

Universidade Estadual de Campinas

Instituto de Química



Dissertação de Mestrado

**“Comparação entre máquinas de vetores de suporte
por mínimos quadrados (LS-SVM) e métodos lineares
para transferência de calibração”**

Autor: Danilo Althmann Maretto

Orientador: Prof. Dr. Ronei Jesus Poppi

Campinas

2007

M335c	<p>Mareto, Danilo Althmann. Comparação entre máquinas de vetores de suporte por mínimos quadrados (LS-SVM) e métodos lineares de transferência de calibração . -- Campinas, SP: [s.n], 2007.</p> <p>Orientador: Ronei Jesus Popi.</p> <p>Mestrado- Universidade Estadual de Campinas, Instituto de Química.</p> <p>1.Cachaça. 2. Ração. 3.LS-SVM. 4.Calibração. I. Poppi. II. Universidade Estadual de Campinas. Instituto de Química. III. Título.</p>
-------	---

Título em inglês: Comparison between Least-Square support vector machines and linear methods for calibration transfer.

Palavras-chaves em inglês: Cachaça, Dog food, LS-SVM, Calibration.

Área de concentração: Química Analítica.

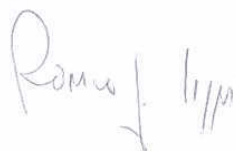
Titulação: Mestre em Química Analítica.

Banca examinadora: Prof. Dr. Ronei Jesus Popi (orientador), Prof. Dr. José Alberto Fracassi da Silva (IQ-Unicamp), Prof. Dr. César Alexandre de Mello (UNIFRAN).

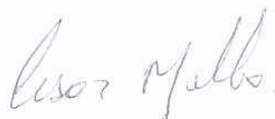
Data de defesa: 27/02/2007

BANCA EXAMINADORA

Prof. Dr. Ronei Jesus Poppi (Orientador)



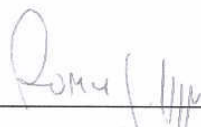
Prof. Dr. Cesar Alexandre de Mello (UNIFRAN)



Prof. Dr. José Alberto Fracassi da Silva (IQ-UNICAMP)



Este exemplar corresponde à redação final da
Dissertação de Mestrado defendida pelo aluno **DANILO
ALTHMANN MARETTO**, aprovada pela Comissão
Julgadora em 27 de fevereiro de 2007.



Prof. Dr. Ronei Jesus Poppi
(Presidente da Banca)

“Existe uma teoria que diz que, se um dia alguém descobrir exatamente para que serve o Universo e por que ele está aqui, ele desaparecerá instantaneamente e será substituído por algo ainda mais estranho e inexplicável”

“Existe uma segunda teoria que diz que isso já aconteceu”

Douglas Adams

Agradecimentos

Ao prof. Dr. Ronei Jesus Poppi pela oportunidade, paciência, orientação e amizade.

Aos membros do grupo LAQQA pelo apoio e amizade: Alessandra, Gilmore, Luciana, Jez, Luiz, Marcello, Patrícia, Paulo Henrique, Thiago, Renato, Marcos, Bruno, Diórginis e Werickson.

À Rosy, ex-aluna do LAQQA e funcionária da Nutron, pelo fornecimento das amostras, material bibliográfico, apoio e amizade.

À Central Analítica do Instituto de Química da Unicamp pelo empréstimo do densímetro digital Anton Paar DMA 4500 para as análises do teor alcoólico em amostras de cachaça.

À técnica do Instituto de Química, Cláudia Martelli pela amizade e pelo auxílio nas análises de infravermelho.

À todos os professores e funcionários do Instituto de Química que direta ou indiretamente contribuíram para a realização deste trabalho.

À todos os meus amigos, os quais seria impossível nomeá-los todos aqui, mas que sempre foram fundamentais.

E principalmente à minha família, que sempre me apoiou e sem a qual nada seria possível.

Currículo

1. Dados pessoais

Danilo Althmann Maretto

Nascimento: 22/02/1979
Naturalidade: Campinas/SP
E-mail: althmann@yahoo.com.br

2. Formação Acadêmica

Bacharelado em Química – Modalidade Tecnológica

Instituto de Química - Universidade Estadual de Campinas

Técnico em Bioquímica

Escola Técnica Estadual “Conselheiro Antonio Prado”

3. Produção científica

3.1. Resumo dos trabalhos científicos apresentados em congresso

3.1.1 **Determinação de etileno glicol em óleo lubrificante por espectroscopia no infravermelho e PLS**, Borges Neto, W.; Maretto, D.A.; Poppi, R.J.; Poster, 28ª Reunião Anual da Sociedade Brasileira de Química, Poços de Caldas, 02 de junho de 2005

3.1.2 **Espectroscopia no Infravermelho e seleção de variáveis para determinar adulteração de lubrificantes por combustíveis**, da Silva, G. A.; Borges Neto, W.; Maretto, D.A.; Poppi, R.J.; 13º Encontro Nacional de Química Analítica, Poster, Universidade Federal Fluminense, Niterói, RJ, 12 a 16 de setembro de 2005

3.1.3 **Determinação de fraude em leite por infravermelho próximo e quimiometria**, Borin, A.; da Silva, G. A.; Maretto D.A.; Poppi, R.J.; Poster e apresentação oral, 13º Encontro Nacional de Química Analítica, Universidade Federal Fluminense, Niterói, RJ, 12 a 16 de setembro de 2005

3.2. Publicações

3.3.1. Artigo publicado – Borin, A; Ferrão, M. F.; Mello, C.; Maretto, D. A.; Poppi, R.J.; **Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk**, Analytica Chimica Acta 579 (2006) 25 –32

3.3.2. Artigo no prelo – 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 (2006), QN 180/06

4. Outros

4.1. Curso Científico

Curso Básico de Fluorescência Molecular

Instituto de Química

Universidade Estadual de Campinas

03, 04, 08 e 10 de agosto de 2005, com carga horária de 15 horas

5. Estágios

Centro Integrado de Pesquisas Oncohematológicas na Infância

Universidade Estadual de Campinas

Campinas/SP

2001-2003

Centro de Biologia Molecular e Engenharia Genética

Universidade Estadual de Campinas

Campinas/SP

1999-2000

6. Experiência profissional

Técnico em Química

Instituto de Química

Universidade Estadual de Campinas

2004-atualmente

Auxiliar Técnico em Biologia Molecular

Allelyx Applied Genomics

Campinas/SP

2003-2004

Resumo

Autor: Danilo Althmann Maretto

Orientador: Ronei Jesus Poppi

Este trabalho teve como objetivo comparar os métodos lineares de calibração "mínimos quadrados parciais" (PLS) e "padronização direta por partes" (PDS) e um método não-linear "máquina de vetores de suporte por mínimos quadrados" (LS-SVM) na transferência de calibração para modelos de calibração multivariada onde se determinou porcentagem de etanol em cachaça a cinco temperaturas diferentes e para modelos onde se determinou a porcentagem de proteína e gordura em ração para cães em três diferentes granulometrias através de espectroscopia na região do infravermelho próximo.

Foram preparadas 50 amostras de cachaça entre 20,86 e 46,48% v/v através de diluição com água Milli-Q ou adição de etanol P.A. (Merck) à cachaça original. A porcentagem alcoólica foi obtida através de um densímetro digital Anton Paar DMA 4500 e os espectros a 5 temperaturas diferentes (15, 20, 25, 30 e 35°C) foram obtidos na faixa de 850 a 1050 nm em um equipamento Agilent 8453.

Um total de 38 amostras de ração moídas foi fornecido pela empresa Nutron Alimentos Ltda a qual realizou testes padrão para determinação de porcentagem de proteína e gordura nas mesmas. As amostras foram então peneiradas, sendo divididas em 3 grupos com tamanhos de partícula diferentes. Os espectros foram obtidos para todos os grupos de partículas de todas as amostras na faixa de 1000 a 2400 nm em um equipamento Varian Cary 5G.

Foram feitas quatro propostas diferentes para se fazer a transferência de calibração para cada uma das três aplicações (determinação do teor de etanol em cachaça, e do teor de proteína e gordura em ração). Na grande maioria delas o LS-SVM foi quem apresentou modelos mais bem ajustados.

Abstract

Author: Danilo Althmann Mareto

Adviser: Ronei Jesus Poppi

The aim of this work was to compare the linear methods of calibration “Partial Least Squares” (PLS) and “Piece-wise Direct Standardization” (PDS) and a nonlinear method “Least-Squares Support Vector Machines” (LS-SVM) on calibration transfer to multivariate calibration models to the determination of alcoholic grade in cachaça in five different temperatures and to determination of protein and fat content in dog food in three different particule sizes by using near infrared spectroscopy.

It has been prepared 50 cachaça samples between 20.86 and 46.48% v/v through dilution with Milli-Q water or adding etanol P.A.(Merck) to the original cachaça. The alcoholic grade has been obtained through a Anton Paar DMA 4500 digital densimeter and the spectra in five different temperatures (15, 20, 25, 30 and 35°C) has been obtained between 850 and 1050 nm in a Agilent 8453 equipament.

The 38 grinded dog food samples were supplied by Nutron Alimentos Ltda wich has realized the standard tests to determination of protein and fat mass porcentage in them. The samples have been bolted, been divided in three groups with different particle sizes. The spectra have been obtained to all the particle groups of all samples between 100 and 24000 nm in a Varian Cary 5G equipament.

It has been done four different proposals to do the calibration transfer to each one of the three applications (etanol grade in cachaça, and protein and fat in dog food). In the most of them LS-SVM has gotten better adjusted models.

Índice de Figuras

Figura 1: diagrama de energia potencial de um oscilador anarmônico.....	11
Figura 2: Espectros na região do infravermelho de ondas curtas para uma mesma amostra de cachaça a várias temperaturas (15, 20, 25, 30 e 35°C).....	12
Figura 3: Espalhamento de radiação incidida em amostra sólida.....	13
Figura 4: Efeito da granulometria sobre o espectro de reflectância difusa na região do infravermelho próximo de amostra de ração para cães.....	14
Figura 5: Representação esquemática da decomposição por PCA.....	18
Figura 6: Representação esquemática do PDS.....	26
Figura 7: Mudança de dimensão dos dados.....	30
Figura 8: Superfície de otimização dos parâmetros σ^2 e γ em função do erro (RMSECV) obtida na previsão de amostras de cachaça a 15°C utilizando amostras nas demais temperaturas no desenvolvimento do modelo de calibração.....	31
Figura 9: Espectros na região do infravermelho de ondas curtas para amostras de cachaça a 20°C.....	39
Figura 10: Representação esquemática da proposta 1 para transferência de calibração para determinação da porcentagem de álcool em cachaça a diferentes temperaturas.....	40
Figura 11: Representação esquemática da proposta 2 para transferência de calibração para determinação da porcentagem de álcool em cachaça a diferentes temperaturas.....	41
Figura 12: Representação esquemática da proposta 3 para transferência de calibração para determinação da porcentagem de álcool em cachaça a diferentes temperaturas.....	42
Figura 13: Representação esquemática da proposta 4 para transferência de calibração para determinação da porcentagem de álcool em cachaça a diferentes temperaturas.....	42
Figura 14: Exemplo de gráfico de erro médio quadrático de calibração cruzada por número de variáveis latentes utilizado na determinação do número ideal de variáveis latentes na construção dos modelos PLS e PDS.....	43

Figura 15: Erros relativos na previsão da porcentagem de álcool em amostras de cachaça obtidos nos modelos construídos para a proposta 1, quando as amostras de validação tiveram seus espectros obtidos nas temperaturas: A. 15°C; B. 20°C; C. 25°C; D. 30°C e E. 35°C.....	46
Figura 16: Erros relativos na previsão da porcentagem de álcool em amostras de cachaça obtidos nos modelos construídos para a proposta 2, quando as amostras de validação tiveram seus espectros obtidos nas temperaturas: A. 15°C; B. 20°C; C. 25°C; D. 30°C e E. 35°C.....	49
Figura 17: Erros relativos na previsão da porcentagem de álcool em amostras de cachaça obtidos nos modelos construídos para a proposta 3, quando as amostras de validação tiveram seus espectros obtidos nas temperaturas: A. 15°C; B. 25°C; C. 30°C e D. 35°C.....	52
Figura 18: Erros relativos na previsão da porcentagem de álcool em amostras de cachaça obtidos nos modelos construídos para a proposta 4.....	54
Figura 19: Diferentes tamanhos de partícula de uma mesma amostra de ração para cães, onde: A. partículas maiores que 0,50 mm; B. partículas com tamanho entre 0,50 e 0,35 mm e C. partículas menores que 0,35 mm.....	59
Figura 20: Espectros na região do infravermelho próximo de amostras de ração para cães com tamanho de partículas maior do que 0,50 mm.....	61
Figura 21: Primeira derivada dos espectros de ração obtidos.....	62
Figura 22: Representação esquemática da proposta 1 para transferência de calibração para determinação do teor de proteína e gordura em ração para cães em diferentes granulometrias.....	63
Figura 23: Representação esquemática da proposta 2 para transferência de calibração para determinação do teor de proteína e gordura em ração para cães em diferentes granulometrias.....	63
Figura 24: Representação esquemática da proposta 3 para transferência de calibração para determinação do teor de proteína e gordura em ração para cães em diferentes granulometrias.....	64

Figura 25: Representação esquemática da proposta 4 para transferência de calibração para determinação do teor de proteína e gordura em ração para cães em diferentes granulometrias.....	64
Figura 26: Exemplo de gráfico de erro médio quadrático de calibração cruzada por número de variáveis latentes utilizado na determinação do número ideal de variáveis latentes na construção dos modelos PLS e PDS.....	65
Figura 27: Erros relativos na previsão do teor de proteína em amostras de ração para cães obtidos nos modelos construídos para a proposta 1, quando as amostras de validação tiveram seus espectros obtidos nas granulometrias: A, B e C.....	67
Figura 28: Erros relativos na previsão do teor de proteína em amostras de ração para cães obtidos nos modelos construídos para a proposta 2, quando as amostras de validação tiveram seus espectros obtidos nas granulometrias: A, B e C.....	69
Figura 29: Erros relativos na previsão do teor de proteína em amostras de ração para cães obtidos nos modelos construídos para a proposta 3, onde: 1. Cal: A e Val: B; 2. Cal:A e Val C; 3. Cal:B e Val: A; 4. Cal: B e Val: C; 5. Cal: C e Val A e 6. Cal: C e Val: B.....	72
Figura 30: Erros relativos na previsão do teor de proteína em amostras de ração para cães obtidos nos modelos construídos para a proposta 4.....	74
Figura 31: Erros relativos na previsão do teor de gordura em amostras de ração para cães obtidos nos modelos construídos para a proposta 1, quando as amostras de validação tiveram seus espectros obtidos nas granulometrias: A, B e C.....	76
Figura 32: Erros relativos na previsão do teor de gordura em amostras de ração para cães obtidos nos modelos construídos para a proposta 2, quando as amostras de validação tiveram seus espectros obtidos nas granulometrias: A, B e C.....	78

Figura 33: Erros relativos na previsão do teor de gordura em amostras de ração para cães obtidos nos modelos construídos para a proposta 3, onde: 1. Cal: A e Val: B; 2. Cal:A e Val C; 3. Cal: B e Val: A; 4. Cal: B e Val: C; 5. Cal: C e Val A e 6. Cal: C e Val: B.....81

Figura 34: Erros relativos na previsão do teor de gordura em amostras de ração para cães obtidos nos modelos construídos para a proposta 4.....83

Índice de Tabelas

Tabela 1: variação de composição (em mg/100mL de álcool anidro) de aguardentes de cana.....	37
Tabela 2: porcentagem alcoólica (V/V) nas amostras de cachaça.....	38
Tabela 3: Atribuição das bandas do espectro na região do infravermelho para cachaça.....	39
Tabela 4: RMSEPs obtidos e parâmetros utilizados na construção dos modelos da proposta 1 para transferência de calibração na determinação do teor de álcool em cachaça.....	44
Tabela 5: Teste F para comparação dos modelos obtidos na proposta 1 para determinação do teor de álcool em cachaça.	45
Tabela 6: RMSEPs obtidos e parâmetros utilizados na construção dos modelos da proposta 2 para transferência de calibração na determinação do teor de álcool em cachaça.....	47
Tabela 7: Teste F para comparação dos modelos obtidos na proposta 2 para determinação do teor de álcool em cachaça.....	48
Tabela 8: RMSEPs obtidos e parâmetros utilizados na construção dos modelos da proposta 3 para transferência de calibração na determinação do teor de álcool em cachaça.....	50
Tabela 9: Teste F para comparação dos modelos obtidos na proposta 3 para determinação do teor de álcool em cachaça.....	51
Tabela 10: RMSEPs obtidos e parâmetros utilizados na construção dos modelos da proposta 4 para transferência de calibração na determinação do teor de álcool em cachaça.....	53
Tabela 11: Teste F para comparação dos modelos obtidos na proposta 4 para determinação do teor de álcool em cachaça.....	53
Tabela 12: Teor de proteína e gordura em % (m/m) nas amostras de ração para cães.....	60
Tabela 13: Atribuição das bandas do espectro na região do infravermelho para as amostras de ração.....	61

Tabela 14: RMSEPs obtidos e parâmetros utilizados na construção dos modelos da proposta 1 para transferência de calibração na determinação do teor proteína em ração para cães.....	66
Tabela 15: Teste F para comparação dos modelos obtidos na proposta 1 para determinação do teor de proteína em ração para cães.....	67
Tabela 16: RMSEPs obtidos e parâmetros utilizados na construção dos modelos da proposta 2 para transferência de calibração na determinação do teor proteína em ração para cães.....	68
Tabela 17: Teste F para comparação dos modelos obtidos na proposta 2 para determinação do teor de proteína em ração para cães.....	69
Tabela 18: RMSEPs obtidos e parâmetros utilizados na construção dos modelos da proposta 3 para transferência de calibração na determinação do teor proteína em ração para cães.....	70
Tabela 19: Teste F para comparação dos modelos obtidos na proposta 3 para determinação do teor de proteínas em ração para cães.....	71
Tabela 20: RMSEPs obtidos e parâmetros utilizados na construção dos modelos da proposta 4 para transferência de calibração na determinação do teor proteína em ração para cães.....	73
Tabela 21: Teste F para comparação dos modelos obtidos na proposta 4 para determinação do teor de proteína em ração para cães.....	73
Tabela 22: RMSEPs obtidos e parâmetros utilizados na construção dos modelos da proposta 1 para transferência de calibração na determinação do teor de gordura em ração para cães.....	74
Tabela 23: Teste F para comparação dos modelos obtidos na proposta 1 para determinação do teor de gordura em ração para cães.....	75
Tabela 24: RMSEPs obtidos e parâmetros utilizados na construção dos modelos da proposta 2 para transferência de calibração na determinação do teor de gordura em ração para cães.....	77
Tabela 25: Teste F para comparação dos modelos obtidos na proposta 2 para determinação do teor de gordura em ração para cães.....	77

Tabela 26: RMSEPs obtidos e parâmetros utilizados na construção dos modelos da proposta 3 para transferência de calibração na determinação do teor de gordura em ração para cães.....	79
Tabela 27: Teste F para comparação dos modelos obtidos na proposta 3 para determinação do teor de gordura em ração para cães.....	80
Tabela 28: RMSEPs obtidos e parâmetros utilizados na construção dos modelos da proposta 4 para transferência de calibração na determinação do teor de gordura em ração para cães.....	82
Tabela 29: Teste F para comparação dos modelos obtidos na proposta 4 para determinação do teor de gordura em ração para cães.....	82

Índice Geral

Introdução e objetivos	1
Capítulo 1: Espectroscopia no Infravermelho	5
1.1 Espectroscopia no Infravermelho.....	7
1.2 Teoria.....	9
1.3 Efeito da temperatura em espectros vibracionais.....	11
1.4 Efeito de granulometria no espectro.....	13
Capítulo 2 – Quimiometria	15
2.1 Quimiometria.....	17
2.2 Análise de Componentes Principais (PCA).....	18
2.3 Regressão por Mínimos Quadrados Parciais (PLS).....	20
2.4 Transferência de calibração.....	22
2.4.1 Padronização direta por partes (PDS).....	24
2.5 Máquinas de vetores de suporte por mínimos quadrados (LS-SVM).....	26
2.6 Parâmetros de qualidade.....	31
Capítulo 3: Determinação do teor de álcool em cachaça	33
3.1 Cachaça.....	35
3.2 Experimental.....	37
3.3 Proposta 1.....	44
3.4 Proposta 2.....	47
3.5 Proposta 3.....	50
3.6 Proposta 4.....	52
Capítulo 4. Determinação do teor de proteína e gordura em ração para cães	55
4.1 Ração para cães.....	57
4.2 Experimental.....	58

4.3 Determinação do teor de proteína em ração para cães.....	66
4.3.1 Proposta 1.....	66
4.3.2 Proposta 2.....	68
4.3.3 Proposta 3.....	70
4.3.4 Proposta 4.....	73
4.4 Determinação do teor de gordura em ração para cães.....	74
4.4.1 Proposta 1.....	74
4.4.2 Proposta 2.....	76
4.4.3 Proposta 3.....	78
4.4.4 Proposta 4.....	82
5. Conclusões.....	85
6. Referências bibliográficas.....	89

Introdução e objetivos

Introdução e objetivos

A utilização de calibração multivariada para extração de informações quantitativas de dados químicos está sendo cada vez mais empregada, especialmente para dados obtidos através de espectroscopia na região do infravermelho, já que esse tipo de técnica analítica fornece análises simples, rápidas, baratas e não destrutivas, sendo, em muitos casos, uma boa substituta a técnicas clássicas de análise.

Entretanto, para a construção de modelos de calibração multivariada para espectroscopia no infravermelho é necessário trabalhar com uma grande quantidade de amostras, obtendo seus espectros e realizando o método padrão necessário para determinação da propriedade de interesse, o que pode demandar muito tempo, custo e esforço para seu completo desenvolvimento. Além disso, esses modelos complexos de calibração podem ser inúteis se houver variações entre novas medidas de previsão e as amostras utilizadas para a calibração.

Duas variações podem ser distinguidas, sendo a primeira observada quando o modelo de calibração é desenvolvido utilizando espectros obtidos em um equipamento e a previsão de uma nova amostra é feita utilizando espectros obtidos em outro equipamento. A segunda fonte de variação vem de diferenças físicas entre amostras de calibração e de previsão.

Essas diferenças físicas podem ser variações de temperatura ou tamanho de partícula entre as fases de calibração e previsão. Independente do caso é necessário desenvolver um novo modelo de calibração, ou seja, coletar inúmeras amostras, analisá-las através do método de referência e desenvolver novamente o modelo de calibração.

Existem alguns meios de tornar a primeira calibração útil para a previsão de amostras em outras condições físicas, de forma a não se necessitar esta nova calibração. Dentre esses meios encontram-se a calibração robusta e padronização de espectros

A transferência de calibração pode em certas situações ser realizada pela utilização de métodos de calibração multivariada não-lineares que possam tratar com as diferenças espectrais ocasionadas pelas variações das condições físicas presentes

nas medidas. Dentre esses métodos destaca-se a máquina de vetores de suporte por mínimos quadrados (LS-SVM).

Neste trabalho foi feita a comparação entre dois métodos de transferência de calibração lineares (calibração robusta com PLS e padronização de espectros com PDS) e a calibração através do LS-SVM.

Foram propostos modelos para determinação do teor de etanol em cachaça e de proteína e gordura em ração para cães através de espectroscopia no infravermelho, sendo as diferenças físicas entre as amostras de calibração e previsão a temperatura (no caso da cachaça) ou a granulometria (no caso da ração).

A apresentação deste trabalho foi dividida da seguinte forma: dois capítulos introdutórios, um tratando sobre espectroscopia no infravermelho próximo e consequências das mudanças físicas das amostras sobre a obtenção dos espectros, e outro sobre quimiometria, calibração multivariada e transferência de calibração.

Nos dois capítulos seguintes serão discutidas as três aplicações, com um capítulo focando as amostras de cachaça e outro focando as amostras de ração para cães, sendo que cada um traz uma breve introdução sobre a amostra em questão, seguido do procedimento experimental e da apresentação e discussão dos resultados. A finalização se dá nas conclusões e no índice de referências bibliográficas.

Capítulo 1

Capítulo 1: Espectroscopia no Infravermelho

1.1 Espectroscopia no Infravermelho

No início do século IXX o astrônomo alemão, então vivendo na Inglaterra, Frederick William Herschel, conduziu um experimento com um termômetro e um prisma de vidro, que ele utilizou para decompor a luz do sol. Ele posicionou o termômetro em cada uma das cores e nada aconteceu, até o termômetro chegar na região após o vermelho, onde nenhuma cor podia ser vista, mas ele começou a indicar um aumento de temperatura, indicando claramente a presença de energia. Essa energia recebeu o nome de infravermelho (IV).

O espectro no IV foi somente uma curiosidade por muito tempo devido a dificuldade de se fazer as medições. Por volta de 1900, W. W. Coblentz notou certos padrões no espectro. Compostos que tinham grupos químicos similares tinham espectros similares. Esta observação abriu caminho para os químicos conseguirem informações sobre a estrutura dos compostos, desde que fosse possível obter seus espectros de IV.¹

Do ponto de vista tanto da aplicação, quanto o da instrumentação, o espectro no infravermelho foi dividido em três regiões, sendo estas: a região do infravermelho próximo, a região do infravermelho médio e a região do infravermelho distante.² Os sobretons ocorrem com comprimento de onda de 780 a 1100 nm e as bandas de combinação ocorrem com comprimento de onda de 1800 à 2500 nm.^{3,4}

Antes da segunda guerra mundial, a região do infravermelho próximo não era considerada importante, pois suas bandas sobrepostas, de difícil resolução eram de pouca utilidade para interpretação de compostos orgânicos e as dificuldades de se utilizar essa região para análises quantitativas eram consideradas enormes, pois o espectro nesta região é bastante influenciado por mudanças nas ligações de hidrogênio, provocadas por efeitos na amostra, como diferença de temperatura e umidade. A utilização da espectroscopia no infravermelho próximo (NIR) para medidas qualitativas começou nos anos 60.^{5,6}

Foi somente a partir da década de 80, com a possibilidade de interfacear os equipamentos de IV a computadores de alta velocidade, e a utilização da análise multivariada de dados químicos que a metodologia NIR se desenvolveu e se firmou como uma alternativa para análises qualitativas ou quantitativas de amostras reais.

A maior parte das aplicações têm sido realizada no infravermelho médio e próximo, amplamente utilizadas para análises qualitativas e quantitativas. Um dos usos mais importantes tem sido na identificação de compostos orgânicos.

Esta utilização vem crescendo devido ao fato de que praticamente todos os compostos orgânicos absorvem no infravermelho próximo, e usualmente seguem a Lei de Beer-Lambert, além de possuir como vantagens em relação a outras técnicas o menor tempo de processamento da análise, baixo custo (quando muitas amostras são analisadas), não necessidade de tratamento químico da amostra (reduzindo gastos com reagentes e tratamento de resíduos), pequenas quantidades de amostra utilizadas, estado físico de agregação indiferente, entre outros.⁷

As medidas feitas no NIR em sólidos podem ser por reflectância especular ou reflectância difusa. O espectro de reflectância não é idêntico ao de absorção, mas carrega informações similares.

A reflectância difusa não requer grande preparo da amostra, apenas moagem para transforma-la em pó, mas deve ter tamanho de partícula uniforme. Sua obtenção é dada pela penetração do feixe de radiação infravermelha na superfície da amostra, ocorrendo então a interação com as partículas sólidas que permite transferir as informações sobre a amostra para a radiação que será refletida de forma difusa.⁸

Davis⁹ sugeriu uma divisão do infravermelho próximo em duas regiões. Essa divisão foi baseada no tipo de detector utilizado para as medidas. A primeira região, chamada de infravermelho próximo de ondas curtas (SW-NIR), é situada entre 780 a 1100 nm, e detectores de silício podem ser utilizados; na região seguinte que vai de 1100 a 2500 nm, chamada apenas de infravermelho próximo, detectores fotocondutores são utilizados (comumente PbS).

Tanto no SW-NIR quanto no NIR são observados essencialmente sinais provenientes de sobretons e combinações vibracionais. As ligações C-H, N-H e O-H

apresentam alta anarmonicidade das vibrações, motivo pelo qual sinais oriundos de sobretons e combinações são detectáveis.

A espectroscopia no infravermelho próximo de ondas curtas tem despertado grande interesse em química analítica, uma vez que a instrumentação é muito simples, pois acessórios comuns como fonte, óptica e detectores para a região do visível podem ser utilizados, acarretando em menor custo com relação aos tradicionais equipamentos para a região do infravermelho médio e próximo.⁹

1.2 Teoria

Para absorver radiação infravermelha, uma molécula precisa possuir um momento de dipolo permanente. Apenas nesta circunstância o campo elétrico alternado da radiação pode interagir com a molécula e causar variações na amplitude de seus movimentos. O momento dipolar é determinado pela magnitude da diferença de carga e a distância de dois centros de carga.²

Se a frequência da radiação coincidir exatamente com a frequência vibracional natural da molécula, ocorre uma transferência de energia efetiva e resulta em uma variação da amplitude da vibração molecular, a absorção é a consequência.²

A rotação de moléculas assimétricas em torno de seus centros de massa resulta em uma variação periódica do dipolo que também pode interagir com a radiação, já moléculas homonucleares não possuem nenhuma variação do momento de dipolo ocorrendo durante suas vibrações ou rotações, de forma que apenas poucos compostos desse tipo podem absorver radiação infravermelha.

Os níveis de energia vibracionais são quantizados e, para a maioria das moléculas, as diferenças de energia entre estados quânticos correspondem a região do infravermelho médio.

As vibrações caem nas categorias de estiramentos e de deformações angulares. Uma vibração de estiramento envolve uma variação contínua na distância interatômica ao longo de um eixo da ligação entre dois átomos. As vibrações de deformação angular são características pela variação do ângulo entre duas ligações.²

Muitas das frequências fundamentais podem ser calculadas assumindo-se que as bandas de energia são provenientes da vibração de um oscilador harmônico ideal, obedecendo a lei de Hooke:¹

$$\nu = 1/2\pi(k/u)^{-1} \quad \text{Eq. 1}$$

onde ν é a frequência vibracional

k é a constante de força

u é a massa reduzida dos dois átomos

Mas diferentemente de um modelo clássico aqui não há um número infinito de níveis de energia, mas sim níveis de energia discretos descritos pela teoria quântica.¹ Podemos, no entanto, usar o conceito do oscilador harmônico simples para desenvolver as equações de onda da mecânica quântica. As soluções dessas equações para as energias potenciais têm a forma:

$$E = \left(\nu + \frac{1}{2} \right) \frac{h}{2\pi} \sqrt{\frac{k}{\mu}} \quad \text{Eq. 2}$$

onde h é a constante de Planck e ν é o número quântico vibracional, que pode tomar apenas valores positivos e inteiros (incluindo zero).²

A teoria quântica indica que são permitidas apenas as transições vibracionais que tem $\Delta\nu = \pm 1$.⁷

Embora o tratamento mecânico-quântico do oscilador harmônico explique as bandas de absorção observadas no IV devido aos modos fundamentais da vibração molecular, ela não explica a presença de bandas de sobretom no NIR. Essas bandas surgem de transições com $\Delta\nu = \pm 2$, $\Delta\nu = \pm 3$ etc. e por isso são proibidas pela regra acima. A resposta para isso é que moléculas reais não obedecem exatamente as leis do movimento harmônico simples e ligações reais, embora elásticas, não obedecem a lei de Hooke exatamente.

Quando dois átomos se aproximam a repulsão coulombiana entre seus núcleos aumenta mais rapidamente do que a previsão do modelo harmônico, e quando a

distancia interatômica se aproxima daquela onde ocorre a dissociação dos átomos, há um decréscimo da energia potencial.

Assim, as curvas de energia mostram forma anarmônica, como pode ser visto na Figura 1.²

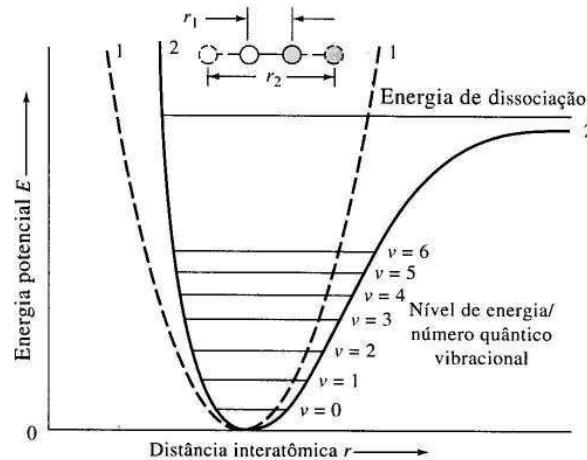


Figura 1: Diagrama de energia potencial de um oscilador anarmônico.

1.3 Efeito da temperatura em espectros vibracionais

Espectros vibracionais de líquidos e sólidos não mostram apenas características moleculares, como estrutura e grupos funcionais, mas também características inter e intramoleculares como ligação de hidrogênio.

Variações na intensidade de ligação manifestam-se nas constantes de forças das ligações. Estas variações provocam alterações dos espectros vibracionais, como deslocamento em relação à frequência e alargamento de banda.^{10,11} Diversos artigos descrevem mudanças nos espectros vibracionais de amostras aquosas provocadas pelo aumento na temperatura, devido à influência desta variável sobre as interações moleculares, sendo que as ligações O-H são as mais afetadas.¹⁰ A razão para este comportamento é a presença de dois osciladores relacionados ao estiramento de grupos O-H “terminais livres”, com máximo de absorvância em 960 nm, e outro é provocado pelo estiramento O-H em pontes de hidrogênio, com máximo de absorvância em 1000 nm.

Luck¹² e Walrafen¹³ encontraram uma maneira conveniente de calcular a entalpia de ligação de hidrogênio em amostras aquosas. Eles consideram um equilíbrio entre O-H livres e O-H ligados. Portanto, a influência da ligação de hidrogênio resulta em deslocamento do equilíbrio para a direção da estrutura de O-H ligados e, experimentalmente, observou-se que este fenômeno provoca o deslocamento da banda para maiores comprimentos de onda. Assim, um decréscimo na quantidade de ligações de hidrogênio, seja devido à diluição ou pelo aumento da temperatura, provoca deslocamento de banda para menores comprimentos de onda. A magnitude do deslocamento pode ser de poucos nm até 50 nm.

Uma vez que bandas de combinação são a soma de duas ou mais bandas fundamentais e sobretons são múltiplos destas mesmas bandas, o efeito da temperatura é mais acentuado na região do infravermelho próximo (incluindo SW-NIR) em relação ao infravermelho médio.

Com o aumento da temperatura, ocorrem mudanças no perfil dos espectros vibracionais^{10,14}, conforme pode ser comprovado na Figura 3, onde estão os espectros de uma amostra de cachaça com teor de etanol de 32,09 %v/v a 15, 20, 25, 30 e 35°C.

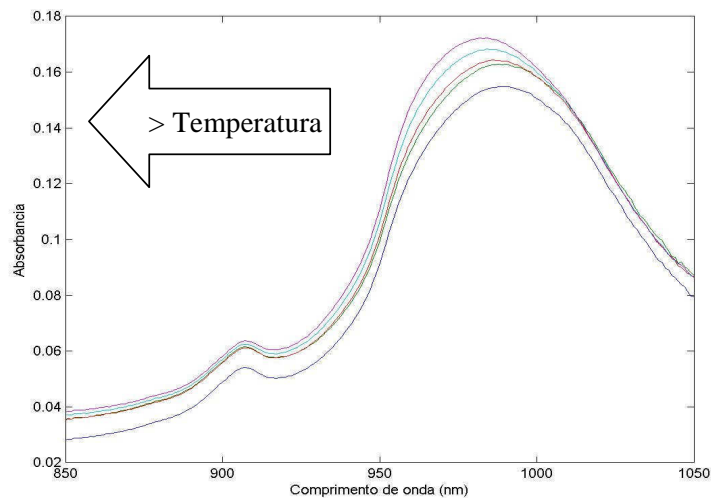


Figura 2: Espectros na região do infravermelho de ondas curtas para uma mesma amostra de cachaça a várias temperaturas (15, 20, 25, 30 e 35°C).

1.4 Efeito de granulometria no espectro

A teoria relacionada à reflectância difusa na região do infravermelho, relatada por Kubelka e Munk ¹⁵, é baseada no fato da radiação que incide sobre uma amostra em pó ser dispersa em todas as direções (Figura 3), podendo ser coletada com uma instrumentação óptica adequada e direcionada para um detector.

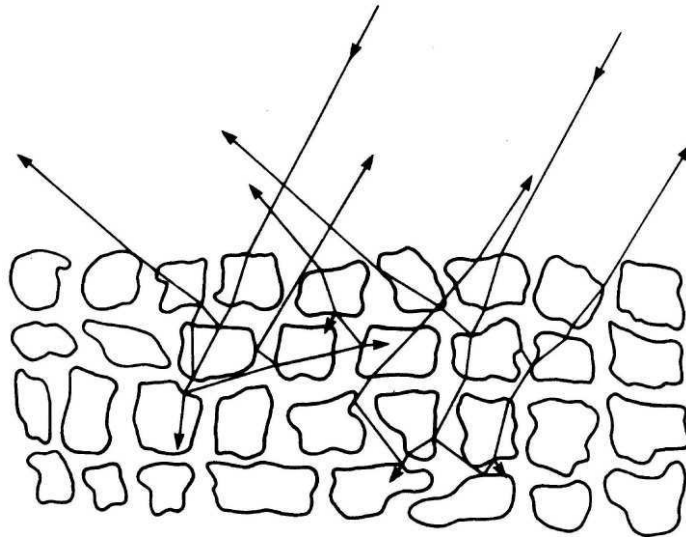


Figura 3: Espalhamento de radiação incidente em amostra sólida

A teoria de Kubelka-Munk (K-M) define o espectro de reflectância como a razão da concentração da amostra com a intensidade do espalhamento de radiação pela amostra. A equação K-M está descrita a seguir:

$$f(R_{\infty}) = (1 - R_{\infty})^2 / (2R_{\infty}) = K'/s \quad \text{Eq. 3}$$

em que R_{∞} é a reflectância de uma camada de espessura infinitesimal do sólido, s é o coeficiente de espalhamento e K' o coeficiente de absorção molar.

Fuller e Griffiths¹⁶ discutiram o uso da teoria K-M para a análise quantitativa de algumas amostras. A teoria diz que há uma relação linear entre o coeficiente de absorção molar, o valor máximo e $f(R_{\infty})$ para cada pico se o coeficiente de

espalhamento s for constante. Este coeficiente de espalhamento é dependente do tamanho da partícula, diminuindo quando o tamanho da partícula aumenta¹⁵, e precisa ser constante para se obter resultados quantitativos. Assim a relação K- M pode ser escrita como:

$$F(R_{\infty}) = (1 - R_{\infty})^2 / (2R_{\infty}) = 2,303kc/s \quad \text{Eq. 4}$$

onde c é a concentração da amostra e k é o coeficiente de absorção convencional da lei de Beer. Esta relação se torna não-linear em altas concentrações¹⁶.

A Figura 4 apresenta um espectro de reflectância difusa na região do infravermelho próximo de uma amostra de ração a diferentes granulometrias (partículas maiores do que 0,50 mm, partículas entre 0,50 e 0,35 mm e partículas menores que 0,35 mm).

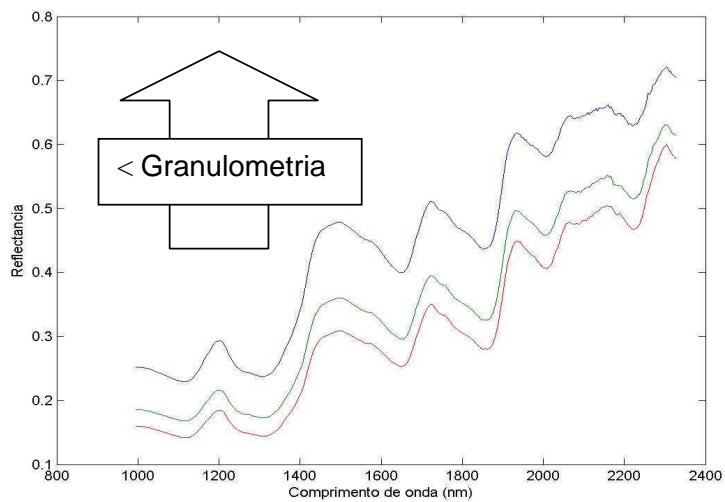


Figura 4: Efeito da granulometria sobre o espectro de reflectância difusa na região do infravermelho próximo de amostra de ração para cães.

Capítulo 2

Capítulo 2 – Quimiometria

2.1 Quimiometria

Nos anos 70 surgiu dentro da química analítica uma área de pesquisa que buscava extrair de uma grande quantidade de dados químicos complexos resultados analíticos interpretáveis. O termo Quimiometria é utilizado hoje para a análise de dados, com finalidade específica dentro de um estudo químico como a otimização de um processo, classificação de dados, modelagem e monitoramento de processos multivariados, construção de modelos de regressão e desenvolvimento de inteligência artificial, entre outros.¹⁷

Dentro da Quimiometria, uma área muito aplicada atualmente é a calibração multivariada. A calibração multivariada, em geral, é uma operação que relaciona uma grandeza de saída com uma grandeza de entrada para um sistema em determinadas condições. A calibração multivariada consiste na execução de três passos principais: calibração, validação e previsão.

Para se fazer a calibração multivariada é obtida uma matriz **X**, constituída de inúmeras medidas instrumentais de mesma natureza (como por exemplo, muitos espectros), obtidas para inúmeros padrões de uma ou mais espécies de interesse. Um modelo matemático que melhor correlacione a matriz de resposta **Y** (concentração, por exemplo) a partir dos dados obtidos para a matriz **X** é alcançado.

A validação é o passo seguinte, onde amostras com concentrações conhecidas tem a mesma propriedade prevista para avaliar se o modelo desenvolvido está adequado. Existem dois métodos para a sua condução: validação interna (quando as próprias amostras de calibração são usadas para a validação) e externa (quando é usado um conjunto distinto, mas com valores **Y** ainda conhecidos).

Na etapa de previsão, a resposta de interesse para uma amostra desconhecida é obtida utilizando o modelo matemático construído na etapa de calibração e validado posteriormente.

O método de calibração multivariada mais utilizado e atualmente considerado como padrão na área, é o método dos mínimos quadrados parciais ou PLS¹⁸ (do inglês “Partial Least Squares”).

Este método de calibração foi desenvolvido por Herman Wold¹⁹ na década de 70. As informações espectrais e as informações das concentrações são usadas ao mesmo tempo na fase de calibração. A matriz de espectros é decomposta em matrizes de variações de espectros (loadings ou pesos) e a posição das amostras (escores) em relação ao novo espaço criado. Os espectros podem ser considerados como combinações lineares das variações dos mesmos (pesos) onde os escores representam essas combinações.

Para uma melhor explicação do PLS será inicialmente introduzida a análise de componentes principais (PCA), pois o PCA é base para o PLS.

2.2 Análise de Componentes Principais (PCA)

A análise de componentes principais (PCA) é um tipo de análise exploratória de dados que visa extrair o máximo de informações de uma tabela de dados convertendo-a em gráficos que mostram a relação entre amostras (linhas de uma matriz) e as variáveis (colunas de uma matriz).

O PCA faz uma aproximação da tabela de dados, ou seja, uma matriz X , em termos da soma de várias matrizes M_i de posto 1, na qual posto significa um número que expressa a verdadeira dimensionalidade da matriz, como mostra a Figura 5.

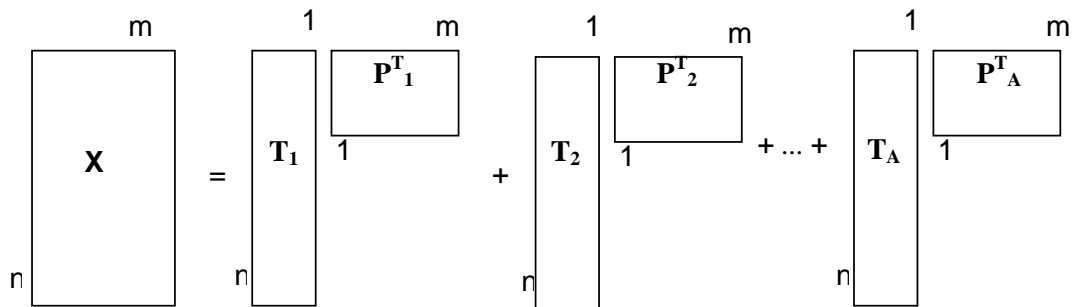


Figura 5: Representação esquemática da decomposição por PCA.

Essa matriz pode ser escrita como produto de dois vetores, escore t_h e peso p_h .

$\mathbf{X} = t_1\mathbf{p}_1 + t_2\mathbf{p}_2 + \dots t_a\mathbf{p}_a$, para “a” componentes principais.

Outra maneira de escrever a equação anterior é:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T \qquad \text{Eq. 5}$$

Onde \mathbf{X} é o conjunto original dos dados com n linhas (amostras) e p colunas (variáveis); \mathbf{T} é a matriz dos escores com n linhas e d colunas (número de componentes principais, novas coordenadas no novo sistema de eixos) \mathbf{P} é a matriz dos pesos (informação do peso de cada variável original na formação dos novos eixos) com d linhas e p colunas.

O novo conjunto de variáveis (componentes principais) é a combinação linear das variáveis originais. Os novos eixos são ortogonais entre si e são constituídos em ordem decrescente da quantidade de variância que descrevem. Assim, podemos dizer que o PCA tenta agrupar aquelas variáveis que estão altamente correlacionadas numa nova variável chamada componente principal.

Como usualmente a grande fração da variância é descrita nos primeiros componentes principais, é possível visualizarmos dados pelo gráfico dos escores de um componente contra o outro. Os algoritmos usados para os cálculos com o PCA em química analítica são o NIPALS²⁰ (do inglês, Nonlinear Iterative Partial Least Squares) e o SVD²⁰ (do inglês, Singular Value Decomposition).

Os principais objetivos desta técnica são o de encontrar relações entre objetos e classificá-los de acordo com suas similaridades, o que torna possível a detecção de amostras anômalas, ou seja, que não pertencem a nenhuma das categorias conhecidas.

Outro objetivo importante é a redução dos dados, que se torna muito útil quando grandes quantidades de informação necessitam ser manipuladas.^{19,20,21,22}

2.3 Regressão por Mínimos Quadrados Parciais (PLS)

Esse método de calibração multivariada foi desenvolvido por Herman Wold^{19,21,22} na década de 70 baseado em uma relação linear entre as variáveis instrumentais (**X**) e as variáveis de interesse (**Y**). As informações da matriz **X** e da matriz **Y** são usadas ao mesmo tempo na fase de calibração. A matriz dos espectros é decomposta em matrizes de variações dos espectros (loadings ou pesos) e a posição das amostras (escores). Os espectros originais podem ser considerados como combinações lineares dos espectros (pesos) onde os escores representam suas contribuições¹⁹.

As matrizes **X** e **Y** podem ser representadas pela Análise de Componentes Principais:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad \text{Eq. 6}$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F} \quad \text{Eq. 7}$$

Onde **T** e **U** são as matrizes escores de **X** e **Y**, respectivamente; **P** e **Q** são as matrizes peso de **X** e **Y**; **E** e **F** são os resíduos.

Uma relação linear entre os dois blocos pode ser realizada correlacionando os escores para cada componente de cada vez, utilizando o modelo linear.

$$\mathbf{U}_h = \mathbf{b}_h \mathbf{T}_h \quad \text{Eq.8}$$

onde “h” é o numero de componentes principais.

Para que a covariância de **T** e **U** seja maximizada deve-se buscar um modelo onde as matrizes dos resíduos **E** e **F** sejam as menores possíveis e , ao mesmo tempo, conseguir uma relação linear ótima entre **t** e **u**.

No PLS isto é obtido por uma leve mudança nos valores dos escores, de forma a produzir a melhor relação possível. Nesta etapa as componentes principais deixam de possuir este nome sendo chamadas então de variáveis latentes, pois elas não mais descrevem a máxima variância dos dados.

O algoritmo utilizado para a estimativa dos vetores coeficientes de regressão \mathbf{b} foi o NIPALS ou SIMPLS (do inglês, straight forward implementation of statistically inspired modification of PLS). Chamando A_{\max} do número Máximo de fatores a serem computados pelo algoritmo, executa-se os seguintes passos para cada um dos fatores $a=1,2,\dots,A_{\max}$.

1-encontrar vetor peso \mathbf{w}_a para maximização da covariância entre a combinação entre a combinação linear $\mathbf{x}_{a-1}\mathbf{w}_a$ e \mathbf{y} , com condição que $\mathbf{w}_a^T\mathbf{w}_a=1$. Isto corresponde encontrar o vetor unitário \mathbf{w}_a que maximiza $\mathbf{w}_a^T\mathbf{x}_{a-1}^T\mathbf{y}_{a-1}$, ou seja, a variância escalada entre \mathbf{x}_{a-1} e \mathbf{y}_{a-1} .

2-encontrar os escore, \mathbf{t}_a como projeção de \mathbf{X}_{a-1} em \mathbf{w}_a , isto é,

$$\mathbf{t}_a=\mathbf{X}_{a-1}\mathbf{W}_a;$$

3-realizar a regressão de \mathbf{X}_{a-1} em \mathbf{t}_a para encontrar os vetores pesos

$$\mathbf{p}_a^T=\mathbf{x}_{a-1}^T\mathbf{t}_a/\mathbf{t}_a^T\mathbf{t}_a$$

4-realizar regressão de \mathbf{y}_{a-1} em \mathbf{t}_a para encontrar vetores pesos

$$\mathbf{q}_a=\mathbf{y}_{a-1}\mathbf{t}_a/\mathbf{t}_a^T\mathbf{t}_a;$$

5-subtrair $\mathbf{t}_a\mathbf{p}_a^T$ de \mathbf{X}_{a-1} e chamar esta nova matriz de \mathbf{X}_a e subtrair $\mathbf{t}_a\mathbf{q}_a$ de \mathbf{y}_{a-1} e chamar esta nova matriz de \mathbf{y}_a ;

6-otimizar o número de fatores por validação cruzada, e considerar as seguintes matrizes

$$\mathbf{W}=\{\mathbf{w}_a\};$$

$$\mathbf{P}=\{\mathbf{p}_a\};$$

Para a número de fatores otimizados;

7-calcula-se, então, os coeficientes de regressão \mathbf{b} através da seguinte relação:

$$\mathbf{b}=\mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}\mathbf{q}$$

em que \mathbf{W} é a matriz de pesos do PLS.

Em seguida é feita a validação do modelo com novas amostras. A validação cruzada é baseada na avaliação da grandeza dos erros de previsão comparando com os valores das variáveis dependentes das amostras do conjunto de calibração com as respectivas previsões, quando as mesmas não participam na construção do modelo de regressão. Na validação cruzada “um por vez” o PLS separa uma das amostras de

calibração e a usa como validação. Isto é feito até que todas as amostras de calibração sejam usadas como amostra de validação. Em paralelo é realizada a validação com um conjunto externo que deve apresentar amostras com valores que compreendam o intervalo de dados do conjunto de calibração e que devem apresentar performance muito semelhante em todos os parâmetros de avaliação dos modelos de calibração.

2.4 Transferência de calibração

Para a construção de modelos de calibração multivariada para espectroscopia no infravermelho é necessário trabalhar com uma grande quantidade de amostras, obtendo seus espectros e realizando o método padrão necessário para determinação da propriedade de interesse, o que pode demandar muito tempo, custo e esforço para seu completo desenvolvimento. Entretanto, esses modelos complexos de calibração podem ser inúteis se houver variações entre novas medidas de previsão e as amostras utilizadas para a calibração.

Duas variações podem ser distinguidas, sendo a primeira observada quando o modelo de calibração é desenvolvido utilizando um equipamento e a previsão de uma nova amostra é feita em outro equipamento. A segunda fonte de variação vem de diferenças físicas entre amostras de calibração e de previsão.

Essas diferenças físicas podem ser, dentre outras, variações de temperatura²³ ou tamanho de partícula²⁴ entre as fases de calibração e previsão. Independente do caso é necessário desenvolver um novo modelo de calibração, ou seja, coletar inúmeras amostras, analisá-las através do método de referência, obter novos espectros e desenvolver novamente o modelo de calibração.

Existem alguns meios de tornar a primeira calibração útil para a previsão de amostras em outras condições físicas, de forma a não se necessitar esta nova calibração. Dentre esses meios encontram-se a calibração robusta e padronização de espectros²⁵.

A estratégia da calibração robusta é incluir no modelo de calibração todas as possíveis variações que serão encontradas durante o uso do modelo, como por exemplo, toda a faixa de temperatura em que poderão se encontrar as amostras de

previsão. Isto faz com que o algoritmo usado para a calibração seja forçado a encontrar um modelo robusto levando-se em consideração as variações impostas ao espectro. A desvantagem deste método está na enorme quantidade de amostras necessárias para se construir o modelo e no conhecimento e previsão das possíveis fontes de erro.

Na padronização de espectros^{26,27}, o espectro obtido em determinada situação é reconstruído como se tivesse sido obtido na mesma situação a qual foram obtidos os espectros na etapa de calibração.

Esse procedimento envolve três passos principais: (1) estimativa da diferença entre os espectros da fase de calibração e da fase de previsão, (2) cálculo dos parâmetros de padronização capazes de estimar as diferenças, (3) validação dos parâmetros de padronização.

Para a transferência de calibração, um modelo matemático $g()$ reproduz o espectro x de uma amostra em uma condição de referência (MESTRE), a partir de um espectro x' , da mesma amostra medida em uma condição diferente (ESCRAVO), utilizando amostras de transferência que devem ser padrões das espécies a se determinar:

$$x=g(x') \qquad \text{Eq.9}$$

Para isso as amostras de transferência devem ser representativas, ou seja, reproduzir, com boa fidelidade as condições da amostra de calibração. A escolha das amostras para a transferência de calibração deve obedecer a um dos critérios abaixo:

- seleção de um grupo de amostras que foram utilizadas para calibração na situação MESTRE e obter a propriedade de interesse na fase de previsão da situação ESCRAVO. A principal vantagem deste método é que as amostras de calibração e transferência são as mesmas. A representatividade da transferência está na escolha apropriada das amostras.

- seleção de um grupo de amostras entre as novas amostras medidas na fase de previsão (ESCRAVO) e, então, tomar novas medidas nas condições da

situação MESTRE. A maior vantagem deste método é que permite a escolha de amostras de transferência com boa representatividade, sempre quando os padrões de calibração não são armazenados por causa de sua instabilidade física ou química.

- seleção de amostras independentes em relação a ambas condições, servindo para estimar as diferenças entre as duas respostas instrumentais. A principal vantagem deste procedimento é que amostras mais estáveis, como padrões genéricos, podem ser usadas para a transferência.

2.4.1 Padronização direta por partes (PDS)

Wang e colaboradores²⁸ desenvolveram um método de padronização multivariada no qual um conjunto de dados obtidos a partir de um instrumento escravo (\mathbf{X}_e) é transformado de modo que reproduza um conjunto de dados obtidos a partir do instrumento mestre (\mathbf{X}_m), utilizando-se uma matriz de transformação \mathbf{F} , que descreve a relação existente entre essas duas situações:

$$\mathbf{X}_m = \mathbf{X}_e \mathbf{F} + \mathbf{E} \quad \text{Eq. 10}$$

Sendo \mathbf{E} o resíduo não modelado da equação.

Para estimar a matriz \mathbf{F} , são necessários padrões que sejam medidos nos dois equipamentos, sendo válida a seguinte relação:

$$\mathbf{F} = \mathbf{X}_{ep}^+ \mathbf{X}_{mp} \quad \text{Eq. 11}$$

sendo \mathbf{X}_{ep} a matriz dos espectros medidos no instrumento escravo, \mathbf{X}_{ep}^+ sua matriz pseudoinversa e \mathbf{X}_{mp} a matriz dos espectros medidos no instrumento mestre. A matriz \mathbf{F} é obtida por modelagem com PCR ou PLS.

Obtendo-se a matriz \mathbf{F} é possível transformar um novo espectro obtido no instrumento escravo em uma matriz (\mathbf{X}_t):

$$\mathbf{X}_t = \mathbf{X}_e \mathbf{F}$$

Eq. 12

Assim a previsão pode ser feita utilizando o modelo de calibração construído no instrumento mestre.

A padronização direta (DS) é uma correção multiplicativa que não corrige, pelo menos com facilidade, diferenças aditivas de linha de base. Além disso, cada variável transformada é calculada utilizando o espectro inteiro, ocasionando grandes riscos de auto-ajuste. Por isso, Wang e seus colaboradores²⁵ introduziram a padronização direta por partes (PDS). A principal diferença entre o DS e o PDS está na maneira de se estimar a matriz de transformação \mathbf{F} .

O modelo PDS é baseado na suposição que uma variável obtida no espectro da situação mestre está relacionada com uma pequena região ao redor desta mesma variável obtida na situação escrava.

Para a construção da matriz de transformação, o método PDS utiliza uma janela móvel (W) para relacionar cada variável do instrumento mestre com as variáveis contidas na janela:

$$\mathbf{W}_i = [\mathbf{X}_{e,i-g}, \mathbf{X}_{e,i-g+1}, \dots, \mathbf{X}_{e,i}, \dots, \mathbf{X}_{e,i+h-1}, \mathbf{X}_{e,i+h}]$$

Eq. 13

Sendo que a janela contém as variáveis entre $i-g$ até $i+h$, não necessariamente simétrico em torno de i .

Os vetores de regressão para cada variável x_i , que relacionam as duas situações, são obtidos da seguinte maneira:

$$\mathbf{b}_i = \mathbf{W}_i^+ \mathbf{x}_{m,i}$$

Eq. 14

onde \mathbf{W}_i^+ é a matrix pseudo inversa de \mathbf{W}_i .

Os vetores \mathbf{b}_i são colocados na diagonal principal da matriz de transformação \mathbf{F} , sendo todos os outros elementos iguais a zero. A reconstrução do espectro de uma amostra desconhecida é executada da seguinte maneira:

$$X_t = X_e F$$

Eq. 15

A previsão da amostra é feita como se o espectro estivesse sido medido na situação de calibração.

Pequenas regiões das extremidades são perdidas neste método, uma vez que as janelas não podem ser construídas para estas regiões.

Como no DS, as diferenças aditivas são de difícil correção, sendo recomendadas correções na linha de base antes da padronização^{25,29,30}.

Uma representação esquemática do PDS pode ser vista na Figura 6.

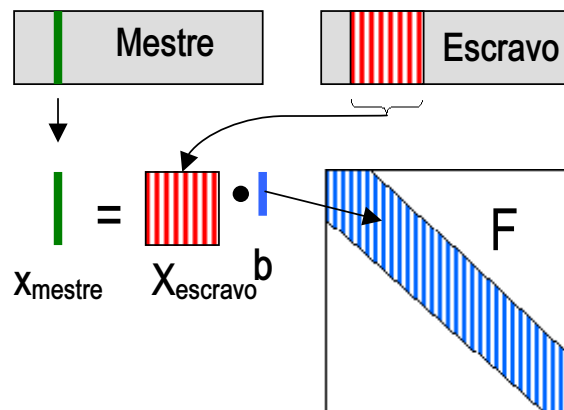


Figura 6: Representação esquemática do PDS

2.5 Máquinas de vetores de suporte por mínimos quadrados (LS-SVM)

A transferência de calibração pode em certas situações ser realizada pela utilização de métodos de calibração multivariada não-lineares que possam tratar com as diferenças espectrais ocasionadas pelas variações das condições físicas presentes nas medidas. Dentre esses métodos destaca-se a máquina de vetor de suporte por mínimos quadrados (LS-SVM).

O algoritmo máquina de vetores de suporte (SVM)³¹ usado para a regressão é uma generalização do algoritmo Generalized Portrait, que começou a ser desenvolvido nos anos 60 por Vapnik e colaboradores; entretanto a forma atual do LS-SVM foi desenvolvida na década de noventa, com maior ênfase nas aplicações práticas.

O SVM é uma poderosa ferramenta, utilizada para resolver problemas de classificação e modelamento, introduzida e aplicada nas áreas de engenharia, mas na área da quimiometria, a qual se referem problemas de quantificação (regressão multivariada) as aplicações encontradas na literatura ainda são escassas.

Modelos matemáticos com capacidade de aproximação universal, como as redes neurais artificiais, ainda não são dotadas de algoritmos de treinamento capazes de maximizar a capacidade de generalização de uma forma sistemática, o que pode levar a um sobreajuste do modelo aos dados.

Como não são conhecidas as não-linearidades presentes e a complexidade intrínseca do problema, os algoritmos de otimização e as ferramentas estatísticas utilizadas para seleção de modelos podem induzir modelos com baixa capacidade de generalização.

Sendo assim, as principais vantagens do SVM em suas aplicações são:

- Elevada capacidade de generalização, evitando o sobreajuste;
- Robustez em grandes dimensões, possibilitando aplicação de SVMs em vetores de características de grandes dimensões;
- Convexidade da função objetivo, a aplicação das SVMs implica na otimização de uma função quadrática, que possui apenas um mínimo;
- Teoria bem estabelecida dentro da Matemática e Estatística.

Entretanto, encontrar o modelo final do SVM pode ser computacionalmente muito difícil porque ele requer a solução de um conjunto de equações não lineares (programação quadrática). Para simplificar esta aproximação, Suykens³² e colaboradores propuseram o uso de máquinas de vetores de suporte por mínimos quadrados (LS-SVM).

As máquinas de vetores de suporte por mínimos quadrados são capazes de tratar de modo relativamente rápido problemas de calibração multivariada lineares ou não lineares.

Em princípio LS-SVM sempre ajusta uma relação linear ($y = wx + b$) entre os regressores (x) e suas variáveis dependentes (y), sendo a melhor relação aquela que minimizar uma função custo (C) contendo uma penalização dos coeficientes de regressão:

$$C = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^N e_i^2 \quad \text{Eq. 16}$$

de modo que:

$$y_i = \mathbf{w}_i^T \mathbf{x}_i + b + e_i \quad \text{Eq. 17}$$

A primeira parte desta função é um decaimento utilizado para regularizar os coeficientes de regressão e penalizar valores altos, para que seus coeficientes sejam os menores possíveis. Já a segunda parte representa os erros de regressão para todos os dados de treinamento. O peso relativo da segunda parte da equação quando comparado com o peso da primeira é fornecido pelo parâmetro γ , que deve ser otimizado pelo usuário. Conforme γ é aumentado, mais ênfase é colocada na redução da magnitude do erro quadrático dos coeficientes do modelo³³. A definição do erro da regressão dado pela diferença entre valores reais e previstos é imposta pela restrição da segunda parte da equação.

O LS-SVM resolve esse problema de otimização restrita utilizando o método dos multiplicadores de Lagrange:

$$L(w, b, e, \alpha) = \frac{1}{2} \|w\|^2 + \gamma \sum_{i=1}^N e_i^2 - \sum_{i=1}^N \alpha_i \{w^T x_i + b + e_i - y_i\} \quad \text{Eq.18}$$

em que

$$y_i = \begin{bmatrix} y \\ y \\ \mathbf{M} \\ y_N \end{bmatrix}, e_{i=} \begin{bmatrix} e_1 \\ e_2 \\ \mathbf{M} \\ e_N \end{bmatrix} \text{ and } \alpha_i = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \mathbf{M} \\ \alpha_N \end{bmatrix}. \quad \text{Eq 19}$$

A função custo definida antes se encontra nos dois primeiros termos desta função de Lagrange (L) (Eq 18), mas essa equação possui multiplicadores de restrição, os chamados multiplicadores de Lagrange (α_i), sendo cada um correspondente á um dado de treinamento. Para obter a solução final as primeiras derivadas parciais são acertadas para zero, e combinações lineares dos dados de treinamento são obtidas. Os

coeficientes de regressão (w) podem ser escritos como uma expansão dos multiplicadores de Lagrange com os respectivos dados de treinamento (x_i):

$$\frac{\partial L(w, b, e, \alpha)}{\partial w} = w - \sum_{i=1}^N \alpha_i \phi(x_i) = 0 \therefore w = \sum_{i=1}^N \alpha_i \phi(x_i) \quad \text{Eq 20}$$

$$\frac{\partial L(w, b, e, \alpha)}{\partial e} = \sum_{i=1}^N \gamma e - \alpha = 0 \therefore \alpha = \gamma e \quad \text{Eq 21}$$

Então:

$$w = \sum_{i=1}^N \alpha_i \phi(x_i) \quad \text{Eq 22}$$

Um importante resultado desta aproximação é que os pesos (w) podem ser escritos como combinações lineares dos multiplicadores de Lagrange com os dados de treinamento correspondentes (x_i). Então, colocando essa expressão na reta de regressão original ($y=wx+b$), o seguinte resultado é obtido:

$$y = \sum_{i=1}^N \alpha_i \phi(x_i)^T \phi(x) + b = \sum_{i=1}^N \alpha_i \langle \phi(x_i)^T, \phi(x) \rangle + b \quad \text{Eq.23}$$

onde o produto interno de x_i e x é indicado por $\langle \phi(x_i)^T, \phi(x) \rangle$.³⁴

Como pode ser visto na equação 22 os multiplicadores de Lagrange são geralmente diferentes de zero, indicando que todos os dados de treinamento contribuem para a solução (esses são os vetores de suporte). Além disso, os dados de calibração que estão localizados longe da reta de regressão influenciam bastante esta reta. Por isso, seus multiplicadores de Lagrange correspondentes são também relativamente altos (proporcionalmente ao erro de previsão).

O modelo final pode ser descrito como uma combinação linear dos produtos internos entre os dados de treinamento e do novo dado (x). Isso é importante por duas razões, sendo elas: a dimensão dos dados não aparece no problema e é mais fácil permitir regressão não-linear como uma extensão da aproximação linear.³²

Por fim o produto interno $\langle \phi(x_i)^T, \phi(x) \rangle$ é substituído por uma função de Kernel $K(x_i, X)$. A função Kernel representa a relação entre o dado de entrada e a propriedade de saída a ser modelada.³³ Essa função determina tanto o mapeamento não-linear, $x \rightarrow \phi(x)$, quanto o produto interno correspondente $\phi(x_i)^T \phi(x)$. Isso leva à seguinte função de regressão não-linear.

$$y = \sum_{i=1}^N \alpha_i K(x_i, x) + b \quad \text{Eq.24}$$

Cada kernel é associado com um parâmetro específico, para as funções polinomiais e de função radial de base (RBF) os parâmetros são o grau do polinômio (d) e a largura da função Gaussiana (σ), respectivamente. Assim, ao invés de calcular um mapeamento específico para cada dimensão dos dados, uma função Kernel apropriada é selecionada e seu parâmetro específico é otimizado³².

A função Kernel transforma o espaço de entrada em um espaço de características de alta dimensão onde a solução do problema pode ser representada como sendo um problema linear, como mostra a Figura 7.³³

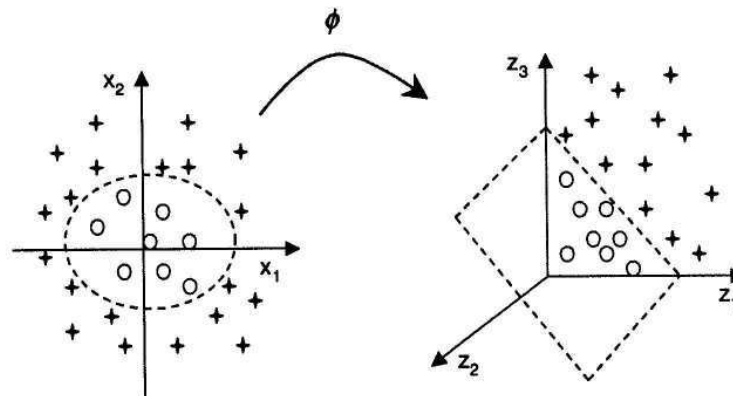


Figura 7: Mudança do espaço dos dados pela função Kernel (ϕ).

A função Kernel mais comumente utilizada é a função radial de base (RBF):

$$\exp\left(-\frac{\|x_i - x\|^2}{2\sigma^2}\right) \quad \text{Eq. 25}$$

Enquanto o produto interno Kernel é uma medida de similaridade entre dois vetores, o Kernel RBF é conceitualmente uma medida não-linear de similaridade. O ajuste do parâmetro de variância (σ^2) muda a largura da gaussiana e o grau da não-linearidade que pode ser modelada. Conforme σ^2 é aumentado o Kernel força o modelo para uma solução linear.³⁵

Em todos os modelos desenvolvidos com LS-SVM, foi necessário se ajustar os parâmetros σ^2 e γ em função do erro de previsão. Um exemplo da superfície de otimização utilizada para se determinar esses valores está na Figura 8.

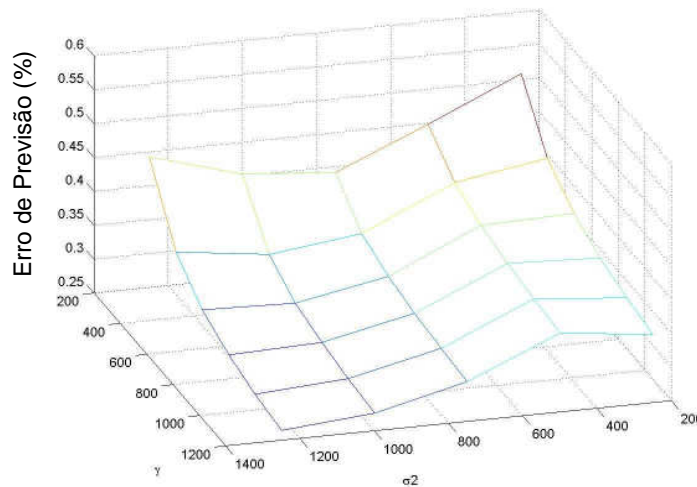


Figura 8: Superfície de otimização dos parâmetros σ^2 e γ em função do erro de previsão.

2.6 Parâmetros de qualidade

Para verificar a performance dos modelos propostos, foi utilizado o RMSE (do inglês Root Mean Square Error), que fornece uma idéia do erro médio absoluto contido em um modelo.

Esse parâmetro pode ser dividido em RMSEC (do inglês Root Mean Square Error of Calibration); RMSECV (Root Mean Square Error of Cross Validation) e RMSEP (Root Mean Square Error of Prediction). O cálculo do RMSE é efetuado da seguinte maneira:

$$\text{RMSE} = \sqrt{\frac{\sum (y_p - y_r)^2}{n}} \quad \text{Eq. 26}$$

Onde y_p é o valor previsto, y_r é o valor real e n o número de amostras.

Para confirmar se os RMSEPs obtidos nos diferentes modelos eram significativamente diferentes foi utilizado o teste F.³⁶

Capítulo 3

Capítulo 3: Determinação do teor de álcool em cachaça

3.1 Cachaça

A produção de aguardente de cana no Brasil teve início no período colonial logo após a introdução da cana-de-açúcar na Capitania de São Vicente e a moagem do primeiro engenho de açúcar. Até a 2ª Guerra Mundial, a produção da aguardente era conduzida por um grande número de pequenos engenhos. Os processos usados eram artesanais e tecnicamente rudimentares. Esses estabelecimentos localizavam-se nas zonas rurais onde as famílias cuidavam da parte agrícola e industrial do processo. A comercialização era feita de forma bastante variada utilizando-se desde os vasilhames de vidro de diversos tamanhos, a barris de madeira de 100 a 200 litros. Os produtos eram armazenados em tonéis ou pipas nos engenhos e a comercialização era feita conforme a demanda. Muitas vezes o produto ficava nesses depósitos por longos períodos, tornando a aguardente involuntariamente envelhecida. Essas, muitas vezes, desenvolviam aromas e sabores mais agradáveis do que as recém destiladas. Não havia, entretanto, estrutura nem características regionais ou padrões de qualidade definidos. A produção por estabelecimento era da ordem de 200 a 1000 litros diários. Com o fim da Guerra, a demanda foi incrementada e começaram a surgir empresas dedicadas as produções industriais de aguardente de cana. Grandes destilarias atuais produzem cerca de 20.000 litros horários.

O aumento da produção industrial foi acompanhado da evolução dos conhecimentos técnicos e científicos desde a cultura da cana-de-açúcar até o engarrafamento e distribuição. No campo agrícola, foram desenvolvidas variedades novas de cana mais produtivas, com teor mais elevado de açúcar e mais adequadas a industrialização. Nas usinas e destilarias, as técnicas de extração do açúcar foram melhoradas para se obter maior rendimento da extração, e na fermentação e destilação o avanço tecnológico permitiu a redução das perdas por infecção e reações adversas.

O Decreto Estadual Nº52.504 (São Paulo) de 28 de julho de 1970 estabeleceu as normas técnicas para aguardente de cana. A lei Nº 5.823 de novembro de 1972 dispõe sobre a padronização, classificação, inspeção e registro de bebidas. As portarias do

Ministério da Agricultura e Reforma Agrária que se seguiram (Nº 371/74 e 879/75) estabeleceram normas para a padronização do produto e dos estabelecimentos produtores e padronizadores de aguardente de cana.

De acordo com a legislação vigente, as bebidas alcoólicas fermento-destiladas são classificadas em: (1) destilados alcoólicos simples com teor alcoólico de 55 a 80 % em volume, compreendendo produtos obtidos a partir de diferentes espécies vegetais e não são destinadas ao consumo direto; (2) bebidas destiladas com 38 a 54 % de etanol em volume, que compreendem aguardente de cana, rum, uísque, conhaque, grapa ou bagaceira, aguardente de frutas, tequila e tiquira; e (3) destilo-retificadas com 38 a 54 % de etanol em volume, que compreendem vodca, genebra, steinhager e aquavit. O destilado alcoólico com teor alcoólico acima de 80% em volume de etanol é denominado álcool etílico potável.

Uma parte do produto fabricado por diversas destilarias no Brasil é comercializada diretamente dos engenhos para os consumidores embalada em garrafas, garrafões e outros recipientes. A quantidade desse produto é bastante variada e não está ao alcance das legislações vigentes.

Análise feita em 185 amostras³⁷ mostrou grande variação na composição das aguardentes de cana, como pode ser visto na Tabela 1.

Tabela 1: Variação de composição (em mg/100mL de álcool anidro) de aguardentes de cana.

Componentes	Mínimo	Máximo
Etanol*	33,43	66,92
Acidez total em ácido acético	1,97	409,0
Acidez fixa em ácido acético	0,60	111,0
Acidez volátil em ácido acético	4,53	125,51
Aldeído acético	0,19	165,99
Acetona	0,05	10,18
Ésteres em acetato de etila	2,05	171,28
Acetal	1,11	24,0
N-propanol	2,60	377,0
L-butanol	4,41	124,0
N-butanol	0,44	7,14
L-amílico	5,43	671,0
N-amílico	0,00	0,95
Álcoois superiores	2,60	825,0
Metanol	0,00	23,27
Componentes secundários	4,78	838,0

* teor de etanol em volume a 20°C

O envelhecimento em tonéis de madeira provoca redução gradual no teor alcoólico, incremento no teor de ésteres e alguns produtos secundários provenientes da madeira e melhora de forma significativa as propriedades sensoriais.³⁷

3.2 Experimental

Foram obtidas 50 amostras de cachaça com porcentagem de teor alcoólico na faixa entre 20,86 e 46,48 % v/v através de diluição com água deionizada ou adição de etanol P.A. (Merck) a cachaça original (Caninha 51, lote 210706, 38,04% v/v). Todas as amostras foram obtidas a partir de uma mesma garrafa de cachaça.

A determinação da porcentagem alcoólica foi realizada com um densímetro Anton Paar DMA 4500 da Central Analítica do Instituto de Química da Unicamp. Os valores obtidos estão na Tabela 2.

Tabela 2: porcentagem alcoólica (V/V) nas amostras de cachaça

<i>AMOSTRA</i>	<i>% Vol:Vol</i>	<i>AMOSTRA</i>	<i>% Vol:Vol</i>	<i>AMOSTRA</i>	<i>% Vol:Vol</i>
1	20,86	18	29,39	35	38,04
2	21,39	19	29,64	36	38,52
3	21,69	20	30,61	37	39,11
4	22,12	21	30,66	38	39,47
5	22,65	22	31,16	39	40,15
6	22,99	23	31,62	40	40,30
7	23,52	24	32,09	41	41,26
8	24,04	25	32,56	42	41,86
9	24,48	26	33,14	43	42,35
10	24,98	27	33,54	44	42,77
11	25,53	28	34,15	45	43,48
12	25,91	29	34,84	46	43,97
13	26,43	30	35,26	47	44,50
14	26,95	31	35,72	48	45,16
15	27,41	32	36,37	49	43,99
16	27,97	33	36,96	50	46,48
17	28,95	34	37,41		

Os espectros na faixa de 850 a 1050 nm foram obtidos em um equipamento Agilent 8453 a cinco temperaturas diferentes, sendo estas 15, 20, 25, 30 e 35°C, controlando-se a temperatura com equipamento Peltier HP 89090A, utilizando-se uma cubeta com tampa rosqueável. Os espectros na região do SW-NIR para amostras de cachaça com diferentes teores de etanol obtidos a 20°C podem ser observados na Figura 9.

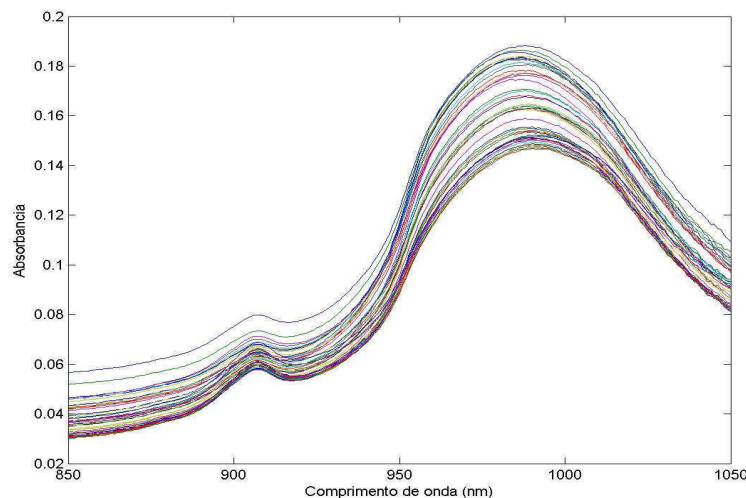


Figura 9: Espectros na região do infravermelho de ondas curtas para amostras de cachaça a 20°C.

As bandas importantes que podem ser observadas nos espectros da figura são especificadas na Tabela 3.

Tabela 3: Atribuição das bandas do espectro na região do infravermelho para cachaça.³⁸

Comprimento de onda (nm)	Designação dos modos vibracionais
905	Terceiro sobretom do estiramento C-H para grupos metil
950 - 980	Segundo sobretom do estiramento de grupos funcionais O-H livres
970 - 1050	Bandas de combinação relacionadas ao estiramento C-H e ao estiramento de grupos O-H

Pode ser observado na Figura 3 que as bandas encontradas na região entre 950 a 1050 são sobrepostas. Ainda assim, é possível chegar a conclusão de que a banda em 905 nm é proveniente os modos vibracionais dos grupos C-H pertencentes a álcoois, já que há uma proporcionalidade direta entre a intensidade da absorvância e a

concentração de etanol da amostra. Em baixas concentrações, as bandas entre 950 e 1050 nm, apresentam altas intensidades e com o máximo se deslocando para maiores comprimentos de onda. A explicação para este fato é que, com a adição de etanol na mistura, o número de ligações O-H livres diminuem, enquanto que a quantidade dos estiramentos OH influenciados por ligações de hidrogênio aumentam, deslocando o máximo para maiores comprimentos de onda.³⁹

Para o tratamento dos dados as amostras foram separadas em dois grupos através do algoritmo de Kennard-Stone,⁴⁰ um para a calibração, com 30 amostras, e outro para validação com 20 amostras. Este algoritmo inicia-se com a seleção da amostra mais próxima (ou mais distante, dependendo da escolha do usuário) do ponto central. Então é selecionada a amostra que possuir maior distância euclidiana em relação à primeira. Esta amostra é então retirada do conjunto e a próxima amostra a ser selecionada é aquela que possuir maior distância euclidiana da anterior, e assim sucessivamente até o número de amostras a ser separado, que foi definido pelo usuário, ser alcançado.

Neste trabalho foram feitas quatro propostas de calibração diferentes. Na proposta 1 os espectros obtidos em quatro temperaturas (situações) diferentes foram utilizados como conjunto de calibração e os espectros obtidos na temperatura restante foram utilizados como dados de validação. Todas as combinações foram feitas, sendo os espectros obtidos em todas as temperaturas utilizados como dados de calibração e validação. Os modelos foram construídos utilizando PLS e LS-SVM.

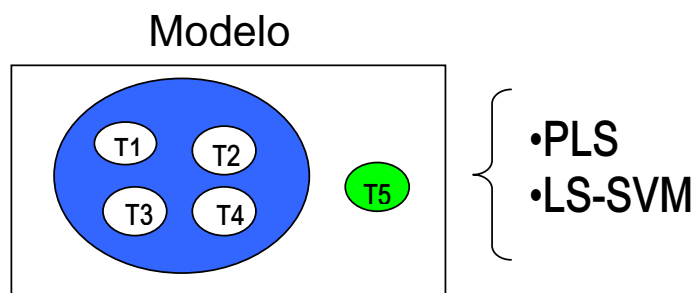


Figura 10: Representação esquemática da proposta 1 para transferência de calibração para determinação da porcentagem de álcool em cachaça a diferentes temperaturas.

Na proposta 2, o conjunto de calibração foi constituído de todos os espectros obtidos a todas temperaturas, enquanto o conjunto de validação era composto de espectros obtidos a apenas uma temperatura. Foram feitas assim cinco validações diferentes, utilizando cada conjunto de espectros obtidos a cada temperatura. Os modelos foram construídos através de PLS e LS-SVM.

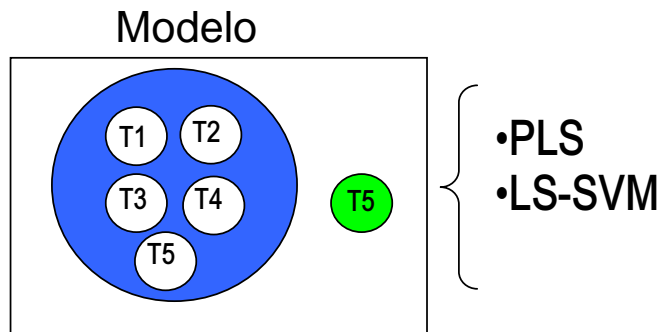


Figura 11: Representação esquemática da proposta 2 para transferência de calibração para determinação da porcentagem de álcool em cachaça a diferentes temperaturas.

Na proposta 3 o algoritmo de Kennard-Stone foi utilizado para separar amostras de calibração e validação dentro de cada conjunto de espectros obtidos em cada temperatura. Os modelos foram construídos utilizando apenas amostras escolhidas como conjunto de calibração cujos espectros foram obtidos a 20°C e a validação foi feita com as amostras de validação cujos espectros foram obtidos nas outras temperaturas. Foram utilizados os conjuntos de validação de todas as temperaturas, um de cada vez. Neste caso os modelos foram construídos através de PLS, LS-SVM e PDS.

Esta foi a única proposta onde foi utilizado o PDS porque foi o único caso onde havia disponibilidade de espectros na situação ESCRAVA obtidos das mesmas amostras utilizadas no conjunto de calibração. Desta forma foi possível a obtenção da matriz de transformação **F**.

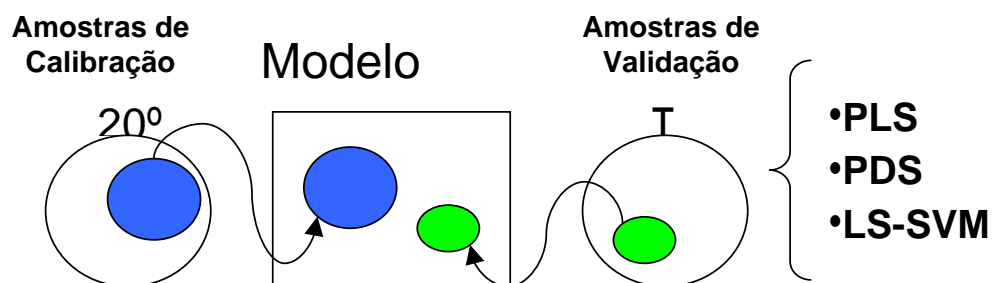


Figura 12: Representação esquemática da proposta 3 para transferência de calibração para determinação da porcentagem de álcool em cachaça a diferentes temperaturas.

A última proposta, a proposta 4, teve seus modelos construídos utilizando todos os espectros obtidos como amostras de calibração para as cinco temperaturas como conjunto de calibração, sendo a validação feita com as amostras de validação de todas as temperaturas. Os modelos foram obtidos com PLS e LS-SVM.

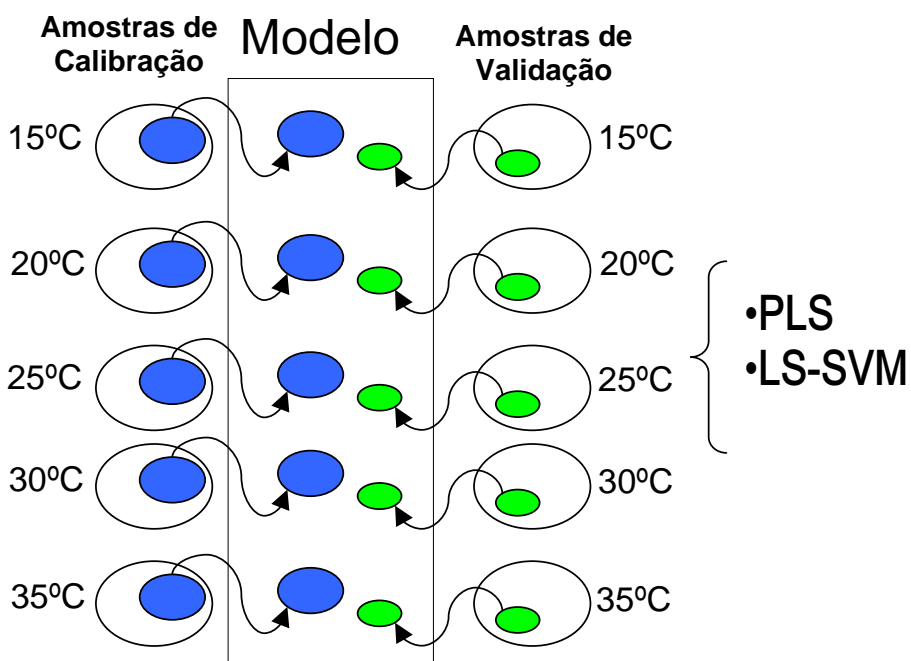


Figura 13: Representação esquemática da proposta 4 para transferência de calibração para determinação da porcentagem de álcool em cachaça a diferentes temperaturas.

Para todos os modelos o pré-processamento adotado foi centrar os dados na média. Os modelos PLS e PDS foram otimizados a partir da validação cruzada “um por vez”. Um exemplo para um gráfico de erro médio quadrático de calibração cruzada por número de variáveis latentes utilizado para auxiliar na escolha do número ideal na construção dos modelos de calibração PLS e PDS está na Figura 14. Neste caso, foram escolhidas quatro variáveis latentes para a construção do modelo pois não existe variação apreciável do erro a partir desse número.

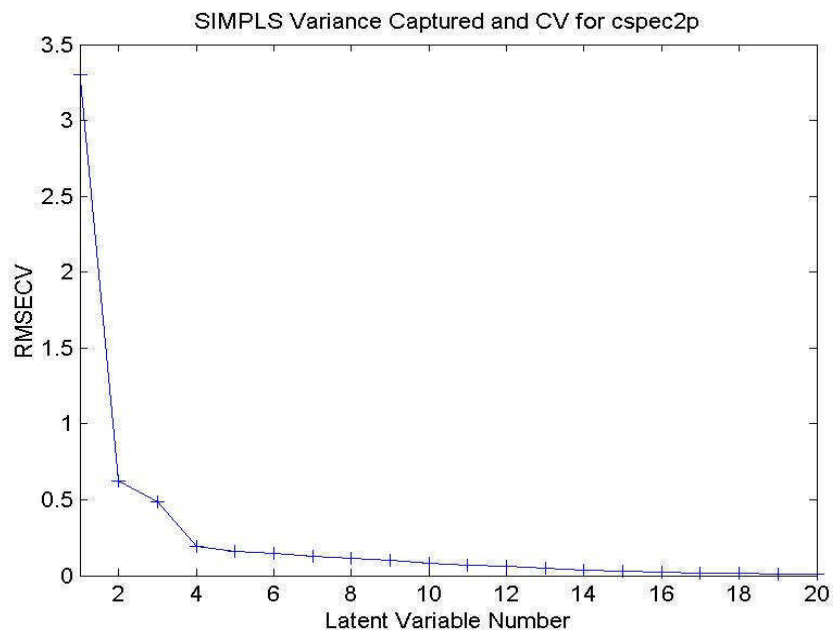


Figura 14: Exemplo de gráfico de erro médio quadrático de calibração cruzada por número de variáveis latentes utilizado na determinação do número ideal de variáveis latentes na construção dos modelos PLS e PDS.

As calibrações foram realizadas através do *PLS-Toolbox* versão 3.5 (Eigenvector Research Inc. ⁴¹) para *Matlab* 6.5 (MathWorks Co.). Os modelos LS-SVM foram obtidos através do *LS-SVM lab toolbox*⁴². Todos os programas foram executados em um microcomputador Intel Pentium 4 CPU 3,00 GHz com 1 GB de RAM.

3.3 Proposta 1

Foram obtidos modelos de calibração para todas as situações da proposta 1 através de PLS e LS-SVM. As diferentes situações eram todas as combinações possíveis da utilização dos espectros obtidos como dado de calibração e validação, sendo utilizados dados obtidos a quatro temperaturas diferentes como conjunto de calibração e os dados obtidos na outra temperatura utilizados como conjunto de validação, de forma que todos os dados fossem utilizados como conjunto de calibração e validação.

A Tabela 4 mostra os RMSEPs obtidos para todos os modelos construídos, as variáveis latentes (utilizadas para a construção dos modelos PLS) e as variáveis γ e σ^2 (utilizadas na construção dos modelos LS-SVM).

Tabela 4: RMSEPs obtidos e parâmetros utilizados na construção dos modelos da proposta 1 para transferência de calibração na determinação do teor de álcool em cachaça.

Temperatura de validação (°C)	PLS		LS-SVM		
	RMSEP (%)	Variáveis latentes	RMSEP (%)	γ	σ^2
15	0,31	4	0,33	1200	1250
20	0,24	4	0,15	600	300
25	0,27	4	0,13	6000	2000
30	0,27	4	0,12	1200	900
35	0,39	4	0,28	7200	1600

Pode ser observado através da Tabela 4, que os modelos construídos por LS-SVM geraram RMSEPs menores em quatro das cinco situações e um RMSEP bastante similar na situação onde os espectros obtidos a 15°C foram utilizados como conjunto de validação. A Tabela 5 traz os resultados obtidos com o teste F, onde se pode confirmar a diferença significativa de previsão entre os modelos de calibração

obtidos com PLS e LS-SVM, nas quatro situações onde o LS-SVM teve um melhor desempenho.

Tabela 5: Teste F para comparação dos modelos obtidos na proposta 1 na determinação do teor de álcool em cachaça.

Temperatura de validação (°C)	F calculado
15	1,13
20	2,56
25	4,31
30	5,06
35	1,94

F crítico = 1,67

O melhor desempenho do LS-SVM em relação ao PLS pode ser mais bem comprovado através da Figura 15 que mostra os erros relativos obtidos para todos os modelos de todas as situações.

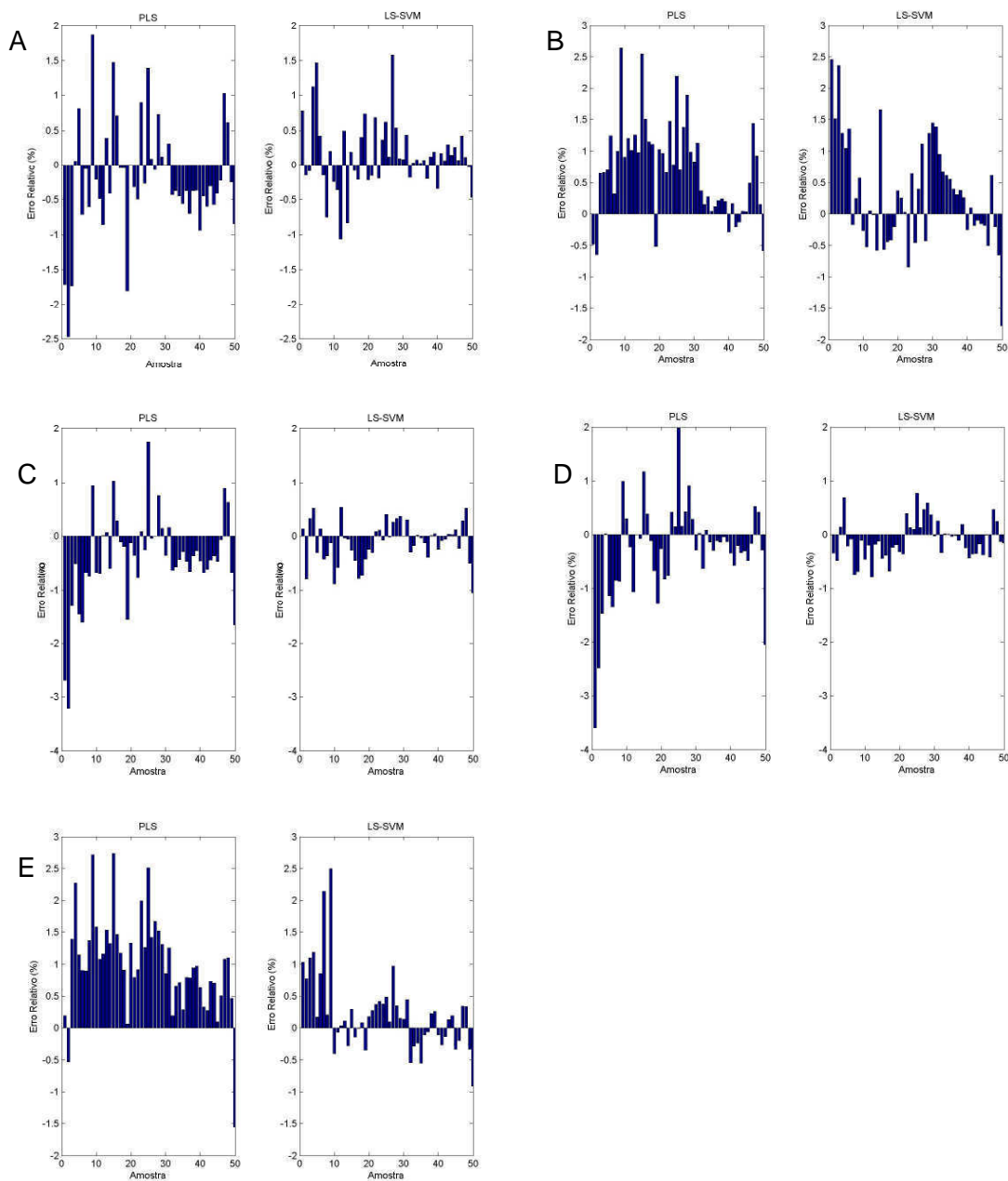


Figura 15: Erros relativos na previsão da porcentagem de álcool em amostras de cachaça obtidos nos modelos construídos para a proposta 1, quando as amostras de validação tiveram seus espectros obtidos nas temperaturas: A. 15°C; B. 20°C; C. 25°C; D. 30°C e E. 35°C.

3.4 Proposta 2

A segunda proposta foi feita construindo-se modelos de PLS e LS-SVM incluindo todos os espectros obtidos como conjunto de calibração e para cada situação foi utilizado o conjunto de validação composto por espectros obtidos a apenas uma temperatura.

Os RMSEPs obtidos para todos os modelos de todas as situações, o número de variáveis latentes utilizadas para a construção dos modelos PLS e os parâmetros γ e σ^2 estão presentes na Tabela 6.

Tabela 6: RMSEPs obtidos e parâmetros utilizados na construção dos modelos da proposta 2 para transferência de calibração na determinação do teor de álcool em cachaça.

Temperatura de validação (°C)	PLS		LS-SVM		
	RMSEP (%)	Variáveis latentes	RMSEP (%)	γ	σ^2
15	0,25	4	0,00085	6000	5
20	0,24	4	0,00083	6000	5
25	0,26	4	0,00070	6000	5
30	0,26	4	0,00074	6000	5
35	0,25	4	0,0010	6000	5

Nesta proposta fica evidente o melhor desempenho do LS-SVM em relação ao PLS. É interessante notar o quanto melhorou o desempenho do LS-SVM em relação a proposta anterior, ao passo que os resultados obtidos com PLS pouco mudaram.

Esse melhor desempenho já era esperado, já que as amostras utilizadas no conjunto de validação também estavam presentes no conjunto de calibração.

A Tabela 7 mostra os valores obtidos no teste F.

Tabela 7: Teste F para comparação dos modelos obtidos na proposta 2 para determinação do teor de álcool em cachaça.

Temperatura de validação (°C)	F calculado (x 10⁵)
15	0,86
20	0,84
25	1,38
30	1,23
35	0,63

F crítico=1,67

A Figura 16 mostra os erros obtidos para todos os modelos construídos nessa proposta. A diferença de grandeza dos erros relativos é tanta que as figuras não puderam ser ajustadas na mesma escala.

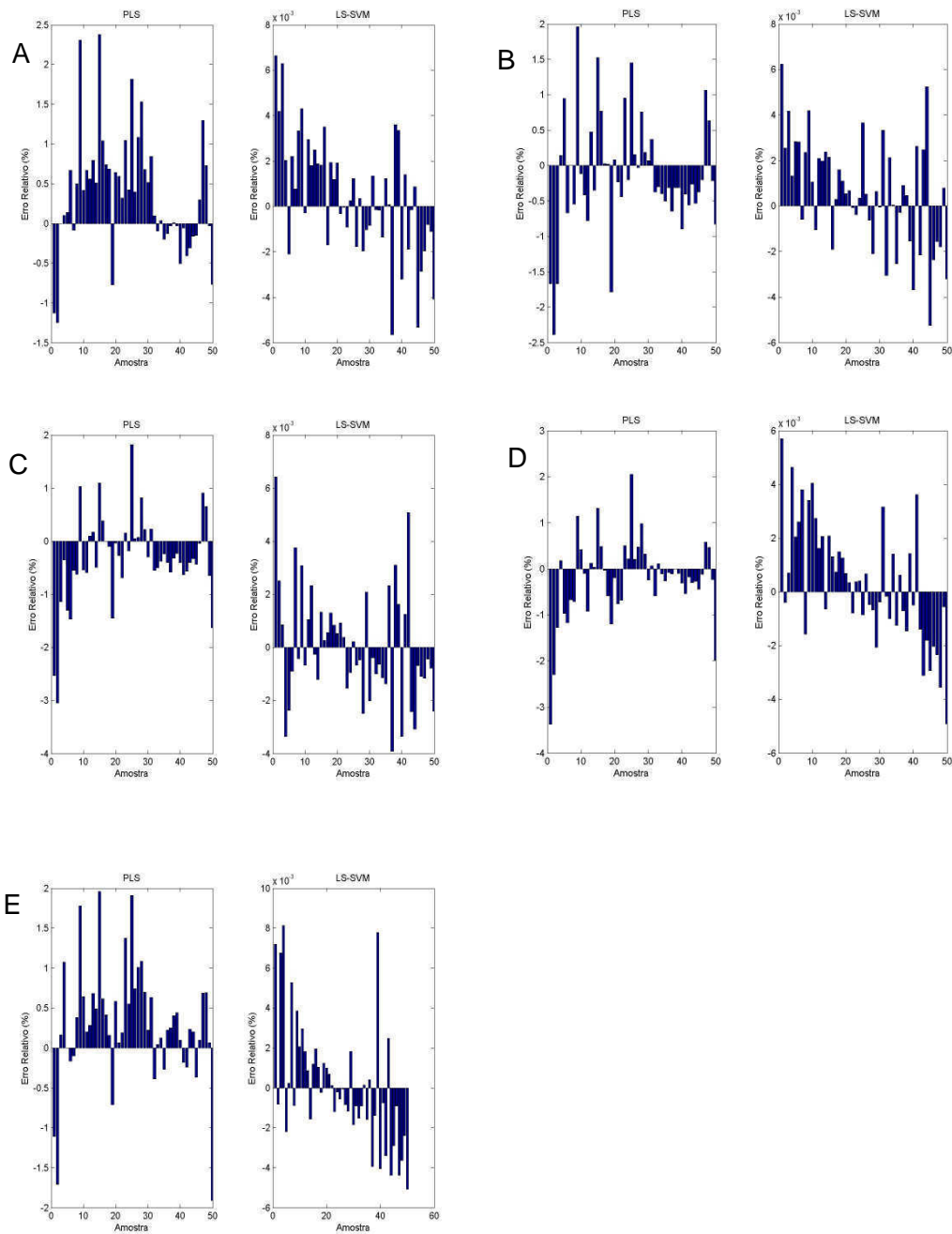


Figura 16: Erros relativos na previsão da porcentagem de álcool em amostras de cachaça obtidos nos modelos construídos para a proposta 2, quando as amostras de validação tiveram seus espectros obtidos nas temperaturas: A. 15°C; B. 20°C; C. 25°C; D. 30°C e E. 35°C.

3.5 Proposta 3

Nesta proposta foram separadas amostras de calibração e validação através do algoritmo kennard-stone, sendo utilizadas as amostras de calibração do conjunto de espectros a 20 °C e para cada situação foram utilizadas as amostras de validação do conjunto de espectros obtidos a uma outra temperatura.

Os RMSEPs dos modelos obtidos, assim como o número de variáveis latentes utilizados para construção dos modelos PLS e parâmetros para construção do LS-SVM estão na Tabela 8.

Tabela 8: RMSEPs obtidos e parâmetros utilizados na construção dos modelos da proposta 3 para transferência de calibração na determinação do teor de álcool em cachaça.

Temperatura de validação (°C)	PLS		PDS	LS-SVM		
	RMSEP (%)	VL*	RMSEP (%)	RMSEP (%)	γ	σ^2
15	0,14	5	0,25	4,78	3000	5000
25	0,31	5	0,23	1,32	3000	3000
30	0,86	5	0,49	2,93	3000	3000
35	0,86	4	0,64	4,77	5000	3000

*VL=variáveis latentes

Nesta proposta, os melhores resultados foram apresentados pelo PDS, enquanto o LS-SVM teve resultados piores do que o PLS. Isso pode ser explicado pelo fato de que os dados de calibração, neste caso, eram bem menos numerosos do que na proposta anterior. Desta forma o algoritmo do LS-SVM não foi capaz de realizar uma generalização adequada, o que foi o motivo essencial na primeira situação pra fazer com que o LS-SVM obtivesse os melhores resultados. Utilizando-se o conjunto de validação da mesma temperatura usada para o desenvolvimento do modelo de calibração (20°C) o RMSEP obtido foi de 0,13.

A Tabela 9 mostra os resultados do teste F obtidos para comparação dos modelos.

Tabela 9: Teste F para comparação dos modelos obtidos na proposta 3 na determinação do teor de álcool em cachaça.

Temperatura de validação (°C)	F calculado		
	PLS/PDS	PLS/LS-SVM (x 10 ³)	PDS/LS-SVM
15	3,19	1,16	365,57
25	1,81	0,018	32,94
30	3,08	0,012	35,76
35	1,81	0,031	55,55

F crítico = 2,12

Observando a Tabela 9 podemos perceber que o PDS produziu previsões significativamente melhores do que o PLS em duas das quatro situações e que o LS-SVM foi o que obteve as piores previsões em todas as situações, quando comparado com o PLS e com o PDS.

A Figura 17 mostra os gráficos de Erros Relativos versus Amostras para todos os modelos obtidos nesta segunda situação quando as amostras de validação estavam a 25°C. Neste caso pode-se notar uma alta tendência de erros sistemáticos em todas as situações apresentadas.

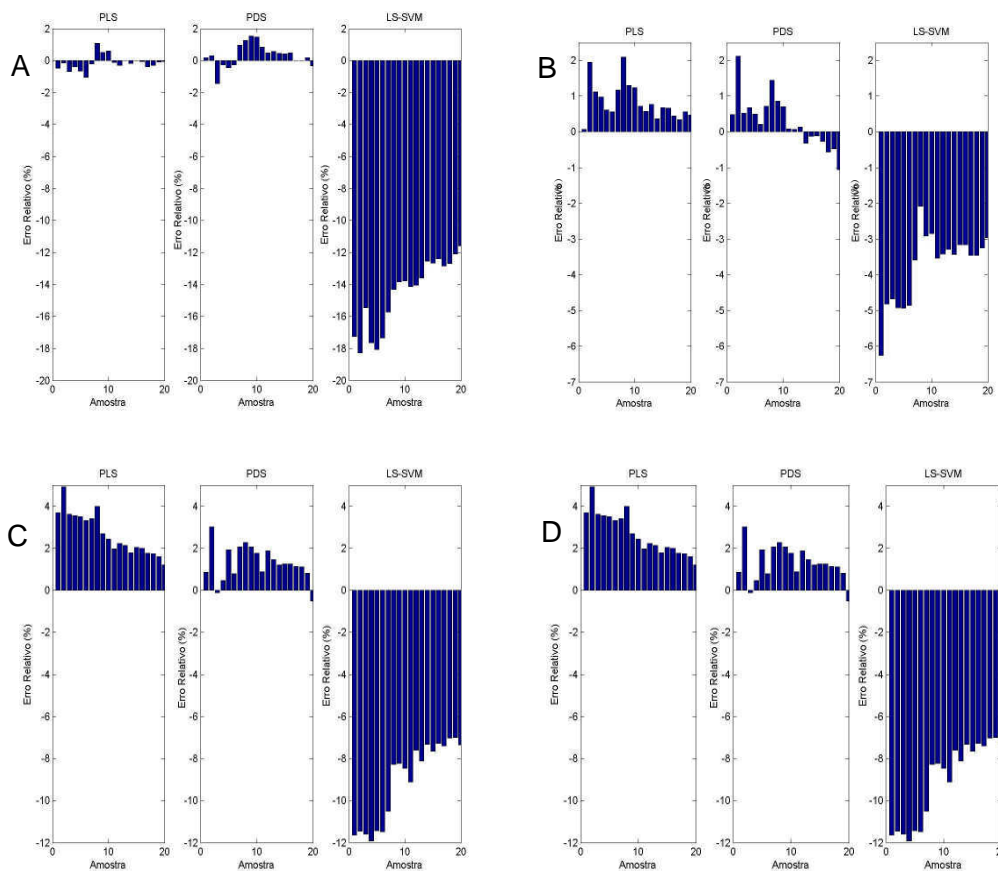


Figura 17: Erros relativos na previsão da porcentagem de álcool em amostras de cachaça obtidos nos modelos construídos para a proposta 3, quando as amostras de validação tiveram seus espectros obtidos nas temperaturas: A. 15°C; B. 25°C; C. 30°C e D. 35°C.

3.6 Proposta 4

Nesta proposta todas as amostras dos conjuntos de calibração separadas anteriormente foram agrupadas em um único conjunto de calibração. O mesmo foi feito com as amostras de validação, formando um único conjunto de validação.

Os RMSEPs e parâmetros do PLS e LS-SVM estão presentes na Tabela 10.

Tabela 10: RMSEPs obtidos e parâmetros utilizados na construção dos modelos da proposta 4 para transferência de calibração para determinação do teor de álcool em cachaça.

PLS		LS-SVM		
RMSEP (%)	Variáveis latentes	RMSEP (%)	γ	σ^2
0,22	4	0,13	600	500

Novamente o LS-SVM produziu um modelo mais ajustado do que o PLS. A inclusão de mais amostras no conjunto de calibração em relação à proposta anterior fez com que tanto PLS quanto LS-SVM gerassem modelos que possuem melhores previsões do que naquela proposta, em especial o LS-SVM. O resultado do teste F está na Tabela 11.

Tabela 11: Teste F para comparação dos modelos obtidos na proposta 4 para determinação do teor de álcool em cachaça.

F calculado
2,86
F crítico = 1,39

O resultado do teste F confirma que houve uma melhora significativa nas previsões quanto utilizado o LS-SVM. Os erros relativos dos dois modelos estão na Figura 18.

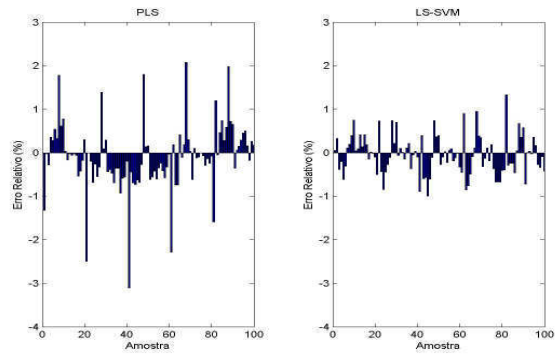


Figura 18: Erros relativos na previsão da porcentagem de álcool em amostras de cachaça obtidos nos modelos construídos para a proposta 4.

Capítulo 4

Capítulo 4. Determinação do teor de proteína e gordura em ração para cães.

Neste trabalho foram propostas duas aplicações envolvendo ração para cães. Foram realizadas comparações entre métodos lineares (PLS e PDS) e um não-linear (LS-SVM) na transferência de calibração de espectros de reflectância difusa na região do infravermelho próximo, para determinação teor de gordura e proteína em ração para cães com diferentes tamanhos de partícula.

Para ambas aplicações foram utilizados os mesmos espectros, obtidos através das mesmas amostras. Devido a isto, após uma introdução em comum este capítulo será dividido em duas partes, cada uma referente a uma aplicação.

4.1 Ração para cães

Segundo a Associação Nacional dos Fabricantes de Alimentos para animais, existem no Brasil cerca de 21 milhões de cães com endereço fixo, a segunda maior população do planeta. Destes, 34% são alimentados com ração industrializada.⁴³

As primeiras notícias da presença da indústria de rações para animais no Brasil reportam-se aos anos da 2ª Guerra Mundial, mencionando-se uma firma de origem alemã, a Pró-Pecuária em São Paulo, o Moinho da Luz, Fluminense e São Cristóvão no Rio de Janeiro.⁴⁴ No ano de 2002, a produção de alimentos para animais de estimação foi da ordem de 1.234.000 toneladas. Existem atualmente no país mais de 300 marcas registradas, produzidas por mais de 70 fabricantes, com preço e qualidade variáveis.

No Brasil a responsabilidade da regulamentação das rações para cães e gatos é do Ministério da Agricultura, Pecuária e Abastecimento, prevista pelo Decreto nº 76.986 de 6 de janeiro de 1976, que regulamentou a Lei nº 6.198, de 26 de dezembro de 1974, que dispõe sobre a inspeção e a fiscalização animal. O Decreto é instituído por meio de Instruções Normativas que são atualizadas e publicadas periodicamente.

Os alimentos industrializados são divididos de acordo com a segmentação comercial instituída pela própria indústria, não caracterizada ou contida na Instrução Normativa nº 8, e baseia-se na qualidade e no tipo de matéria prima, concentração de

nutrientes, características do rótulo e preço, sendo normalmente aceita pelos consumidores como um critério qualitativo que norteia decisões de compra.⁴³

As rações para cães são compostas por uma fonte de proteína, ingredientes amiláceos como cereais e amido, gorduras, açúcar, um adesivo proteínico, e um agente plastificante. A porcentagem em massa da fonte de proteína deve ser de 0 a 35%, a porcentagem em massa de gordura deve ser de 3 a 15% e a porcentagem em massa de açúcar deve ser entre 0 e 35%. Vitaminas, sais minerais, colorizantes, aromatizante, e outros suplementos podem ser usados na fabricação de rações.

Geralmente as rações são produzidas através da mistura dos ingredientes e água suficiente para o processamento e posterior cozimento. A mistura cozida é resfriada e cortada.⁴⁵

4.2 Experimental

As 38 amostras fornecidas pela empresa Nutron Alimentos Ltda, a qual realizou as análises padrão para determinação do teor de proteína⁴⁶ e gordura⁴⁷, foram peneiradas com peneiras Granutest de 0,50 e 0,35 mm de abertura, separando-se assim, três diferentes granulometrias para cada amostra: A = partículas maiores do que 0,50 mm, B = partículas entre 0,50 e 0,35 mm, C = partículas menores que 0,35 mm.

Um exemplo de separação das partículas de granulometrias diferentes de uma mesma amostra está na Figura 19.



Figura 19: Diferentes tamanhos de partícula de uma mesma amostra de ração para cães, onde: A. partículas maiores que 0,50 mm; B. partículas com tamanho entre 0,50 e 0,35 mm e C. partículas menores que 0,35 mm.

As amostras possuíam teor de proteína entre 19,25 e 37,98 % m/m e teor de gordura entre 7,55 e 22,10 % m/m. Os teores de proteína e gordura não eram relacionados, de forma que uma amostra com alto teor de proteína poderia ter alto ou baixo teor de gordura. Os teores de proteína e gordura de cada amostra são apresentados na Tabela 12.

Tabela 12: teor de proteína e gordura em % (m/m) nas amostras de ração para cães.

AMOSTRA	Proteína (%)	Gordura (%)	AMOSTRA	Proteína (%)	Gordura (%)
1	20,10	10,22	20	22,73	10,52
2	23,01	9,98	21	22,46	10,10
3	26,72	10,25	22	35,44	17,23
4	28,14	11,89	23	20,40	9,92
5	22,82	10,90	24	28,22	10,15
6	24,49	8,13	25	30,54	14,76
7	27,66	10,99	26	27,26	9,90
8	20,10	8,70	27	31,07	12,77
9	24,58	10,67	28	21,78	8,58
10	21,69	11,41	29	22,40	10,70
11	31,90	16,64	30	30,04	14,93
12	29,20	9,55	31	22,84	9,90
13	36,14	18,88	32	36,65	19,85
14	20,85	8,34	33	35,13	11,34
15	22,56	11,39	34	20,24	8,06
16	20,52	7,55	35	28,01	16,21
17	36,16	16,67	36	22,70	10,49
18	19,25	8,38	37	22,77	11,41
19	37,98	22,10	38	21,95	9,98

Os espectros foram obtidos para todas as granulometrias de todas as amostras em um equipamento Varian Cary 5G na faixa de 1000 a 2400 nm. A Figura 20 mostra os espectros na região do infravermelho próximo obtidos para as amostras com granulometria A.

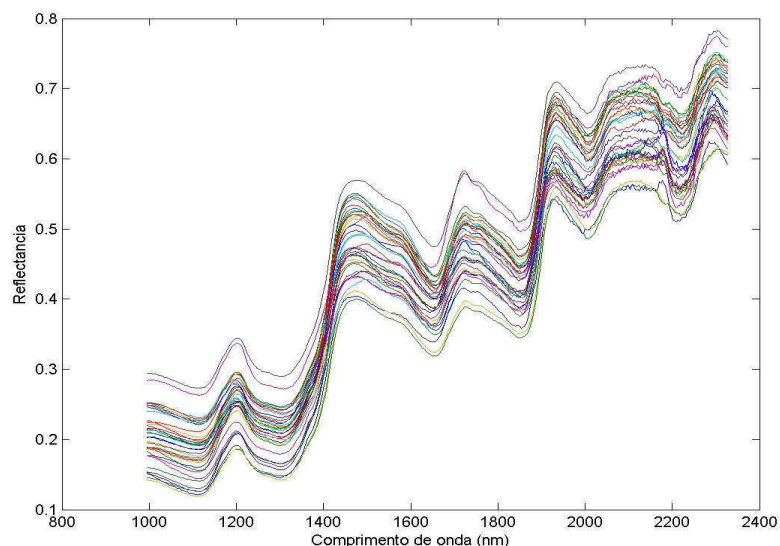


Figura 20: Espectros na região do infravermelho próximo de amostras de ração para cães com tamanho de partículas maior do que 0,50 mm.

As bandas importantes que podem ser observadas são mostradas na Tabela 13.

Tabela 13: Atribuição das bandas do espectro na região do infravermelho para as amostras de ração.⁷

Comprimento de onda (nm)	Designação dos modos vibracionais
1492	Primeiro sobretom de estiramento N-H
1980	Estiramento N-H assimétrico
2050	Estiramento N-H simétrico
1720	Primeiro sobretom do estiramento C-H
1760	Primeiro sobretom do estiramento C-H
2140	Banda de combinação do estiramento =C-H com estiramento C=C
2352	Segundo sobretom da deformação C-H

Para a realização dos modelos, os dados foram centrados na média e os melhores resultados foram obtidos através da primeira derivada dos espectros (Figura 21). Os modelos de PLS e PDS foram otimizados a partir da validação cruzada “um por vez”.

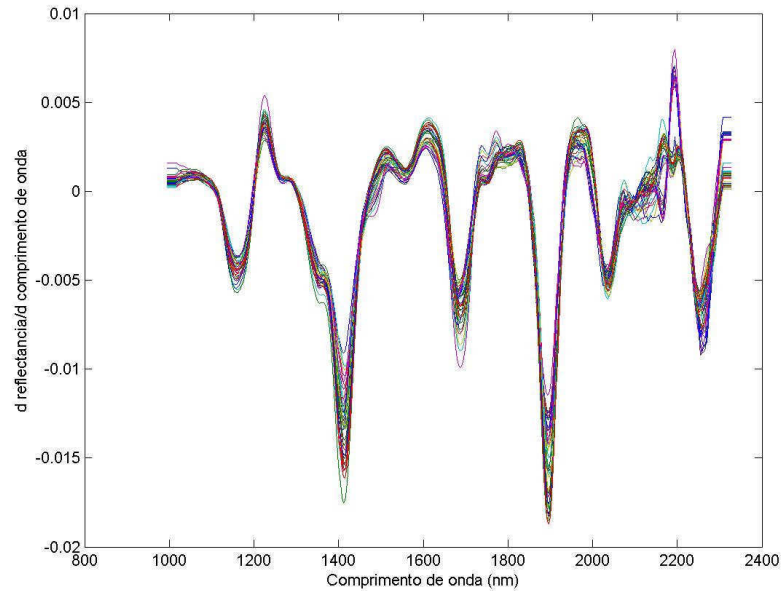


Figura 21: Primeira derivada dos espectros de razão obtidos.

Para o tratamento dos dados as amostras (com todas suas granulometrias) foram separadas em dois grupos, um para a calibração, com 28 amostras, e outro para validação com 10 amostras, através do algoritmo de Kennard-Stone.⁴⁰

Assim como na determinação do teor de álcool em cachaça, foram feitas quatro propostas para construção dos modelos de calibração, tanto para determinação do teor de proteína quanto para determinação do teor de gordura. Na proposta 1 os espectros obtidos de amostras em duas granulometrias diferentes foram utilizados como conjunto de calibração e os espectros obtidos das amostras com a granulometria restante foram utilizados como dados de validação. Todas as combinações foram feitas, sendo os espectros obtidos em todas as granulometrias utilizados como dados de calibração e validação. Os modelos foram construídos utilizando PLS e LS-SVM.

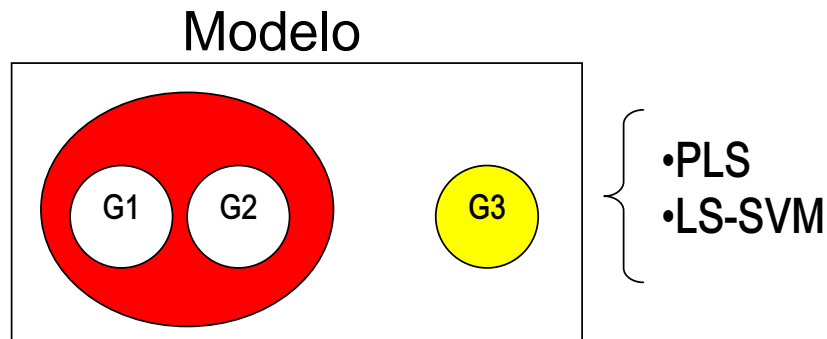


Figura 22: Representação esquemática da proposta 1 para transferência de calibração para determinação do teor de proteína e gordura em ração para cães em diferentes granulometrias.

Na proposta 2, o conjunto de calibração era constituído de todos os espectros obtidos das amostras com todas granulometrias, enquanto o conjunto de validação era composto de espectros obtidos de amostras de apenas uma granulometria. Foram feitas assim três validações diferentes, utilizando cada conjunto de espectros obtidos a cada granulometria. Os modelos foram obtidos através de PLS e LS-SVM.

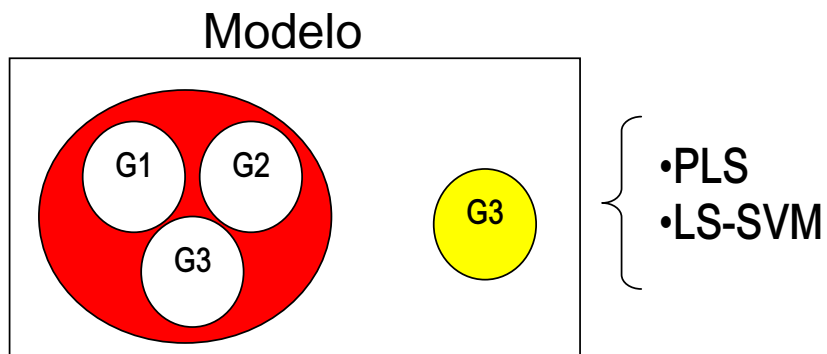


Figura 23: Representação esquemática da proposta 2 para transferência de calibração para determinação do teor de proteína e gordura em ração para cães em diferentes granulometrias.

A proposta 3 foi construída utilizando apenas espectros de amostras de calibração de uma única granulometria e a validação foi feita com os espectros de amostras de validação obtidos em outra granulometria. Todas as combinações foram feitas, de modo

que todos os conjuntos de calibração e validação foram utilizados. Neste caso os modelos eram construídos através de PLS, LS-SVM e PDS.

Este foi o único modelo onde foi utilizado o PDS porque foi o único caso onde havia disponibilidade de espectros na situação ESCRAVA obtidos das mesmas amostras utilizadas no conjunto de calibração. Desta forma era possível a obtenção da matriz de transformação **F**.

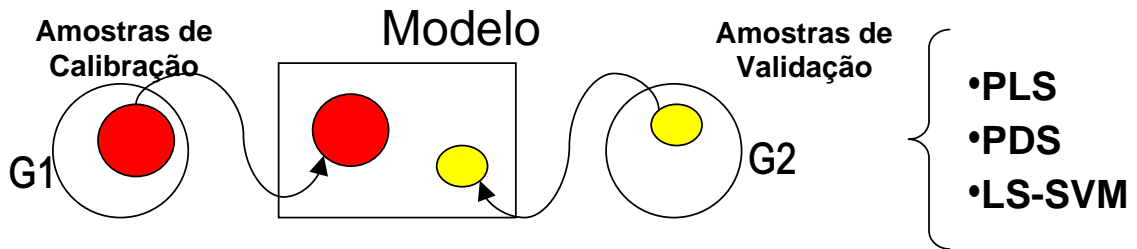


Figura 24: Representação esquemática da proposta 3 para transferência de calibração para determinação do teor de proteína e gordura em ração para cães em diferentes granulometrias.

A proposta 4, foi construída utilizando todos os espectros tidos como amostras de calibração para as três granulometrias, sendo a validação feita com as amostras de validação de todas as granulometrias. Os modelos foram obtidos com PLS e LS-SVM.

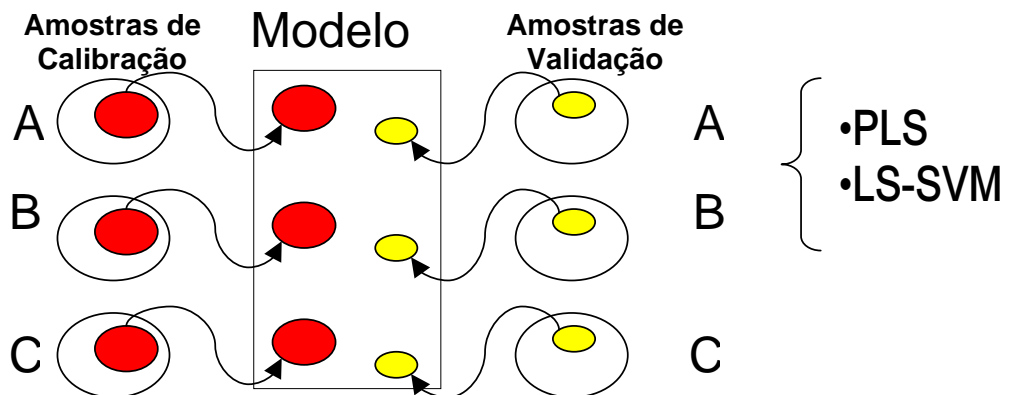


Figura 25: Representação esquemática da proposta 4 para transferência de calibração para determinação do teor de proteína e gordura em ração para cães em diferentes granulometrias.

Os modelos PLS e PDS foram otimizados a partir da validação cruzada “um por vez”. Um exemplo para um gráfico de erro médio quadrático de calibração cruzada por número de variáveis latentes utilizado para definir o número de variáveis latentes a serem utilizados na construção dos modelos PLS e PDS está na Figura 26. Neste caso foram escolhidas oito variáveis latentes para o desenvolvimento do modelo de calibração.

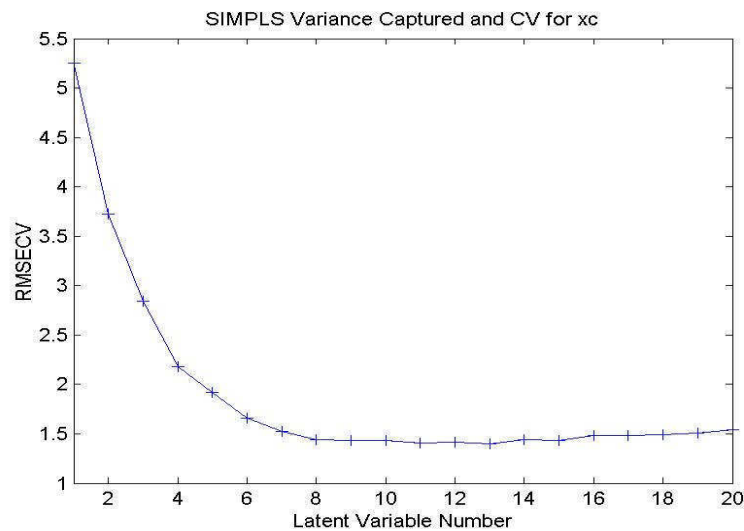


Figura 26: Exemplo de gráfico de erro médio quadrático de calibração cruzada por número de variáveis latentes utilizado na determinação do número ideal de variáveis latentes na construção dos modelos PLS e PDS.

As calibrações foram realizadas através do *PLS-Toolbox* versão 3.5 (Eigenvector Research Inc. ⁴¹) para *Matlab* 6.5 (MathWorks Co.). Os modelos LS-SVM foram obtidos através do *LS-SVM lab toolbox*⁴². Todos os programas foram executados em um microcomputador Intel Pentium 4 CPU 3,00 GHz com 1 GB de RAM.

4.3 Determinação do teor de proteína em ração para cães.

4.3.1 Proposta 1

A Tabela 14 traz os valores de RMSEP e parâmetros dos modelos PLS e LS-SVM para todas as situações da proposta 1 de transferência de calibração para determinação do teor de proteína em ração para cães.

Tabela 14: RMSEPs obtidos e parâmetros utilizados na construção dos modelos da proposta 1 para transferência de calibração na determinação do teor proteína em ração para cães.

Granulometria de validação	PLS		LS-SVM		
	RMSEP (%)	Variáveis latentes	RMSEP (%)	γ	σ^2
A	3,05	5	1,77	3000	2000
B	1,66	5	1,48	1000	10000
C	2,11	5	2,08	1000	10000

É possível notar que o LS-SVM produziu melhores resultados e seu melhor desempenho em relação ao PLS foi quando o modelo foi construído com as granulometrias B e C e a previsão foi feita com a granulometria A (com o RMSEP passando de 3,05 para 1,77%). Isso pode ser explicado pelo motivo de que as amostras B e C são mais parecidas entre si (ou seja, possuíam granulometrias mais próximas) do que com a A. Desta forma a capacidade maior de generalização do LS-SVM foi capaz de prever as amostras de granulometria A com menores erros.

O resultado do teste F presente na Tabela 15 mostra que há uma diferença significativa nos modelos PLS e LS-SVM nos modelos onde foi feita a previsão do teor de proteína das amostras com granulometria A.

Tabela 15: Teste F para comparação dos modelos obtidos na proposta 1 para determinação do teor de proteína em ração para cães.

Granulometria de validação F calculado	
A	2.97
B	1.26
C	1.03

F crítico = 1,69

A Figura 27 mostra os erros relativos obtidos para todos os modelos construídos.

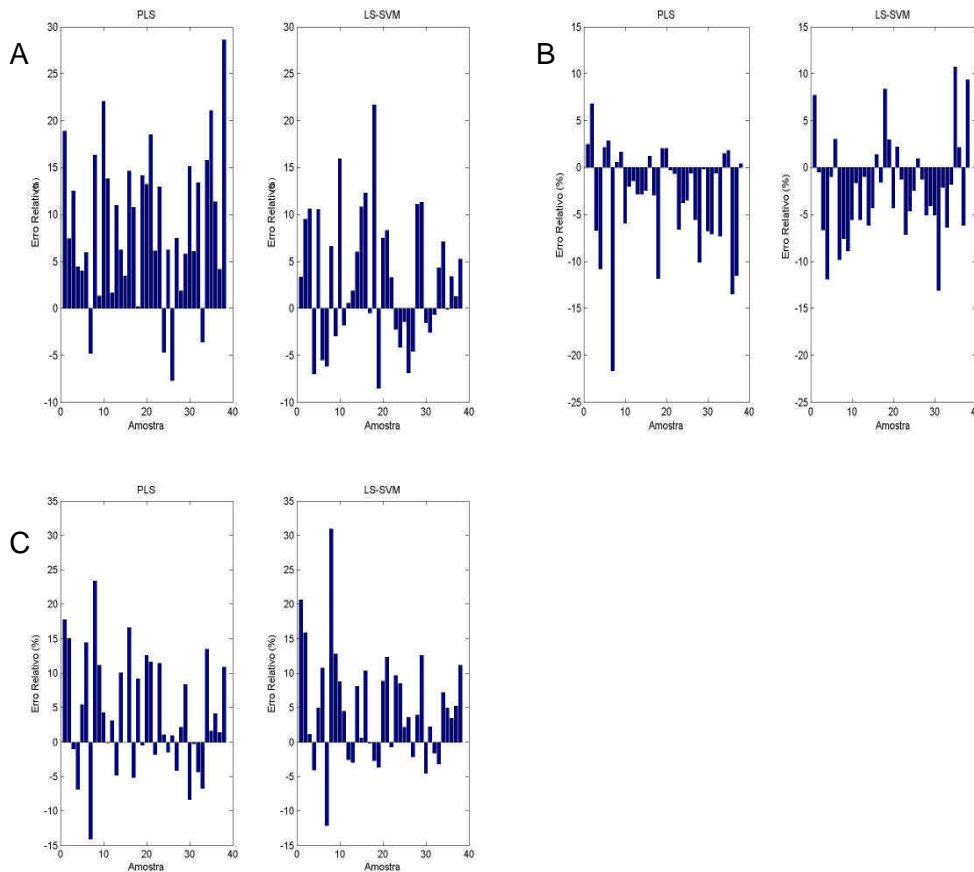


Figura 27: Erros relativos na previsão do teor de proteína em amostras de ração para cães obtidos nos modelos construídos para a proposta 1, quando as amostras de validação tiveram seus espectros obtidos nas granulometrias: A, B e C.

4.3.2 Proposta 2

Na tabela abaixo estão os valores de RMSEP e condições do PLS e LS-SVM utilizados na construção dos modelos para todas as situações da proposta 2 para transferência de calibração de espectros na determinação do teor de proteína em ração. Nesta proposta os espectros obtidos a todas granulometrias foram utilizados como um único conjunto de calibração. A validação foi feita com cada conjunto de espectros obtidos a uma determinada granulometria.

Tabela 16: RMSEPs obtidos e parâmetros utilizados na construção dos modelos da proposta 2 para transferência de calibração para determinação do teor proteína em ração para cães.

Granulometria de validação	PLS		LS-SVM		
	RMSEP (%)	Variáveis latentes	RMSEP (%)	γ	σ^2
A	1,43	5	0,00077	6000	50
B	1,30	5	0,00069	6000	50
C	2,78	5	0,00073	6000	50

É notável o melhor desempenho do LS-SVM em relação ao PLS nesta proposta, diminuindo de forma acentuada o valor dos RMSEPs obtidos para cada situação. A diminuição dos erros foi tanta que não permitiu que a Figura 28 mostrasse na mesma escala os erros relativos de previsão. Pode ser observado também que a inclusão dos espectros obtidos das amostras com mesma granulometria das amostras de validação no conjunto de calibração fez com que o modelo LS-SVM conseguisse prever bem melhor amostras com essa granulometria em relação à proposta anterior.

A Tabela 17 traz os resultados obtidos com o teste F na comparação dos modelos de calibração.

Tabela 17: Teste F para comparação dos modelos obtidos na proposta 2 para determinação do teor de proteína em ração para cães.

Granulometria de validação F calculado ($\times 10^7$)	
A	0.34
B	0.36
C	1.45

F crítico = 1,69

É possível notar na Tabela 17 que em todos os modelos houve diferença significativa dos resultados obtidos com PLS e LS-SVM.

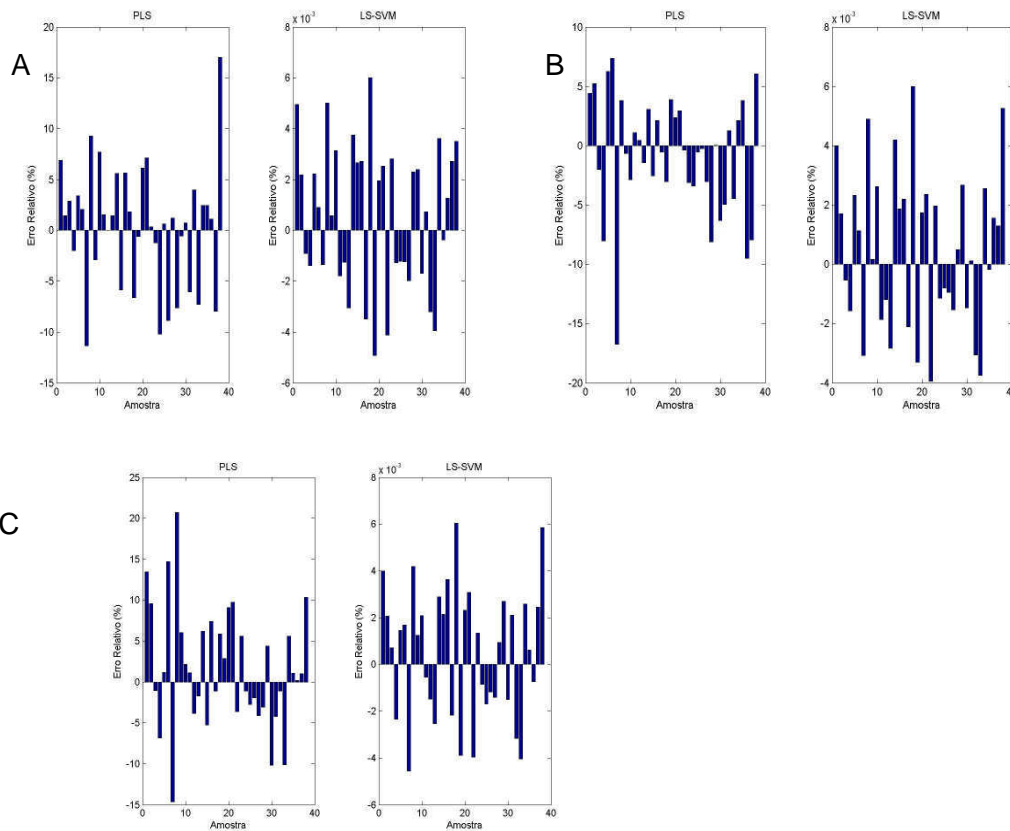


Figura 28: Erros relativos na previsão do teor de proteína em amostras de ração para cães obtidos nos modelos construídos para a proposta 2, quando as amostras de validação tiveram seus espectros obtidos nas granulometrias: A, B e C.

4.3.3 Proposta 3

Na Tabela 18 estão apresentados os valores de RMSEP e parâmetros utilizados na construção dos modelos construídos para todas as situações da proposta 3. Nessa proposta foram separadas amostras de calibração e validação para todas as granulometrias e foram feitas todas as combinações possíveis dos conjuntos de calibração e validação.

Tabela 18: RMSEPs obtidos e parâmetros utilizados na construção dos modelos da proposta 3 para transferência de calibração para determinação do teor proteína em ração para cães.*

Granulometria		PLS		PDS	LS-SVM		
Cal ^a	Val ^b	RMSEP (%)	VL ^c	RMSEP (%)	RMSEP (%)	γ	σ^2
A	B	12,93	10	10,37	10,89	1000	2000
A	C	26,46	10	8,43	19,72	1000	2000
B	A	1,82	3	1,78	1,86	6000	10000
B	C	3,67	3	1,71	2,89	500	2000
C	A	7,54	3	1,99	3,50	500	2000
C	B	4,54	3	1,38	3,40	500	1000

*onde (a) Cal é a granulometria do conjunto de calibração, (b) Val é a granulometria do conjunto de validação e (c) VL significa variáveis latentes.

Observando esses dados é possível notar que os piores resultados são aqueles onde o conjunto de calibração possuía a granulometria A, o que pode ser novamente explicado pela diferença maior de granulometria entre este conjunto e os demais tamanhos de partícula. Como neste caso foram utilizadas amostras de apenas uma granulometria para a calibração o LS-SVM não foi capaz de fazer a generalização necessária para a previsão.

A Tabela 19 mostra os resultados obtidos com o teste F para comparação dos modelos obtidos

Tabela 19: Teste F para comparação dos modelos obtidos na proposta 3 para determinação do teor de proteína em ração para cães.

Granulometria		F calculado		
Cal ^a	Val ^b	PDS/PLS	PDS/LS-SVM	PLS/LS-SVM
A	B	1.55	1.10	1.41
A	C	9.85	5.47	1.80
B	A	1.05	1.09	1.04
B	C	4.60	2.86	1.61
C	A	14.36	3.09	4.64
C	B	10.82	6.07	1.783

F crítico = 2,98

*onde (a) Cal é a granulometria do conjunto de calibração, (b) Val é a granulometria do conjunto de validação e (c) VL significa variáveis latentes.

Como se pode conferir pela Tabela 19 na comparação entre os modelos obtidos com PLS e PDS apenas dois não tiveram diferenças significativas; comparando-se os modelos obtidos com PDS e LS-SVM é possível notar que houve diferenças significativas entre três dos seis modelos e, se comparando os modelos PLS e LS-SVM nota-se que em apenas um caso houve diferenças significativas entre os dois algoritmos. Os erros relativos de todos os modelos construídos estão na Figura 29.

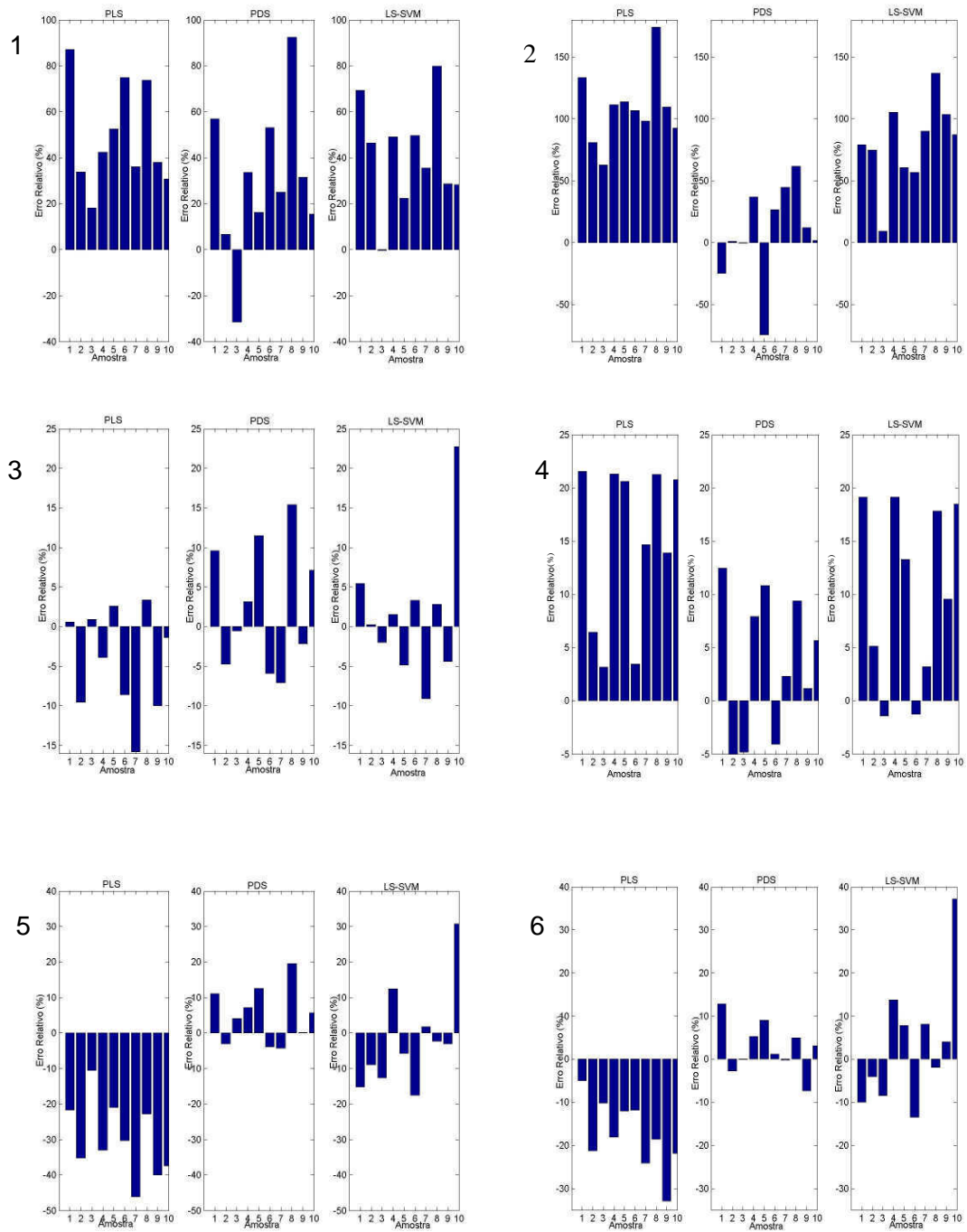


Figura 29: Erros relativos na previsão do teor de proteína em amostras de ração para cães obtidos nos modelos construídos para a proposta 3, onde: 1. Cal: A e Val: B; 2. Cal:A e Val C; 3. Cal:B e Val: A; 4. Cal: B e Val: C; 5. Cal: C e Val A e 6. Cal: C e Val: B.

4.3.4 Proposta 4

Na Tabela 20 estão presentes os valores de RMSEP e parâmetros necessários para a construção dos modelos PLS e LS-SVM para esta última proposta de transferência de calibração para modelos de determinação do teor de proteínas em ração para cães.

Tabela 20: RMSEPs obtidos e parâmetros utilizados na construção dos modelos da proposta 4 para transferência de calibração para determinação do teor proteína em ração para cães.

PLS		LS-SVM		
RMSEP (%)	Variáveis latentes	RMSEP (%)	γ	σ^2
3,41	10	2,27	5	2000

Nessa proposta o LS-SVM obteve novamente um melhor RMSEP do que o PLS. Observando-se a Figura 30 é possível notar que no PLS houve dez amostras com erros relativos acima de 15%, enquanto no modelo do LS-SVM foram apenas três. Observando-se o valor obtido com o teste F (na Tabela 21) pode-se confirmar que houve diferença significativa entre os modelos.

Nota-se também que a inclusão de mais amostras no conjunto de calibração melhora bastante a performance tanto do PLS quanto do LS-SVM em relação à proposta anterior.

Tabela 21: Teste F para comparação dos modelos obtidos na proposta 4 para determinação do teor de proteína em ração para cães.

F calculado
2.26
F crítico=1,84

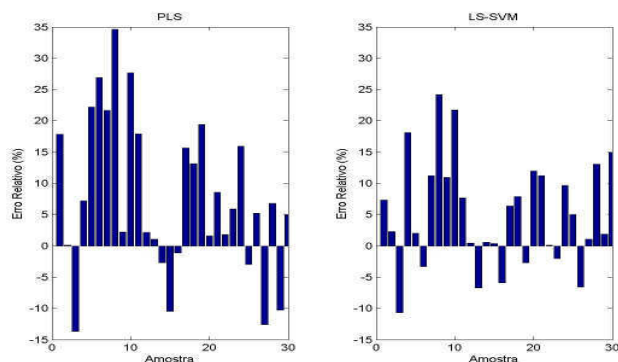


Figura 30: Erros relativos na previsão do teor de proteína em amostras de ração para cães obtidos nos modelos construídos para a proposta 4.

4.4 Determinação do teor de gordura em ração para cães.

4.4.1 Proposta 1

Os RMSEPs e parâmetros para construção dos modelos PLS e LS-SVM para as três situações da proposta 1 para transferência de calibração na determinação de gordura em ração através de NIR e quimiometria estão na Tabela 22.

Tabela 22: RMSEPs obtidos e parâmetros utilizados na construção dos modelos da proposta 1 para transferência de calibração para determinação do teor de gordura em ração para cães.

Granulometria de validação		PLS		LS-SVM		
va	RMSEP (%)	Variáveis Latentes	RMSEP (%)	γ	σ^2	
A	1,62	5	1,07	1000	2000	
B	1,01	5	0,85	1000	2000	
C	1,01	5	0,89	1000	10000	

Novamente o LS-SVM obteve resultados melhores do que o PLS. Como era esperado, o melhor desempenho do LS-SVM em relação ao PLS foi na previsão das

amostras com granulometria A, devido a maior diferença de tamanho entre essas partículas e as demais. A Tabela 23 mostra os resultados do teste F realizados na comparação dos modelos obtidos nesta proposta.

Tabela 23: Teste F para comparação dos modelos obtidos na proposta 1 para determinação do teor de gordura em ração para cães.

Granulometria de validação F calculado	
A	2,29
B	1,41
C	1,29

F crítico = 1,69

Observando-se a Tabela 23 nota-se que apenas na primeira situação o LS-SVM obteve um modelo significativamente melhor.

A Figura 31 mostra os erros relativos de todos os modelos construídos.

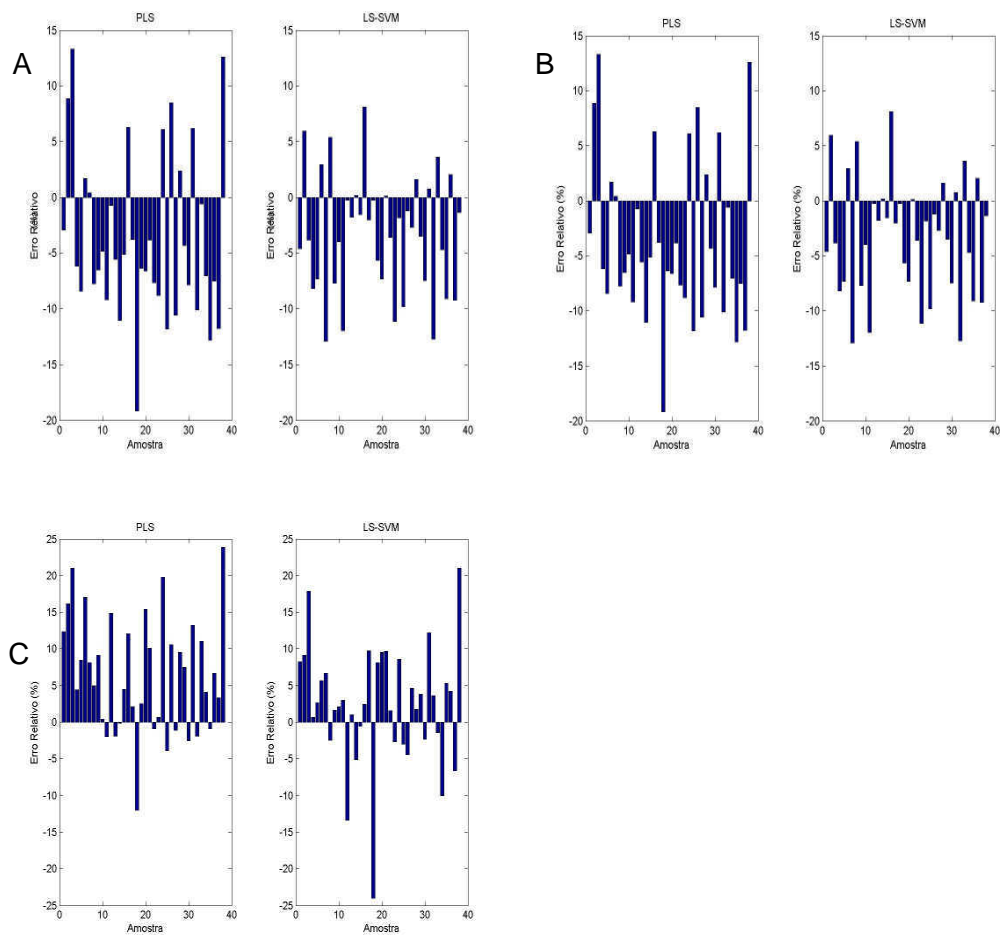


Figura 31: Erros relativos na previsão do teor de gordura em amostras de ração para cães obtidos nos modelos construídos para a proposta 1, quando as amostras de validação tiveram seus espectros obtidos nas granulometrias: A, B e C.

4.4.2 Proposta 2

Os RMSEPs e parâmetros usados para construção dos modelos PLS e LS-SVM para transferência de calibração na determinação do teor de gorduras em ração para cães estão na Tabela 24.

Tabela 24: RMSEPs obtidos e parâmetros utilizados na construção dos modelos da proposta 2 para transferência de calibração para determinação do teor de gordura em ração para cães.

Granulometria de validação	PLS		LS-SVM		
	RMSEP (%)	Variáveis latentes	RMSEP (%)	γ	σ^2
A	0,82	6	0,00048	6000	50
B	0,73	6	0,00041	6000	50
C	0,69	6	0,00045	6000	50

Nesta proposta a melhoria da previsão fica evidente quando se utiliza o LS-SVM ao invés do PLS para se construir os modelos. Novamente ficou impossível se colocar as figuras de erro na mesma escala, tamanha a diferença de grandeza entre os erros relativos obtidos nos diferentes modelos, como pode ser comprovado na Figura 32. Os RMSEPs obtidos nos modelos desta proposta foram menores do que na proposta anterior. Os resultados obtidos com o teste F estão apresentados na Tabela 25.

Tabela 25: Teste F para comparação dos modelos obtidos na proposta 2 para determinação do teor de gordura em ração para cães.

Granulometria de validação	F calculado ($\times 10^6$)
A	2,92
B	3,17
C	2,35

F crítico = 1,69

Em todas as situações houve diferença significativa entre os modelos obtidos com o PLS e LS-SVM.

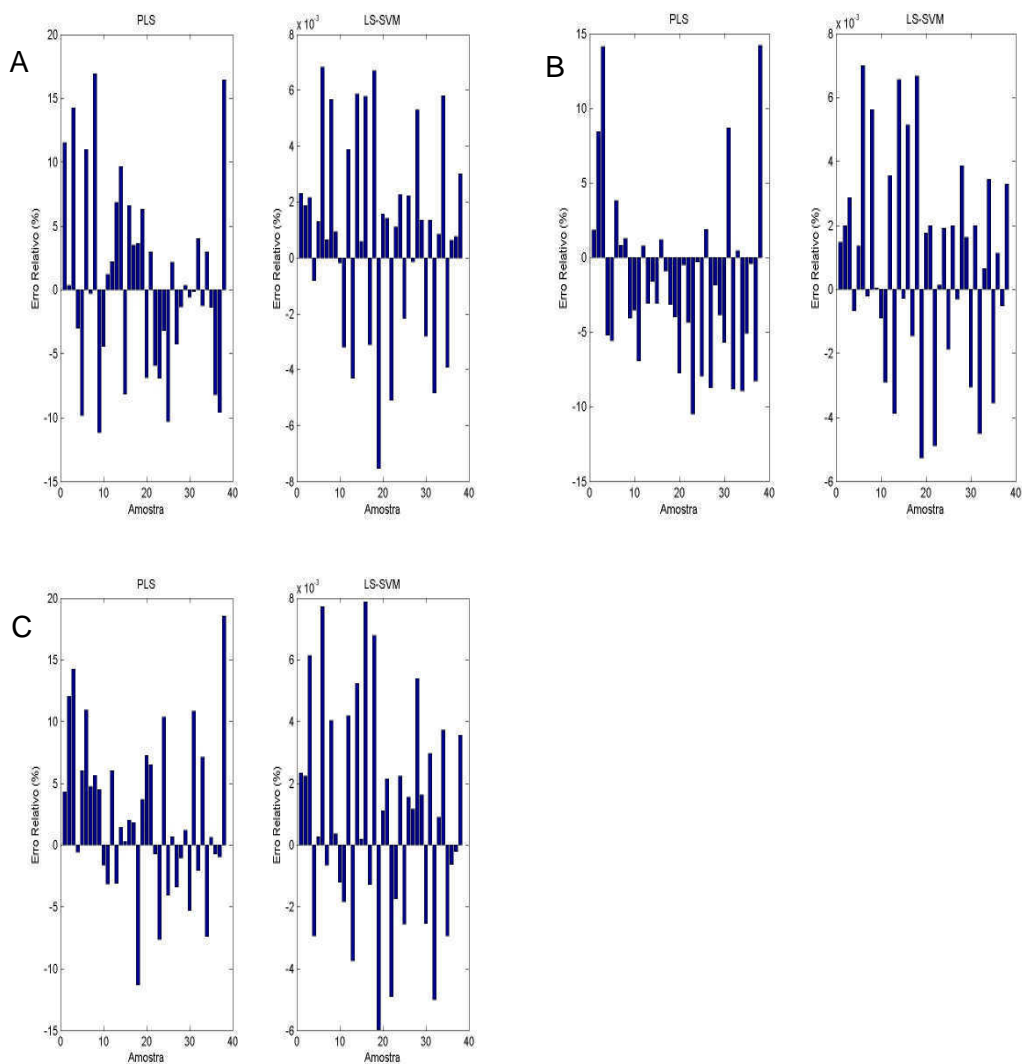


Figura 32: Erros relativos na previsão do teor de gordura em amostras de ração para cães obtidos nos modelos construídos para a proposta 2, quando as amostras de validação tiveram seus espectros obtidos nas granulometrias: A, B e C.

4.4.3 Proposta 3

A Tabela 26 mostra os RMSEPs e parâmetros utilizados para construção dos modelos PLS e LS-SVM para todas as seis situações da proposta 3 para transferência de calibração na determinação do teor de gorduras em ração para cães.

Tabela 26: RMSEPs obtidos e parâmetros utilizados na construção dos modelos da proposta 3 para transferência de calibração para determinação do teor de gordura em ração para cães.*

Granulometria		PLS		PDS	LS-SVM		
Cal ^a	Val ^b	RMSEP (%)	VL ^c	RMSEP (%)	RMSEP (%)	γ	σ^2
A	B	11,82	7	2,98	1,91	3000	5000
A	C	1,73	7	1,40	3,45	3000	5000
B	A	1,17	6	1,95	1,40	500	3000
B	C	1,87	6	1,81	1,58	500	5000
C	A	3,84	7	1,93	2,46	3000	2000
C	B	1,51	7	1,63	1,26	3000	2000

*onde (a) Cal é a granulometria do conjunto de calibração, (b) Val é a granulometria do conjunto de validação e (c) VL significa variáveis latentes.

Essa foi a proposta onde foram obtidos maiores erros para determinação do teor de gordura em ração. A falta de ajuste dos modelos pode ser explicada pela pequena quantidade de amostras de calibração.

A Tabela 27 traz os resultados obtidos com o teste F na comparação dos modelos obtidos nessa proposta.

Tabela 27: Teste F para comparação dos modelos obtidos na proposta 3 para determinação do teor de gordura em ração para cães.

Granulometria		F calculado		
Cal ^a	Val ^b	PLS/LS-SVM	PDS/LS-SVM	PDS/PLS
A	B	38,30	2,43	15,73
A	C	3,98	6,07	1,52
B	A	1,43	1,94	2,78
B	C	1,40	1,31	1,07
C	A	2,44	1,62	3,96
C	B	1,44	1,67	1,17

F crítico = 2,98

*onde (a) Cal é a granulometria do conjunto de calibração, (b) Val é a granulometria do conjunto de validação e (c) VL significa variáveis latentes.

Observando-se a Tabela 27, notamos que não houve muitas diferenças significativas entre os modelos.

Os erros relativos de todos os modelos construídos para esta proposta estão na Figura 33.

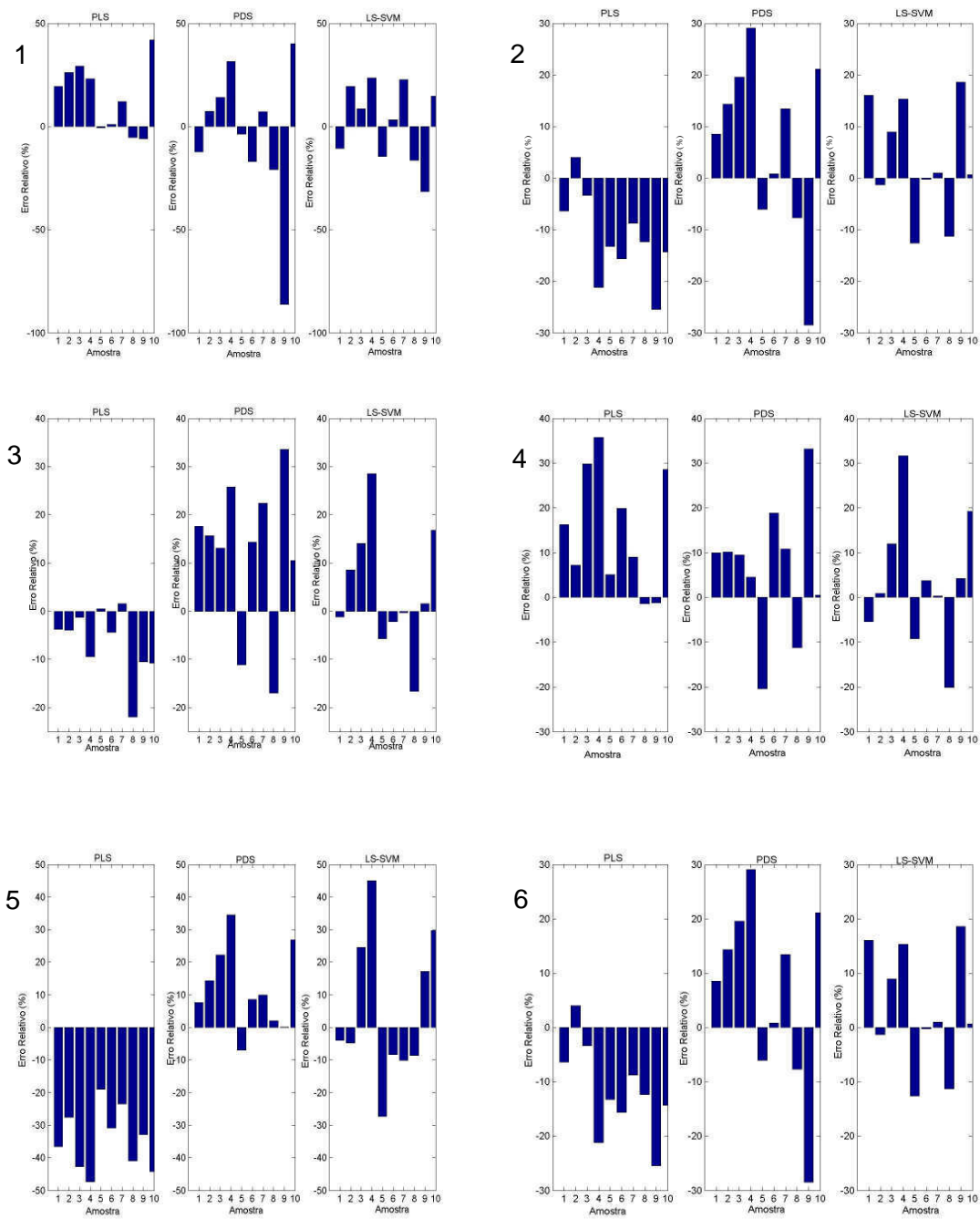


Figura 33: Erros relativos na previsão do teor de gordura em amostras de ração para cães obtidos nos modelos construídos para a proposta 3, onde: 1. Cal: A e Val: B; 2. Cal:A e Val C; 3. Cal: B e Val: A; 4. Cal: B e Val: C; 5. Cal: C e Val A e 6. Cal: C e Val: B.

4.4.4 Proposta 4

Os valores de RMSEP obtidos e os parâmetros utilizados na construção dos modelos da proposta 4 para determinação do teor de gordura em ração para cães estão na Tabela 28.

Tabela 28: RMSEPs obtidos e parâmetros utilizados na construção dos modelos da proposta 4 para transferência de calibração para determinação do teor de gordura em ração para cães.

PLS		LS-SVM		
RMSEP (%)	Variáveis latentes	RMSEP (%)	γ	σ^2
1,45	8	1,10	600	400

O LS-SVM fez previsões ligeiramente melhores do que o PLS. Observando a Figura 34 é possível notar que o PLS produziu 10 previsões com mais de 15% de erro relativo, enquanto que no modelo obtido através da utilização do LS-SVM houve seis amostras com erros de previsão acima de 15%.

Neste caso o maior número de amostras no conjunto de validação não alterou muito o desempenho dos modelos PLS e LS-SVM em relação à proposta anterior.

Tabela 29: Teste F para comparação dos modelos obtidos na proposta 4 para determinação do teor de gordura em ração para cães.

$$\frac{\mathbf{F\ calculado}}{\mathbf{F\ crítico = 1,84}}$$
$$\frac{1.74}{\mathbf{F\ crítico = 1,84}}$$

Observando-se a Tabela 29 é possível confirmar que os modelos PLS e LS-SVM não apresentaram diferenças significativas de previsão entre si.

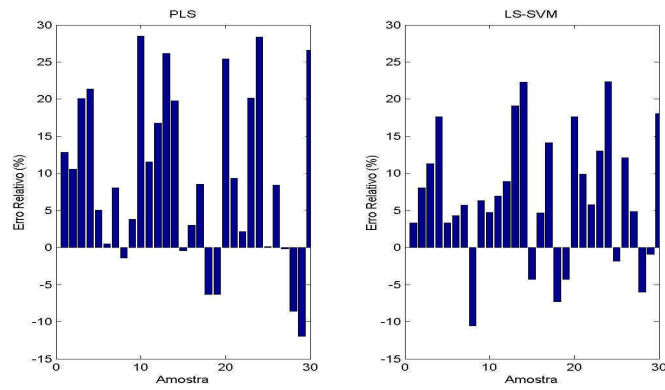


Figura 34: Erros relativos na previsão do teor de gordura em amostras de ração para cães obtidos nos modelos construídos para a proposta 4.

Conclusões

5. Conclusões

Os resultados obtidos mostraram que a utilização da espectroscopia na região do infravermelho próximo, combinada com a calibração multivariada é bastante eficiente na construção de modelos quantitativos para a determinação do teor de álcool em cachaça e para determinar teor de proteína e gordura em ração para cães, sendo um excelente método alternativo para a determinação destes fatores, já que é um método rápido, não destrutivo e que não necessita do processamento das amostras, o que faz com que não haja necessidade do uso de reagentes e de tratamento de resíduos.

Observou-se também que a utilização do LS-SVM para transferência de calibração é um procedimento possível, trazendo, na grande maioria dos casos resultados melhores do que outras formas de transferência de calibração tradicionais.

Nas amostras de cachaça, a determinação de etanol a diferentes temperaturas foi mais eficaz através do LS-SVM em três das quatro propostas apresentadas, sendo a melhor performance aquela em que todos espectros obtidos em todas as temperaturas foram utilizados para calibração. A proposta 3 foi aquela em que LS-SVM não teve uma boa performance. Isto pode ser explicado pela pouca quantidade de amostras no conjunto de calibração nesta proposta.

Uma possível aplicação seria o monitoramento on-line utilizando espectroscopia no infravermelho próximo do teor alcoólico da cachaça durante a produção da mesma, sem ter necessidade de se controlar a temperatura das amostras.

Para o caso da determinação do teor de proteína e gordura em ração para cães a diferentes granulometrias o LS-SVM também se mostrou como a melhor alternativa em três das quatro propostas apresentadas. Assim como na aplicação anterior, a proposta onde se obteve os melhores resultados foi a 2 através do LS-SVM, tanto para determinação do teor de proteína quanto para determinação do teor de gordura, e a proposta com os piores resultados foi a 3.

A utilização dessas duas últimas aplicações poderia tornar a determinação desses parâmetros (teor de proteína e gordura) mais práticas para a indústria de rações, já que são métodos rápidos, que não requerem utilização de reagentes e

tratamento de resíduos e não seria necessário se padronizar o tamanho de partículas antes da obtenção dos espectros.

Referências Bibliográficas

6. Referências Bibliográficas

1. Burns, D.A., Ciurczak, E.W., **'Handbook of near-infrared analysis'**, Marcel Dekker, New York, (1992).
2. Skoog, D.A., **'Princípios de análise instrumental'**; 5ªEdição, Bookman, Porto Alegre (2002).
3. Miller, J.N., Miller, J.C., **'Statistics and chemometrics for analytical chemistry'**,: 4nd Edition, Prentice Hall,Chinchester (2000).
4. Vandegniste, B.G., Massat, D.L., Buydens, L.M.C., Jong,S.,Lewi, P.J.,Verbeque, J.S., **'Handbook of chemometrics and qualimetrics: part b.'**, Amsterdam: Elsevier (1998).
5. Massie, D.R., Norris, K.H., **'Spectral reflectance and transmittance properties of grain in the visible and near infrared'** Trans. Am. Soc. Agric. Eng., 8 (1965) 568.
6. Norris, K.H., Hart, J.R., **'Principles and methods of measuring moisture in liquids and solids'**, Vol 4, Reinhold, New York (1965).
7. Osborne, B.G.,Fearn, T., Hindle, P.H., **'Practical NIR spectroscopy with applications in food and beverage analysis'**, Longman Scientific and Technical, UK, (1993).
8. Massart, B.G.M., Vandegniste, S., Deming, S.N., **'Chemometrics: a text book'**, Amsterdam: Elsevier (1988).
9. Davies, A. M. C., **'Subdivisions of the infrared region'**, Appl. Spectrosc., 44 (1990) 14A
10. Libnau, F.O, Kvalheim, O.M, Christy, A, Toft, J., **'Spectra of water in the near- and mid-infrared region'** Vib. Spectrosc.,7(1994) 243.
11. Finch, J.N., Lippincott, E.R., **'Hydrogen bond systems: temperature dependence of OH frequency shifts and OH band intensities'**, J. Chem. Phys.,24 (1956) 908
12. Luck, W.A.P., **'Structure of water and aquous solution'**, Verlag Chemie, Weinhein, (1974).
13. Walrafen, G.E., **'Water a comprehensive treatise'**,Vol.1,Plenum Press, New York, (1972).

14. Mayes, M. D., Callis, J.B., '**A photodiode-array-based near-infrared spectrophotometer for the 600-1100 nm wavelength region**', Appl.Spectrosc., 43 (1989) 27.
15. Coleman, P.B., '**Practical sampling techniques for infrared analysis**', CRC Press, Boca Raton (1993).
16. Mackenzie, M.W., '**Advances in applied fourier transform infrared spectroscopy**', St. Edmundsbury Press, Great Britain (1988).
17. Brereton, R.G., '**Introduction to multivariate calibration in analytical chemistry**', Analyst, (2000) 2125.
18. Thomas, E.V., '**A primer on multivariate calibration**', Anal. Chem., 66 (1994) 795A
19. Geladi, P., Kowalski, B.R., '**Partial least square regression: a tutorial**', Anal. Chim. Acta, 185 (1986) 17.
20. Otto, M., '**Chemometrics**', wiley –VCH, New York, (1999).
21. Custódio, R., Andrade, J.C., Augusto, F., '**O ajuste de funções matemáticas a dados experimentais**', Química nova, 20 (1987) 219.
22. Martens, N., Naes, T., '**Multivariate calibration**', John Wiley & Sons, New York, (1989).
23. Wulfert, F., Kok, W.T., Noord, O.E. and Smilde, '**Correction of temperature-induced spectral variation by continuous piecewise direct standardisation**', Anal Chem 72 , 7 (2000) 1639.
24. Szalay, A., Antal, I., Zsigmond, Z., Marton, S., Erős, I., Regdon jr., G., Pintye-Hódi, K., '**Study on the relationship between particle size and near infrared diffuse reflectance spectroscopic data**', Part. Part. Syst. Charact 22 (2005) 219.
25. Bouveresse, E., Massart, D.L., '**Standardisation of near-infrared spectrometric instruments**', Vibr. Spectrosc., 11 (1996) 3.
26. Despagne, F., Massart, D.L., Jansen, M., Daalen, H.V., '**Intersite transfer of industrial calibration models**', Anal. Chim. Acta., 406 (2000) 233.
27. Lin, J., Lo, S.C., Brown, C.W., '**Calibration transfer from a scanning near-IR spectrophotometer to a FT-near-IR spectrophotometer**', Anal. Chim. Acta, 349 (1997) 263

28. Wang, Y., Veltkamp, D.J., Kowalski, B.R., **'Multivariate instrument standardization'** Anal Chem, 63 (1991) 2750.
29. Noord, O.E., **'Multivariate calibration standardization'**, Chemom. Int. Lab. Systems., 25 (1994) 85.
30. Bouveresse, E., Massart, D.L., **'Improvement of the piecewise direct standardisation procedure for the transfer of NIR spectra for multivariate calibration'**, Chemom. Int. Lab. Systems, 32 (1996) 201.
31. 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 (2007), no prelo.
32. Thissen, U., Üstün, B., Melssen, W.J., Buydens, L.M.C., **'Multivariate Calibration with least-squares support vector machines'** Anal. Chem. 76, (2004), 3099.
33. Üstün, B.; Melssen, W.J.; Buydens, L.M.C.; **'Facilitating the application of support vector regression by using a universal Pearson VII function based kernel'** Chemom. Int. Lab. Systems 81 (2006) 29.
34. Borin, A., Ferrão, M.F., Mello, C. Maretto, D.A., Poppi, R.J., **'Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk'**, Anal. Chim. Acta, 579 (2006) 25-32
35. Cogdill, R.P.; Dardenne, P.; **'Least-squares support vector machines for chemometrics: an introduction and evaluation'**, J.Near Infrar. Spectrosc. 12, 93-100 (2004)
- 36 Otto, M., J. Anal. Chem., 359 (1997) 123.
37. Yokaya, F., **'Fabricação de aguardente de cana'**, Fundação Tropical de Pesquisas e Tecnologia "André Tosello", Campinas (1995)
38. Workman, J.J., **'Review of chemometrics applied to spectroscopy: 1985-95 .2.'**, Appl. Spectrosc. Rev., 31 (1996) 251.
39. Smilde, A.K., Kok, W.T., Wulfert, F., **'Influence of temperature on vibrational spectra and consequences for the predictive ability of multivariate models'**, Anal. Chem., 70 (1998) 1761.
40. Kennard, R. W.; Stone, L. A.; **'Computer aided design of experiments'**, Technometrics, 11 (1969), 137

41. Wise, B.M., Gallagher, N.B., Bro, R., Shaver, J.M., Windig, W., Koch, R.S., **'PLS Toolbox 3.5 for use with Matlab™'**, Eigenvector Research Inc., Manson (2005).
42. Suykens, J.A.K., van Gestel, T., de Brabanter, J., de Moor, B., Vanderwalle, J., **'Least-Squares Support Vector Machines'**, World Scientific, Singapore (2002).
43. Carciofi, A. C.; Vasconcellos, R.S. Borges, N.C.; Moro, J.V.; Fraga, V. O. ; **'Composição nutricional e avaliação de rótulo de rações secas para cães comercializadas em Jaboticabal-SP'** Arq. Bras. Méd. Vet. Zootec., v.58, n.3, (2006) 421
44. José Eduardo Butolo, **'Qualidade de Ingredientes na Alimentação Animal'**, OESP Gráfica S/A, Campinas, (2002)
45. Sodano, C.S., **'Animal feeds and pet foods: recent developments'**, Noyes Data Corporation, New Jersey. (1979)
46. AACC Method 46-16 **'Approved Methods of the American Association of Cereal Chemists'**, 9th Edition, The Association, St. Paul (1995).
47. Método 52, **'Compêndio brasileiro de alimentação animal'** Guia de métodos Analíticos. Sindirações, São Paulo. (2005) 120