por mínimos quadrados parciais (PLS), que são métodos de calibração lineares (FERREIRA *et al.*, 1999; FERRÃO, 2000). Existem também redes neurais que podem ser empregadas como calibração multivariada para dados lineares ou não. Um exemplo de rede neural é a máquina de vetor de suporte com mínimos quadrados (LS-SVM) (BORIN, 2007; MARETTO, 2007). Nesse trabalho foi utilizado o algoritmo *i*PLS (uma extensão do PLS) para selecionar as regiões espectrais mais representativas para a quantificação de biodiesel em blendas biodiesel/diesel e posteriormente a construção de modelos multivariados por LS-SVM nessas regiões espectrais selecionadas.

### 2.2.1 Mínimos quadrados parciais

A regressão por mínimos quadrados parciais (PLS) é um método de calibração multivariada, onde é desenvolvido um modelo de regressão linear permitindo a previsão de uma desejada característica (variável dependente) a partir de um espectro medido (variáveis independentes) (NORGAARD *et al.*, 2000). Uma das principais vantagens da calibração multivariada é

permitir determinações simultâneas de mais de um tipo de analito de interesse, permitir determinações, mesmo na presença de interferentes e apresentar uma diminuição do erro estimado no modelo devido ao fato de ser um método que utiliza múltiplas variáveis. (SANTOS, 2005, p.27).

O alicerce da regressão por PLS reside na decomposição da matriz  $\mathbf{X}$ , entendida como a soma de várias matrizes  $\mathbf{M}$ , mais outra matriz de resíduos  $\mathbf{E}$ , correspondendo ao resíduo, representada pela equação 1:

$$\mathbf{X} = \mathbf{M}_1 + \mathbf{M}_2 + \dots + \mathbf{M}_n + \mathbf{E} \quad (1)$$

onde:

$n$  é o número de fatores utilizados para truncar a igualdade (as variáveis latentes).

$$\mathbf{M}_1 = \mathbf{t}_1 \mathbf{p}_1^t$$

$$\mathbf{M}_2 = \mathbf{t}_2 \mathbf{p}_2^t$$

$$\mathbf{M}_n = \mathbf{t}_n \mathbf{p}_n^t$$

E como  $\mathbf{t}_1 \mathbf{p}_1^t + \mathbf{t}_2 \mathbf{p}_2^t + \mathbf{t}_n \mathbf{p}_n^t = \mathbf{TP}^t$

Logo  $\mathbf{X} = \mathbf{TP}^t + \mathbf{E}$  e de forma análoga para  $\mathbf{y}$ , equação 2:

$$\mathbf{y} = \mathbf{UQ}^t + \mathbf{F} \quad (2)$$

onde:

$\mathbf{X}$  é a matriz de dados (espectros);

$\mathbf{y}$  é o vetor com as concentrações das soluções lidas no espectrofotômetro;

$\mathbf{T}$  e  $\mathbf{U}$  são os escores de  $\mathbf{X}$  e  $\mathbf{y}$ , respectivamente;

$\mathbf{P}$  e  $\mathbf{Q}$  são os pesos de  $\mathbf{X}$  e  $\mathbf{y}$ , respectivamente;

$\mathbf{E}$  e  $\mathbf{F}$  são os resíduos de  $\mathbf{X}$  e  $\mathbf{y}$ , respectivamente.

A relação entre  $\mathbf{X}$  e  $\mathbf{y}$ , resulta na equação 3 representando a curva de calibração:

$$\mathbf{U} = \mathbf{bT} + \mathbf{e} \quad (3)$$

onde  $\mathbf{U}$  é uma matriz contendo as propriedades (variáveis dependentes) de todas as amostras;  $\mathbf{b}$  é um vetor contendo os parâmetros do modelo;  $\mathbf{T}$  é uma matriz resposta (assim como um conjunto de espectros) para uma série de amostras de calibração e  $\mathbf{e}$  é um vetor que representa o resíduo do espectro não explicados pelo modelo e os erros do modelo (KONZEN *et al.*, 2003).

Uma etapa fundamental na modelagem dos espectros através da regressão por PLS é a determinação do número de variáveis latentes para a construção de um modelo de calibração eficiente. Para ser eficiente o modelo tem que ser capaz de estimar a situação real a que foi proposto, da melhor maneira, abrangendo o máximo possível de variações. Pode-se utilizar para a determinação do número de variáveis latentes a validação externa, ou a validação cruzada. Na primeira situação, divide-se o conjunto de amostras em 3 grupos: calibração, validação e previsão e para a seleção do número de variáveis latentes calcula-se o RMSEV (*Root Mean Square Error Validation*), equação 4, para o conjunto de validação:

$$RMSEV = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (4)$$

onde  $y_i$  é o valor de referência e  $\hat{y}_i$  é o valor estimado e  $n$  o número total de amostras do conjunto utilizado no cálculo.

Para saber o erro de calibração do modelo calcula-se o RMSEC (*Root Mean Square Error Calibration*), representado na equação 5, para o conjunto de calibração.

$$RMSEC = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (5)$$

onde  $y_i$  é o valor de referência e  $\hat{y}_i$  é o valor estimado e  $n$  o número total de amostras do conjunto utilizado no cálculo.

Para conhecer o erro de previsão do modelo calcula-se o RMSEP (*Root Mean Square Error Prediction*), disposto na equação 6, para o c conjunto de dados de previsão.

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (6)$$

onde  $y_i$  é o valor de referência e  $\hat{y}_i$  é o valor estimado e  $n$  o número total de amostras do conjunto utilizado no cálculo.

Já na validação cruzada divide-se o conjunto de amostras em 2 grupos: calibração e previsão, sendo que a calibração é repetida tantas vezes quanto for o número de amostras que compõem o modelo (conjunto de calibração), onde para cada uma das amostras excluídas um modelo é construído com as demais e realiza-se a previsão da amostra excluída, visando identificar o número adequado de variáveis latentes para obter um bom modelo. Para selecionar o número mais apropriado, calcula-se o RMSECV equação 7 (GARCIA *et al.*, 2004).

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (7)$$

onde  $y_i$  é o valor de referência e  $\hat{y}_i$  é o valor estimado e  $n$  o número total de amostras do conjunto utilizado no cálculo.

E para conhecer os erros de previsão do modelo calcula-se o RMSEP, equação 6.

### 2.2.2 Mínimos quadrados parciais por intervalo

O método de regressão por mínimos quadrados parciais por intervalo (iPLS) é uma extensão do PLS, no qual o conjunto de dados (espectros) é dividido em um determinado número de intervalos equidistantes. Para cada intervalo é construído um modelo PLS, além do modelo construído para o espectro inteiro, sendo os resultados apresentados em um gráfico para facilitar a comparação, como o da Figura 5.



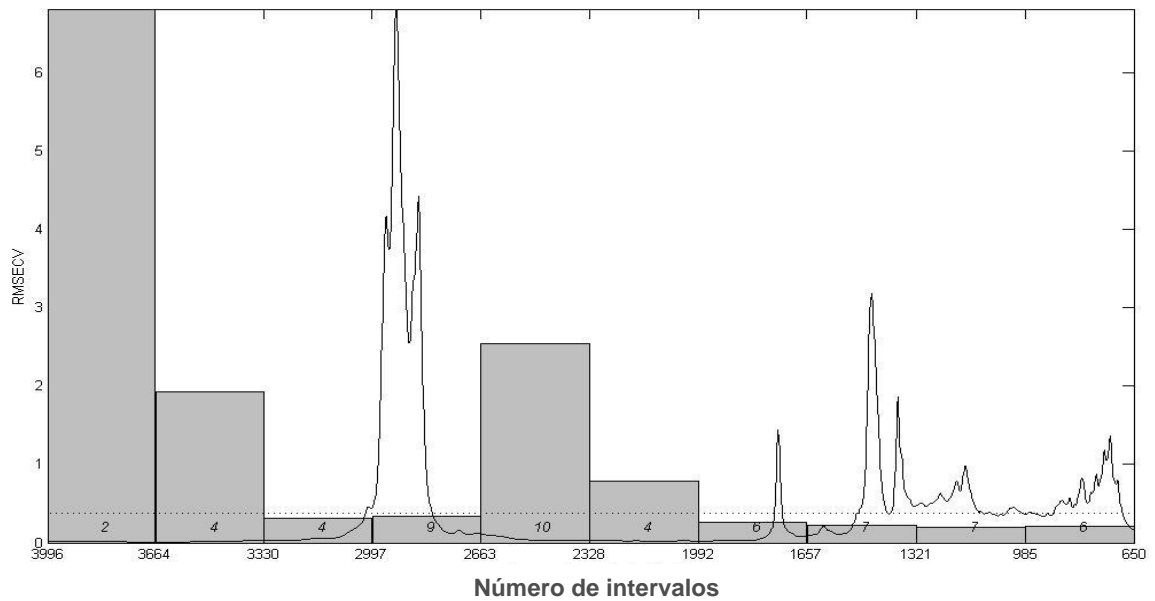


Figura 5 – Gráfico gerado pelo algoritmo *i*PLS, com os valores de RMSECV para o modelo construído com o espectro inteiro (linha pontilhada) e para os modelos construídos com os intervalos (barras). Os números dentro das barras correspondem ao número de variáveis latentes para cada intervalo. Os números abaixo das barras são os números de onda.

O método é planejado para dar uma visão geral dos dados e pode ser útil na interpretação de quais sinais (ou regiões) do espectro são mais representativos na construção de um bom modelo de calibração multivariada. Entretanto deve-se cuidar o tamanho do intervalo, pois, se ele for muito grande o modelo pode utilizar informações que não são representativas para o problema, ao passo que, se for muito pequeno, pode acontecer o inverso, isto é, não conseguir utilizar as informações necessárias para prever a propriedade de interesse (NORGAARD *et al.*, 2000).

Para eleger o modelo *i*PLS, com a região mais adequada, verifica-se qual o intervalo que produz o modelo com menor RMSECV, Figura 5. Para determinação do número de variáveis latentes, o RMSECV é calculado para valores crescentes de variáveis latentes. Como pode ser observado na Figura 6, há um determinado valor em que o incremento de variáveis latentes não resulta em um melhor desempenho de modelo, ou seja, não reduz o RMSECV.

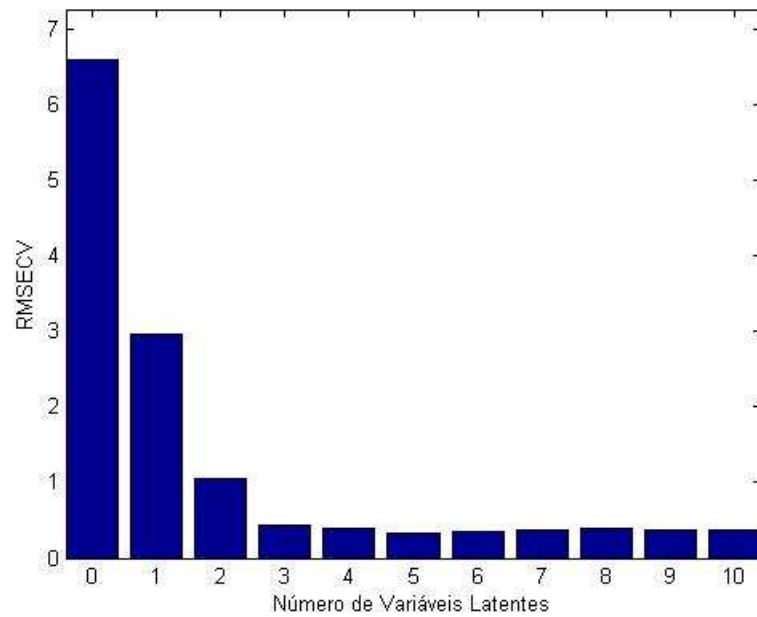


Figura 6 – Representação do comportamento das variáveis latentes de um modelo.

Além destes métodos lineares de calibração que são empregados na modelagem multivariada, pode-se utilizar a máquina de vetor de suporte com mínimos quadrados, que é um método não-linear, como uma ferramenta quimiométrica.

### 2.2.3 Máquina de vetor de suporte com mínimos quadrados

A máquina de vetor de suporte com mínimos quadrados (LS-SVM), que é um tipo de rede neural, conforme Figura 7, pode ser empregada como um método de calibração multivariada para conjuntos de dados lineares e não-lineares.

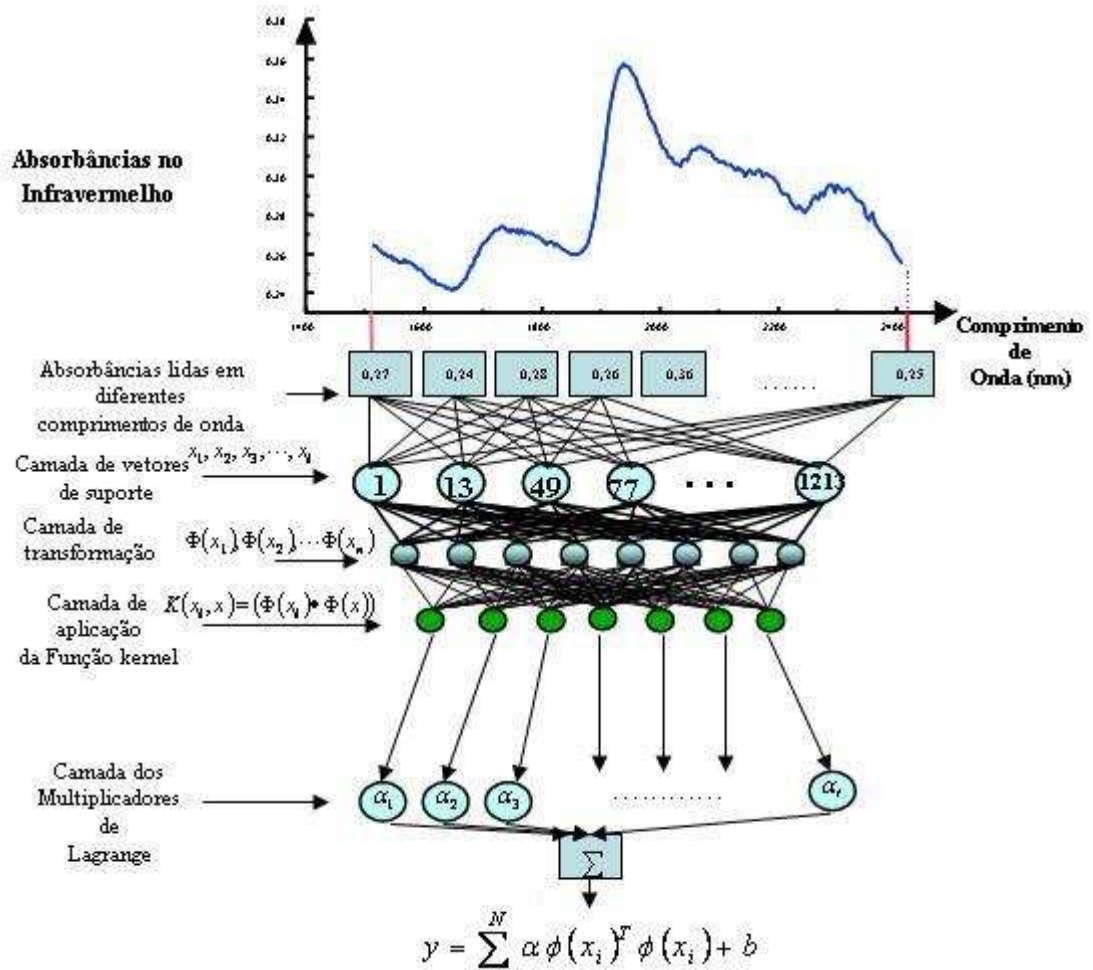


Figura 7 – Representação da LS-SVM como uma rede neural.  
Fonte: Adaptado de FERRÃO *et al.*, 2007.

A base da LS-SVM é um conjunto de dados de treinamento (vetores de suporte) com entradas  $n$ -dimensionais do vetor  $\mathbf{x}_i \in R^n$  e as saídas  $y_i \in R$ . Quando se empregam sinais espectroscópicos o modelo de regressão é construído, usando uma função de mapeamento não-linear, equações 8 ou 9, que mapeia os dados de entrada, para um espaço de característica de dimensão mais elevada (LIU *et al.*, 2008; WU *et al.*, 2008; AN *et al.*, 2007; BISHOP, 2006), buscando um ajuste da relação linear entre os dados espectrais e as propriedades de interesse, sendo o melhor ajuste aquele que minimizar a função custo descrita na equação 10:

$$N(x_i, x_j) = e^{\left( \frac{-\|x_i - x_j\|^2}{\sigma^2} \right)} \quad (8)^3$$

<sup>3</sup> A distância entre vetores, também conhecida por norma, é assim representada  $\|x_i - x_j\|$ .

$$N(x_i, x_j) = ((x_i x_j) + \theta)^d \quad (9)$$

Nas equações 8 e 9 os parâmetros  $\sigma^2$ ,  $\theta$  e  $d$  são ajustados pelo usuário (CHEN *et al.*, 2004). A variável  $\sigma^2$  controla o valor de erro de regressão da função e influencia diretamente no número inicial de autovalores e autovetores. Pequenos valores de  $\sigma^2$  geram um grande número de regressores e pode ocasionar um sobre-ajuste no modelo, e o contrário para grandes valores de  $\sigma^2$ . Esse parâmetro reflete a sensibilidade da LS-SVM (WU *et al.*, 2008).

$$C = \mathbf{w}^T \mathbf{w} + \frac{1}{2} \gamma \sum_{i=1}^N \mathbf{e}_i^2 \quad (10)$$

A otimização da rede é baseada na função custo. Sua primeira parte regulariza os pesos, já que pesos elevados deterioram a capacidade de generalização da rede. A segunda parte representa os erros de regressão para todos os dados de treinamento,  $\gamma$  é ajustado pelo usuário e fornece o peso relativo da segunda metade da equação 7, além de penalizar os erros altos,  $\mathbf{w}$  é o vetor de pesos,  $\mathbf{e}_i$  o vetor de mínimos quadrados (HAYKIN, 2001; BISHOP, 2006; CHEN *et al.*, 2004; FERRÃO, *et al.*, 2007, SCHÖLKOPF *et al.*, 1998).

Entretanto, existe uma restrição à função custo apresentada na equação 11:

$$y_i = \mathbf{w}^T \phi(x_i) + b + \mathbf{e}_i \text{ com } i = 1, \dots, N \quad (11)$$

onde  $\phi$  representa a função de linearização dos dados,  $\mathbf{w}$  o vetor de pesos e  $\mathbf{e}_i$  o vetor de erros. Para resolver esse problema de otimização convexa<sup>4</sup> é utilizado a função de Lagrange com seus multiplicadores (WU *et al.*, 2008; FERRÃO *et al.*, 2007), equação 12:

$$L(\mathbf{w}, b, \mathbf{e}_i, \boldsymbol{\alpha}_i) = \frac{1}{2} \|\mathbf{w}\|^2 + \gamma \sum_{i=1}^N \mathbf{e}_i^2 - \sum_{i=1}^N \boldsymbol{\alpha}_i \{ \mathbf{w}^T \phi(x_i) + b + \mathbf{e}_i - \mathbf{y}_i \} \quad (12)$$

onde  $\mathbf{y}_i$  corresponde ao vetor de respostas,  $\mathbf{e}_i$  o vetor de erros e  $\boldsymbol{\alpha}_i$  o vetor de multiplicadores de Lagrange (chamados vetores de suporte). Cada um dos multiplicadores corresponde a um dado de treinamento. Para obter a solução da equação 12 faz-se a derivada parcial de cada variável, equações 13, 14, 15 e 16.

$$\frac{\partial L(\mathbf{w}, b, \mathbf{e}_i, \boldsymbol{\alpha}_i)}{\partial \mathbf{w}} = 0 \therefore \mathbf{w} = \sum_{i=1}^N \boldsymbol{\alpha}_i \phi(x_i) \quad (13)$$

$$\frac{\partial L(\mathbf{w}, b, \mathbf{e}_i, \boldsymbol{\alpha}_i)}{\partial b} = 0 \therefore \sum_{i=1}^N \boldsymbol{\alpha}_i = 0 \quad (14)$$

---

<sup>4</sup> A otimização convexa não apresenta vários mínimos locais, mas sim uma única solução, um mínimo global.

$$\frac{\partial L(\mathbf{w}, b, \mathbf{e}_i, \boldsymbol{\alpha}_i)}{\partial \mathbf{e}_i} = 0 \therefore \boldsymbol{\alpha}_i = \gamma \mathbf{e}_i \quad i = 1, \dots, N \quad (15)$$

$$\frac{\partial L(\mathbf{w}, b, \mathbf{e}_i, \boldsymbol{\alpha}_i)}{\partial \boldsymbol{\alpha}_i} = 0 \therefore \mathbf{w}^T \phi(x_i) + b + \mathbf{e}_i - \mathbf{y}_i = 0 \quad i = 1, \dots, N \quad (16)$$

Os pesos ( $\mathbf{w}$ ) podem ser escritos como uma combinação linear dos multiplicadores de Lagrange correspondentes aos dados de treinamento. Como a maioria dos multiplicadores de Lagrange possui valores diferentes de zero, a maioria dos dados de treinamento contribui para a solução. Aplicando a equação 13 numa equação de reta ( $y = w^T \phi(x) + b$ ) se obtém a reta de regressão, equação 17 (WU *et al.*, 2008; FERRÃO *et al.*, 2007):

$$y = \sum_{i=1}^N \alpha_i \phi(x_i)^T \phi(x_i) + b \quad (17)$$

A determinação de um subconjunto ótimo da característica de entrada, a função núcleo apropriada e os parâmetros ótimos do núcleo são os elementos cruciais para construir um bom modelo de regressão de LS-SVM com exatidão e estabilidade elevadas para predição (YU *et al.*, 2009). O melhor treinamento garante o melhor resultado de predição (CHEN *et al.*, 2004).

A seguir é apresentado o pseudocódigo da superfície de otimização dos parâmetros  $\sigma^2$  e  $\gamma$  para um conjunto de espectros de infravermelho, utilizado para a construção de um modelo de calibração LS-SVM.

*Início*

*Forneça intervalo de valores para gama;*

*Forneça intervalo de valores para sigma;*

*Para 1 até número de intervalos para sigma faça*

*Para 1 até número de intervalos para gama faça*

*Para 1 até número de amostras faça*

*Deixe uma amostra dos espectros de fora;*

*Deixe a concentração dessa amostra de espectro de fora;*

*Construa o modelo com demais amostras e concentrações;*

*Treine a rede;*

*FimPara;*

*Estime a concentração da amostra;*

*Calcule o RMSECV;*

*FimPara;*

*FimPara;*

*Faça o gráfico da superfície de otimização (sigma, gama, RMSECV);*

*Fim.*

Fonte: BORIN *et al.*, 2006.

Na Figura 8 tem-se um exemplo de superfície de otimização dos parâmetros  $\sigma^2$  e  $\gamma$  em função do erro (RMSECV).

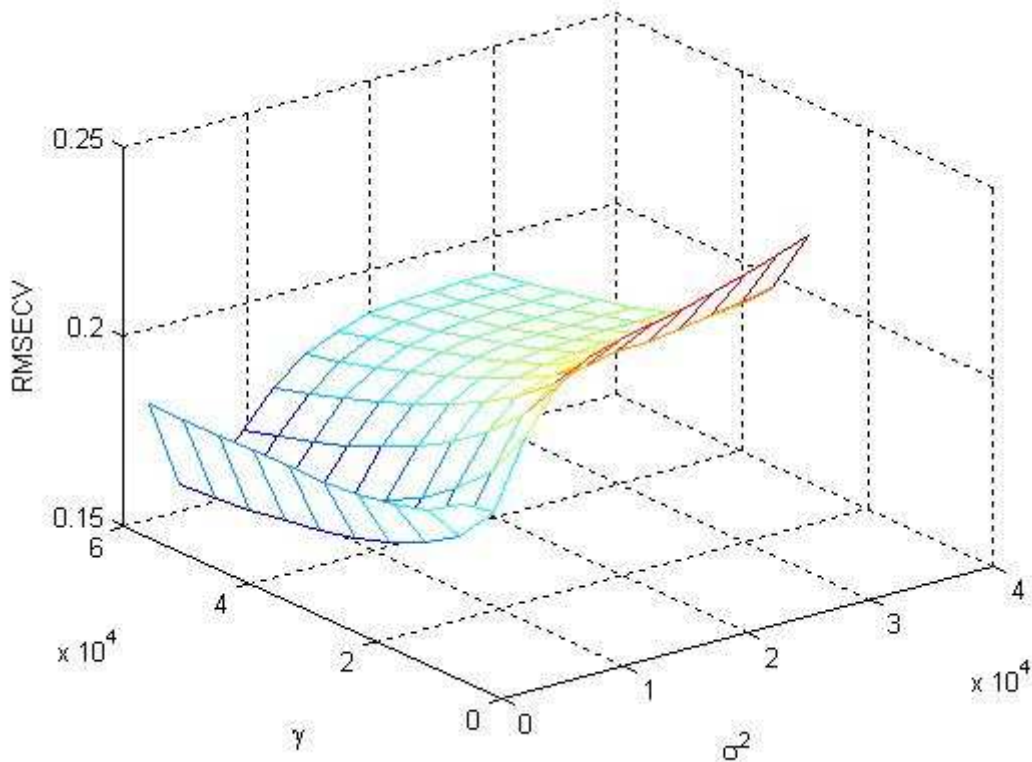


Figura 8 – Exemplo de superfície de otimização dos parâmetros  $\gamma$  e  $\sigma^2$  em função do erro (RMSECV).

Os valores dos erros de calibração localizam-se no plano vertical (RMSECV) e os parâmetros ( $\gamma$  e  $\sigma^2$ ) da máquina de vetor de suporte com mínimos quadrados estão situados no plano horizontal.

### 2.3 Transformada discreta *wavelets*

Tanto a transformada de Fourier quanto a transformada discreta *wavelets* (TDW) são funções matemáticas utilizadas no tratamento de sinais. A transformada de Fourier teve uma contribuição muito importante na química analítica. Essa técnica realiza tratamentos matemáticos nos sinais para serem usados na instrumentação analítica e no processamento computacional dos dados químicos. Contudo, em determinadas situações onde o sinal apresenta-se de forma descontínua ou se possui pontos muito agudos, assim como, no estudo de sinais não estacionários e na variação da largura das janelas espectrais, tanto na frequência como no tempo, a transformada de Fourier pode apresentar resultados não satisfatórios.

Já a transformada *wavelets* pode realizar o tratamento do sinal, mantendo as características originais, permitindo uma redução da dimensão do sinal de forma bastante considerável, independente do tempo e da frequência. Atuando diretamente na frequência do sinal analítico, o que contribui para aumentar a diferenciação entre sinal e ruído, e na conservação de picos estreitos com a respectiva informação, possibilitando resultados mais adequados (DAUBECHIES, 1992).

Do ponto de vista matemático, a função *wavelet*-mãe (*mother wavelet*) gera as funções *wavelets*-filhas que irão se decompor hierarquicamente conforme esquema apresentado na Figura 9. Para cada decomposição são gerados dois subsinais de comprimentos  $n/2$ , onde  $n$  representa o sinal inteiro. O primeiro subsinal corresponde aos subsinais *scalings* ( $A_1$ ), enquanto que o segundo aos subsinais *wavelets* ( $D_1$ ).

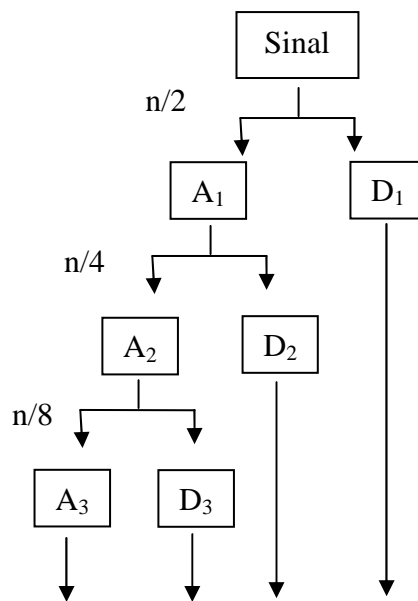


Figura 9 – Esquema do processo de decomposição das transformadas discretas *wavelets*.  
Fonte: Adaptado de MALLAT, 1989.

Para criar as funções *wavelets*-filhas, além de escolher o tipo de *wavelets* é necessário utilizar um banco de filtros, compostos por coeficientes especiais. Para tal são empregados filtros digitais, de passa-baixa, os quais produzem os coeficientes de aproximação do sinal ( $A_1$ ), e de passa-alta que produzem os coeficientes de detalhamento do sinal ( $D_1$ ). O ruído do sinal estará mais concentrado nessa segunda metade. Processando os coeficientes de aproximação ( $A_1$ ) novamente como dados de entrada com o mesmo banco de filtros, gera-se assim outro conjunto de coeficientes de aproximação e de detalhe com um nível de resolução inferior (DAUBECHIES, 1992).

Para ilustrar este processo, na Figura 10 (a) é apresentado um sinal de uma função, e na Figura 10 (b), o sinal da mesma função num primeiro nível de decomposição (transformada). Com a junção adequada dos subsinais *scalings* e *wavelets* da primeira transformada, pode-se recuperar o sinal original, implicando na preservação da informação desse sinal (DAUBECHIES, 1992).

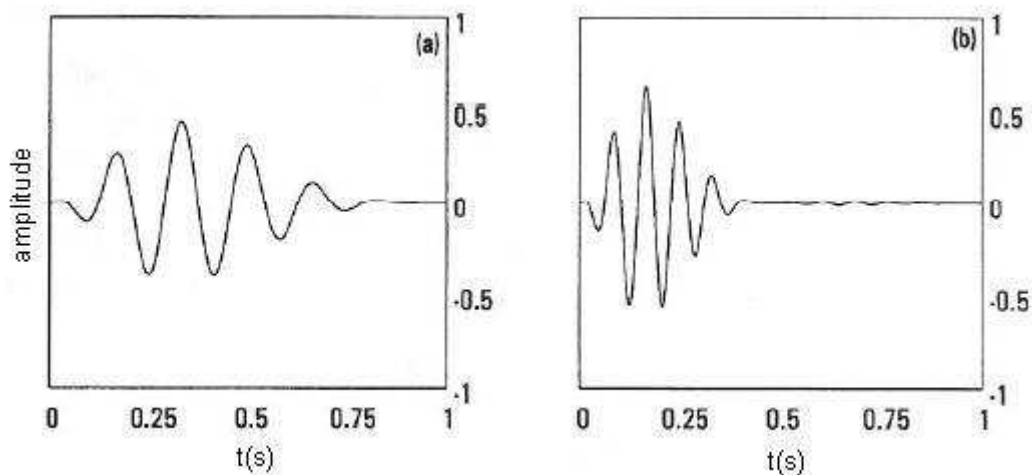


Figura 10 – Representação do sinal de uma função (a); representação do sinal da função com um nível de decomposição (b).

Fonte: WALKER, 1999.

Uma família de *wavelets* bastante comum é a TDW Daubechies com 4 coeficientes (db4). Para aplicar a TDW db4 em um sinal (S) são necessários esses coeficientes:

$$h_0 = \frac{1 + \sqrt{3}}{4\sqrt{2}} \quad h_1 = \frac{3 + \sqrt{3}}{4\sqrt{2}} \quad h_2 = \frac{3 - \sqrt{3}}{4\sqrt{2}} \quad h_3 = \frac{1 - \sqrt{3}}{4\sqrt{2}}$$

$$g_0 = h_3 \quad g_1 = -h_2 \quad g_2 = h_1 \quad g_3 = -h_0$$

que são dispostos numa matriz quadrada como a que segue:

$$\begin{bmatrix} h_0 & h_1 & h_2 & h_3 & 0 & 0 & 0 & 0 \\ g_0 & g_1 & g_2 & g_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & h_0 & h_1 & h_2 & h_3 & 0 & 0 \\ 0 & 0 & g_0 & g_1 & g_2 & g_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & h_0 & h_1 & h_2 & h_3 \\ 0 & 0 & 0 & 0 & g_0 & g_1 & g_2 & g_3 \\ h_2 & h_3 & 0 & 0 & 0 & 0 & h_0 & h_1 \\ g_2 & g_3 & 0 & 0 & 0 & 0 & g_0 & g_1 \end{bmatrix}$$

Desde 1989, a TDW tem sido aplicada em estudos químicos de análise por injeção em fluxo, cromatografia líquida de alta performance, espectrometria no infravermelho,



espectrometria de massa, espectrometria de ressonância magnética nuclear, espectrometria no UV-visível e voltametria para dados compactados e alisados (CHAU *et al.*, 2004) e espectroscopia NIR (infravermelho próximo) (CAI *et al.*, 2008; CAI *et al.*, 2010).

### 3 MÉTODO

Nesse capítulo é apresentado o método que foi utilizado para a realização desta pesquisa quantitativa. Para facilitar o entendimento, na Figura 11 é apresentado um fluxograma com as etapas que foram executadas.

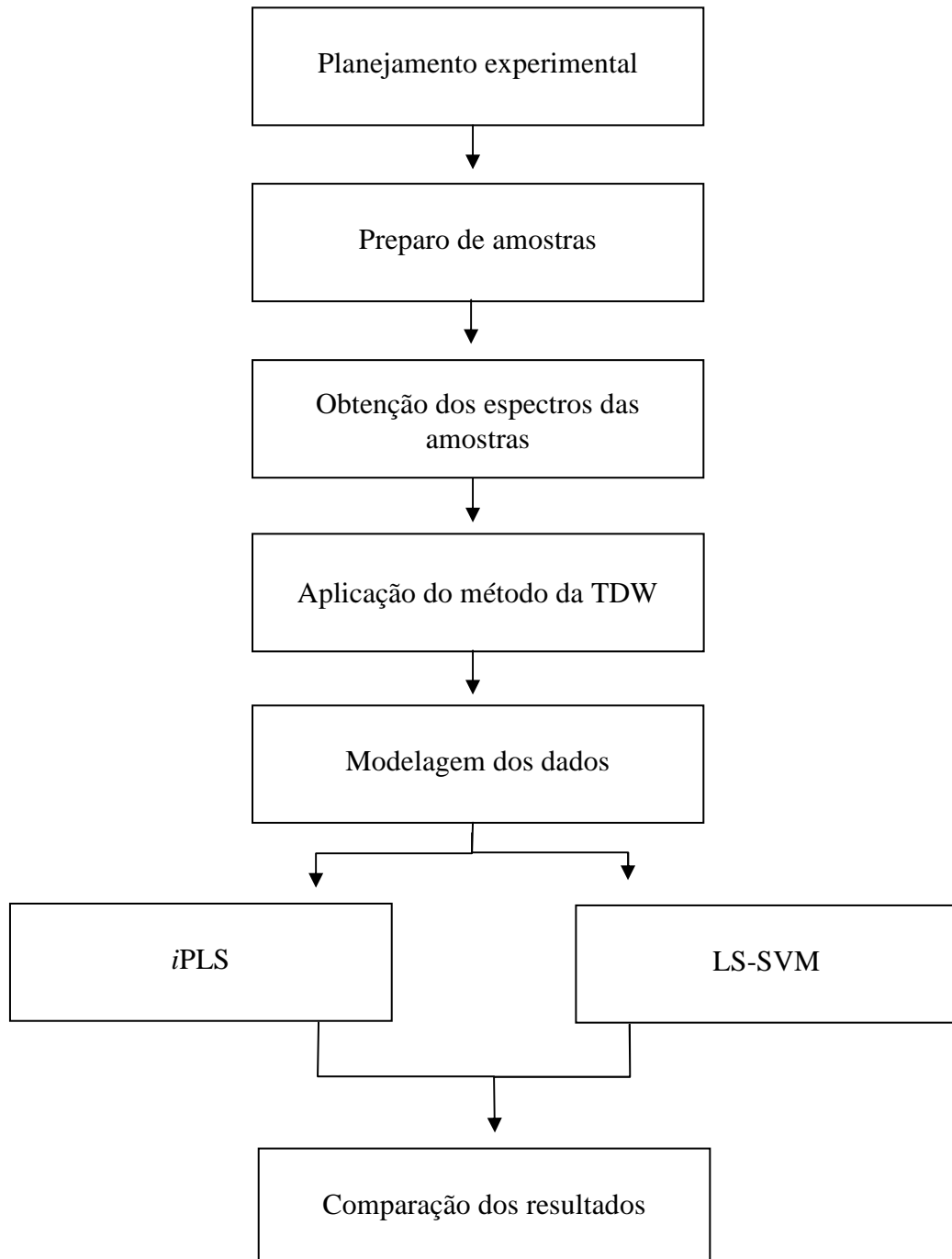


Figura 11 – Fluxograma das etapas executadas.

### 3.1 Planejamento experimental

Foram formuladas 45 blendas biodiesel/diesel com concentrações de biodiesel de soja que variaram de 8 – 30% (v/v), Tabela 2, com base na Norma NBR 15.568<sup>5</sup>. No presente estudo foi utilizada somente a faixa de 8 – 30% (v/v) devido ao fato de existirem poucos estudos na literatura para esta faixa e como apresentado anteriormente, a maioria dos países da América Latina, vêm ampliando a quantidade adicionada do biodiesel ao diesel. Adicionalmente, no Brasil existem condições de suprir a demanda para utilização do B20 na região metropolitana das principais cidades nos próximos anos e em todo o país em 2020 (UBRABIO – União Brasileira do Biodiesel, 2010). A ANP, por sua vez, já autoriza o uso de B6 a B20 em frota cativa e em equipamentos instalados em empresas, bem como, teores acima de 20% de biodiesel desde que o fabricante da frota mantenha a garantia dos motores (DE FREITAS, 2010). Em Porto Alegre, até 2016, os ônibus do transporte coletivo deverão utilizar o B100, segundo um projeto de lei da Câmara Municipal aprovado no fim de 2010. Tendo um prazo de 5 anos, a partir da sanção do texto, para a implantação gradual do biodiesel (DE FREITAS, 2011). O biodiesel utilizado na composição das blendas foi cedido pela OLEOPLAN S.A. Os óleos diesel, interior e metropolitano (este possui menor teor de enxofre), empregados nas formulações das 45 blendas foram obtidos na Refinaria Alberto Pasqualini – REFAP S/A. Os espectros dessas blendas foram coletadas no espectrofotômetro de infravermelho por transformada de Fourier Shimadzu modelo IR Prestige-21, localizado no laboratório de química da UFRGS, e no espectrofotômetro de infravermelho por transformada de Fourier Nicolet Magna 550, localizado no laboratório de espectroscopia da UNISC.

---

<sup>5</sup> Biodiesel – Determinação do teor de biodiesel em óleo diesel por espectroscopia na região do infravermelho médio. Ela prevê o emprego de duas regiões espectrais  $1.100\text{ cm}^{-1} - 1.200\text{ cm}^{-1}$  e  $1.735\text{ cm}^{-1} - 1.750\text{ cm}^{-1}$ , para a construção dos modelos multivariados, aceitando erros de previsão até 1% para as concentrações de 8 % (v/v) à 30 % (v/v).

Tabela 2 – Concentrações das blendas biodiesel/diesel.

<b>Biodiesel (%v/v)</b>	<b>Diesel (%v/v)</b>
8,04	91,96
8,46	91,54
8,95	91,05
9,54	90,46
10,03	89,97
10,53	89,47
11,03	88,97
11,54	88,46
11,97	88,03
12,43	87,57
12,95	87,05
13,45	86,55
13,95	86,05
14,42	85,58
14,98	85,02
15,44	84,56
16,03	83,97
16,53	83,47
17,09	82,91
17,60	82,40
18,06	81,94
18,56	81,44
18,88	81,12
19,38	80,62
20,05	79,95
20,55	79,45
21,06	78,94
21,57	78,43
21,93	78,07
22,40	77,60
22,91	77,09
23,41	76,59
23,91	76,09
24,41	75,59
24,88	75,12
25,62	74,38
26,03	73,97
26,53	73,47
27,13	72,87
27,38	72,62
28,01	71,99
28,51	71,49
28,74	71,26
29,23	70,77
30,08	69,92

Fonte: Elaborada pela autora.

### 3.2 Preparo de amostras

Os balões volumétricos de 25 ml utilizados como recipientes passaram por uma lavagem e após foram secos. Posteriormente foi colocada água até o menisco de cada balão e realizada a pesagem do mesmo. Para terminar a aferição do balão foi realizado um cálculo de correção com base na temperatura e variação da densidade.

### 3.3 Obtenção dos espectros das amostras

As amostras foram submetidas à homogeneização antes da aquisição dos espectros, sendo os mesmos obtidos em duplicata na faixa de  $4.000\text{ cm}^{-1}$  –  $650\text{ cm}^{-1}$ , com resolução de  $4\text{ cm}^{-1}$  e 32 varreduras, utilizando um acessório de reflexão total atenuada horizontal da Pike (HATR) com cristal de ZnSe, acoplado num espectrofotômetro da Nicolet Magma 550 e num espectrofotômetro Shimadzu IR Prestige-21. Para a construção da matriz de dados, os espectros foram normalizados e posteriormente efetuou-se a média das duplicatas desses espectros.

### 3.4 Aplicação da transformada discreta *wavelet*

Foi empregada a transformada discreta *wavelet* Daubechies com 4, 6 e 8 coeficientes para decompor a matriz de dados, em 2 níveis, e para a remoção de ruídos, sendo utilizados somente os subsinais *scalings* para posterior construção dos modelos multivariados. O computador utilizado foi um ACER – ASPIRE 4732Z-4620 com processador Pentium(R) Dual-Core CPU T4300 2.10 GHz 2.09 GHz, 2,92 GB de RAM.

### 3.5 Modelagem dos dados

As matrizes de dados espectrais transformados e não transformados foram divididas em dois grupos: calibração e previsão. Nessa divisão 2/3 dos dados integrou o conjunto de calibração, incluindo dados com valores extremos. No conjunto de dados de previsão foram colocados os dados restantes.

Antes da construção dos modelos multivariados os dados foram auto-escalados, onde cada valor de absorção (ou  $\log(1/R)$ ), de cada número de onda, é subtraído da média, e dividido pelo desvio-padrão, respectivo.

Para o tratamento dos dados, tanto para aplicação do método da TDW, como a modelagem dos dados, foi empregado o *software* MATLAB versão 6.1 (The Math Works, Natick, USA). Para desenvolver os modelos *i*PLS foi aplicado o pacote *iToolbox* (NORGAARD *et al.*, 2000) e para os modelos LS-SVM foi empregado o pacote LS-SVMlab1.5 (SUYKENS *et al.*, 2002).

Com a identificação das principais regiões dos espectros através do *i*PLS, foram construídos modelos multivariados utilizando a regressão LS-SVM. Para validar os modelos LS-SVM foi empregado o grupo das amostras de previsão. Isto é, foi avaliado se o comportamento desse grupo de dados é semelhante ao do grupo de calibração, utilizado para construir o modelo. Foram construídos modelos utilizando sinais não transformados e transformados com *wavelets*. O computador utilizado foi um ACER – ASPIRE 4732Z-4620 com processador Pentium(R) Dual-Core CPU T4300 2.10 GHz 2.09 GHz, 2,92 GB de RAM.

### **3.8 Comparação dos resultados**

Os resultados dos modelos multivariados obtidos via *i*PLS foram comparados aos obtidos pela regressão por LS-SVM modelados nas mesmas faixas espectrais para os dados não transformados. Uma comparação similar foi realizada entre os modelos multivariados obtidos por *i*PLS e LS-SVM, para as diferentes TDW empregadas na decomposição dos dados. Posteriormente foi realizada uma avaliação entre resultados obtidos neste trabalho com outros estudos relacionados já realizados.

## 4 RESULTADOS

### 4.1 Comparação dos resultados obtidos com aplicação do método da transformada discreta *wavelet*

Na redução do número de variáveis espectrais originais foi empregada a transformada discreta *wavelet* (TDW) Daubechies. Foi utilizado 3 tipos de transformadas Daubechies: com 4, 6 e 8 coeficientes. Para cada um deles foi realizado 2 níveis de decomposição, isto é, os dados com 1 nível de decomposição foram novamente decompostos.

A seguir é apresentada uma sequência de Figuras com os dados espectrais originais (sem aplicação das decomposições) e decompostos com 1 e 2 níveis de decomposição por 4, 6 e 8 coeficientes. Além do percentual de informação retida por cada um deles.

Na Figura 12 tem-se os espectros originais e, portanto com 100% de informação.

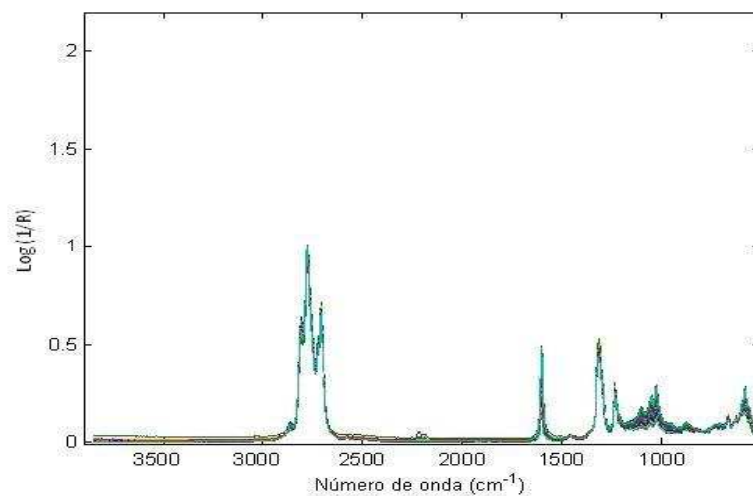


Figura 12 – Sinal dos espectros das blendas biodiesel/diesel originais.

Nas Figuras 13 e 14, respectivamente, tem-se os dados espectrais com 1 e 2 níveis de decomposição obtidos com a aplicação de 4 coeficientes.

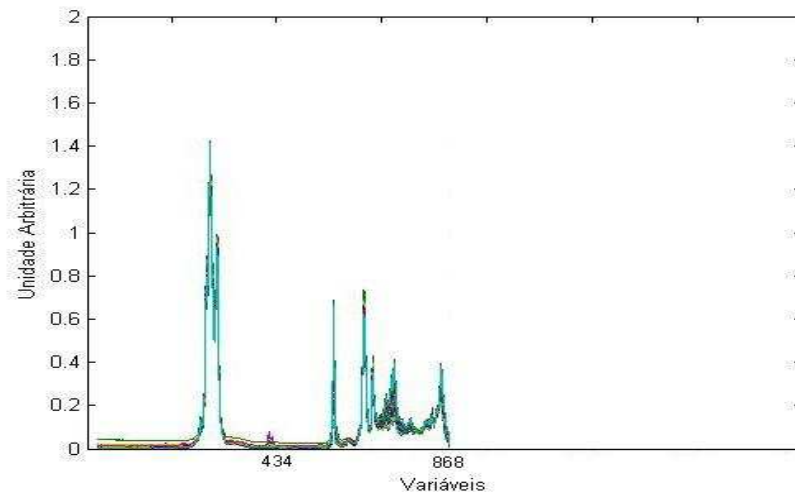


Figura 13 – Sinal dos espectros de amostras de blendas biodiesel/diesel com aplicação de 4 coeficientes, obtendo 1 nível de decomposição, contendo 868 novas variáveis, e retendo 99,93% de informação.

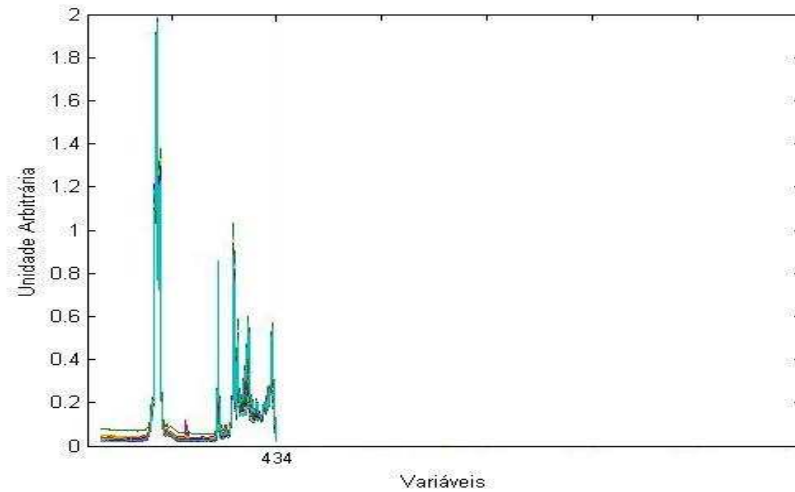


Figura 14 – Sinal dos espectros de amostras de blendas biodiesel/diesel com aplicação de 4 coeficientes obtendo o segundo nível de decomposição, contendo 434 novas variáveis, e retendo 99,61% de informação.

Nas Figuras 15 e 16, respectivamente, tem-se os dados espectrais com 1 e 2 níveis de decomposição obtidos com a aplicação de 6 coeficientes.



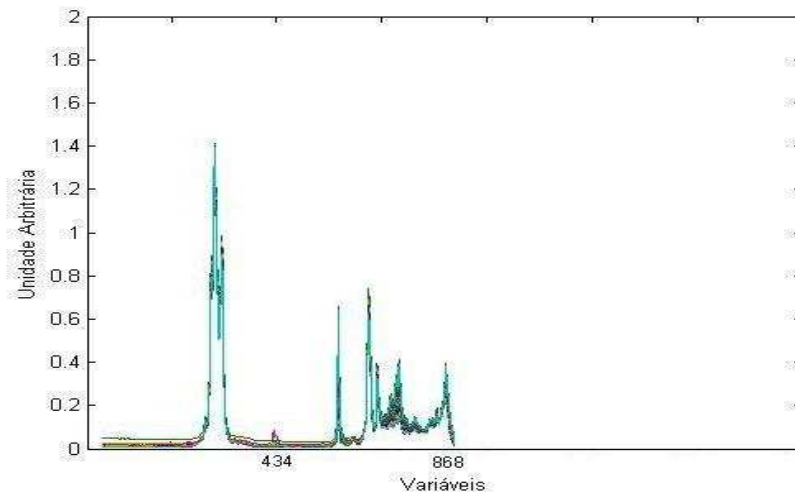


Figura 15 – Sinal dos espectros de amostras de blendas biodiesel/diesel com aplicação de 6 coeficientes obtendo 1 nível de decomposição, contendo 868 novas variáveis, retendo 99,36% de informação.

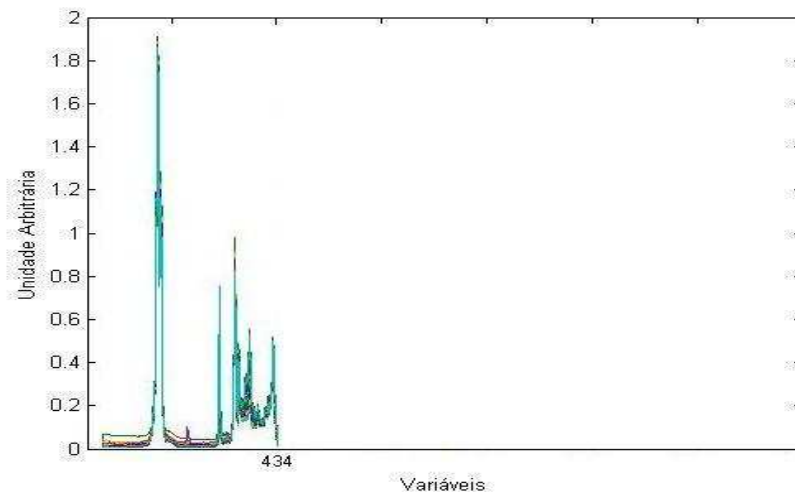


Figura 16 – Sinal dos espectros de amostras de blendas biodiesel/diesel com aplicação de 6 coeficientes obtendo o segundo nível de decomposição, contendo 434 novas variáveis, retendo 97,42% de informação.

Nas Figuras 17 e 18, respectivamente, tem-se os dados espectrais com 1 e 2 níveis de decomposição obtidos com a aplicação de 8 coeficientes.

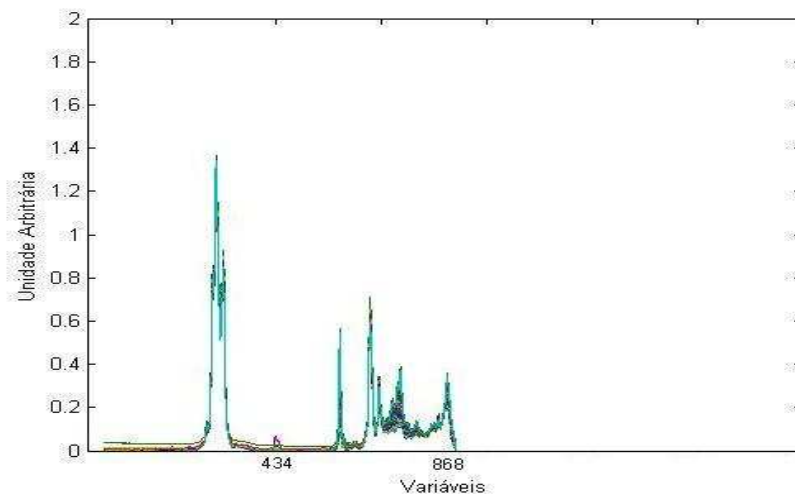


Figura 17 – Sinal dos espectros de amostras de blendas biodiesel/diesel com aplicação de 8 coeficientes obtendo 1 nível de decomposição, contendo 868 novas variáveis, retendo 98,31% de informação.

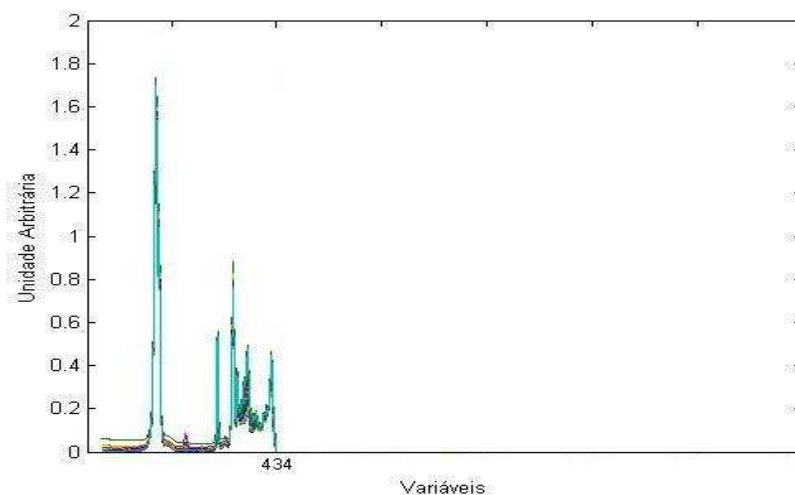


Figura 18 – Sinal dos espectros de amostras de blendas biodiesel/diesel com aplicação de 8 coeficientes obtendo o segundo nível de decomposição, contendo 434 novas variáveis, retendo 94,04% de informação.

Observa-se que ao aplicar o método da TDW no sinal, seja em 1 ou 2 níveis, com 4, 6 ou 8 coeficientes, ele fica compactado e com sua amplitude aumentada. Ao observar apenas os sinais resultantes das segundas decomposições, seja com 4, 6 ou 8 coeficientes, verifica-se que ao aumentar o número de coeficientes a amplitude do sinal fica menor.

#### 4.2 Resultados para determinação do biodiesel em blendas biodiesel/diesel

Neste estudo foi utilizada somente a faixa de 8 – 30% (v/v) em virtude de existirem poucos estudos na literatura para teores de biodiesel acima de 5% (v/v), e também por muitos países, assim como o Brasil, estarem ampliando a quantidade adicionada do biodiesel ao

diesel. Além disso, no Brasil existem condições de suprir a demanda para utilização do B20 na região metropolitana das principais cidades nos próximos anos e em todo o país em 2020 (UBRABIO – União Brasileira do Biodiesel, 2010). A ANP, por sua vez, já autoriza o uso de B6 a B20 em frota cativa e em equipamentos instalados em empresas, bem como, teores acima de 20% de biodiesel desde que o fabricante da frota mantenha a garantia dos motores (DE FREITAS, 2010). Em Porto Alegre, até 2016, os ônibus do transporte coletivo deverão utilizar o B100, segundo um projeto de lei da Câmara Municipal aprovado no fim de 2010. Tendo um prazo de 5 anos, a partir da sanção do texto, para a implantação gradual do biodiesel (DE FREITAS, 2011).

Para a determinação do biodiesel em blendas biodiesel/diesel foram gerados modelos com os dados originais e também com os dados decompostos com 4, 6 e 8 coeficientes da *wavelet* Daubechies. Para a construção dos modelos os espectros foram subdivididos em 10, 15 ou 20 intervalos eqüidistantes a fim de identificar qual a melhor região espectral para determinar o biodiesel em blendas biodiesel/diesel, com o pré-processamento de auto-escalar os dados.

Para melhor compreensão do leitor sobre a subdivisão dos espectros para a construção dos modelos *i*PLS são apresentados na sequência três Figuras. Nas Figuras 19, 20 e 21 tem se a subdivisão dos espectros em 10, 15 e 20 intervalos, respectivamente, onde para cada intervalo foi construído um modelo de PLS (regressão por mínimos quadrados), além do modelo construído empregando todo o espectro. Na regressão PLS é construído um modelo de regressão linear buscando estimar a concentração de biodiesel a partir dos espectros medidos. A subdivisão dos espectros é realizada para poder identificar qual faixa espectral é mais indicada para a construção de um bom modelo preditivo.

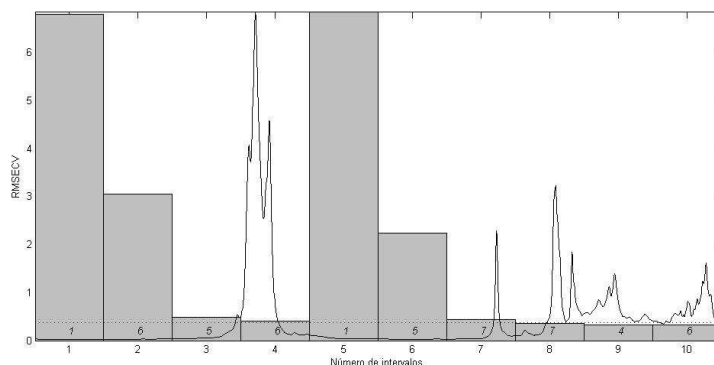


Figura 19 – Sinal dos espectros das blendas biodiesel/diesel subdivididos em 10 intervalos. Para esse caso os intervalos 9 e 10 apresentam os menores erros (RMSECV) indicado pela altura da barra, a linha pontilhada corresponde ao erro (RMSECV) para o modelo que emprega o espectro inteiro.

Quando o espectro foi subdividido em 10 intervalos (Figura 19) os intervalos 9 e 10 apresentaram menores RMSECV quando são empregadas 4 e 6 variáveis latentes, respectivamente.

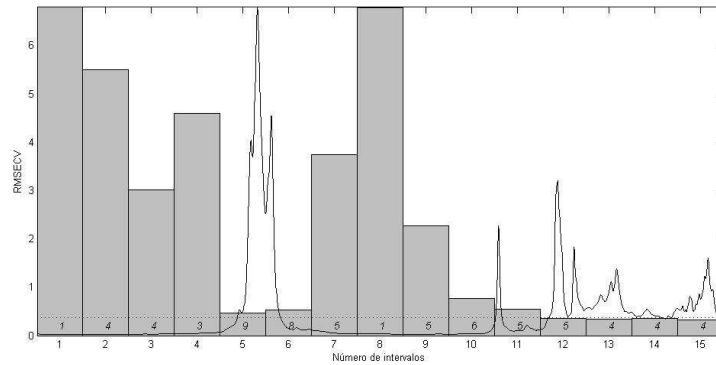


Figura 20 – Sinal dos espectros das blendas biodiesel/diesel subdivididos em 15 intervalos. Para esse caso os intervalos 13 e 15 apresentam os menores erros (RMSECV) indicados pela altura das barras, a linha pontilhada corresponde ao erro (RMSECV) para o modelo que emprega o espectro inteiro.

Para a subdivisão em 15 intervalos (Figura 20) os intervalos 13 e 15, quando empregaram 4 variáveis latentes, apresentam os menores EMSECV.

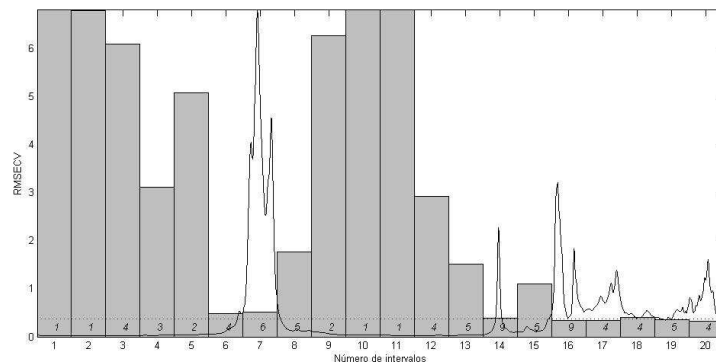


Figura 21 – Sinal dos espectros das blendas biodiesel/diesel subdivididos em 20 intervalos. Para esse caso os intervalos 16, 17, 19 e 20 apresentam os menores erros (RMSECV) indicados pela altura das barras, a linha pontilhada corresponde ao erro (RMSECV) para o modelo que emprega o espectro inteiro.

Já para a subdivisão do espectro em 20 intervalos (Figura 21) os intervalos 16, 17, 19 e 20 apresentam os menores RMSECV empregando 9, 4, 5 e 4 variáveis latentes, respectivamente.

Na Tabela 3 são apresentados os resultados dos melhores modelos *i*PLS, empregando dados originais e com aplicação da TDW, além do modelo que emprega o espectro inteiro.

Tabela 3 – Resultados dos melhores modelos *i*PLS para a quantificação de biodiesel em blendas biodiesel/diesel com espectros originais e decompostos em até 2 níveis com 4, 6 e 8 coeficientes.

N	C	IR	ND	NVT	NI	QVI	MODELOS <i>i</i> PLS								
							FE (cm <sup>-1</sup> )	EC1	EC2	EC3	VL	EP1	EP2	EP3	R <sup>2</sup> <sub>cal</sub>
A	0	100	0	1736	1	1736	4000 – 650	0,335	0,390	0,269	5	0,445	0,524	0,368	0,999
B	0	100	0	1736	10	174	1321– 985	0,333	0,389	0,265	4	0,332	0,352	0,311	0,999
C	0	100	0	1736	15	116	1321– 1097	0,340	0,402	0,263	4	0,359	0,397	0,316	0,999
D	0	100	0	1736	20	87	1321– 1153	0,348	0,414	0,265	4	0,332	0,343	0,320	0,999
E	4	99,93	1	868	20	44	1331 – 1161	0,348	0,409	0,273	5	0,330	0,313	0,347	0,999
F	4	99,61	2	434	20	22	1165 – 995	0,356	0,413	0,288	2	0,383	0,422	0,339	0,999
G	6	99,36	1	868	20	44	1331 – 1161	0,347	0,407	0,276	5	0,328	0,312	0,343	0,999
H	6	97,42	2	434	20	22	1165 – 995	0,368	0,440	0,278	2	0,372	0,410	0,330	0,998
I	8	98,31	1	868	20	44	1331 – 1161	0,349	0,407	0,270	6	0,333	0,329	0,337	0,999
J	8	94,04	2	434	20	22	1165 – 995	0,367	0,440	0,274	5	0,370	0,413	0,321	0,998

N – nome do modelo.

C – número de coeficientes.

IR – informação retida em percentual.

ND – número de decomposição.

NVT – número de variáveis totais.

NI – número de intervalos.

QVI – quantidade de variáveis por intervalo.

FE – faixa espectral.

EC1 – erros de validação cruzada (RMSECV), em percentual, com dados lidos nos dois espectrofotômetros.

EC2 – erros de validação cruzada (RMSECV), em percentual, com dados lidos no espectrofotômetro Shimadzu.

EC3 – erros de validação cruzada (RMSECV), em percentual, com dados lidos no espectrofotômetro Nicolet.

VL – variáveis latentes.

EP1 – erros de previsão (RMSEP), em percentual, com dados lidos nos dois espectrofotômetros.

EP2 – erros de previsão (RMSEP), em percentual, com dados lidos no espectrofotômetro Shimadzu.

EP3 – erros de previsão (RMSEP), em percentual, com dados lidos no espectrofotômetro Nicolet.

R<sup>2</sup><sub>cal</sub> – correlação dos modelos de calibração.

Fonte: Elaborado pela autora.

Uma análise inicial permite constatar que todos os erros de calibração são menores para os modelos construídos com dados obtidos pelo espectrofotômetro Nicolet em comparação aos Shimadzu. Em relação aos erros de previsão, a maioria são menores para os modelos *i*PLS construídos com os espectros do espectrofotômetro Nicolet em comparação aos do Shimadzu. Isso deve estar relacionado, provavelmente, à realização das leituras das amostras em espectrofotômetros diferentes e por operadores diferentes, sendo conseqüentemente gerados erros aleatórios e sistemáticos de forma diferenciada. Além disso, estas diferenças também podem ser atribuídas as diferentes fontes de ruído, já que os equipamentos encontravam-se em laboratórios distintos em distintas instituições.

Na modelagem *i*PLS em conjunto dos espectros Shimadzu e Nicolet verifica-se uma uniformidade entre os erros de validação cruzada e os erros de previsão. Indicando a capacidade preditiva dos modelos. Também é possível verificar que a decomposição dos

espectros em novos conjuntos de dados, com dimensões menores, seja com 1 ou 2 níveis de decomposição empregando 4, 6 ou 8 coeficientes, não prejudica a capacidade dos modelos construídos a partir destes dados. Dessa forma é possível construir modelos em menor tempo e com a mesma capacidade preditiva dos modelos construídos a partir dos dados originais. Para identificar qual espectrofotômetro teria melhor desempenho frente ao outro quando modelados seus espectros em conjunto, calculou-se em separado os erros de validação cruzada e de previsão para ambos e percebeu-se que de maneira geral os espectros Nicolet apresentaram menores erros, tanto de validação cruzada quanto de previsão, que os espectros Shimadzu, indicando que de uma forma geral os espectros Nicolet tem melhor desempenho frente aos espectros Shimadzu quando modelados em conjunto.

Pode ser constatado nos resultados apresentados que a utilização de uma determinada faixa do espectro para a construção de um modelo produz melhores resultados do que a utilização do espectro inteiro.

Na sequência têm-se as Figuras 22, 23 e 24 com as curvas de correlação para os valores de referência e os valores previstos pelos modelos *i*PLS construídos com o espectro inteiro, sendo a primeira utilizando os dados originais de ambos os espectrofotômetros, a segunda utilizando somente os dados do espectrofotômetro Shimadzu e a terceira utilizando os dados do espectrofotômetro Nicolet.

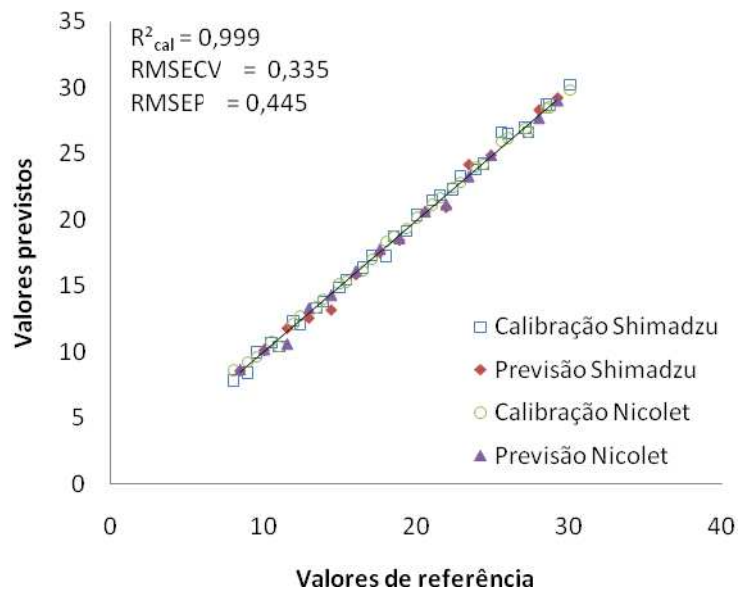


Figura 22 – Modelo construído com o espectro inteiro (faixa espectral de  $4.000\text{ cm}^{-1} - 650\text{ cm}^{-1}$ ) empregando 5 variáveis latentes com os espectros obtidos nos dois espectrofotômetros.

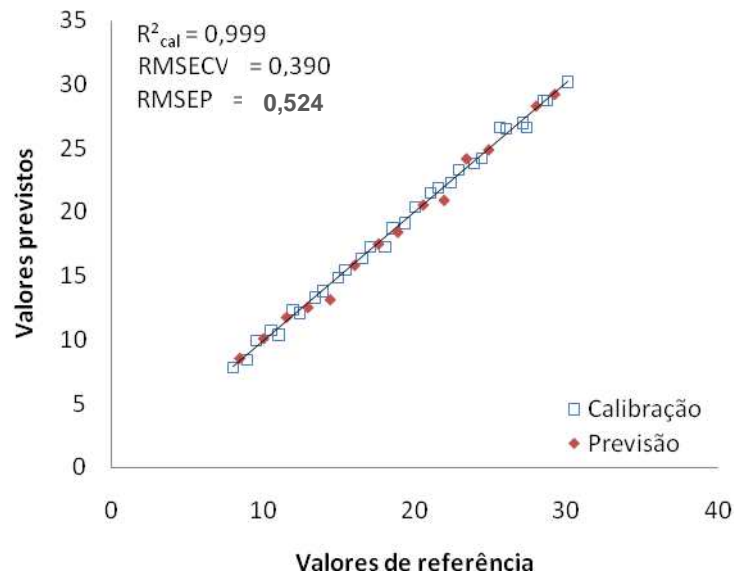


Figura 23 – Modelo construído com o espectro inteiro (faixa espectral de  $4.000\text{ cm}^{-1} - 650\text{ cm}^{-1}$ ) empregando 5 variáveis latentes com espectros obtidos no espectrofotômetro Shimadzu.

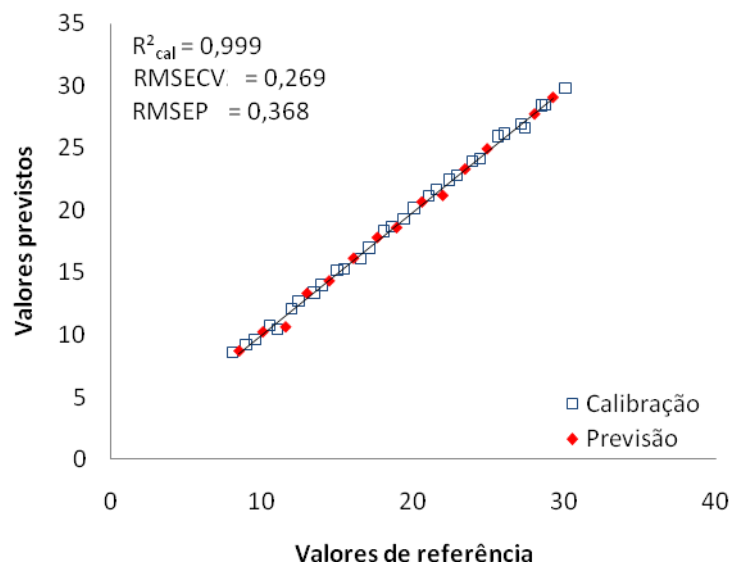


Figura 24 – Modelo construído com o espectro inteiro (faixa espectral de  $4.000\text{ cm}^{-1} - 650\text{ cm}^{-1}$ ) empregando 5 variáveis latentes com espectros obtidos no espectrofotômetro Nicolet.

Nas Figuras 25, 26 e 27, respectivamente, têm-se as curvas de correlação entre os valores de referência e os valores previstos pelos modelos *i*PLS que empregam a subdivisão dos espectros originais em 10 intervalos. Sendo que a Figura 25 emprega os espectros dos dois espectrofotômetros, já as Figuras 26 e 27 empregam os dados dos espectrofotômetros Shimadzu e Nicolet, respectivamente.

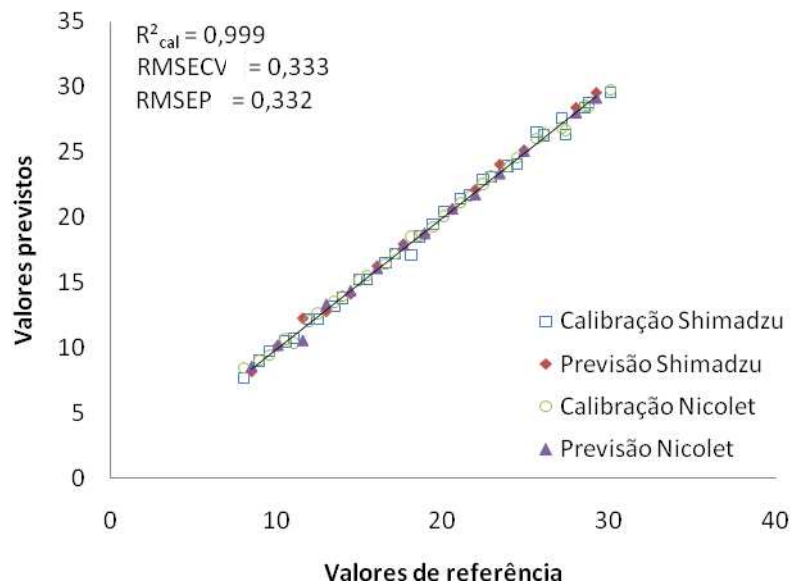


Figura 25 – Modelo *i*PLS subdividindo os espectros originais em 10 intervalos, empregando faixa espectral de  $1.321\text{ cm}^{-1} - 985\text{ cm}^{-1}$ , 4 variáveis latentes e com dados lidos nos dois espectrofotômetros.

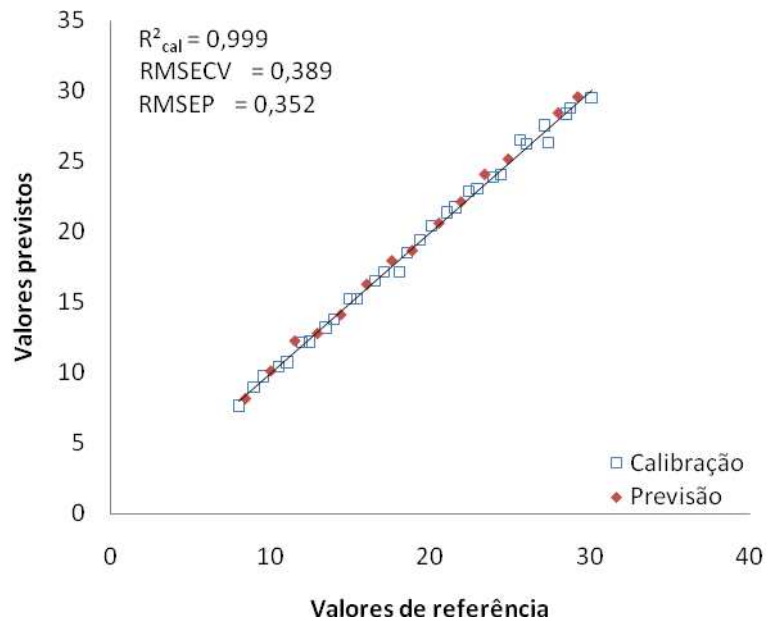


Figura 26 – Modelo *i*PLS subdividindo os espectros originais em 10 intervalos, empregando faixa espectral de  $1.321\text{ cm}^{-1} - 985\text{ cm}^{-1}$ , 4 variáveis latentes e com dados lidos no espectrofotômetro Shimadzu.



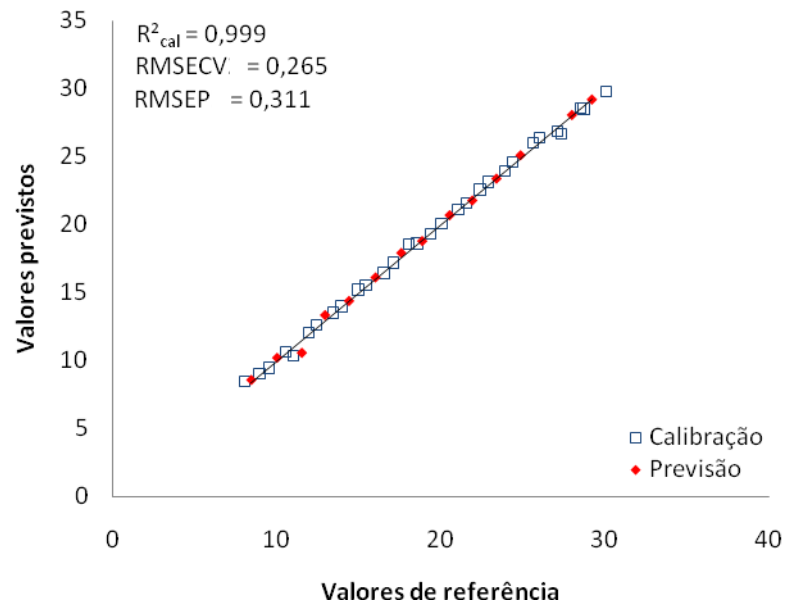


Figura 27 – Modelo *i*PLS subdividindo os espectros originais em 10 intervalos, empregando faixa espectral de  $1.321\text{ cm}^{-1} - 985\text{ cm}^{-1}$ , 4 variáveis latentes e com dados lidos no espectrofotômetro Nicolet.

Nas Figuras 28, 29 e 30 têm-se as curvas de correlação entre os valores de referência e os valores previstos pelos modelos *i*PLS que empregam a subdivisão dos espectros em 20 intervalos e com 2 níveis de decomposição através da TDW Daubechies com 4 coeficientes, sendo a primeira utilizando os dados de ambos os espectrofotômetros, a segunda utilizando somente os dados do espectrofotômetro Shimadzu e a terceira utilizando os dados do espectrofotômetro Nicolet.

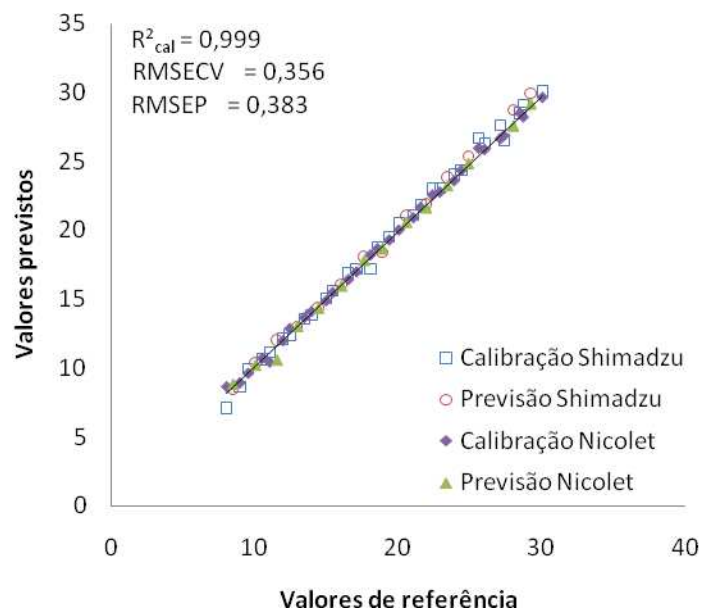


Figura 28 – Modelo *i*PLS com subdivisão dos espectros em 20 intervalos, empregando a faixa espectral de  $1.165\text{ cm}^{-1} - 995\text{ cm}^{-1}$ , 2 variáveis latentes, com 2 níveis de decomposição e construído com dados lidos nos dois espectrofotômetros.

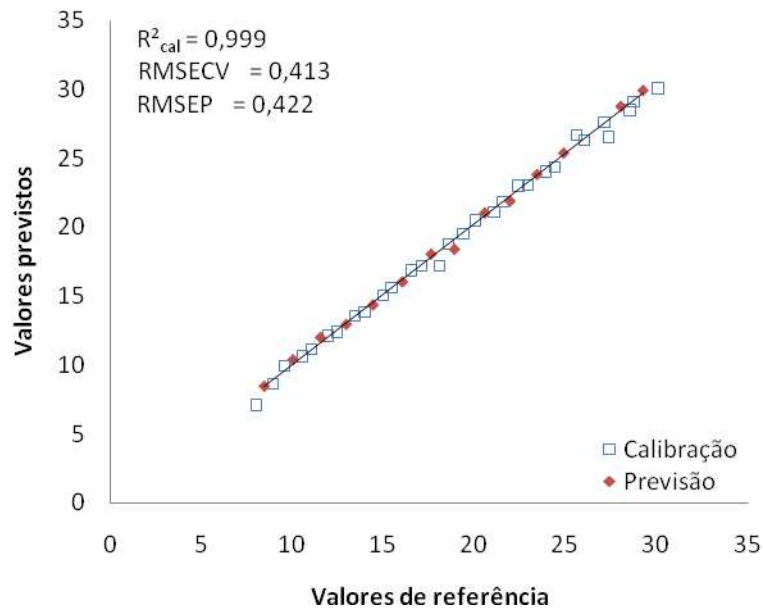


Figura 29 – Modelo *i*PLS com subdivisão dos espectros em 20 intervalos, empregando a faixa espectral de  $1.165\text{ cm}^{-1} - 995\text{ cm}^{-1}$ , 2 variáveis latentes, com 2 níveis de decomposição e construído com dados lidos no espectrofotômetro Shimadzu.

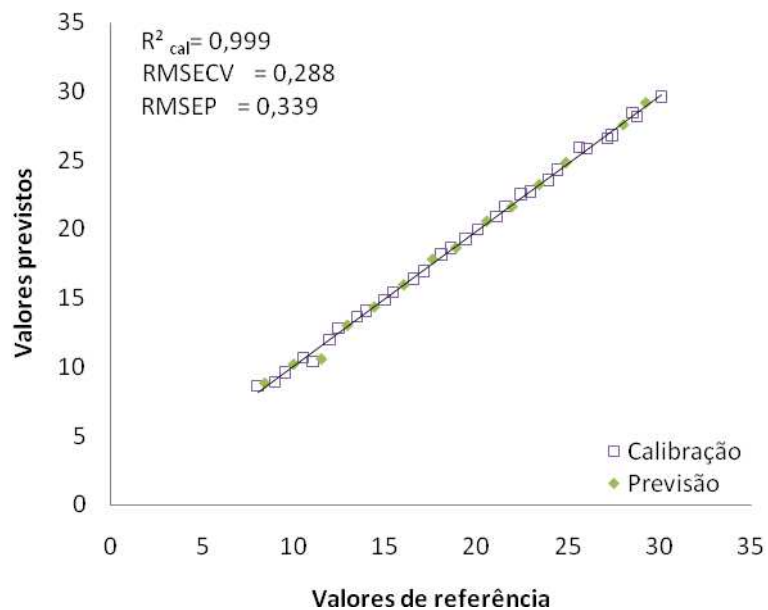


Figura 30 – Modelo *i*PLS com subdivisão dos espectros em 20 intervalos, empregando a faixa espectral de  $1.165\text{ cm}^{-1} - 995\text{ cm}^{-1}$ , 2 variáveis latentes, com 2 níveis de decomposição e construído com dados lidos no espectrofotômetro Nicolet.

Podemos observar na Tabela 3 que independente da quantidade de coeficientes empregados na decomposição dos espectros, bem como, do número de decomposições, além do número de intervalos no qual o espectro foi subdividido, a região selecionada refere-se ao sinal do  $-C-O-C(O)-$  presente nos ésteres (biodiesel). Em recente trabalho (Ferrão *et al.*, 2011) onde o *i*PLS foi aplicado para quantificação do biodiesel, a região selecionada também foi referente ao sinal do  $-C-O-C(O)-$ .

As faixas espectrais selecionadas pelo *i*PLS como mais representativas na construção de um bom modelo *i*PLS, seja subdividindo os espectros em 10, 15 ou 20 intervalos ou aplicando a TDW com 4, 6 ou 8 coeficientes em 1 ou 2 níveis de decomposição, também foram empregadas na construção dos modelos LS-SVM, além do espectro inteiro, empregando o mesmo pré-processamento: auto-escalar os dados. Na Tabela 4 são apresentados os melhores resultados dos modelos LS-SVM.

Tabela 4 – Resultados dos melhores modelos LS-SVM para a quantificação de biodiesel em blendas biodiesel/diesel com espectros originais e decompostos em até 2 níveis com 4, 6 e 8 coeficientes.

N	C	IR	ND	NVT	NI	QVI	MODELOS LS-SVM									
							FE (cm <sup>-1</sup> )	$\gamma$	$\sigma^2$	EC1	EC2	EC3	EP1	EP2	EP3	R <sup>2</sup> <sub>cal</sub>
K	0	100	0	1736	1	1736	4000 – 650	15003	3654000	0,353	0,400	0,299	0,764	0,961	0,533	0,998
L	0	100	0	1736	10	174	1321 – 985	2505	8024	0,314	0,379	0,248	0,367	0,384	0,350	0,998
M	0	100	0	1736	15	116	1321 – 1097	15010	4032	0,334	0,404	0,255	0,360	0,371	0,348	0,998
N	0	100	0	1736	20	87	1321 – 1153	15010	4032	0,336	0,435	0,256	0,389	0,423	0,351	0,998
O	4	99,93	1	868	20	44	1331 – 1161	22505	4032	0,340	0,427	0,265	0,378	0,412	0,341	0,998
P	4	99,61	2	434	15	29	1327 – 1103	15010	8024	0,336	0,409	0,281	0,421	0,482	0,348	0,998
Q	6	99,36	1	868	20	44	1331 – 1161	30002	2008	0,330	0,420	0,260	0,377	0,398	0,354	0,998
R	6	97,42	2	434	15	29	1327 – 1103	30002	2008	0,325	0,414	0,280	0,378	0,400	0,355	0,998
S	8	98,31	1	868	20	44	1331 – 1161	40000	2008	0,330	0,419	0,263	0,376	0,397	0,355	0,998
T	8	94,04	2	434	15	29	1327 – 1103	30001	2008	0,334	0,419	0,284	0,386	0,416	0,354	0,998

N – nome do modelo.

C – número de coeficientes.

IR – informação retida em percentual.

ND – número de decomposição.

NVT – número de variáveis totais.

NI – número de intervalos.

QVI – quantidade de variáveis por intervalo.

FE – faixa espectral.

EC1 – erros de validação cruzada (RMSECV), em percentual, com dados lidos nos dois espectrofotômetros.

EC2 – erros de validação cruzada (RMSECV), em percentual, com dados lidos no espectrofotômetro Shimadzu.

EC3 – erros de validação cruzada (RMSECV), em percentual, com dados lidos no espectrofotômetro Nicolet.

EP1 – erros de previsão (RMSEP), em percentual, com dados lidos nos dois espectrofotômetros.

EP2 – erros de previsão (RMSEP), em percentual, com dados lidos no espectrofotômetro Shimadzu.

EP3 – erros de previsão (RMSEP), em percentual, com dados lidos no espectrofotômetro Nicolet.

R<sup>2</sup><sub>cal</sub> – correlação dos modelos de calibração.

Fonte: Elaborado pela autora.

Assim como na modelagem *i*PLS, na modelagem LS-SVM os espectros Nicolet e Shimadzu também foram modelados em conjunto, e observou-se uma uniformidade entre os erros de validação cruzada e os erros de previsão, exceto para o modelo que emprega todas as variáveis. Neste, o erro de previsão é mais que o dobro do erro de validação cruzada, indicando um sobre-ajuste do modelo. Sendo que os espectros Shimadzu e Nicolet participaram do grupo de calibração e previsão, portanto espera-se que os erros, de validação cruzada e de previsão, tenham flutuações suaves; o que aconteceu para os demais modelos.

Também é possível verificar que a decomposição dos espectros em novos conjuntos de dados, com dimensões menores, seja com 1 ou 2 níveis de decomposição empregando 4, 6 ou 8 coeficientes, reduz muito o número de variáveis, mas sem perder informação, o que resulta na construção de modelos com desempenho equivalentes a modelos com dados originais. Tendo como um dos principais benefícios a economia de tempo na construção dos modelos. Para identificar qual espectrofotômetro teria melhor desempenho frente ao outro quando modelados seus espectros em conjunto, calculou-se em separado os erros de validação cruzada e de previsão para ambos e percebeu-se que de maneira geral os espectros Nicolet apresentaram menores erros, tanto de validação cruzada quanto de previsão, que os espectros Shimadzu, indicando que de uma forma geral os espectros Nicolet tem melhor desempenho frente aos espectros Shimadzu quando modelados em conjunto.

Na sequência são apresentadas: a superfície de otimização dos parâmetros  $\gamma$  e  $\sigma^2$  em função do erro (RMSECV) do modelo LS-SVM construído com os dados originais e subdivididos em 10 intervalos, Figura 31, e as curvas de correlação entre os valores de referência e os valores previstos pelos modelos LS-SVM. Acrescenta-se ainda que a Figura 32 emprega os espectros dos dois espectrofotômetros, e as Figuras 33 e 34 empregam os dados dos espectrofotômetros Shimadzu e Nicolet, respectivamente.

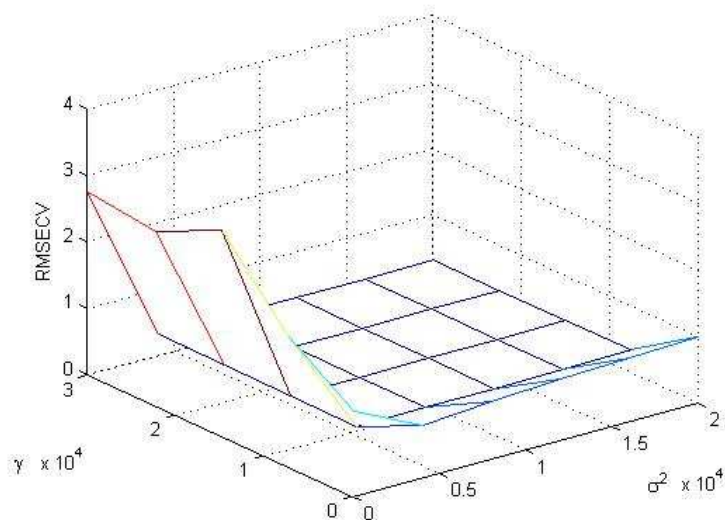


Figura 31 – Superfície de otimização dos parâmetros  $\gamma$  e  $\sigma^2$  em função do erro (RMSECV) para o modelo LS-SVM com dados originais e subdivididos em 10 intervalos.

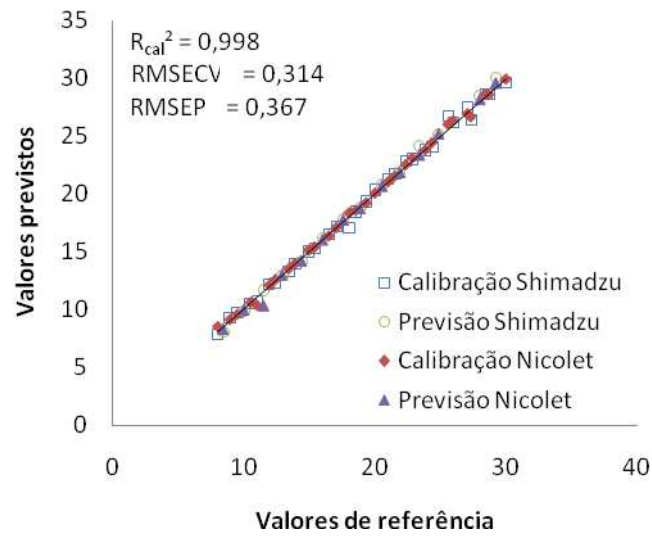


Figura 32 – Modelo LS-SVM subdividindo os espectros originais em 10 intervalos, empregando a faixa espectral de  $1.321\text{ cm}^{-1} - 985\text{ cm}^{-1}$  construído com dados lidos em ambos espectrofotômetros.

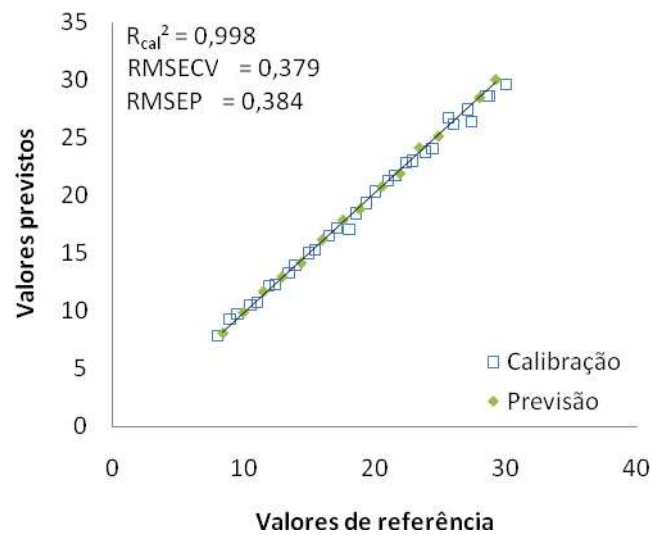


Figura 33 – Modelo LS-SVM subdividindo os espectros originais em 10 intervalos, empregando a faixa espectral de  $1.321\text{ cm}^{-1} - 985\text{ cm}^{-1}$  construído com dados lidos no espectrofotômetro Shimadzu.

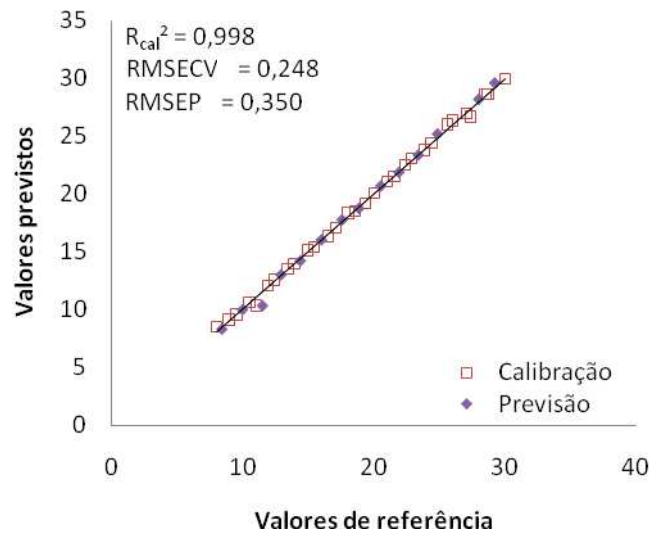


Figura 34 – Modelo LS-SVM subdividindo os espectros originais em 10 intervalos, empregando a faixa espectral de  $1.321\text{ cm}^{-1} - 985\text{ cm}^{-1}$  construído com dados lidos no espectrofotômetro Nicolet.

Analisando os resultados das Figuras 32, 33 e 34 verifica-se que os modelos são capazes de relacionar adequadamente os teores de biodiesel reais nas blendas biodiesel/deisel e os teores de biodiesel previstos, já que a correlação é 0,998.

Na Figura 35 é apresentada a superfície de otimização dos parâmetros  $\gamma$  e  $\sigma^2$  em função do erro (RMSECV) para os modelos LS-SVM que teve os espectros decompostos em 2 níveis pela TDW Daubechies com 6 coeficientes e subdividido em 15 intervalos. Nas Figuras 36, 37 e 38 têm-se as curvas de correlação entre os valores de referência e os valores previstos por este modelo LS-SVM, sendo a primeira utilizando os dados de ambos os espectrofotômetros, a segunda utilizando somente os dados do espectrofotômetro Shimadzu e a terceira utilizando os dados do espectrofotômetro Nicolet.

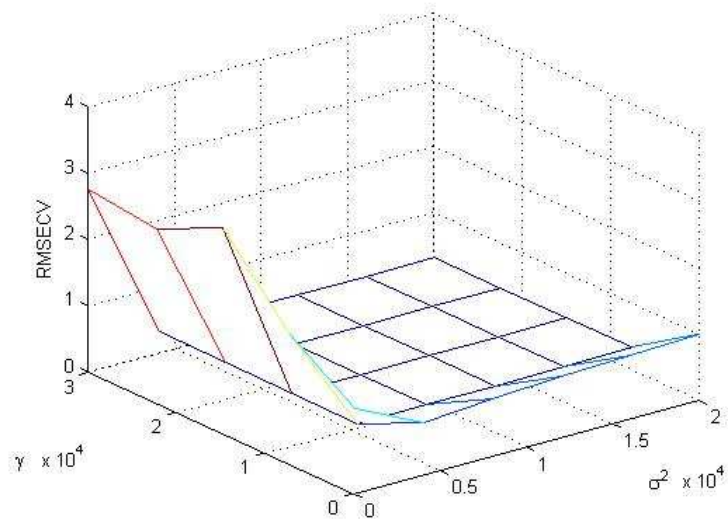


Figura 35 – Superfície de otimização dos parâmetros  $\gamma$  e  $\sigma^2$  em função do erro (RMSECV) para o modelo LS-SVM com dados decompostos em 2 níveis, com 6 coeficientes e subdivididos em 15 intervalos.

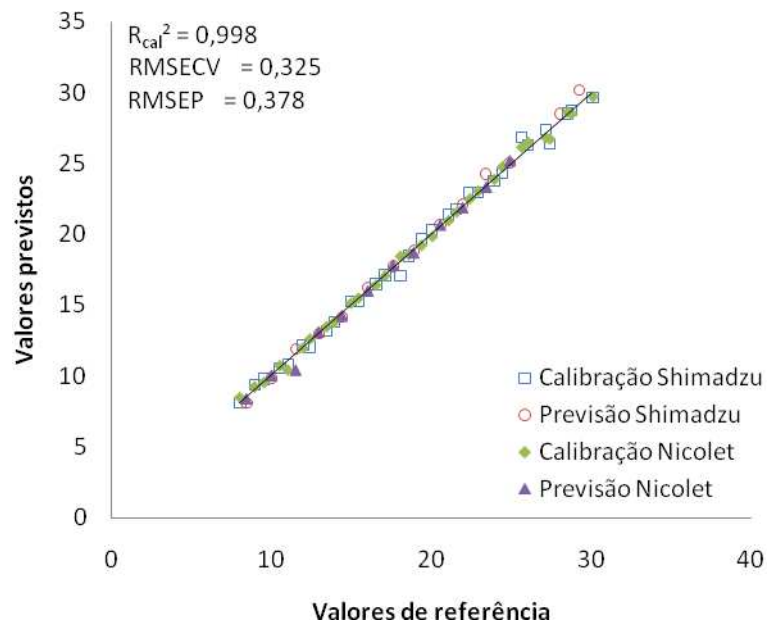


Figura 36 – Modelo LS-SVM com dados decompostos em 2 níveis, com 6 coeficientes e subdividido em 15 intervalos, empregando a faixa espectral de  $1.326 \text{ cm}^{-1} - 1.103 \text{ cm}^{-1}$ , construído com dados lidos em ambos espectrofotômetros.

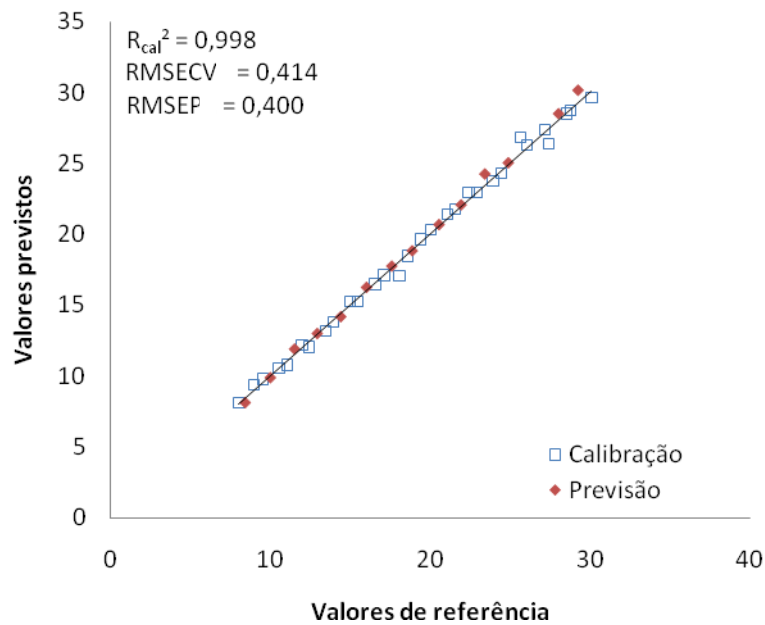


Figura 37 – Modelo LS-SVM com dados decompostos em 2 níveis, com 6 coeficientes e subdividido em 15 intervalos, empregando a faixa espectral de  $1.326\text{ cm}^{-1}$  –  $1.103\text{ cm}^{-1}$ , construído com dados lidos no espectrofotômetro Shimadzu.

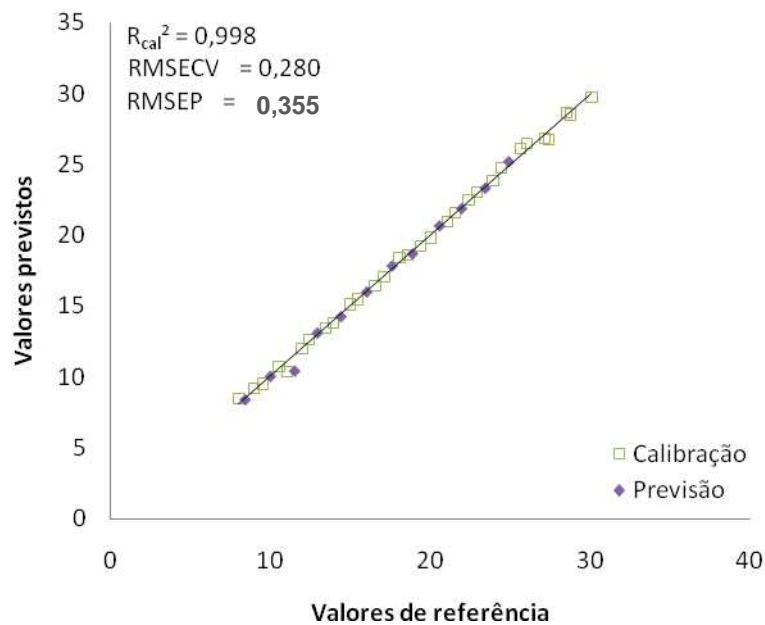


Figura 38 – Modelo LS-SVM com dados decompostos em 2 níveis, com 6 coeficientes e subdividido em 15 intervalos, empregando a faixa espectral de  $1.326\text{ cm}^{-1}$  –  $1.103\text{ cm}^{-1}$ , construído com dados lidos no espectrofotômetro Nicolet.

Para salientar a uniformidade entre os modelos construídos para a quantificação de biodiesel em blendas biodiesel/diesel, apresentados nesse capítulo foi aplicado o teste F, equação 15, com 95% de confiança.



$$F_{(i,j)} = \left( \frac{RMSEP_i}{RMSEP_j} \right)^2 \quad (15)$$

onde  $i$  e  $j$  representam o número de amostras da previsão empregadas em cada modelo construído com os espectros em conjunto, no caso 30 amostras. Para esse número de amostras o valor crítico da tabela F é de 1,84 (FREUND *et al.*, 2000). Quando o valor calculado for acima do valor crítico, os modelos comparados serão considerados significativamente diferentes e o contrário para valores abaixo do valor crítico.

Na Tabela 5 são apresentados os resultados obtidos através do teste F para avaliação da uniformidade entre os modelos *i*PLS apresentados na Tabela 3. Para isso foram empregados os erros de previsão dos modelos *i*PLS que utilizaram os espectros de ambos os espectrofotômetros.

Tabela 5 – Resultados do teste F para os modelos *i*PLS.

	A	B	C	D	E	F	G	H	I	J
A	1,00	–	–	–	–	–	–	–	–	–
B	1,80	1,00	–	–	–	–	–	–	–	–
C	1,54	1,17	1,00	–	–	–	–	–	–	–
D	1,80	1,00	1,17	1,00	–	–	–	–	–	–
E	1,82	1,01	1,18	1,01	1,00	–	–	–	–	–
F	1,35	1,33	1,14	1,33	1,35	1,00	–	–	–	–
G	1,84	1,02	1,20	1,02	1,01	1,36	1,00	–	–	–
H	1,43	1,26	1,07	1,26	1,27	1,06	1,29	1,00	–	–
I	1,79	1,01	1,16	1,01	1,02	1,32	1,03	1,25	1,00	–
J	1,45	1,24	1,06	1,24	1,26	1,07	1,27	1,01	1,23	1,00

Fonte: Elaborado pela autora.

Como pode ser observado na Tabela 5, com exceção da coluna A linha G, todos os valores estão abaixo do valor crítico, 1,84, dessa forma todos os modelos apresentam mesmo desempenho para a quantificação do biodiesel nas blendas, menos os modelos *i*PLS A e G que são significativamente diferentes entre si. Ao observar a Tabela 3, verifica-se que o maior erro de previsão é do modelo *i*PLS A (0,445), que emprega o espectro inteiro, e que o menor erro de previsão é do modelo *i*PLS G (0,328).

Na Tabela 6 são apresentados os resultados obtidos através do teste F para avaliação da uniformidade entre os modelos LS-SVM apresentados na Tabela 4. Para isso foram

empregados os erros de previsão dos modelos que utilizaram os espectros de ambos os espectrofotômetros.

Tabela 6 – Resultados do teste F para os modelos LS-SVM.

	K	L	M	N	O	P	Q	R	S	T
K	1,00	–	–	–	–	–	–	–	–	–
L	4,33	1,00	–	–	–	–	–	–	–	–
M	4,50	1,04	1,00	–	–	–	–	–	–	–
N	3,86	1,12	1,17	1,00	–	–	–	–	–	–
O	4,09	1,06	1,10	1,06	1,00	–	–	–	–	–
P	3,29	1,32	1,37	1,17	1,24	1,00	–	–	–	–
Q	4,11	1,06	1,10	1,06	1,01	1,25	1,00	–	–	–
R	4,09	1,06	1,10	1,06	1,00	1,24	1,01	1,00	–	–
S	4,13	1,05	1,09	1,07	1,01	1,25	1,01	1,01	1,00	–
T	3,92	1,11	1,15	1,02	1,04	1,19	1,05	1,04	1,05	1,00

Fonte: Elaborado pela autora.

Nos modelos LS-SVM a maioria dos valores calculados são inferiores ao valor crítico, 1,84, exceto a coluna K, conforme pode ser observado na Tabela 6. Logo, os modelos LS-SVM apresentam mesmo desempenho para a quantificação de biodiesel nas blendas, excetuando-se o modelo LS-SVM que emprega espectro inteiro, o qual registrou o maior erro de previsão como pode ser observado na Tabela 4. Na modelagem LS-SVM, o emprego do espectro inteiro também reduz a capacidade preditiva do modelo para a quantificação de biodiesel em blendas biodiesel/diesel.

Na Tabela 7 são apresentados os resultados obtidos através do teste F para avaliação da uniformidade entre modelos *i*PLS e LS-SVM apresentados nas Tabelas 3 e 4. Para isso foram empregados os erros de previsão dos modelos que utilizaram os espectros de ambos os espectrofotômetros.

Na comparação entre modelos *i*PLS e LS-SVM, Tabela 7, verifica-se que todos os modelos *i*PLS são equivalentes aos modelos LS-SVM, uma vez que os valores calculados foram menores que 1,84, excetuando os modelos *i*PLS e LS-SVM que empregam o espectro inteiro.

Tabela 7 – Resultados do teste F para os modelos *i*PLS e LS-SVM.

	K	L	M	N	O	P	Q	R	S	T
A	2,95	–	–	–	–	–	–	–	–	–
B	–	1,22	–	–	–	–	–	–	–	–
C	–	–	1,01	–	–	–	–	–	–	–
D	–	–	–	1,44	–	–	–	–	–	–
E	–	–	–	–	1,31	–	–	–	–	–
F	–	–	–	–	–	1,21	–	–	–	–
G	–	–	–	–	–	–	1,32	–	–	–
H	–	–	–	–	–	–	–	1,03	–	–
I	–	–	–	–	–	–	–	–	1,27	–
J	–	–	–	–	–	–	–	–	–	1,09

Fonte: Elaborado pela autora.

### 4.3 Descrição dos espectrofotômetros

O espectrofotômetro de infravermelho por transformada de Fourier Shimadzu modelo IR Prestige-21, utiliza uma fonte de luz cerâmica de alta energia, uma ótica de alta eficiência, um detector DLATGS de alta sensibilidade. A otimização dos sistemas ótico, eletrônico e de sinal, minimizam o ruído, garantindo análise de alta sensibilidade.

Possui também divisor de feixes, detectores e fonte de radiação que cobrem as regiões de infravermelho próximo, médio e distante; além de três tipos de detectores para diferentes aplicações: DLATGS / MCT / InGaAs; correção atmosférica para eliminação da influência de vapor d'água e CO<sub>2</sub> do meio externo e ampla variedade de acessórios, incluindo microscópio FTIR.

O espectrofotômetro de infravermelho por transformada de Fourier Nicolet Magna 550 apresenta o alinhamento contínuo e dinâmico, fonte de laser e detectores pré-alinhados reduzindo drasticamente manutenção geral, o compartimento de amostra é grande, tem dois espelhos de divisor de feixe para detector: espelho parabólico simples de divisor de feixe para amostra e espelho elipsoidal único a partir de amostra ao detector, rápida digitalização interferômetro e excelente relação sinal/ruído.

### 4.4 Resultados da quantificação de biodiesel em blendas biodiesel/diesel seguindo a Norma NBR 15.568

Nesse capítulo são apresentados os resultados obtidos para a quantificação de biodiesel em blendas biodiesel/diesel seguindo as orientações da Norma NBR 15.568. Para tal, empregaram-se as faixas espectrais  $1.200\text{ cm}^{-1} - 1.100\text{ cm}^{-1}$  e  $1.750\text{ cm}^{-1} - 1.735\text{ cm}^{-1}$  e o pré-processamento de centrar os dados na média, onde cada valor, de cada número de onda, é subtraído da média, de cada número de onda.

Foram construídos 7 modelos para a quantificação do biodiesel. Em um deles não foi aplicado o método da TDW, nos demais houve a aplicação com 4, 6 e 8 coeficientes com 1 e 2 níveis de decomposição. Os resultados desses modelos estão apresentados na Tabela 8.

Esses resultados mostram que os erros de calibração dos modelos são menores que os erros de previsão. Além disso, verifica-se uma uniformidade de desempenho entre os modelos, o que significa que a aplicação do método da TDW é capaz de reter informação. Garantindo economia de tempo na construção de modelos.

Tabela 8 – Resultados dos modelos para a determinação de biodiesel em blendas biodiesel/diesel com espectros originais e decompostos em até 2 níveis com 4, 6 e 8 coeficientes, seguindo as regiões espectrais e o pré-processamento da Norma NBR 15.568.

N	C	IR	ND	NVT	MODELOS			
					RMSECV	VL	RMSEP	R <sup>2</sup> <sub>cal</sub>
AA	0	100	0	60	0,524	2	0,818	0,994
EE	4	99,93	1	30	0,530	2	0,840	0,994
FF	4	99,61	2	15	0,547	2	0,876	0,994
GG	6	99,36	1	30	0,540	2	0,878	0,994
HH	6	97,42	2	15	0,541	2	0,871	0,994
II	8	98,31	1	30	0,540	2	0,893	0,994
JJ	8	94,04	2	15	0,564	2	0,932	0,993

N – nome do modelo.

C – número de coeficientes.

IR – informação retida em percentual.

ND – número de decomposição.

NVT – número de variáveis totais.

RMSECV – erros de validação cruzada, em percentual, com dados lidos nos espectrofotômetros Shimadzu e Nicolet.

VL – variáveis latentes.

RMSEP – erros de previsão, em percentual, com dados lidos nos espectrofotômetros Shimadzu e Nicolet.

R<sup>2</sup><sub>cal</sub> – correlação dos modelos de calibração.

Fonte: Elaborado pela autora.

Para verificar a equivalência da capacidade preditiva entre os modelos que empregam as regiões sugeridas pela Norma NBR 15.568 e os modelos *i*PLS apresentados na Tabela 3 efetuou-se o teste F empregando os erros de previsão dos respectivos modelos. Os resultados estão apresentados na Tabela 9.

Tabela 9 – Resultados do teste F para os erros de previsão dos modelos *i*PLS das Tabelas 3 e 8.

N	N	TESTE F
A	AA	3,38
E	EE	6,48
F	FF	5,23
G	GG	7,17
H	HH	5,48
I	II	7,19
J	JJ	6,34

N – nome do modelo.

Fonte: Elaborado pela autora.

Como pode ser observado, na Tabela 9, todos os valores obtidos são superiores ao valor crítico de 1,84. Como os modelos que empregam as regiões da Norma, são significativamente diferentes dos modelos *i*PLS, apresentados na Tabela 3, e com erros de previsão bem superiores (a diferença oscila entre 0,4 e 0,6), logo os modelos *i*PLS apresentados na Tabela 3 são melhores, isto é, a utilização da carbonila ( $1.750\text{ cm}^{-1} - 1.750\text{ cm}^{-1}$ ) não contribui para melhorar a capacidade preditiva dos modelos que empregam as regiões espectrais sugeridas pela Norma NBR 15.568.

#### 4.5 Resultados comparativos dos tempos necessários para a aplicação do método da TDW

Pelos resultados apresentados nas Tabelas 5, 6 e 7 pode-se observar que a aplicação da transformada discreta *wavelet* (TDW) Daubechies com diferentes coeficientes (4, 6 e 8) produz resultados equivalentes nos modelos. Ou seja, há uma uniformidade nos erros dos modelos multivariados construídos. Na Tabela 10 são apresentados os tempos necessários para a aplicação da TDW com 4, 6 e 8 coeficientes para 1 e 2 níveis de decomposição. Para construir a Tabela 10 foi empregada uma matriz espectral de ordem  $90 \times 1736$ .

Tabela 10 – Resultados do tempo de execução da transformada discreta *wavelet* Daubechies com 4, 6 e 8 coeficientes.

Transformada Discreta <i>Wavelet</i> Daubechies			
	4 coeficientes	6 coeficientes	8 coeficientes
Tempo de execução (seg) – 1º. Nível	2	2	3

Tempo de execução (seg) – 2º. Nível	3	4	5
-------------------------------------	---	---	---

Fonte: Elaborada pela autora.

Fica claro pelos resultados da Tabela 10 que independente da quantidade de coeficientes é muito rápida a aplicação da TDW.

#### 4.6 Resultados comparativos dos tempos necessários para a construção de modelos *i*PLS com e sem aplicação do método da TDW

Para comparar os tempos necessários na construção de modelos *i*PLS com e sem aplicação do método da TDW empregou-se três conjuntos de dados espectrais. Um deles é o conjunto original de espectros e os outros conjuntos tem os espectros decompostos em 1 e 2 níveis de decomposição pela TDW Daubechies com 4 coeficientes. Os resultados do tempo de execução estão apresentados na Tabela 11.

Pode ser observado que a aplicação da TDW em 2 níveis reduz muito o tempo necessário para a execução da modelagem *i*PLS quando se utiliza o pré-processamento de auto-escalar os dados. Já para o pré-processamento de auto-escalar os dados e também realizar a correção do espalhamento de luz a redução do tempo para construir um modelo *i*PLS não é tão expressiva, mas mesmo assim existe uma redução do tempo.

Tabela 11 – Resultados do tempo de execução dos modelos *i*PLS com e sem aplicação da transformada discreta *wavelet* Daubechies com 4 coeficientes empregando 2 pré-processamentos e com subdivisão espectral de 10 intervalos.

	<i>i</i> PLS com os dados auto-escalados	<i>i</i> PLS com os dados auto-escalados e com correção do espalhamento de luz
	Tempo em segundos	Tempo em segundos
Sem decomposição	7	59
1 nível de decomposição	4	52
2 níveis de decomposição	2	48

Fonte: Elaborado pela autora.

#### 4.7 Considerações sobre os resultados para a quantificação de biodiesel em blendas biodiesel/diesel

Os resultados dos modelos apresentados nas Tabelas 3 e 4, foram construídos empregando parte de uma das regiões espectrais sugeridas pela Norma NBR 15.568 ( $1.200\text{ cm}^{-1} - 1.100\text{ cm}^{-1}$ )<sup>6</sup>, porém nunca foi empregada a outra região proposta por ela ( $1.750\text{ cm}^{-1} - 1.735\text{ cm}^{-1}$ ). Embora utilizando um pré-processamento não sugerido pela Norma NBR 15.568, os dados autoescalados, a grande maioria dos erros de previsão obtidos foram abaixo do limite estabelecido pela Norma NBR 15.568 (1%), quando a concentração de biodiesel varia de 8% (v/v) a 30% (v/v). Com o emprego das duas faixas espectrais ( $1.200\text{ cm}^{-1} - 1.100\text{ cm}^{-1}$  e  $1.750\text{ cm}^{-1} - 1.735\text{ cm}^{-1}$ ) e pré-processamento sugeridos pela Norma NBR 15.568, se obtém modelos, Tabela 8, com desempenho inferior aos modelos apresentados nas Tabelas 3 e 4. Dessa forma, verifica-se que a faixa  $1.750\text{ cm}^{-1} - 1.735\text{ cm}^{-1}$  não é necessária para a construção de bons modelos e que o sinal espectral referente ao  $-\text{O}-\text{C}-(\text{O})-$  é mais seletivo para a construção modelos visando a quantificação do biodiesel em blendas biodiesel/diesel.

Também foi verificado que o emprego de toda a faixa espectral resultou em modelos com desempenhos inferiores aos modelos que empregaram parte do espectro. Isso ocorre porque determinadas faixas espectrais não possuem informações ou contém muito ruído, e outras onde as informações não são representativas para a quantificação do analito em questão. Assim, a utilização destas, deteriora a capacidade preditiva do modelo.

Com base nos resultados do teste F, apresentados na Tabela 7, neste trabalho verificou-se que 2 ferramentas independentes e equivalentes podem ser utilizadas com excelente desempenho para a quantificação de biodiesel em blendas biodiesel/diesel. Além disso, ambas associadas à espectroscopia no infravermelho médio são ambientalmente corretas. Foi constatado também, mediante os resultados apresentados nesse trabalho, que os algoritmos *i*PLS e LS-SVM são capazes de quantificar de maneira satisfatória o biodiesel, até mesmo quando modelam simultaneamente espectros oriundos de espectrofotômetros diferentes, no caso Nicolet e Shimadzu.

Também com base nos resultados do teste F, Tabelas 5 e 6, ficou comprovado que a aplicação do método da TDW contribui para uma visível redução do número de variáveis sem

---

<sup>6</sup> - A faixa espectral  $1.200\text{ cm}^{-1} - 1.100\text{ cm}^{-1}$  contém uma parte da faixa de absorção do estiramento C-O dos ésteres presentes no biodiesel, localizada entre  $1.160\text{ cm}^{-1} - 1.050\text{ cm}^{-1}$  (BARBOSA, 2007).

perder a capacidade de prever a concentração do biodiesel nos modelos. Por isso o método da TDW reduz o custo computacional na construção de modelos multivariados.

Poucos trabalhos foram publicados na literatura que quantificam o biodiesel em blendas para teores acima de 5%. No trabalho de Pimentel *et al.* (2006), o RMSEP foi de 0,25% (v/v) para a faixa de 0 – 8% (v/v) de concentração de biodiesel, sendo modelada a região que corresponde ao sinal da carbonila, assim, como, a faixa abaixo de  $1.470\text{ cm}^{-1}$  que contém o sinal correspondente ao estiramento C-O. Já no trabalho de Soares e colaboradores (2008) o RMSEP variou entre 0,65 – 1,40% (v/v) para a faixa de 1 – 40% (v/v) de biodiesel nas blendas. Neste caso o método *Forward* foi empregado para selecionar os comprimentos de onda contidos nas faixas de  $650\text{ cm}^{-1}$  –  $1.800\text{ cm}^{-1}$  e  $2.760\text{ cm}^{-1}$  –  $4.000\text{ cm}^{-1}$  que apresentam melhor capacidade preditiva. Ainda em 2009 Mahamuni e Adewuyi obtiveram RMSEP de 0,12% (v/v) para previsão de amostras da faixa de 1 – 100% (v/v) modelando através de Mínimos Quadrados Clássicos (CLS). Outras metodologias espectroscópicas também foram usadas como no trabalho de Camila *et al.* (2008) onde a espectrofluorimetria foi empregada para amostras contendo biodiesel na faixa de 0 – 25% (v/v), porém com valores de RMSEP mais elevados variando de 1,6% a 2,6% (v/v).

Neste contexto, tanto os modelos obtidos pelo *iPLS*, assim como, os obtidos pela LS-SVM, com ou sem aplicação da TDW, apresentam RMSEP comparáveis aos descritos na literatura.



## 5 CONCLUSÕES

Esse trabalho mostrou a eficiência do método da TDW na redução significativa do número de variáveis e ao mesmo tempo a retenção das informações. Possibilitando a construção de modelos multivariados mais robustos em menor tempo. Além disso, mostrou o quão rápido é a aplicação do método da TDW.

Também utilizou dois métodos de regressão multivariada independentes e equivalentes, estatisticamente, (*i*PLS e LS-SVM) para a quantificação de biodiesel em blendas biodiesel/diesel com teores de 8% – 30% (v/v). Sendo, ambas, capazes de quantificar adequadamente o biodiesel até mesmo quando os espectros de infravermelho são oriundos de equipamentos diferentes.

A aplicação do algoritmo *i*PLS foi eficiente na seleção da faixa espectral do infravermelho médio mais relevante para a quantificação do biodiesel em blendas biodiesel/diesel, independente do número de intervalos nos quais os espectros foram subdivididos.

Por fim, nesse trabalho mostraram-se modelos preditivos mais robustos para a quantificação de biodiesel em blendas biodiesel/diesel, que os da Norma NBR 15.568 utilizando uma metodologia rápida, precisa e ambientalmente correta.

A partir dos estudos aqui apresentados, abrem-se algumas possibilidades de trabalhos futuros como a ampliação do percentual de biodiesel nas blendas biodiesel/diesel, além de uma maior variabilidade nas origens de biodiesel e a utilização de vários espectrofotômetros a fim de ter um modelo capaz de abranger a máxima variabilidade possível.

## REFERÊNCIAS

- ALMEIDA, J.C.; PORTELA, F.M.; SANTOS, D.Q.; HERNANDÉZ-TERRONES, M.G.; BORGES NETO, W. Determinação do teor de biodiesel metílico de soja e soja residual AM diesel através de espectroscopia no infravermelho médio e PLS. *Anais do 4º Congresso da Rede Brasileira de Tecnologia de Biodiesel e 7º Congresso Brasileiro de Plantas Oleaginosas, Óleos, Gorduras e Biodiesel*, p.1313-1314, 2010.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 15.568: Biodiesel – Determinação do teor de biodiesel em óleo diesel por espectroscopia na região do infravermelho médio. Rio de Janeiro, 2008.
- AN, S.; LIU, W.; VENKATESH, S. Fast cross-validation algorithms for least squares support vector machine and kernel ridge regression. *Pattern Recognition*, v.40, p.2154-2162, 2007.
- ARAÚJO, R.G.O.; MACEDO, S.M.; KORN, M.G.A.; PIMENTEL, M.F.; BRUNS, R.E.; FERREIRA, S.L.C. Mineral composition of wheat flour consumed in brasilian cities. *Journal of Brazilian Chemical Society*, v.19, n.5, p.935-942, 2008.
- BARBOSA, L.C.A. *Espectroscopia no infravermelho na caracterização de compostos orgânicos*. Viçosa: UFV, 2007.
- BISHOP, C.M. *Pattern Recognition and Machine Learning*. Nova York: Springer, 2006.
- BORGES NETO, W. *Parâmetros de qualidade de lubrificantes e óleo de oliva através de espectroscopia vibracional, calibração multivariada e seleção de variáveis*. 2005. 142 f. Tese. (Programa de Pós-Graduação em Química – Doutorado) – Universidade de Campinas, Campinas, 2005.
- BORIN, A. *Aplicações de Máquinas de vetores de suporte por mínimos quadrados (LS-SVM) na quantificação de parâmetros de qualidade de matrizes lácteas*. 2007. 128 f. Tese. (Programa de Pós-Graduação em Química – Doutorado) – Universidade de Campinas, Campinas, 2007.
- BORIN, A.; FERRÃO, M.F.; MELLO, C.; CORDI, L.; PATACA, L.C.M.; DURÁN, N.; POPPI, R.J. Quantification of Lactobacillus in fermented milk by multivariate image analysis with least-squares support-vector machines *Analytical and Bioanalytical Chemistry*, v.387, n.3, p.1105-1112, 2007.
- BORIN, A.; MARETTO, D.A.; FERRÃO, M.F.; MELLO, C.; POPPI, R.J. Least- Squares Support Vector Machines and Near Infrared Spectroscopy for Quantification of Common Adulterants in Powdered Milk. *Analytica Chimica Acta*, v.579, p.25-32, 2006.
- BRASIL. Lei nº 11097, de 13 de janeiro de 2005. Disponível em: <[http://www.planalto.gov.br/ccivil/\\_Ato2004-2006/2005/Lei/L11097.htm](http://www.planalto.gov.br/ccivil/_Ato2004-2006/2005/Lei/L11097.htm)>. Acesso em: 6 ago. 2009.
- BUENO, A.W. *Manual de Espectroscopia Vibracional*. São Paulo: McGraw-Hill, 1989.
- CAI, C.B.; HAN, Q.J.; TANG, L.J.; NIE, J.F.; OUYANG, L.Q.; YU, R.Q. Treating NIR data with orthogonal discrete wavelet transform: Predicting concentrations of a multi-component system through a small-scale calibration set. *Talanta*, v.77, p.822-826, 2008.

- CAI, C.B.; XU, L.; HAN, Q.J.; WU, H.L.; NIE, J.F.; FU, H.Y.; YU, R.Q. Combining the least correlation design, wavelet packet transform and correlation coefficient test to reduce the size of calibration set for NIR quantitative analysis in multi-component systems. *Talanta*, v.81, p.799-804, 2010.
- CHAU, Y.; LIANG, J.G.; XUE-GUANG, S. *Chemometrics: from basics to wavelet transform*. New Jersey – Canada: John Wiley & Sons Inc. v.164, 2004.
- CHEN, N.; WENCONG, L.; YANG, J.; LI, G. *Support Vector Machine in Chemistry*. Singapura: World Scientific Publishing Company, 2004.
- CORDEIRO, G.A. *Desenvolvimento de metodologias espectroscópicas multivariadas para quantificação de fármacos em formas farmacêuticas*. 2006. 118 f. Dissertação (Programa de Pós-Graduação em Química – Mestrado) – Universidade Federal do Paraná, Curitiba, 2006.
- CORGOZINHO, C.N.C.; PASA, V.M.D.; BARBEIRA, P.J.S. Determination of residual oil in diesel oil by spectrofluorimetric and chemometric analysis. *Talanta*, v.76, p.479-484, 2008.
- CORREIA, P.R.M.; FERREIRA, M.M.C. Reconhecimentos de padrões por métodos não supervisionados: explorando procedimentos quimiométricos para tratamento de dados analíticos. *Química Nova*, v.30, n.2, p.481-487, 2007.
- 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, n.1, p.46-52, 2002.
- DANTAS FILHO, H.A. *Desenvolvimentos de técnicas quimiométricas de compressão de dados e redução de ruído aplicadas a óleo diesel e madeira de eucalipto empregando espectroscopia NIR*. 2007. 158 f. Tese. (Programa de Pós-Graduação em Química – Doutorado) – Universidade de Campinas, Campinas, 2007.
- DA SILVA, H.R.G.; TANAJURA, A.S.; MEIRA, M.; DA COSTA NETO, P.R.; PEPE, I.M.; QUINTELLA, C.M. Predição da concentração e viscosidade de misturas de biodiesel em diesel por espectrofluorimetria e quimiometria. *Anais do 4º Congresso da Rede Brasileira de Tecnologia de Biodiesel e 7º Congresso Brasileiro de Plantas Oleaginosas, Óleos, Gorduras e Biodiesel*, p.1163-1164, 2010.
- DAUBECHIES, I. *Ten lectures on wavelets*. Philadelphia: SIAM, 1992.
- DE FREITAS, R.C. Disponível em: <<http://www.biodieselbr.com>>. Acesso em: 03 nov. 2010.
- DE FREITAS, R.C. Disponível em: <<http://www.biodieselbr.com>>. Acesso em: 12 jan. 2011.
- DENNEY, R.C.; SINCLAIR, R. *Visible and Ultraviolet Spectroscopy*. London: John Willey, 1993.
- FERRÃO, M.F. *Aplicação de técnicas espectroscópicas de reflexão no infravermelho no controle de qualidade de farinha de trigo*. 2000. 219 f. Tese (Programa de Pós-Graduação em Química – Doutorado) – Universidade Estadual de Campinas, Campinas, 2000.
- \_\_\_\_\_. Técnicas de reflexão no infravermelho aplicadas na análise de alimentos. *Tecno-Lógica*, v.5, n.1, p.63-85, 2001.

FERRÃO, M.F.; MELLO, C.; BORIN, A.; MARETTO, D.A.; POPPI, R.J. LS-SVM: uma nova ferramenta quimiométrica para regressão multivariada. Comparação de modelos de regressão LS-SVM e PLS na quantificação de adulterantes em leite em pó empregando NIR. *Química Nova*, v.30, n.4, p.852-859, 2007.

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 HART-FTIR spectra and PLS, *i*PLS or *si*PLS regressions. *FUEL*, v.90, p.701-706, 2011.

FERREIRA, M.C.; ANTUNES, A.M.; MELGO, M.S.; VOLPE, P.L.O. Quimiometria I: Calibração Multivariada, um Tutorial. *Química Nova*, v.22, p.724-731, 1999.

FREUND, J.E.; SIMON, G.A. *Estatística aplicada: economia, administração e contabilidade*. 9.ed. Porto Alegre: Bookman, 2000.

FONSECA, F.S.; FERNANDES, T.D.; FERNANDES, D.M.; SANTOS, D.Q.; HERNADÉZ-TERRONES, M.G.; BORGES NETO, W. Quantificação de biodiesel metílico e etílico de girassol em diesel por calibração multivariada e infravermelho médio. *Anais do 4º Congresso da Rede Brasileira de Tecnologia de Biodiesel e 7º Congresso Brasileiro de Plantas Oleaginosas, Óleos, Gorduras e Biodiesel*, p.1317-1318, 2010.

GARCIA, M.H.F.; SIMONE, B.F.; FERREIRA, B.G. Determinação quantitativa da concentração de silicone em antiespumantes por espectroscopia FT-IR/ATR e calibração multivariada. *Polímeros: Ciência e Tecnologia*, v.14, n.5, p.322-325, 2004.

GERBASE, A.E.; FERRÃO, M.F.; FACHINI, D.; VIERA, M.S.; NASCIMENTO, A.M.R.; FRANCESQUETT, J.J.; PAZOS, R.E.P.; WERLANG, M. Determinação simultânea do teor de biodiesel e da massa específica em blendas biodiesel/diesel. *Anais do III Congresso da rede brasileira de tecnologia de biodiesel*, 2009.

GERIS, R.; DOS SANTOS, N.A.C.; AMARAL, B.A.; MAIA, I.S., CASTRO, V.D.; CARVALHO, J.R.M. Biodiesel de soja – reação de transesterificação para aulas práticas de química orgânica. *Química Nova*, v.30, n.5, p.1369-1373, 2007.

GODINHO, M.S.; PEREIRA, R.O.; RIBEIRO, K.O.; SCHIMIDT, F.; DE OLIVEIRA, A.E. Classificação de refrigerantes através de análise de imagens e análise de componentes principais (PCA). *Química Nova*, v.31, n.6, p.1485-1489, 2008.

HAYKIN, S. *Redes Neurais: Princípios e Prática*/ Simon Haykin; trad. Paulo Martins Engel. 2.ed. Porto Alegre: Bookman, 2001.

HIND, A. R.; BHARGAVA, S. K.; MCKINNON, A. At the solid/liquid interface: FTIR/ATR - the tool of choice. *Advances in Colloid and Interface Science*, v.93 p.91-114, 2001.

KONZEN, P.H.A.; FURTADO, J.C.; CARVALHO, C.W.; FERRÃO, M.F.; MOLZ, R.F.; BASSANI, I.A. & HÜNING, S.L. Otimização de métodos de controle de qualidade de fármacos usando algoritmo genético e busca tabu. *Pesquisa Operacional*, v.23, n.1, p.189-206, 2003.

LENARDÃO, E.J; FREITAG, R.A.; DABDOUB, M.J.; BATISTA, A.C.F.; SILVEIRA, C.C. “Green Chemistry” – Os 12 princípios da química verde e sua inserção nas atividades de ensino e pesquisa. *Química Nova*, v.26, n.1, p.123-129, 2003.

- LIU, F.; HE, Y.; WANG, L. Comparison of calibrations for the determination of soluble solids content and pH of rice vinegars using visible and short-wave near infrared spectroscopy. *Analytica Chimica Acta*. v.610, n.2, p.196-204, 2008.
- MAHAMUNI, N.N.; ADEWUYI, Y.G. Fourier transform infrared spectroscopy (FTIR) method to monitor soy biodiesel and soybean oil in transesterification reactions, petrodiesel – biodiesel blends, and blend adulteration with soy oil. *Energy & Fuels*, v.23, p.3773-3782, 2009.
- MALLAT, S.G. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. v.2, n.7, 1989.
- MARETTO, D.A. *Comparação entre máquinas de vetores de suporte por mínimos quadrados (LS-SVM) e métodos lineares para transferência de calibração*. 2007. 103 f. Dissertação. (Programa de Pós-Graduação em Química – Mestrado) – Universidade de Campinas, Campinas, 2007.
- MCMURRY, J. *Química orgânica*. 6.ed. v.1. São Paulo: Pioneira Thomson Learning, 2005.
- 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*, v.54, n.3, p.413-419, 2000.
- OLIVEIRA, J. S.; MONTALVÃO, R.; DAHER, L.; SUAREZ, P. A. Z.; RUBIM, J. C. Determination of methyl ester contents in biodiesel blends by FTIR-ATR and FTNIR spectroscopies. *Talanta*, v.69, p.1278-1284, 2006.
- PATACA, L.C.M.; LEAL, L.H.Q.; MARCUCCI, M.C.; POPPI, R.J. Determinação da tipagem da própolis utilizando espectroscopia no infravermelho e PLS-DA. *29ª Reunião Anual da Sociedade Brasileira de Química*, 2006.
- PIMENTEL, M.F.; RIBEIRO, G.M.G.S.; DA 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.201-206, 2006.
- RIZZI, M. *Aplicação do método enxame de partículas na otimização de modelos de regressão multivariada aplicados na quantificação da mistura biodiesel/óleo vegetal/diesel*. 2008. 113 f. . Dissertação (Programa de Pós-Graduação em Sistemas e Processos Industriais – Mestrado) – Universidade de Santa Cruz do Sul, Santa Cruz do Sul, 2008.
- SANTOS, M.C.S. *Utilização da espectroscopia na região do infravermelho (FTIR) e quimiometria na identificação do café torrado e moído adulterado*. 2005. 77 f. Dissertação (Programa de Pós-Graduação em Agronomia – Mestrado) – Universidade Federal de Lavras, Lavras, 2005.
- SCHERER, M.D.; OLIVEIRA, S.L.; ANDRADE, L.H.C.; LIMA, S.M.; CAÍRES, A.R.L. Caracterização óptica das blendas diesel/biodiesel, diesel/óleo e biodiesel/óleo. *Anais do 4º Congresso da Rede Brasileira de Tecnologia de Biodiesel e 7º Congresso Brasileiro de Plantas Oleaginosas, Óleos, Gorduras e Biodiesel*, p.1275-1276, 2010.

SCHÖLKOPF, B.; BURGESS, C.J.C.; SMOLA, A.J. *Advances in Kernel Methods: Support Vector Learning*. London: Cambridge University Press, 1998.

SILVERSTEIN, R.M.; WEBSTER, F.X.; KIEMLE, D.J. *Identificação espectrométrica de compostos orgânicos*. 7.ed. Rio de Janeiro: LTC, 2007.

SKOOG, D.A. et al. *Fundamentos de química analítica*. São Paulo: Thomson, 2006.

SOARES, I.P.; REZENDE, T.F.; SILVA, R.C.; CASTRO, E.V.R.; FORTES, I.C.P. Multivariate calibration by variable selection for blends of raw soybean oil/biodiesel from different sources using Fourier transform infrared spectroscopy (FTIR) spectra data. *Energy & Fuels*, v.22, p.2079-2083, 2008.

SUYKENS, J.A.K.; VAN GESTEL, T.; DE BRABANTER, J.; DE MOOR, B.; VANDEWALLE, J. *Least-Squares Support Vector Machines*, Singapura: World Scientific, 2002.

VIERA, M.S.; NASCIMENTO, A.M.R.; FRANCESQUETT, J.J.; WERLANG, M.; FERRÃO, M.F.; PAZOS, R.E.P.; FACHINI, D.; GERBASE, A.E.; Aplicação das transformadas discretas wavelets e infravermelho (FTIR-ATR) na quantificação de biodiesel em blendas biodiesel/diesel. *Anais do III Congresso da rede brasileira de tecnologia de biodiesel*, 2009.

UNIÃO BRASILEIRA DO BIODIESEL (UBRABIO). O biodiesel e sua contribuição ao desenvolvimento brasileiro, 2010.

WALKER, J.S. *A Primer on Wavelets and Their Scientific Applications*. New York: Chapman & Hall, 1999.

WU, D.; HE, Y.; FENG, S.; SUN, D.W. Study on infrared spectroscopy technique for fast measurement of protein content in milk powder based on LS-SVM. *Journal of Food Engineering*, v.84, p.124-131, 2008.

YU, H.Y.; NIU, X.Y.; LIN, H.J.; YING, Y.B.; LI, B.B.; PAN, X.X. A feasibility study on on-line determination of rice wine composition by Vis-NIR spectroscopy and least-squares support vector machines. *Food Chemistry*, v.113, p.291-296, 2009.

**ANEXO A** Tabela 1 – Resultados dos modelos *i*PLS para dados espectrais originais (sem aplicação da TDW), para os conjuntos com os 2 níveis de decomposição pela *wavelet* Daubechies com 4 coeficientes, para os conjuntos com os 2 níveis de decomposição pela *wavelet* Daubechies com 6 coeficientes, para os conjuntos com os 2 níveis de decomposição pela *wavelet* Daubechies com 8 coeficientes.

C	IR	ND	NVT	NI	QVI	MODELOS <i>i</i> PLS								
						FE (cm <sup>-1</sup> )	EC1	EC2	EC3	VL	EP1	EP2	EP3	R <sup>2</sup> <sub>cal</sub>
4	99,93	1	868	10	87	1322– 978	0,335	0,390	0,268	4	0,321	0,346	0,307	0,999
4	99,93	1	868	15	58	1322 – 1099	0,344	0,405	0,268	4	0,350	0,385	0,312	0,999
4	99,61	2	434	10	44	1334 – 995	0,333	0,388	0,267	4	0,324	0,340	0,307	0,999
4	99,61	2	434	15	29	1326 – 1103	0,350	0,410	0,277	4	0,359	0,398	0,317	0,999
6	99,36	1	868	10	87	1322 – 978	0,334	0,389	0,267	4	0,326	0,344	0,307	0,999
6	99,36	1	868	15	58	1322 – 1099	0,341	0,405	0,263	4	0,355	0,392	0,312	0,999
6	97,42	2	434	10	44	1334 – 995	0,340	0,403	0,262	4	0,326	0,348	0,303	0,999
6	97,42	2	434	15	29	1326 – 1103	0,357	0,416	0,287	4	0,372	0,414	0,325	0,999
8	98,31	1	868	10	87	1322 – 978	0,334	0,391	0,266	4	0,326	0,344	0,307	0,999
8	98,31	1	868	15	58	1322 – 1099	0,347	0,408	0,271	4	0,358	0,396	0,315	0,999
8	94,04	2	434	10	44	1673 – 1334	0,391	0,352	0,275	8	0,418	0,494	0,323	0,999
8	94,04	2	434	15	29	1326 – 1103	0,359	0,418	0,290	4	0,374	0,420	0,323	0,998

C – número de coeficientes.

IR – informação retida em percentual.

ND – número de decomposição.

NVT – número de variáveis totais.

NI – número de intervalos.

QVI – quantidade de variáveis por intervalo.

FE – faixa espectral.

EC1 – erros de validação cruzada (RMSECV) em percentual com dados lidos nos dois espectrofotômetros.

EC2 – erros de validação cruzada (RMSECV) em percentual com dados lidos no espectrofotômetro Shimadzu.

EC3 – erros de validação cruzada (RMSECV) em percentual com dados lidos no espectrofotômetro Nicolet.

VL – variáveis latentes.

EP1 – erros de previsão (RMSEP) em percentual com dados lidos nos dois espectrofotômetros.

EP2 – erros de previsão (RMSEP) em percentual com dados lidos no espectrofotômetro Shimadzu.

EP3 – erros de previsão (RMSEP) em percentual com dados lidos no espectrofotômetro Nicolet.

R<sup>2</sup><sub>cal</sub> – correlação dos modelos de calibração.

Fonte: Elaborado pela autora.

**ANEXO B** Tabela 1 – Resultados dos modelos LS-SVM para dados espectrais originais (sem aplicação da TDW), para os conjuntos com os 2 níveis de decomposição pela *wavelet* Daubechies com 4 coeficientes, para os conjuntos com os 2 níveis de decomposição pela *wavelet* Daubechies com 6 coeficientes, para os conjuntos com os 2 níveis de decomposição pela *wavelet* Daubechies com 8 coeficientes.

C	IR	ND	NVT	NI	QVI	MODELOS LS-SVM							
						FE (cm <sup>-1</sup> )	EC1	EC2	EC3	EP1	EP2	EP3	R <sup>2</sup> <sub>cal</sub>
4	99,93	1	868	10	87	1322 – 978	0,312	0,382	0,259	0,401	0,421	0,380	0,995
4	99,93	1	868	15	58	1322 – 1099	0,321	0,405	0,264	0,384	0,431	0,330	0,995
4	99,61	2	434	10	44	1334 – 995	0,318	0,390	0,267	0,470	0,519	0,416	0,995
4	99,61	2	434	20	22	1164 – 995	0,371	0,437	0,297	0,804	0,978	0,581	0,993
6	99,36	1	868	10	87	1322 – 978	0,311	0,372	0,246	0,417	0,439	0,393	0,995
6	99,36	1	868	15	58	1322 – 1099	0,314	0,397	0,251	0,354	0,365	0,355	0,995
6	97,42	2	434	10	44	1334 – 995	0,315	0,395	0,258	0,531	0,602	0,448	0,995
6	97,42	2	434	20	22	1164 – 995	0,353	0,447	0,283	0,831	1,008	0,605	0,992
8	98,31	1	868	10	87	1322 – 978	0,310	0,373	0,243	0,442	0,463	0,420	0,995
8	98,31	1	868	15	58	1322 – 1099	0,315	0,399	0,252	0,361	0,368	0,354	0,995
8	94,04	2	434	10	44	1673 – 1334	0,340	0,413	0,244	1,188	1,458	0,836	0,983
8	94,04	2	434	20	22	1164 – 995	0,348	0,437	0,289	0,833	0,993	0,635	0,993

C – número de coeficientes.

IR – informação retida em percentual.

ND – número de decomposição.

NVT – número de variáveis totais.

NI – número de intervalos.

QVI – quantidade de variáveis por intervalo.

FE – faixa espectral.

EC1 – erros de validação cruzada (RMSECV) em percentual com dados lidos nos dois espectrofotômetros.

EC2 – erros de validação cruzada (RMSECV) em percentual com dados lidos no espectrofotômetro Shimadzu.

EC3 – erros de validação cruzada (RMSECV) em percentual com dados lidos no espectrofotômetro Nicolet.

EP1 – erros de previsão (RMSEP) em percentual com dados lidos nos dois espectrofotômetros.

EP2 – erros de previsão (RMSEP) em percentual com dados lidos no espectrofotômetro Shimadzu.

EP3 – erros de previsão (RMSEP) em percentual com dados lidos no espectrofotômetro Nicolet.

R<sup>2</sup><sub>cal</sub> – correlação dos modelos de calibração.

Fonte: Elaborado pela autora.



# ANEXO C – Artigo publicado nos anais do 4º Congresso da Rede Brasileira de Tecnologia em Biodiesel e do 7º Congresso Brasileiro de Plantas Oleaginosas, Óleos, Gorduras e Biodiesel

4º Congresso da Rede Brasileira de Tecnologia de Biodiesel  
7º Congresso Brasileiro de Plantas Oleaginosas, Óleos, Gorduras e Biodiesel

## Quantificação de biodiesel em blendas biodiesel/diesel: comparação de diferentes modelos de regressão multivariada

Márcia Werlang (UNISC, werlangm@yahoo.com.br), Rubén E. P. Pazos (UNISC, rpazos@unisc.br), Mariela S. Viera (UNISC, marielaviera@hotmail.com), Daniela Fachini (UFRGS, danifachini@ig.com.br), Annelise E. Gerbase (UFRGS, agerbase@ufrgs.br) e Marco F. Ferrão (UFRGS, ferrao@iq.ufrgs.br)

**Palavras Chave:** Blendas biodiesel/diesel, quantificação, *wavelets*, infravermelho, regressão multivariada.

### 1 - Introdução

Para a comercialização do biodiesel em mistura com diesel em percentuais crescentes se faz necessário o estudo de metodologias precisas, rápidas e com custo baixo, que quantifiquem o percentual estabelecido pela legislação. Nesse contexto, os métodos de regressão multivariada aplicados à espectroscopia no infravermelho permitem o desenvolvimento de metodologias ambientalmente corretas, já utilizadas em diferentes áreas.

No desenvolvimento dos modelos multivariados um grande volume de dados espectroscópicos é utilizado tornando o custo computacional elevado. Buscando reduzir este custo o presente estudo propôs a utilização das transformadas discretas *wavelet*, seguida de modelagem por dois métodos multivariados: regressão por mínimos quadrados parciais (PLS)<sup>1</sup> ou regressão por máquina de vetores de suporte (LS-SVM)<sup>2</sup>.

Além disso, avaliou-se a influência de dados oriundos de diferentes equipamentos, buscando gerar um único modelo capaz de quantificar o biodiesel presente nas blendas, independentemente do equipamento utilizado.

### 2 - Material e Métodos

Foram formuladas 45 blendas biodiesel/diesel com concentrações de biodiesel que variaram de 8% (v/v) a 30% (v/v), com base na Norma ABNT NBR número 15568 de 2008. O biodiesel utilizado na composição das blendas foi doado pela OLEOPLAN S.A. e os óleos diesel, interior e metropolitano, foram obtidos na Refinaria Alberto Pasqualini – REFAP S/A. Estas blendas foram homogeneizadas antes da aquisição dos espectros. Os espectros foram obtidos em dois equipamentos distintos (um espectrofotômetro da Nicolet Magma 550 e um espectrofotômetro Shimadzu IR Prestige-21), em duplicata, na faixa de 4000-650 cm<sup>-1</sup>, com resolução de 4 cm<sup>-1</sup> e 32 varreduras, utilizando um acessório de reflexão total atenuada (ATR) com cristal de seleneto de zinco (ZnSe). Um total de 180 espectros foram assim obtidos.

Para a construção da matriz de dados os espectros foram normalizados e posteriormente efetuou-se a média das duplicatas desses espectros. Foi empregado o método da transformada discreta *wavelet* (TDW)<sup>3</sup> Daubechies com 4 coeficientes para decomposição da matriz de dados, em 2 níveis, reduzindo-se para ½ e ¼ dos sinais originais, sendo posteriormente construídos os modelos multivariados no ambiente MATLAB 7.0. As matrizes de dados espectrais com e sem decomposição foram divididas em dois grupos:

2/3 para a calibração (60 espectros médios) e 1/3 para a previsão (30 espectros médios). Para avaliar os modelos construídos foram calculados e comparados os valores das raízes quadradas dos erros médios de calibração cruzada (RMSECV) e dos erros médios de previsão (RMSEP).

### 3 - Resultados e Discussão

Para a construção dos modelos, *i*PLS e LS-SVM, foram empregados os espectros originais (sem aplicação da TDW), os espectros com 1 nível de decomposição e os espectros com 2 níveis de decomposição, além da subdivisão dos espectros em 10, 15 ou 20 intervalos, afim de identificar qual a melhor faixa espectral para realizar a quantificação de biodiesel em blendas biodiesel/diesel. Os resultados dos modelos estão apresentados na Tabela 1.

**Tabela 1.** Resultados dos modelos *i*PLS e LS-SVM para a quantificação de biodiesel em blendas biodiesel/diesel empregando os espectros originais e decompostos com TDW, sendo subdivididos em 10, 15 ou 20 intervalos.

ND	IR	NI	Modelos <i>i</i> PLS			
			RMSECV (% v/v)	RMSEP (% v/v)	VL	R
0	100	10	0,33	0,33	4	0,99
0	100	15	0,34	0,36	4	0,99
0	100	20	0,35	0,33	4	0,99
1	99,93	10	0,36	0,32	4	0,99
1	99,93	15	0,34	0,35	4	0,99
1	99,93	20	0,35	0,33	5	0,99
2	99,61	10	0,33	0,32	4	0,99
2	99,61	15	0,35	0,36	4	0,99
2	99,61	20	0,36	0,38	2	0,99
Modelos LS-SVM						
0	100	10	0,31	0,37	-	0,99
0	100	15	0,33	0,36	-	0,99
0	100	20	0,34	0,39	-	0,99
1	99,93	10	0,31	0,40	-	0,99
1	99,93	15	0,32	0,38	-	0,99
1	99,93	20	0,34	0,38	-	0,99
2	99,61	10	0,32	0,47	-	0,99
2	99,61	15	0,34	0,42	-	0,99
2	99,61	20	0,37	0,80	-	0,99

ND – Número de decomposições;

IR – Informação retida (valores em %);

NI – Número de intervalos;

VL – Variáveis latentes;

R – correlação.

Em uma primeira análise pode-se observar que há uma uniformidade entre os erros de calibração (RMSECV) e previsão (RMSEP) dos modelos *i*PLS, indicando que esses modelos tem boa capacidade preditiva, tendo,

4º Congresso da Rede Brasileira de Tecnologia de Biodiesel  
7º Congresso Brasileiro de Plantas Oleaginosas, Óleos, Gorduras e Biodiesel

inclusive, erros de previsão inferiores ao que está estabelecido na Norma ABNT NBR número 15568. O mesmo é verificado nos modelos LS-SVM, com exceção do último modelo da Tabela 1. Estes resultados mostram a equivalência dos modelos *i*PLS e LS-SVM desenvolvidos.

Também é possível constatar na Tabela 1 que a decomposição dos espectros em 2 níveis de transformada (Figuras 1 e 2), ou seja a redução de 1736 variáveis para 434 variáveis, não prejudica a capacidade de previsão dos modelos quando utilizado o algoritmo *i*PLS, o que confere menor custo e rapidez na construção dos modelos multivariados.

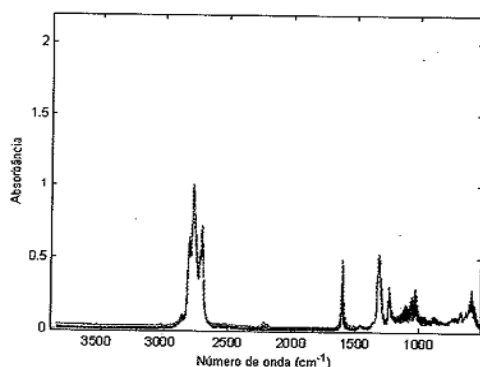


Figura 1. Espectros originais das blendas.

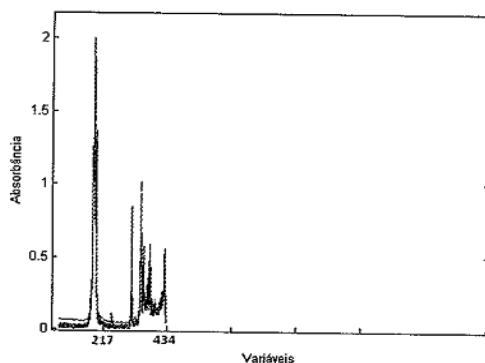


Figura 2. Espectros com 2 níveis de decomposição.

Ainda dentre os modelos obtidos podemos observar que a região do espectro correspondente ao intervalo 18 da Figura 3 representa a melhor faixa do espectral para a determinação do biodiesel nas amostras estudadas, visto que 2 variáveis latentes são suficientes para obter valores de RMSECV e RMSEP de 0,36 % (v/v) e 0,38 % (v/v), respectivamente. Na Figura 4 é apresentada a curva de correlação entre os valores de referência e os valores previstos por este modelo *i*PLS com 2 níveis de decomposição via TDW. Pode-se verificar a adequada precisão na quantificação do biodiesel, a partir dos espectros das amostras de previsão obtidas tanto do espectrofotômetro da Nicolet Magma 550, quanto da Shimadzu IR Prestige-21.

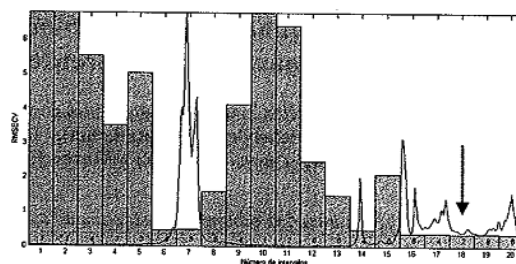


Figura 3. Espectros com 2 níveis de decomposição e subdivididos em 20 intervalos, onde o intervalo indicado pela seta foi selecionado pelo *i*PLS como sendo o mais representativo para quantificar biodiesel em blendas biodiesel/diesel.

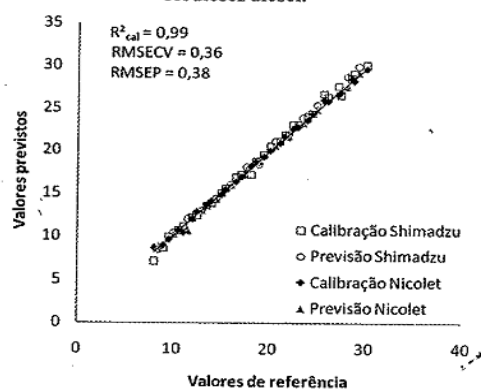


Figura 4. Curva de correlação entre os valores de referência e os valores previstos pelo modelo *i*PLS com 2 níveis de decomposição e com subdivisão dos espectros em 20 intervalos.

Os resultados mostram a eficiência da TDW na redução do número de variáveis sem perda das informações espectrais, o que possibilita a construção de modelos multivariados robustos em menor tempo. Pode-se também verificar que o emprego de modelos *i*PLS para a quantificação de biodiesel em blendas biodiesel/diesel é tão eficiente quanto os modelos de regressão obtidos por LS-SVM. Com isso, foi possível quantificar o biodiesel em blendas através de uma metodologia rápida e não destrutiva, mesmo quando são modelados simultaneamente dados de diferentes equipamentos.

#### 4 - Agradecimentos

FAPERGS, CAPES, CNPq, REFAP, Oleoplan, LABCOM/CECOM-IQ/UFRGS, FAP-UNISC.

#### 5 - Bibliografia

- <sup>1</sup> Norgaard, L.; Saudland, A.; Wagner, J.; Nielsen, J.P.; Munck, L. & Engelsen, S.B.; *Applied spectroscopy*, 2000, 54, 413.
- <sup>2</sup> Borin, A.; Ferrão, M.F.; Mello, C.; Cordi, L.; Pataca, L.C.M.; Durán, N.; Poppi, R.J.; *Analytical and Bioanalytical Chemistry*, 2007, 387, 1105.
- <sup>3</sup> Daubechies, I.; *Ten lectures on wavelets*. Philadelphia: SIAM (1992).