

Universidade Estadual de Campinas

Instituto de Química

LAQQA - Laboratório de Quimiometria em Química Analítica

“Aplicações de Máquinas de vetores de suporte por mínimos quadrados (LS-SVM) na quantificação de parâmetros de qualidade de matrizes lácteas”

Tese de Doutorado

**Alessandra Borin**

**Orientador:** Prof. Dr. Ronei Jesus Poppi

Campinas, Julho de 2007.

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO INSTITUTO DE  
QUÍMICA DA UNICAMP

B645a	<p>Borin, Alessandra.</p> <p>Aplicações de máquinas de vetores de suporte por mínimos quadrados (LS-SVM) na quantificação de parâmetros de qualidade de matrizes lácteas / Alessandra Borin. -- Campinas, SP: [s.n], 2007.</p> <p>Orientador: Ronei Jesus Poppi.</p> <p>Tese - Universidade Estadual de Campinas, Instituto de Química.</p> <p>1. Leite. 2. Leite humano. 3. Infravermelho. 4. Máquina de vetores de suporte. I. Poppi, Ronei Jesus. II. Universidade Estadual de Campinas. Instituto de Química. III. Título.</p>
-------	--

**Título em inglês:** Application to least square support vector machines to quantification of quality parameters in milky matrices

**Palavras-chaves em inglês:** Milk, Human milk, Infrared, Support vector machine

**Área de concentração:** Química Analítica

**Titulação:** Doutor em Ciências

**Banca examinadora:** Ronei Jesus Poppi (orientador), Adriana Vitorino Rossi (UNICAMP-IQ), Maria Izabel Maretti Silveira Bueno (UNICAMP-IQ), Edenir Rodrigues Pereira Filho (UFSCAR), Marco Flôres Ferrão (UNISC)

**Data de defesa:** 13/07/2007

*Dedico este trabalho aos meus pais,  
Mauro e Fátima, por me darem o melhor deles e pelas constantes orações que me protegem sempre.  
E a Renato, meu amor, que encontrei durante a realização desta tese.*

## **AGRADECIMENTOS**

- Ao prof. Dr. Ronei Jesus Poppi pela orientação sempre presente e generosa, pela paciência serena em todos estes anos.
- Aos membros do grupo LAQQA pela amizade e pelos bons momentos: Renato, Gilmare, Danilo, Jez, Patrícia, Paulo Henrique, Luiz, Waldomiro, Marcello, Luciana e Joana e outras grandes amizades de outros laboratórios: Rafaela, Renata, Lílian, Fabíola, Tatiana e Lívia.
- A Dra. Cláudia Maria Monteiro Sampaio, médica responsável pelo Banco de Leite Humano da Maternidade Municipal de Campinas pela atenção e acompanhamento na coleta das amostras de Leite Humano.
- Ao prof. Dr. Nelson Duran e Lívia Cordi do laboratório de Química Biológica do Instituto de Química da UNICAMP pela amizade e parceria no nosso trabalho.
- Ao prof. Dr. Marco F. Ferrão (UNISC) pela amizade e fundamental participação neste trabalho.
- À Luiz Carlos M. Pataca e Danilo A. Maretto pela amizade e parceria neste trabalho.
- Ao prof. Dr. César Mello (UNIFRAN) pela amizade e pela participação neste trabalho.
- A Maternidade Municipal de Campinas e os funcionários do Banco de Leite Humano.
- A Empresa Tangará Laticínios pelo fornecimento de amostras.
- À técnica do Instituto de Química, Cláudia Martelli pela amizade e pelo auxílio nas análises no infravermelho.

- Ao órgão financiador CNPq que permitiu a realização deste projeto.
- Estendo meus sinceros agradecimentos a todos meus queridos amigos, funcionários e professores do Instituto de Química da UNICAMP, e todas as pessoas que direta ou indiretamente contribuíram para a realização deste trabalho e na conclusão de mais uma etapa na minha vida. Muito Obrigada!

## CURRÍCULO

ALESSANDRA BORIN

Data de nascimento: 26/05/78

Naturalidade: Valinhos - SP

### Formação Acadêmica

---

#### Pós Graduação

- Doutora em Ciências - Área de concentração: Química Analítica – UNICAMP – Jul/2007
- Mestre em Química Analítica – UNICAMP – Fev/ 2003

#### Graduação

- Bacharelado em Química Tecnológica - UNICAMP – Dez/ 2001
- Licenciatura em Química – UNICAMP – Dez/ 2000

#### Ensino Médio

- Magistério – E.E.S.G. Prof Cyro de Barros Resende – Dez/ 1995

### Experiência em Ensino Superior

---

- Professora responsável da disciplina ST108C – Química Aplicada  
Curso: Tecnologia de Saneamento Ambiental - CESET / Limeira – UNICAMP  
Período: 1ºSemestre/ 2007.
- Monitora da disciplina QA582 – Química Analítica Instrumental I  
Curso: Química Tecnológica - Instituto de Química – UNICAMP  
Período: 1º Semestre/ 2006.
- Auxiliar Didático da disciplina QG102-Química Geral Experimental I  
Curso: Engenharia Química - Instituto de Química – UNICAMP  
Período: 1ºSemestre/2001.

### Prêmios

---

- Projeto de Mestrado premiado pela CTPetro – processo nº 460261/01-0.

- Prêmio Lavoisier - CRQ IV de melhor desempenho no curso de Química Tecnológica no período de 1998- 2001.

### **Artigos Científicos**

---

- Borin A, Ferrão MF, MelloC, Cordi L, Pataca LM, Poppi RJ, Quantification of Lactobacillus in fermented milk by multivariate image analysis and least squares support vector machines, *Analytical and Bioanalytical Chemistry*, 387, **2007**, 1105.
- Ferrão MF, MelloC, Borin A, Maretto DA, Poppi RJ, 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*, 30 (4), **2007**, 852 .
- Borin A, Maretto DA, Ferrão MF, Mello C, Poppi RJ, Least-Squares Support Vector Machines and Near Infrared Spectroscopy for Quantification of Common Adulterants in Powdered Milk, *Analytica Chimica Acta*, 579, **2006**, 25.
- Borin A, Poppi, R J, Application of Mid infrared spectroscopy and i-PLS in quantification of contaminants in lubricating oil , *Vibrational spectroscopy*, 37, **2005**, 27
- Borin A, Poppi, RJ, Multivariate quality control of lubricating oils using Fourier transform infrared spectroscopy , *Journal Brazilian Chemical Society*, 15(4), **2004**, 570.
- Fava A R, Estudo aperfeiçoa controle de qualidade de lubrificantes, *Jornal da UNICAMP*, 211, Maio **2003**.

### **Capítulo de Livro**

---

- Borin A, Barbieri C V, Rovere G M, Butin M R, Lima M L, “Cassino Pyxis” em Cardieri T, “Como nossa sociedade realmente funciona?” Cultrix-Amaná-key, São Paulo, **2007**.

## Resumo

Autor: **Alessandra Borin**

Orientador: **Dr. Ronei Jesus Poppi**

Neste trabalho foram comparados modelos desenvolvidos utilizando as máquinas de vetores de suporte por mínimos quadrados (LS-SVM) e regressão por mínimos quadrados parciais (PLS) para três aplicações em determinação de parâmetros de qualidade em matrizes lácteas. Na primeira aplicação determinaram-se simultaneamente três adulterantes em leite em pó (amido, soro de leite e sacarose), empregando espectroscopia no infravermelho próximo por reflectância difusa. Em todos os modelos desenvolvidos, os resultados obtidos com LS-SVM foram superiores ao PLS, apresentando erros de previsão na faixa de 1,10% para amido, 1,10% para soro e 0,60% para sacarose, comparados aos erros de previsão obtidos pelo PLS: 2,40% para amido, 4,00% para soro e 3,00% para sacarose. Na segunda aplicação, quantificou-se *Lactobacillus* em leite fermentado a partir de imagens escaneadas de placas de crescimento microbiológico. A imagem foi convertida em histograma de cores, para ser usada no desenvolvimento dos modelos PLS e LS-SVM. O modelo desenvolvido pelo LS-SVM apresentou erros de previsão da ordem de 22,16 ufc e o PLS apresentou erros de previsão de 41,46 ufc. Na última aplicação, foi realizada a determinação do índice de gordura em leite humano, provenientes de diferentes doadoras, empregando a espectroscopia no infravermelho médio por reflectância total atenuada. Novamente o modelo LS-SVM apresentou melhores resultados em relação ao PLS, onde para a LS-SVM foram obtidos erros de previsão da ordem de 41,00 Kcal/L e, para o PLS, erros de previsão de 103,00 Kcal/L.

## Abstract

**Author:** Alessandra Borin

**Adviser:** Dr. Ronei Jesus Poppi

In this work calibration models by using Least-square support vector machine (LS-SVM) and Partial least square (PLS) regression were compared for three applications in determination of quality parameters in milky matrices. In the first application, three adulterants in powdered milk (starch, whey and sucrose) had been simultaneously determined employing near infrared spectroscopy by diffuse reflectance. In all developed models the results obtained with LS-SVM were superior to PLS, presenting prediction errors of 1,10% for starch, 1,10% for whey and 0,60% for sucrose compared with the errors obtained by the PLS of 2,40% for starch, 4,00% for whey and 3,00% for sucrose. In the second application, *Lactobacillus* in fermented milk were quantified from digital images of microbiological plates of growth. The images was converted into histogram of colors to be used in the development of PLS and LS-SVM models. The model developed for the LS-SVM presented errors of 22,16 cfu an the PLS presented prediction errors of 41,46 cfu. In the last application, the fat index in human milk was determined, proceeding from different mothers, employing Mid spectroscopy by attenuated total reflectance. Again, LS-SVM model presented better reulted than PLS, with prediction error 41,00 Kcal/L against PLS the prediction error of 103,00 Kcal/L.

## Índice

<b>Introdução e Objetivo</b>	<b>3</b>
------------------------------	----------

### **Capítulo 1 – Máquinas de vetores de suporte por mínimos quadrados (LS-SVM) empregadas para regressão multivariada**

1.1. A Calibração Multivariada	7
1.1.1. Regressão por Mínimos Quadrados Parciais – PLS	7
1.2. Calibração Não-linear	10
1.3. Máquinas de vetores de suporte (SVMs) e Máquinas de vetores de suporte por mínimos quadrados (LS-SVMs)	12
1.4. Regressão multivariada empregando Máquinas de vetores de suporte por mínimos quadrados (LS-SVM)	13
1.4.1. Função Núcleo (Kernel) para mapeamento do espaço de entrada para o espaço de características	13
1.4.2. Função Custo para regularização do modelo para um resultado global e único do modelo LS-SVM	15
1.4.3. Otimização dos parâmetros para o desenvolvimento do modelo LS-SVM com função núcleo RBF (rede de função de base radial)	19
1.5. Referências Bibliográficas	21

### **Capítulo 2 – Aplicação 1: Análise exploratória de diferentes tipos de leite em pó e determinação simultânea de adulterantes em leite em pó por Reflectância difusa no infravermelho próximo e Máquinas de vetores de suporte por mínimos quadrados (LS-SVM).**

2.1. Introdução	27
2.1.1. Fraude em Leite em pó	27
2.1.2. Metodologias analíticas de referência para determinação de Amido, Soro e Sacarose em leite	28
2.1.3. A técnica de Espectroscopia no Infravermelho próximo por Reflectância Difusa	30
2.2. Parte Experimental	33

2.2.1. Amostras para análise qualitativa de tipos de leite em pó	32
2.2.2. Amostras para a quantificação dos adulterantes em leite em pó	31
2.2.3. Instrumental	36
2.2.4. Pré-tratamento dos espectros	36
2.2.5. Escolha das amostras para calibração e previsão	37
2.2.6. Modelagem dos dados	37
2.2.7. Parâmetros para a avaliação dos modelos	38
2.3. Resultados e Discussões	39
2.3.1. Análise qualitativa de tipos de leite em pó	39
2.3.2. Análise qualitativa de leite em pó com diferentes adulterantes	40
2.3.3. Modelos para quantificação simultânea de adulterantes para conjunto de amostras contendo um e dois adulterantes	43
2.3.4. Modelos para quantificação simultânea de adulterantes para o conjunto de amostras puras (sem adulterante) e amostras contendo um e dois adulterantes	49
2.4. Referências Bibliográficas	55

**Capítulo 3 - Aplicação 2: Quantificação de Lactobacillus em leite fermentado por imagem escaneada e Máquinas de vetores de suporte por mínimos quadrados (LS-SVM).**

3.1. Introdução	59
3.1.1. Os Lactobacillus	59
3.1.2. Métodos padrões de contagem de bactérias	59
3.1.3. Interface entre Química Analítica e Microbiologia	60
3.1.4. Imagem colorida RGB	61
3.1.5. Convertendo cores de RGB para HSI	62
3.2. Parte Experimental	65
3.2.1. Preparação da amostra e diluições seriadas	65
3.2.3. Inoculação	66
3.2.3. Incubação	66
3.2.4. Obtenção e tratamento das imagens das placas	66
3.2.5. Escolha das amostras para calibração e previsão	72

3.2.6. Modelagem dos dados	73
3.2.7. Avaliação do desempenho dos modelos	73
3.3. Resultados e Discussões	73
3.3.1. Histograma e Contagem das colônias na placa de crescimento	73
3.3.2. Resultados dos modelos PLS e LS-SVM	74
3.3.3. Conversão de ufc por quadrado de cada quadrante para ufc/mL por placa	79
3.4. Referências Bibliográficas	81

**Capítulo 4 - Aplicação 3: Determinação do índice de gordura em leite humano por Reflectância Total Atenuada no Infravermelho Médio (ATR) e Máquinas de Vetores de suporte por mínimos quadrados (LS-SVM).**

4.1. Introdução	85
4.1.1. Leite Humano	85
4.1.2. A importância do Controle de qualidade nos Bancos de Leite Humano (BLHs) para Saúde pública	87
4.1.3. A Espectroscopia no Infravermelho médio (FTIR) por Reflectância Total Atenuada (ATR)	88
4.1.4. Comitê de Ética de Pesquisa	91
4.2. Parte Experimental	92
4.2.1. Amostras para a separação de tipos de leites fluidos por Análises de componentes principais (PCA)	92
4.2.2. Amostras e homogeneização	92
4.2.3. Técnica do Crematócrito – método de referência para determinação de índice de gordura em leite humano	93
4.2.4. Instrumental	94
4.2.5. Pré-tratamento dos espectros	94
4.2.6. Escolha das amostras para calibração e previsão	94
4.2.7. Modelagem dos dados	95
4.2.8. Avaliação do desempenho dos modelos	95
4.3. Resultados e Discussão	95

4.3.1. Separação de tipos de leites por Análise de componentes principais	95
4.3.2. Determinação do índice de gordura no leite humano	98
4.4. Referências Bibliográficas	103
<b>Conclusões</b>	<b>109</b>
<b>Perspectivas</b>	<b>113</b>
<b>Anexos</b>	<b>117</b>

## Índice de Figuras

Figura 1. Representação geométrica do PLS com uma variável latente modelando os dados de X (variáveis independentes) e Y (variáveis dependentes)	9
Figura 2. A) Relação linear entre x e y; B) relação com fraca não-linearidade; C) relação com forte não-linearidade; D) relação com aproximações locais	11
Figura 3. Representação do mapeamento realizado no espaço de entrada par ao espaço de características n-dimensional	14
Figura 4. Representações de funções (a): sobreajustada; (b) subajustada; (c) adequadamente ajustada	16
Figura 5. Representação dos erros mínimos em relação à reta de regressão	17
Figura 6. Representação de uma função complexa com presença de mínimos locais e de uma função convexa com uma única solução	18
Figura 7. Exemplo de superfície de resposta da otimização dos parâmetros $\sigma^2$ e $\gamma$ em função do erro do modelo	20
Figura 8. Representação do espalhamento da reflectância difusa em amostra sólida	30
Figura 9. Representação do esquema de uma esfera de integração	31
Figura 10. Superfície gerada pela LS-SVM para encontrar o ponto mínimo entre os parâmetros $\sigma^2$ e $\gamma$	37
Figura 11. Espectros no Infravermelho próximo (NIR) de amostras de diferentes tipos de leite: A- sem o pré-tratamento com MSC; B- com pré-tratamento MSC	38
Figura 12. Gráfico de escores gerado pela Análise de componentes parciais (PCA) com a separação dos diferentes tipos de leites em pó	39
Figura 13. Espectro do leite em pó usado para adulteração e espectros dos contaminantes em pó: amido, soro de leite e sacarose	40
Figura 14. Espectros NIR das amostras de leite em pó puras e adulteradas	41
Figura 15. Gráfico dos escores do conjunto das 306 amostras constituído de leite em pó sem adulteração e adulterado com diferentes adulterantes	42
Figura 16. Gráfico com valores de RMSECV em função do número de variáveis latentes	43

Figura 17. Superfície de resposta para a otimização dos parâmetros $\sigma^2$ e $\gamma$ a serem aplicados no modelo LS-SVM	44
Figura 18. Gráficos de correlação para a calibração e previsão da % de amido adicionado como adulterante	46
Figura 19. Gráficos de correlação para a calibração e previsão da % de sacarose adicionado como adulterante	47
Figura 20. Gráficos de correlação para a calibração e previsão da % de soro de leite adicionado como adulterante	47
Figura 21. Gráficos com valores de RMSECV em função do número de variáveis latentes	48
Figura 22. Superfície de resposta para a otimização dos parâmetros $\sigma^2$ e $\gamma$ a serem aplicados no modelo LS-SVM	49
Figura 23. Gráficos de correlação para a calibração e previsão da % de amido adicionado como adulterante	51
Figura 24. Gráficos de correlação para a calibração e previsão da % de soro de leite adicionado como adulterante	51
Figura 25. Gráficos de correlação para a calibração e previsão da % de sacarose adicionado como adulterante	52
Figura 26. Representação de um tensor com matrizes de valores de R, G e de B	62
Figura 27. Representação gráfica das cores em valores de HSI, a partir de RGB	64
Figura 28. Valores RGB indicada pela flecha, referente à cor indicada neste exemplo	65
Figura 29. Imagem da placa de crescimento bacteriano na diluição $10^4$ ufc/mL	68
Figura 30. Imagem da placa de crescimento bacteriano na diluição $10^3$ ufc/mL com indicação das áreas escolhidas (quadrados) em cada quadrante da placa	69
Figura 31. Imagem do quadrado de um dos quadrantes da placa de crescimento na diluição $10^3$ ufc/ mL	70
Figura 32. Esquema de obtenção do histograma de cores a partir das imagens escaneadas das placas microbiológicas	72

Figura 33. Histograma de cores _____	74
Figura 34. Valores de RMSECV contra o número de variáveis latentes _____	75
Figura 35. Superfície de resposta da otimização dos parâmetros para modelagem LS-SVM _____	76
Figura 36. Gráfico dos valores de ufc de referência versus valores previstos pelo modelo PLS _____	77
Figura 37. Gráfico dos valores de ufc de referência versus valores previstos pelo modelo LS-SVM _____	78
Figura 38. Representação do acessório de Reflexão Total Atenuada para obtenção do espectro de amostra de leite humano _____	90
Figura 39. Espectros de 30 amostras de diferentes tipos de leite com tratamento MSC _____	96
Figura 40. Gráfico da separação dos grupos das 30 amostras de diferentes tipos de leite por Análise de componentes principais _____	97
Figura 41. Análise de componentes principais das 15 amostras de leites de origem animal _____	98
Figura 42. Espectros das 62 amostras de leite humano na região do infravermelho de 980 a 3000 $\text{cm}^{-1}$ com pré-tratamento MSC _____	99
Figura 43. Espectros de 38 amostras de leite humano na região do infravermelho de 1700 a 3000 $\text{cm}^{-1}$ _____	100
Figura 44. Superfície de otimização da LS-SVM para o conjunto de calibração do índice de gordura em leite humano _____	101
Figura 45. Conjunto de calibração LS-SVM para determinação do índice de gordura em leite humano _____	102

## Índice de Tabelas

Tabela 1. Diferentes funções núcleo para o mapeamento dos dados do espaço de entrada para o espaço de características _____	15
Tabela 2. Composição centesimal média do leite de vaca _____	26
Tabela 3. Amostras de tipos de leite em pó para análise qualitativa _____	32
Tabela 4. Planejamento para a adição de dois adulterantes no leite em pó ____	33
Tabela 5. Descrição dos conjuntos de amostras cujos espectros foram obtidos_	34
Tabela 6. Número de espectros empregados na calibração e previsão _____	34
Tabela 7. Número de espectros de calibração e validação para os modelos ____	35
Tabela 8. Comparação em termos de erros, entre os resultados de PLS e LS-SVM _____	45
Tabela 9. Comparação dos resultados entre os modelos PLS e LS-SVM com amostras sem adulterantes, com amostras com um e com dois adulterantes__	50
Tabela 10. Erro absoluto para a previsão das amostras de leite puro para os modelos PLS e LS-SVM _____	53
Tabela 11. Parâmetros para os modelos PLS e LS-SVM usando os quadrantes _____	76
Tabela 12. Valores do conjunto de referência e as previsões para os modelos LS-SVM e PLS _____	80
Tabela 13. Média de alguns macronutrientes na composição do leite humano __	86
Tabela 14. Resultados dos modelos PLS e LS-SVM para determinação do índice de gordura em leite humano _____	103

# **Introdução e Objetivos**

## Introdução e Objetivos

As máquinas de vetores de suporte por mínimos quadrados (LS-SVM) são modificações das máquinas de vetores de suporte (SVMs), que são muito utilizadas para classificação e, em alguns casos para regressão. As LS-SVM são bastante recentes na área da quimiometria, especialmente para aplicações de regressão multivariada. Baseado nisso, o objetivo principal deste trabalho foi a aplicação das LS-SVMs para o uso de regressão e comparação de seu desempenho em relação a regressão por mínimos quadrados parciais (PLS). No Capítulo 1 estão apresentados os conceitos básicos das máquinas de vetores de suporte por mínimos quadrados e sua importância, principalmente, para a modelagem de dados não-lineares.

As três aplicações para o desenvolvimento dos modelos LS-SVM e PLS são apresentados nos capítulos 2, 3 e 4. O Capítulo 2 apresenta uma proposta de metodologia, empregando a espectroscopia no infravermelho próximo por reflectância difusa, para num primeiro momento, a realização de uma análise exploratória de diferentes tipos de leite em pó e a análise exploratória para um leite em pó com três diferentes adulterantes (amido, soro de leite e sacarose). Posteriormente, utilizando as LS-SVM, para a quantificação simultânea dos adulterantes: amido, soro e sacarose em amostras de leite integral em pó.

O Capítulo 3 se remete a proposta de quantificação de *Lactobacillus* em leite fermentado por imagens obtidas por scanner. O scanner não necessita de controle de iluminação, é acessível e um instrumento multifuncional em um laboratório. As imagens obtidas foram de placas de crescimento microbiológico

que apresentam a quantidade de unidades formadoras de colônias (ufc) do microrganismo desejado. Estas imagens foram convertidas em histograma de cores, as quais foram usadas no desenvolvimento dos modelos PLS e LS-SVM para a quantificação das unidades formadoras de colônias em cada imagem da placa de crescimento.

A aplicação apresentada no Capítulo 4 foi motivada pela grande necessidade no desenvolvimento de metodologias analíticas mais rápidas e eficientes para o controle de qualidade de leite humano nos bancos de leite humano espalhadas pelo país. Assim, empregando a espectroscopia no infravermelho médio por reflectância total atenuada, primeiramente foi realizada a análise exploratória de diferentes tipos de leite fluidos e posteriormente a determinação do índice de gordura em amostras de leite humano, provenientes de diferentes doadoras.

# Capítulo 1

**MÁQUINAS DE VETORES DE SUPORTE POR MÍNIMOS QUADRADOS (LS-SVM) EMPREGADAS PARA REGRESSÃO MULTIVARIADA**

## **1.1. A Calibração Multivariada**

A calibração multivariada é um dos tópicos mais importantes dentro da quimiometria. Basicamente, a calibração multivariada é o desenvolvimento de um modelo matemático que relaciona variáveis independentes (como espectros) com um ou mais variáveis dependentes (como concentrações). Após o desenvolvimento do modelo e a partir da projeção de novas variáveis independentes (novas amostras), é possível prever os valores das variáveis dependentes das mesmas<sup>1</sup>.

Entre os métodos de calibração multivariada estão: os mínimos quadrados clássicos (CLS), a regressão linear múltipla (MLR), a regressão por componentes principais (PCR); porém, a mais comumente usada é a regressão por mínimos quadrados parciais (PLS)<sup>2</sup>. O PLS modela os dados multivariados por meio do desenvolvimento de uma regressão numa nova base, que é uma combinação linear das variáveis originais.

### **1.1.1. Regressão por Mínimos Quadrados Parciais – PLS**

Esta técnica multivariada tem sido utilizada para análises quantitativas e algumas vezes capaz de modelar variações não lineares. O cálculo do modelo previsto é realizado diretamente e não requer nem consumo de tempo iterativo nem parâmetros de aprendizagem escolhidos arbitrariamente ou arquiteturas em rede<sup>3</sup>.

A regressão PLS é um método de análises de fatores em que a matriz de espectros é decomposta em matrizes de variações do espectro (pesos) e a posição das amostras (scores). Os espectros originais podem ser considerados como combinações lineares das variações dos espectros (pesos) onde os scores representam a contribuição de cada.

No PLS, tanto a matriz das variáveis independentes  $\mathbf{X}$ , como a das variáveis dependentes  $\mathbf{Y}$  são representadas pela Análise de Componentes Principais<sup>3,4</sup>:

$$\mathbf{X} = \mathbf{TP}^t + \mathbf{E} \quad (\text{Eq.1})$$

$$\mathbf{Y} = \mathbf{UQ}^t + \mathbf{F} \quad (\text{Eq.2})$$

Uma relação entre os dois blocos pode ser realizada correlacionando-se os escores para cada componente de cada vez, utilizando-se um modelo linear:

$$\mathbf{u}_h = \mathbf{b}_n \mathbf{t}_h \quad (\text{Eq.3})$$

onde :

$$\mathbf{b}_n = \mathbf{u}_h \mathbf{t}_h^t / \mathbf{t}_h^t \mathbf{t}_h \quad (\text{Eq.4})$$

para cada  $h = 1, 2, \dots$  a componentes principais

Esse modelo, entretanto, não é o melhor possível. Isto porque a Análise de Componentes Principais é realizada em cada matriz separadamente, podendo resultar numa relação não muito satisfatória (não linear) entre os escores dos dois blocos. Deve-se buscar um modelo onde as matrizes de resíduos  $\mathbf{E}$  e  $\mathbf{F}$  sejam as menores possíveis e , ao mesmo tempo, conseguir uma relação linear entre  $\mathbf{t}$  e  $\mathbf{u}$ .

No PLS isto é realizado por uma leve mudança nos valores dos escores, de forma a produzir a melhor relação possível. Como pode ser notado, há um compromisso entre a habilidade dos componentes principais em descrever as amostras nos espaços individuais (modelagem dos blocos  $\mathbf{X}$  e  $\mathbf{Y}$ ), e o aumento na correlação entre  $\mathbf{t}$  e  $\mathbf{u}$ .

### Previsão

A parte fundamental da regressão é a verificação de sua validade. Ou seja, se o modelo desenvolvido é capaz de prever corretamente ou com pequena margem de erro os valores de novas amostras. Por exemplo, espectros de infravermelho de amostras desconhecidas são dispostos na forma de matriz ( $\mathbf{X}$  teste). Após isto, os dados passam por pré-tratamentos matemáticos assim como os dados de calibração.

O passo seguinte consiste em se obter o vetor de escores ( $t$  teste) e os resíduos ( $e$ ) pelo ajuste das matrizes dos novos espectros,  $\mathbf{X}_{teste}$  para  $p$  (pesos do bloco  $\mathbf{X}$  calculados na fase de treinamento):

$$x_{teste} = \sum_{a=1}^A T_a p_a^t + e \quad (\text{Eq.5})$$

para “A” – variáveis latentes na modelagem

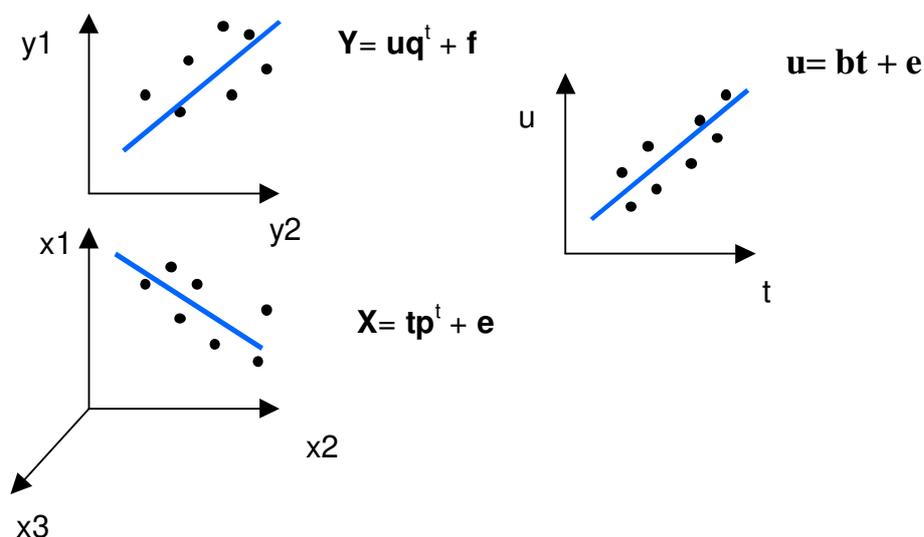
A partir dos coeficientes de regressão  $b$  calculados na fase de calibração, os escores  $\mathbf{T}$  teste produzem uma estimativa do vetor de concentrações, já que:

$$\mathbf{Y} = \sum_{a=1}^A \mathbf{t}_a b_a \mathbf{q}_a^t \quad (\text{Eq.6})$$

para “A” variáveis latentes.

Onde  $q_a$  (os pesos do bloco  $\mathbf{Y}$ ) foi calculado na fase de treinamento.

A figura 1 mostra uma representação geométrica do PLS com uma variável latente modelando os dados do bloco de variáveis independentes ( $\mathbf{X}$ ) e o das variáveis dependentes ( $\mathbf{Y}$ ).



**Figura 1.** Representação geométrica do PLS com uma variável latente modelando os dados de  $\mathbf{X}$  (variáveis independentes) e  $\mathbf{Y}$  (variáveis

independentes). Os dados de X são formados por três variáveis e os dados de Y por duas variáveis.

As vantagens do PLS são: sua simplicidade, rapidez e excelentes resultados para modelos lineares<sup>5</sup>. Em alguns casos, é possível desenvolver o modelo mesmo com presença de suaves não-linearidades<sup>6</sup>, empregando um maior número de variáveis latentes para o modelo; porém, este pode ficar sobreajustado e modelar amostras anômalas e ruídos.

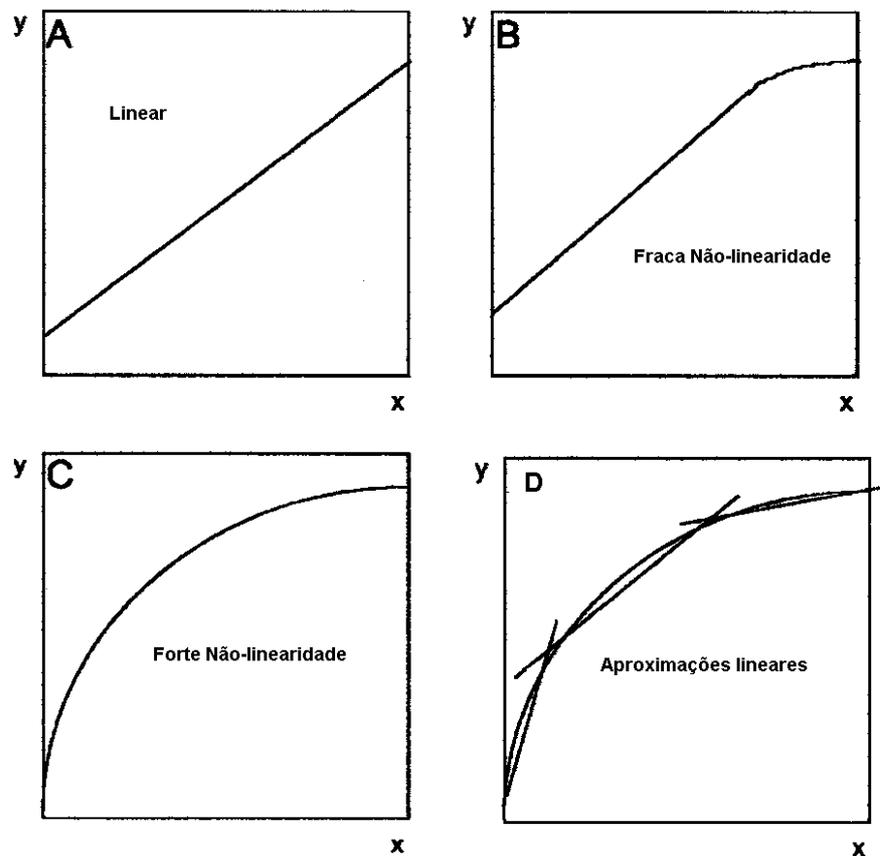
Outra estratégia para contornar o problema com estas suaves não-linearidades é o emprego de métodos de seleção de variáveis, os quais escolhem regiões com informações significativas e com comportamento linear para a modelagem. Entre as seleções de variáveis temos: PLS por intervalos (iPLS)<sup>7</sup>, eliminação de variáveis não-informativas (UVE-PLS)<sup>8</sup>, algoritmos genéticos<sup>9</sup> entre outros. Porém, o PLS falha se fortes não-linearidades estiverem presentes na relação entre variáveis dependentes e independentes.

## **1.2. Calibração Não Linear**

A Figura 1 apresenta possíveis relações entre variáveis independentes X e variáveis dependentes Y. Na figura 2a tem-se uma relação linear.

Uma fraca não linearidade significa que grande parte da relação entre as variáveis dependentes e independentes é linear e que apenas existe não-linearidade nos valores extremos das variáveis (Figura 2b). Uma forte não-linearidade é estabelecida quando em todo espaço amostral a relação não é linear (figura 2c). A Figura 2d, mostra que aproximações lineares podem ser utilizadas para dados não-lineares<sup>1</sup>.

Há diversas razões para um conjunto de dados possuir um perfil não-linear como, por exemplo: desvios da lei de Lambert-Beer devidos à qualidade da luz e homogeneidade da amostra; Não-linearidades dos detectores fotocondutivos; dispersão da luz nos instrumentos ópticos de NIR; não-linearidade na reflectância difusa; composição da matriz e não-linearidade no relacionamento entre a propriedade de interesse e sua medida instrumental.



**Figura 2.** A) Relação linear entre  $x$  e  $y$ ; B) relação com fraca não linearidade; C) relação com forte não linearidade; D) relação com aproximações locais.

Como observado na lista de possíveis não-linearidades, muitos métodos espectroscópicos são comumente usados para previsão de propriedades físicas e químicas. Porém, infelizmente, muitos deles não exibem uma dependência linear com o espectro. Particularmente, a espectroscopia em infravermelho próximo (NIR -near infrared) por reflectância difusa é bastante sujeita a não-linearidades devido às diferenças do efeito multiplicativo no caminho efetivo, aumentando com as diferenças nos tamanhos das partículas das amostras<sup>10</sup>.

Para modelar um conjunto de dados não-lineares são propostos certos tipos de regressão. O principal problema no tratamento destes tipos de conjuntos não-lineares é a forte tendência em desenvolver modelos sobre-ajustados<sup>11,12</sup>.

Como exemplos, têm-se o “PLS spline” que utiliza uma função “spline” (quadrática ou cúbica) para estabelecer a relação entre o bloco dos escores da matriz de variáveis independentes contra o bloco dos escores da matriz de variáveis dependentes. Nesse caso, as seções das polinomiais juntam-se em nós.

O PCR/PLS polinomial também apresenta uma função polinomial para ajustar os escores dos primeiros componentes principais/ variáveis latente à resposta. O maior problema do uso destes métodos, que empregam uma função polinomial, é encontrar o grau do polinômio que forneça o melhor ajuste<sup>13</sup>.

As redes neurais, que são métodos matemáticos e computacionais projetados para imitar o processamento de informação do cérebro humano, são boas alternativas para modelagem de dados não-lineares. Entre as diversas arquiteturas de redes, as mais usadas em calibração multivariada são a retropropagação de erros (back-propagation) e a redes com funções de base radial (RBF). 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 sobre-ajuste do modelo aos dados<sup>14</sup>. Por operar no espaço original dos dados, em que as não-linearidades presentes e a complexidade intrínseca do problema não são conhecidas *a priori*, os algoritmos de otimização para ajuste de parâmetros e as ferramentas estatísticas adotadas para seleção de modelos podem induzir modelos com baixa capacidade de generalização<sup>15</sup>.

### **1.3. Máquinas de vetores de suporte (SVMs) e as máquinas de vetores de suporte por mínimos quadrados (LS-SVMs)**

As máquinas de vetores de suporte (support vector machines, SVMs)<sup>16</sup> propostas por Vapnik<sup>17,18</sup> são máquinas de aprendizagem que podem ser usadas para classificação<sup>19,20,21,22</sup> e para regressão<sup>23,24,25</sup>.

As máquinas de vetores de suporte por mínimos quadrados (least -squares support vector machines, LS-SVMs)<sup>26, 27</sup> propostas por Suykens e colaboradores<sup>28</sup> são modificações das SVMs, que ao invés de usarem a programação quadrática usada nas SVMs (que são de grande complexidade matemática e despedem bastante tempo computacional) usam um sistema de equações lineares e a função custo por mínimos quadrados. Como as SVMs, porém ainda recentes, as LS-SVMs são empregadas tanto para classificação<sup>29</sup>, quanto para regressão<sup>30,31</sup>.

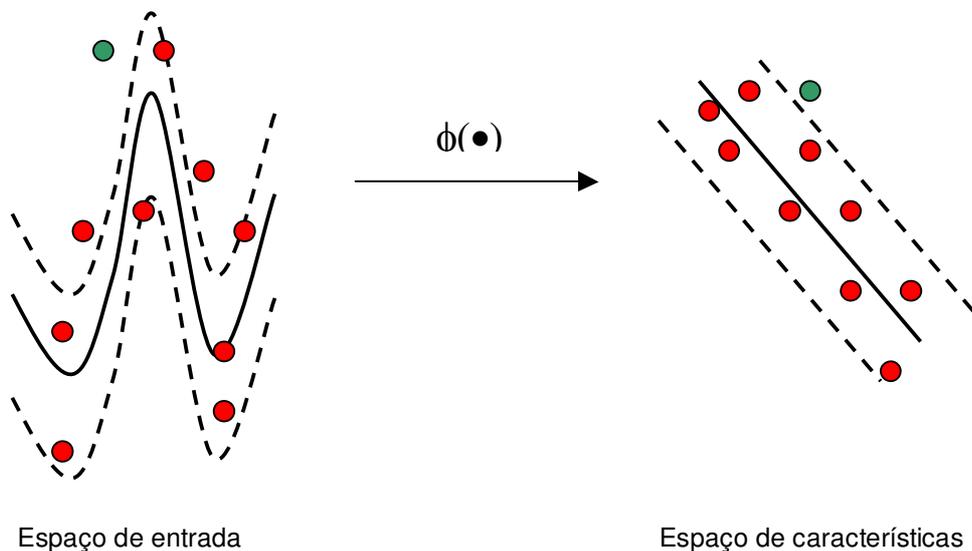
#### **1.4. Regressão multivariada empregando máquinas de vetores de suporte por mínimos quadrados (LS-SVM)**

Como mencionado as LS-SVMs correspondem a modificações das SVMs, através de uma função custo usando mínimos quadrados e da utilização de restrições de igualdade e soluções por equações lineares, proporcionando menor custo computacional em relação às SVMs, sem perda na qualidade das soluções, uma vez que os princípios que ambas se baseiam são os mesmos.

Porém, a desvantagem das LS-SVMs reside no fato de que **todos** os dados de treinamento são considerados **vetores de suporte**, diferentemente das SVMs que detectam apenas uma pequena fração deste conjunto durante a fase de treinamento.

##### **1.4.1. Função Núcleo (Kernel) para mapeamento do espaço de entrada para o espaço de características**

A etapa inicial da LS-SVM é mapear os dados **X** do espaço de entrada para um espaço de características de alta dimensão via um mapeamento linear ou não linear  $\varphi(\bullet)$  e então realizar a regressão neste espaço de característica de alta dimensão, como mostra a figura 3 .



**Figura 3.** Representação do mapeamento realizado no espaço de entrada para o espaço de características n-dimensional. Bolinhas vermelhas representam o conjunto de treinamento e a bolinha verde, a amostra a ser prevista.

Para mapear dados do espaço de entrada para o espaço de características, são introduzidas as funções núcleo (kernel). Dentre os diferentes núcleos, temos, para dados lineares, a função núcleo *Linear*, e para dados não-lineares temos: *Polinomial* (que manipula uma função polinomial, cujo grau pode ser definido durante os treinamentos), *Gaussiano* (corresponde a um espaço de características de dimensão infinita; a utilização desse núcleo permite que uma SVM apresente características de uma rede de função de base radial - RBF) e *Sigmoidal* (permite que a SVM tenha o comportamento semelhante ao de uma rede neural). A correta definição do núcleo e de seus respectivos parâmetros possui forte influência nos resultados obtidos por uma SVM.

As equações matemáticas referentes a cada núcleo são apresentadas na Tabela 1, onde podemos observar que se necessita a otimização dos parâmetros associados ao núcleo escolhido<sup>32</sup>.

**Tabela 1.** Diferentes funções núcleo para o mapeamento dos dados do espaço de entrada para o espaço de características.

Núcleo	Expressão	Parâmetros
Linear	$\mathbf{x}_i^T \cdot \mathbf{x}_j$	-
RBF	$e^{-\frac{\ \mathbf{x}_i - \mathbf{x}_j\ ^2}{2\sigma^2}}$	$\sigma^2$
Polinomial	$(\mathbf{x}_i^T \mathbf{x}_j + a)^b$	a, b
Sigmóide	$\text{Tanh}(\beta_0 \mathbf{x}_i^T \mathbf{x}_j + \beta_1)$	$\beta_0, \beta_1$

Dessa forma, assumindo um conjunto de treinamento :

$$(x_i, y_i), \dots, (x_N, y_N) \in \mathbb{R}^n \times \mathbb{R}$$

A função  $\varphi(x)$  é empregada para mapear o espaço da entrada original  $\mathbb{R}^n$  para um espaço de características de alta dimensão:

$$\varphi(x) = (\varphi(x_1), \varphi(x_2), \dots, \varphi(x_N)) \quad (\text{Eq. 7})$$

neste espaço de características de alta dimensão, é que ocorrerá a regressão linear:

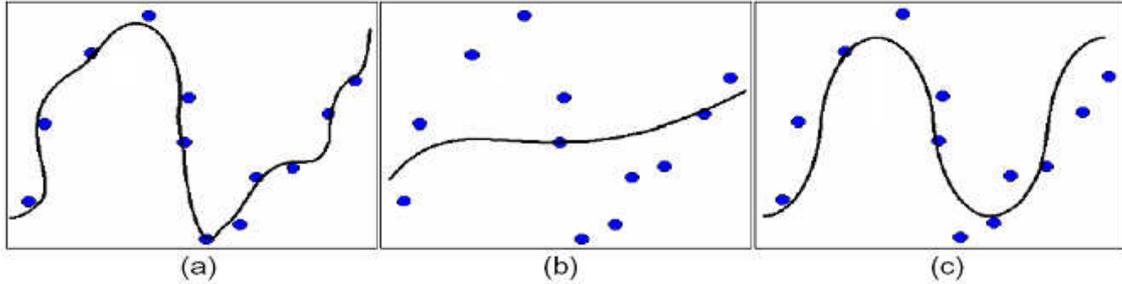
$$y(x_i) = \mathbf{w}^T \varphi(x_i) + b \quad (\text{Eq. 8})$$

onde  $\mathbf{w}$  é o vetor de pesos e  $b$  coeficiente linear da regressão.

#### 1.4.2. Função Custo para regularização do modelo para um resultado global e único do modelo LS-SVM

Para se obter um modelo com boa capacidade de generalização necessita-se evitar o sobre-ajuste deste modelo. Na Figura 4 estão representadas funções

com perfil sobre-ajustada (Figura 4a), sub-ajustada (4b) e adequadamente ajustada (3c).



**Figura 4.** Representações de função (a) sobre-ajustada; (b) sub-ajustada e (c) adequadamente ajustada.

Para evitar o sobre-ajuste do modelo na LS- SVM, é necessário minimizar uma função custo (C) contendo uma penalização dos erros de regressão, conforme equação 9:

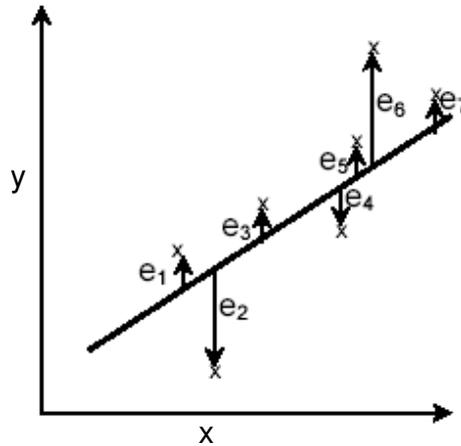
$$C = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^N e_i^2 \quad (\text{Eq.9})$$

onde  $\mathbf{w}$  é o vetor de pesos,  $\gamma$  é o parâmetro que penaliza erros altos e é otimizado pelo usuário,  $\mathbf{e}$  é o vetor de mínimos quadrados.

De modo que a restrição à função custo é:

$$y_i = w^T \phi(x_i) + b + e_i \quad (\text{Eq.10})$$

para todos os valores de  $i = 1, \dots, N$ ; em que  $\phi$  denota a função de linearização dos dados,  $\mathbf{w}$  os pesos e  $e_i$  são os erros mínimos em relação a reta de regressão (Figura 5).



**Figura 5.** Representação dos erros mínimos em relação à reta de regressão.

A primeira parte desta função custo é um decaimento utilizado para regularizar os pesos e penalizar aqueles que são elevados. Devido a esta regularização, os pesos convergem para valores menores. Este procedimento é necessário, pois pesos elevados causam excessiva variância do modelo, deteriorando a capacidade de generalização das LS-SVM e podendo ocorrer sobre-ajustes. A segunda parte da equação 6 representa os erros de regressão para todos os dados de treinamento. O parâmetro de regularização gama ( $\gamma$ ), o qual deve ser otimizado pelo usuário, fornece o peso relativo da segunda parte da equação 6, quando comparado com a primeira parte da mesma. A restrição de igualdade imposta pela equação 9 fornece a definição do erro de regressão.

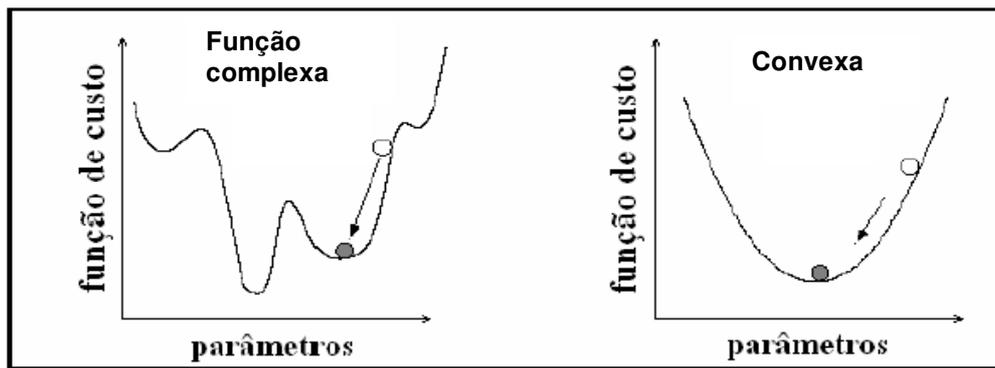
Analisando a equação 9 e sua restrição dada pela equação 10 é possível concluir que temos um típico problema de otimização convexa o qual pode ser resolvido usando o método dos multiplicadores de Lagrange<sup>33</sup>, como segue:

$$L = \frac{1}{2} \|w\|^2 + \gamma \sum_{i=1}^N e_i^2 - \sum_{i=1}^N \alpha_i \{w^T \phi(x_i) + b + e_i - y_i\} \quad (\text{Eq.11})$$

em que  $y_i = \begin{bmatrix} y \\ y \\ \vdots \\ y_N \end{bmatrix}$ ,  $e_i = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}$  e  $\alpha_i = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix}$ .

Onde  $y_i$  é o vetor de respostas (saída);  $e_i$  é o vetor de erros;  $\alpha_i$  é o vetor de multiplicadores de Lagrange que são **os vetores de suporte**. Na LS-SVMs, os vetores de suporte são valores calculados a partir de todos os dados de treinamento.

A otimização convexa da função custo possui a grande vantagem de apresentar apenas um mínimo global, ao invés do que ocorre com outros algoritmos, como as redes neurais, por exemplo, que possuem diferentes mínimos locais (Figura 6).



**Figura 6.** Representação de uma função complexa com presença de mínimos locais e de uma função convexa com uma única solução.

O ponto ótimo da otimização convexa é obtido fazendo-se as derivadas parciais  $\frac{\partial L(w, b, e_i, \alpha_i)}{\partial w}$ ,  $\frac{\partial L(w, b, e_i, \alpha_i)}{\partial b}$ ,  $\frac{\partial L(w, b, e_i, \alpha_i)}{\partial e_i}$ ,  $\frac{\partial L(w, b, e_i, \alpha_i)}{\partial \alpha_i}$  e em seguida igualando todas a zero, como segue:

$$\frac{\partial L(w, b, e, \alpha)}{\partial w} = w - \sum_{i=1}^N \alpha_i \phi(x_i) = 0 \therefore w = \sum_{i=1}^N \alpha_i \phi(x_i) \quad (\text{Eq.12})$$

Pela equação 12, percebe-se que os pesos ( $w$ ) podem ser escritos como uma combinação linear dos multiplicadores de Lagrange (que são os vetores de suporte) correspondentes aos dados de treinamento ( $x_i$ ).

$$\frac{\partial L(w, b, e, \alpha)}{\partial e} = \sum_{i=1}^N \gamma e - \alpha = 0 \therefore \alpha = \gamma e \quad (\text{Eq.13})$$

Pela Equação 13, observa-se que a maioria dos multiplicadores de Lagrange (os vetores de suporte) não possui valor zero, indicando que a maioria dos dados de treinamento contribui para a solução.

Colocando o resultado da equação 12 na reta de regressão original ( $y = w^T \phi(x) + b$ ), o seguinte resultado é obtido:

$$y = \sum_{i=1}^N \alpha_i \phi(x_i)^T \phi(x) + b \quad (\text{Eq.14})$$

Estes vetores  $\alpha$  são obtidos resolvendo-se o sistema linear de equações, seguindo as condições de Karush – Kuhn – Tucker (KKT)<sup>34</sup>, resolvendo-se pela forma matricial da equação 15.

$$\begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & K(x_1, x_1) + 1/\gamma & \cdots & K(x_1, x_N) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & K(x_N, x_1) & \cdots & K(x_N, x_N) + 1/\gamma \end{bmatrix} \begin{bmatrix} b \\ \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} = \begin{bmatrix} 0 \\ y_1 \\ \vdots \\ y_N \end{bmatrix} \quad (\text{Eq. 15})$$

Resultando no modelo LS-SVM:

$$y(x) = \sum_{i=1}^N \alpha_i K(x, x_i) + b \quad (\text{Eq. 16})$$

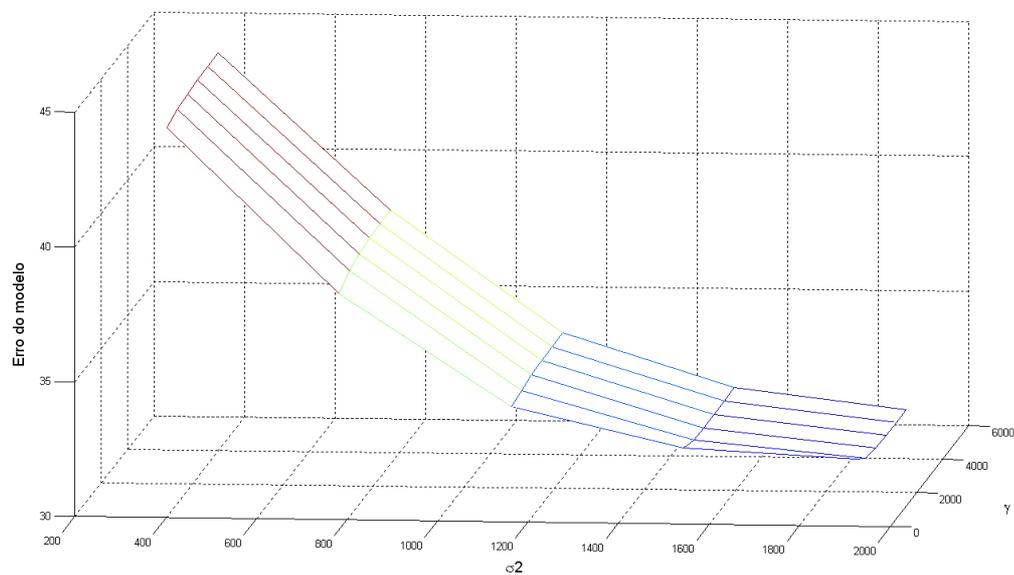
onde  $\alpha_i$  e  $b$  são as soluções do sistema linear, lembrando que, pela solução das equações lineares, nas condições de KKT, se obtém um única resposta<sup>35</sup>.

### 1.4.3. Otimização dos parâmetros para o desenvolvimento do modelo LS-SVM com função núcleo RBF (rede de função de base radial)

Neste trabalho, para o desenvolvimento dos modelos LS-SVM, foi empregada a RBF como função núcleo, como comentado anteriormente. Ela se baseia numa

simples função gaussiana,  $\exp\left(-\|x_i - x_j\|^2 / 2\sigma^2\right)$ , em que  $\sigma^2$  é a largura da função gaussiana, a qual deve ser otimizada pelo usuário.

A otimização do parâmetro da função custo  $\gamma$  e da RBF,  $\sigma^2$ , é realizada pela superfície de resposta de otimização destes parâmetros simultaneamente, em função do erro do modelo como mostra a Figura 7.



**Figura 7.** Exemplo da superfície de resposta da otimização dos parâmetros  $\sigma^2$  e  $\gamma$  em função do erro do modelo.

Observando-se a superfície de resposta da otimização dos parâmetros sigma quadrado ( $\sigma^2$ ) e do gama ( $\gamma$ ), encontra-se o mínimo da superfície, correspondendo ao mínimo do valor do erro do modelo, que nesta tese foi o erro médio quadrático da validação cruzada (RMSECV).

Nas aplicações deste trabalho foram comparados modelos de regressão desenvolvidos por PLS e por LS-SVM, pode-se observar dois pontos importantes da teoria da LS-SVM em relação ao PLS: a LS-SVM onde possibilita o

desenvolvimento de um modelo com um resultado único (pela solução de equações lineares) e global (pela função custo por otimização convexa), além de conseguir desenvolver modelos com dados não-lineares, escolhendo assim uma função núcleo apropriada para o conjunto de dados.

## 1.5. Referências Bibliográficas

- <sup>1</sup> Geladi P, Some recent trends in the calibration literature, *Chemometrics Intelligent Laboratory Systems*, 60, **2002**, 211-224.
- <sup>2</sup> Vandeginste B G M, Massart D L, Buydens L M C, de Jong S, Lewi P J, Smeyers-Verbeke J, *Handbook of Chemometrics and Qualimetrics: Part B*, Elsevier, New York, **1998**, 349-381.
- <sup>3</sup> Geladi P, Kowalski B R, Partial least square regression: a tutorial, *Analytica Chimica Acta*, 185, **1986**, 1-17.
- <sup>4</sup> Luinge H J, Van der Maas J H, Visser T, Partial least squares regression as a multivariate tool for the interpretation of infrared spectra, *Chemometrics and Intelligent Laboratory Systems*, 28, **1995**, 129-138
- <sup>5</sup> Martens H, Naes T, *Multivariate Calibration*, John&Wiley, New York, **1996**.
- <sup>6</sup> Centner V, Noord O E, Massart D L, Detection of nonlinearity in multivariate calibration, *Analytica Chimica Acta*, 376, **1998**, 153-168.
- <sup>7</sup> Borin A, Poppi R J, Application of mid infrared spectroscopy and iPLS for the quantification of contaminants in lubricating oil, *Vibrational Spectroscopy*, 37, **2005**, 27-32
- <sup>8</sup> Centner V, Massart D L, de Noord O E, Jong S, Vandeginste B G M, Sterna C, Elimination of uninformative variables for multivariate calibration, *Analytical Chemistry*, 68, **1996**, 3851 - 3858.
- <sup>9</sup> Jouan-Rimbaud D, Massart D L, Leardi R, Genetic algorithm as a tool for wavelength selection in multivariate calibration, *Analytical Chemistry*, 67, **1995**, 4295 - 4301.
- <sup>10</sup> Blanco M, Coello J, Iturriaga H, Maspocch S, Pagès J, Calibration in non-linear near infrared reflectance spectroscopy a comparison of several methods, *Analytica Chimica Acta*, 384, **1999**, 207-214.
- <sup>11</sup> Despagne F, Massart D L, Developmente of a robust calibration model for nonlinear in-line process data, *Analytical Chemistry*, 72, **2000**, 1657-1665.
- <sup>12</sup> Barzilai I L, Schechter I, Spectroscopic prediction of nonlinear properties by principal component regression, *Analytica Chimica Acta*, 348, **1997**, 345-356.
- <sup>13</sup> Dieterle F, Busche S, Gauglitz G, Different approaches to multivariate calibration of nonlinear sensor data, *Analytical and Bioanalytical Chemistry*, 380, **2004**, 383-396.
- <sup>14</sup> de Cerqueira E O, Andrade J C, Poppi R J, Redes Neurais e suas aplicações em calibração multivariada, *Química Nova*, 24(6), **2001**, 864 - 873.

- 
- <sup>15</sup> Lima C A M, *Comitê de Máquinas: Uma Abordagem Unificada Empregando Máquinas de Vetores-Suporte*. Tese de Doutorado. Faculdade de Engenharia Elétrica. Universidade Estadual de Campinas, Brasil, **2004**.
- <sup>16</sup> Noble S W, What is a support vector machine?, *Nature Biotechnology*, 24(12), **2006**, 1565 – 1567.
- <sup>17</sup> Vapnik V N, *Statistical Learning Theory*, Wiley, New York, **1998**.
- <sup>18</sup> Vapnik V N, *The Nature of Statistical Learning Theory*, Springer, New York, **1995**.
- <sup>19</sup> Trafalis B T, Gilbert R C, Robust classification and regression using support vector machines, *European Journal of Operational Research*, 173, **2006**, 893–909.
- <sup>20</sup> Pierna J A F, Volery P, Besson R, Baeten V, Dardenne P, Classification of Modified Starches by Fourier Transform Infrared Spectroscopy Using Support Vector Machines, *Journal of Agricultural Food Chemistry*, 53, **2005**, 6581-6585.
- <sup>21</sup> Brudzewski K, Osowski S, Markiewicz T, Classification of milk by means of an electronic nose and SVM neural network, *Sensors and Actuators B*, 98, **2004**, 291–298.
- <sup>22</sup> Lima A R G, Máquinas de Vetores Suporte na Classificação de Impressões Digitais. Dissertação de mestrado. Departamento de Computação. Universidade Federal do Ceará, **2002**.
- <sup>23</sup> Ustun B, Melssena W J, Oudenhuizenb M, Buydens L M C, Determination of optimal support vector regression parameters by genetic algorithms and simplex optimization, *Analytica Chimica Acta*, 544, **2005**, 292–305.
- <sup>24</sup> Yanga S, Lua W, Chena N, Hu Q, Support vector regression based QSPR for the prediction of some physicochemical properties of alkyl benzenes, *Journal of Molecular Structure: THEOCHEM*, 719, **2005**, 119–127.
- <sup>25</sup> Xu L, Wencong L, Shengli J, Yawei L, Nianyi C, Support vector regression applied to materials optimization of sialon ceramics, *Chemometrics and Intelligent Laboratory Systems*, 82, **2006**, 8-14.
- <sup>26</sup> Codgill R P, Dardenne P, Least-square support vector machines for chemometrics: an introduction and evaluation, *Journal of Near Infrared Spectroscopy*, 12, **2004**, 93-100.
- <sup>27</sup> Huo H B, Zhu X H, Cão G Y, Nonlinear modeling of a SOFC stack based on a least square support vector machine, *Journal of Power Sources*, 162, **2006**, 1220-1225.
- <sup>28</sup> Suykens J A K, van Gestel T, de Brabanter J, de Moor B, Vandewalle J, *Least-Squares Support Vector Machines*, World Scientific, Singapore, **2002**.
- <sup>29</sup> Suykens J A K, Vandewalle J, Least-Squares Support Vector Machine Classifiers, *Neural Processing Letters*, 9, **1999**, 293–300.
- <sup>30</sup> Thissen U, Ustun B, Melssen W J, Buydens L M C, Multivariate calibration with Least-squares support vector machine, *Analytical Chemistry*, 76, **2004**, 3099-3105.
- <sup>31</sup> Chauchard F, Cogdill R, Roussel S, Roger J M, Maurel V B, Application of LS-SVM to non-linear phenomena in NIR spectroscopy: development of a robust and portable sensor for acidity prediction in grapes, *Chemometrics Intelligent Laboratory Systems*, 71, **2004**, 141-150.

---

<sup>32</sup> Howley T, Madden M G, O'Connell M L, Ryder A G, The effect of principal component analysis on machine learning accuracy with high-dimensional spectral data, *Knowledge-Based Systems*, 19, **2006**, 363-370.

<sup>33</sup> Arfken G B, Weber H J, *Mathematical Methods for Physicists*, 4th, Academic Press, New York, **1995**.

<sup>34</sup> Cristianini N, Shawe-Taylor J, *An introduction to support vector machines*, Cambridge University Press, UK, **2000**.

<sup>35</sup> Ustun B, *A Comparison of Support Vector Machines and Partial Least Squares regression on spectral data*. Departament Analytical Chemistry, Katholieke Universiteit Nijmegen, **2003**.

# Capítulo 2

**APLICAÇÃO 1 – ANÁLISE EXPLORATÓRIA DE DIFERENTES TIPOS DE LEITE EM PÓ E DETERMINAÇÃO SIMULTÂNEA DE ADULTERANTES EM LEITE EM PÓ POR REFLECTÂNCIA DIFUSA NO INFRAVERMELHO PRÓXIMO E MÁQUINAS DE VETORES DE SUPORTE POR MÍNIMOS QUADRADOS (LS-SVM).**

## 2.1. Introdução

### 2.1.1. Fraude em Leite em pó

O leite animal, produto da biosíntese das glândulas mamárias sob controle hormonal, é constituído basicamente de lactose, gordura, proteínas e sais minerais, como apresentado na Tabela 2<sup>1</sup>.

**Tabela 2.** Composição centesimal média do leite de vaca

Constituinte	Teor %	Varição (%)
Água	87,30	85,50- 88,70
Extrato seco desengordurado	8,80	7,90 – 10,00
Gordura	3,90	2,40 – 5,50
Lactose	4,60	3,80 – 5,30
Proteínas	3,25	2,30 – 4,40
Substâncias minerais	0,65	0,53 – 0,80

O leite é um produto sujeito a adulterações. No caso de leite fluido, as adulterações mais comuns são: a adição de água, neutralizadores para mascarar acidez, sal ou açúcar para mascarar água extra ou teor de sólidos elevados e soro de queijo, entre outros<sup>2</sup>. Adultera-se o leite até mesmo com a adição do “leite sintético”, que é uma mistura de óleo vegetal, uréia, detergente em pó, soda cáustica, açúcar, leite em pó e água<sup>3</sup>. Estas adulterações se estendem também para o leite em pó.

O leite em pó, além do lugar de destaque na alimentação da população, possui grande importância do ponto de vista econômico. Em 2002 O Brasil foi o quinto maior produtor de leite no mundo e responsável por 4,3% do consumo global. A indústria de laticínios no Brasil, em 2002, foi a quarta maior no setor nacional de alimentos, gerando aproximadamente 70.000 vagas de emprego<sup>4</sup>.

Desde 2003, o perfil de exportação do Brasil mudou, tendo sido exportadas mais de 55 toneladas de leite em pó em 2004, em comparação com o ano de 2000, quando foram de menos de cinco toneladas<sup>5</sup>. Isso repercute numa maior exigência da qualidade do produto, por exemplo, em análises rápidas em portos e aeroportos entre outras. Além da preocupação com a qualidade do leite exportado, a preocupação também é com o mercado interno onde há um grande consumo de leite em pó, bem como este é largamente usado em merenda escolar.

No caso do leite em pó, recentemente vêm sendo noticiados na imprensa casos de adulterações, como por exemplo, a adição de soro de leite em pó, maltose, sacarose, amido e até soda cáustica<sup>6,7</sup>. Entre as mais comuns, está a adição do soro de leite em pó, que é 90% mais barato do que o leite e, geralmente, é importado da Argentina, Nova Zelândia e Austrália<sup>8</sup>.

A faixa de fraude mais comum é de 20 a 25% pois até este valor, o consumidor não percebe diferenças no sabor. No entanto, a fraude pode chegar até 60%. No Brasil, a qualidade do leite em pó é regulamentada pelo Regulamento técnico de identidade e qualidade de leite em pó<sup>9</sup>.

### **2.1.2. Metodologias analíticas de referência para determinação de Amido, Soro e Sacarose em leite.**

Há diferentes metodologias de doseamento das proteínas do leite como método por fixação de corante, espectrofotométricos, cromatográficos entre outros<sup>10</sup>. No caso da adição de soro de leite, que tem por objetivo a incorporação fraudulenta de proteínas do soro ao leite, é normalmente detectada e quantificada pela determinação de glicomacropéptido (GMP), que também pode ser denominado de caseinomacropéptido (CMP). Este polipeptídeo, sendo um componente específico do soro, resulta da quebra da ligação peptídica da  $\kappa$ -caseína entre os aminoácidos 105-106 Phe-Met (Fenilalanina – Metionina)<sup>11</sup>. Lembrando que a caseína, principal proteína do leite fresco, é uma mistura de fosfoproteínas (proteínas combinadas com ácido fosfórico) muito semelhantes: as  $\alpha$ -,  $\beta$ -,  $\kappa$ - caseína<sup>12</sup>. Assim, desta cisão na  $\kappa$ - caseína forma-se a *para- $\kappa$ -caseína*

(da sequência de aminoácidos de 1-105), que permanece nas micelas de caseína e o glicomacropeptídeo (106-169), que fica no soro.

Os métodos oficiais para detecção de soro em leite foram publicados no Diário Oficial da União em 2005 pela Portaria nº 124<sup>13</sup>, correspondentes à determinação qualitativa de ácido siálico (presente no GMP), empregando o reativo de Ehrlich, e à determinação quantitativa do glicomacropeptídeo, por cromatografia líquida de alta eficiência (HPLC). Outros métodos que também podem quantificar o GMP são: a cromatografia de exclusão molecular, cromatografia de troca iônica<sup>9</sup> e métodos eletroforéticos<sup>14</sup>. Deve-se lembrar aqui que a presença do GMP pode também ser devida à atividade proteolítica de bactérias psicotróficas que degradam a  $\kappa$ -caseína<sup>15,16</sup> durante a estocagem, a detecção de GMP corre o risco de ser um resultado falso positivo de ocorrência de fraude com adição de soro de leite<sup>17</sup>.

Outro adulterante muito comum é um açúcar, a sacarose, que é um dissacarídeo não redutor (ou seja, não reage com a solução de Fehling ou solução de íons de prata), que tem como principais fontes a cana de açúcar e a beterraba, mas pode ser encontrada em todas as plantas que sofrem o processo de fotossíntese<sup>18</sup>. Para a determinação de sacarose em leite, é realizada análise qualitativa através da caramelização deste açúcar em meio ácido<sup>1</sup>.

O amido também é um adulterante bastante freqüente. É constituído por uma mistura de dois polissacarídeos denominados amilose e amilopectina<sup>19</sup>, em proporções que variam entre os amidos procedentes de diferentes espécies vegetais e, mesmo entre amidos provenientes da mesma espécie, de acordo com o grau de maturação das plantas<sup>12</sup>. Para a identificação do amido, realiza-se o teste de iodo onde sob aquecimento (para promover a abertura da cadeia helicoidal da fração amilose, permitindo a adsorção do iodo), adiciona-se solução de iodo<sup>1</sup>.

### 2.1.3. A técnica de Espectroscopia no Infravermelho próximo por Reflectância Difusa

A região do infravermelho próximo (NIR), que compreende a região de 780 a 2500 nm, vem sendo amplamente empregada na análise de alimentos por ser uma técnica rápida (entre 15s a 90s), não destrutiva que exige pequena quantidade de amostra e geralmente sem necessidade de preparo das mesmas. Para absorver radiação infravermelha, uma molécula precisa sofrer uma variação no momento de dipolo, como conseqüência do movimento vibracional ou rotacional. Apenas nestas circunstâncias, o campo elétrico alternado da radiação pode interagir com a molécula e causar variações na amplitude de um de seus movimentos<sup>20</sup>.

Desde 1960, a espectroscopia no infravermelho próximo vem sendo usada para análises de alimentos<sup>21</sup>. Os espectros de NIR de alimentos consistem de bandas, correspondendo às sobreposições e combinações de modos vibracionais, envolvendo ligações químicas como C-H, O-H e N-H<sup>22</sup>.

Há muitos estudos bem sucedidos de aplicação de espectroscopia no infravermelho próximo em análises de amostras de leite<sup>23</sup>, como para a determinação de gordura<sup>24</sup>, lactose<sup>25</sup> e adulteração de leite fluido<sup>26</sup>.

Líquidos, geralmente são medidos por transmitância, enquanto amostras sólidas por reflectância. Na espectroscopia de reflectância no infravermelho próximo, a amostra sólida é finamente moída para então se obter o espectro. Geralmente, no espectro de reflectância, a ordenada é o logaritmo da reflectância R,

$$\log (1/R) = - \log R \quad (\text{Eq. 17})$$

onde R é a razão entre a intensidade de radiação refletida da amostra e a reflectância de um padrão de reflexão, como por exemplo, sulfato de bário e óxido de magnésio.

A reflectância difusa ocorre em superfícies não totalmente planas, podendo o substrato ser contínuo ou fragmentado (na forma de pó). Neste processo de reflexão, o feixe incidente penetra a superfície da amostra interagindo com a

matriz, retornando a superfície da mesma, após absorção parcial e múltiplos espalhamentos<sup>27</sup>, conforme ilustrado na figura 8.

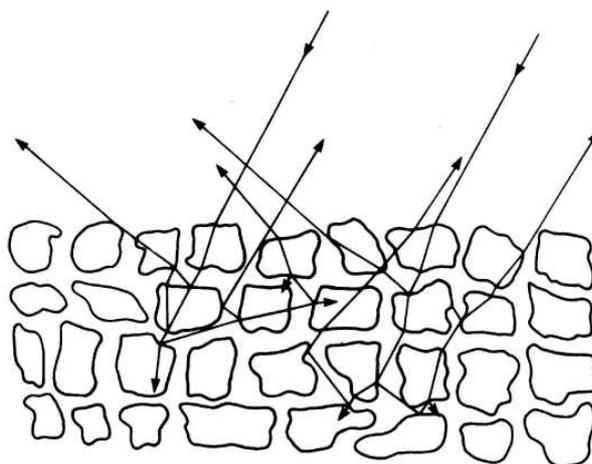


Figura 8. Representação do espalhamento da reflectância difusa em amostra sólida.

Na reflectância difusa, a radiação incidente entra em contato diversas vezes com as partículas da amostra, sendo então conseqüentemente atenuada. Com isto, a radiação que sofre a reflectância difusa fornece informações qualitativas e quantitativas sobre a natureza química da amostra. É importante salientar, que a radiação que retorna de forma difusa de um substrato é geralmente muito inferior em magnitude que a radiação incidente. Em função disso, a maior parte dos acessórios de reflexão difusa apresenta esquemas óticos que visam convergir a radiação, para posteriormente a mesma ser focalizada sobre o sistema de detecção dos instrumentos.

Na figura 9, é mostrado um esquema de uma esfera de integração que é utilizada em medidas de reflectância difusa. As paredes internas da esfera de integração que circunda a amostra são recobertas com um material que reflete difusamente, de modo quase perfeito, como o sulfato de bário. A radiação refletida pela amostra eventualmente chega aos detectores após múltiplas reflexões nas paredes da esfera<sup>21</sup>.

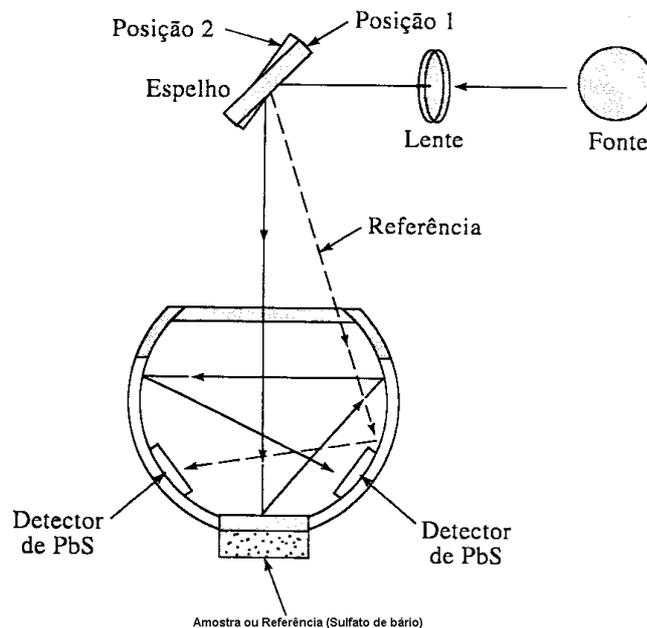


Figura 9. Representação do esquema de uma esfera de integração.

Os espectros de reflectância freqüentemente exibem desvios de linha base pela variação na compactação da amostra, por espalhamento da superfície das partículas e pela variação no tamanho das partículas do material analisado. Para minimizar a influência destas variações, são empregados pré-tratamentos matemáticos como a correção do espalhamento multiplicativo (MSC)<sup>28</sup>, dentre outros.

## 2.2. Parte Experimental

### 2.2.1. Amostras para análise qualitativa de tipos de leite em pó

Para uma análise qualitativa entre diferentes tipos de leite em pó (integral, desnatado, formulação infantil) foi coletado um total de 36 amostras (marcas A, B, C, D, E, F e G) como mostrado na Tabela 3. Deste conjunto, quatro amostras eram do leite integral “E”, coletado de dois lotes de fabricação diferentes, e outras oito amostras do leite “E”, de um mesmo lote, de data de fabricação mais antiga, que foi estocada por cinco meses, em contato com o ar e à temperatura ambiente,

num armário do laboratório. Estas amostras foram nomeadas como Integral degradado. Para as outras amostras do leite integral (A, B e G), desnatado (B, C e G) e formulações infantis (D e F), foram coletadas três amostras de cada marca, de três embalagens diferentes de um mesmo lote de fabricação.

**Tabela 3.** Amostras de tipos de leite em pó para análise qualitativa.

Número da amostra	Tipo do leite em pó	Marca
1 a 8	Integral degradado	E
9 a 12	Integral	E
13 a 15	Integral	A
16 a 19	Integral	B
20 a 22	Desnatado	B
23 a 25	Desnatado	C
26 a 28	Formulação Infantil	D
29 a 31	Formulação Infantil	F
31 a 33	Integral	G
34 a 36	Desnatado	G

### 2.2.2. Amostras para a Quantificação dos adulterantes em leite em pó

Para a quantificação de adulterantes, foi usado o leite em pó integral “E”, vendido a granel por uma distribuidora de alimentos. Para as quantificações de cada adulterante, as amostras de leite em pó “E” foram contaminadas artificialmente com *soro de leite em pó* (Tangará), *amido* (Synth) e *sacarose* (Synth), separadamente, em níveis de concentração que foram de 1% a 50% (m/m). Para as misturas binárias, foi realizado um planejamento com seis níveis de concentração (4, 8, 12, 15, 20 e 25% (m/m)). Um total de sete conjuntos de espectros foram empregados, conforme os limites descritos na Tabela 4, incluindo amostras de leite em pó puro “E”. Deve-se lembrar que em todos estes casos, os pós, por serem muito finos, não foram peneirados.

**Tabela 4.** Planejamento para a adição de dois adulterantes no leite em pó.

<b>mistura</b>	<b>leite em pó (%m/m)</b>	<b>adulterante 1 (%)</b>	<b>adulterante 2 (%)</b>	<b>mistura</b>	<b>leite em pó (%m/m)</b>	<b>adulterante 1 (%)</b>	<b>adulterante 2 (%)</b>
<b>01</b>	92	4	4	<b>18</b>	81	15	4
<b>02</b>	88	4	8	<b>19</b>	77	15	8
<b>03</b>	84	4	12	<b>20</b>	73	15	12
<b>04</b>	81	4	15	<b>21</b>	70	15	15
<b>04</b>	76	4	20	<b>22</b>	65	15	20
<b>05</b>	71	4	25	<b>23</b>	60	15	25
<b>06</b>	88	8	4	<b>24</b>	76	20	4
<b>07</b>	84	8	8	<b>25</b>	72	20	8
<b>08</b>	80	8	12	<b>26</b>	68	20	12
<b>09</b>	77	8	15	<b>27</b>	65	20	15
<b>10</b>	72	8	20	<b>28</b>	60	20	20
<b>11</b>	67	8	25	<b>29</b>	55	20	25
<b>12</b>	84	12	4	<b>30</b>	71	25	4
<b>13</b>	80	12	8	<b>31</b>	67	25	8
<b>14</b>	76	12	12	<b>32</b>	63	25	12
<b>15</b>	73	12	15	<b>33</b>	60	25	15
<b>16</b>	68	12	20	<b>34</b>	55	25	20
<b>17</b>	63	12	25	<b>35</b>	50	25	25

Na Tabela 5, são apresentadas as faixas de concentrações (%) de cada adulterante nas misturas.

**Tabela 5.** Descrição dos conjuntos de amostras cujos espectros foram obtidos.

Experimento	amido (%)	Soro (%)	sacarose (%)	nº de amostras
1	1,05 – 50,05	-	-	45
2	-	1,00 – 49,82	-	45
3	-	-	1,24 – 50,35	45
4	4,22 – 25,17	4,16 – 25,11	-	27
5	4,05 – 25,59	-	4,02 – 25,71	27
6	-	4,13 – 25,21	4,12 – 25,77	27
7*	-	-	-	12

\* leite em pó puro

Na Tabela 6, é apresentado o número de amostras para os conjuntos de calibração e previsão, para cada experimento.

**Tabela 6.** Número de espectros empregados na calibração e previsão.

Experimento	espectros de calibração	espectros de previsão
1	30	15
2	30	15
3	30	15
4	20	7
5	20	7
6	20	7
7*	6	6
<b>total</b>	<b>156</b>	<b>72</b>

\* leite em pó puro

E, por fim, na Tabela 7 são apresentados os dois modelos distintos, desenvolvidos usando LS-SVM e PLS. No primeiro, foram usados espectros dos

experimentos 1 a 6 (Tabela 6). No segundo modelo, os espectros usados foram dos experimentos de 1 a 7 (Tabela 6).

**Tabela 7.** Número de espectros de calibração e validação para os modelos.

<b>modelo</b>	<b>Conjunto sem amostras contendo leite puro</b>	<b>Conjunto com todas as amostras</b>
nº de espectros para calibração	150	156
nº de espectros para previsão	66	72

### **2.2.3. Instrumental**

Para a obtenção dos espectros, as amostras (especialmente as misturas), foram previamente misturadas com a espátula e, em seguida, com o emprego de três esferas metálicas dentro de cada recipiente, de cada amostra, e agitadas em vórtex por um minuto. Os espectros foram obtidos em triplicata usando o Espectrofotômetro dispersivo UV-VIS Cary 5G, com detector de sulfeto de chumbo, lâmpada de tungstênio/halogênio e com o acessório de reflectância difusa baseado em esfera de integração. A região espectral usada foi entre 810 nm a 2400 nm com resolução de 1 nm. O espectro de referência foi obtido usando politetrafluoretileno (teflon).

### **2.2.4. Pré-tratamento dos espectros**

Todos os conjuntos de espectros foram pré-processados pela correção do espalhamento multiplicativo de sinal (MSC), que remove a variação do espectro causado por espalhamento de luz das amostras.

### **2.2.5. Escolha das amostras para calibração e previsão**

O conjunto de calibração deve ser representativo da população para a qual previsões futuras serão realizadas, ou seja, abranja toda uma faixa de concentração desejável deste espaço amostral. Neste caso, a partir da matriz de espectros montada na ordem dos conjuntos da Tabela 6, seguindo o planejamento da Tabela 7, a escolha das amostras para o conjunto de previsão foi pelo critério de escolher uma amostra no intervalo de outras três (ou seja, de três em três amostras) e que também não estivesse nos extremos das faixas de concentração.

### **2.2.6. Modelagem dos dados**

Os modelos multivariados foram obtidos usando-se um microcomputador com processador Intel Pentium 4 3.00 Ghz e 1 Gbytes de memória RAM, empregando-se os pacotes PLS-Toolbox<sup>29</sup> e LS-SVM<sup>30</sup> para MATLAB. Para os modelos PLS, o pré-processamento usado foi o centrado na média, onde ocorre a translação do centro dos dados para a origem. Foi usado o algoritmo NIPALS (Nonlinear Iterative Partial Squares) para encontrar os escores e os pesos. O número de variáveis latentes foi escolhido por validação cruzada, onde todas as amostras de calibração foram validadas uma a uma (leave-one-out).

Para o desenvolvimento do modelo através das máquinas de vetor suporte por mínimos quadrados, a função núcleo escolhida foi a função de base radial (RBF), sendo necessário definir os dois parâmetros gama e sigma quadrático. Os parâmetros gama e sigma quadrático foram combinados, a fim de verificar a tendência do erro de validação cruzada (RMSECV). A identificação da combinação dos parâmetros gama e sigma quadrático é realizada através da localização do ponto mínimo gerada pela superfície onde a diminuição do RMSECV não seja mais significativa, semelhante a que está apresentada na Figura 10.

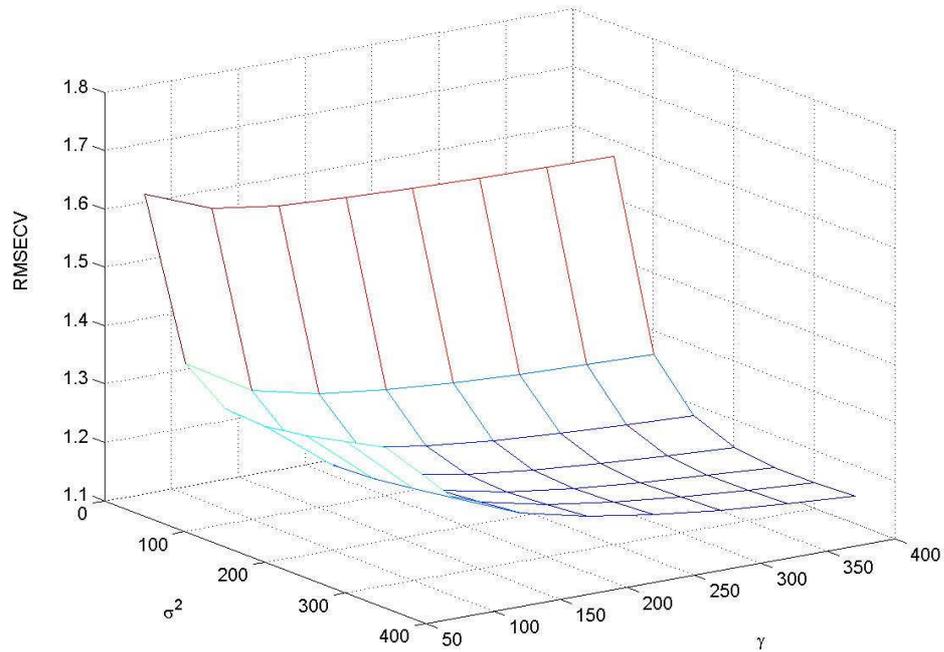


Figura 10. Superfície gerada pela LS-SVM para encontrar o ponto de mínimo entre os parâmetros sigma quadrado e gama.

### 2.2.7. Parâmetros para a avaliação dos modelos

Para a avaliação do desempenho dos modelos desenvolvidos, utilizando PLS e LS-SVM, os parâmetros observados foram os coeficientes de correlação dos dados de referência versus os de previsão ( $R^2_{\text{prev}}$ ), os erros quadráticos médios (RMSE - Root Mean Square Error), de validação cruzada (RMSECV) e de previsão (RMSEP) de cada modelo. Estes erros são calculados conforme equação 18.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (\text{Eq. 18})$$

onde  $\hat{y}_i$  representa o valor previsto pelo modelo multivariado,  $y_i$  representa o valor referência e  $n$  corresponde ao número de amostras.

## 2.3. Resultados e Discussões

### 2.3.1. Análise qualitativa de tipos de leite em pó

A análise qualitativa dos tipos de leite em pó teve como objetivo averiguar se os diferentes leites se dividiram em grupos distintos, se isso ocorresse, saberia-se previamente que para cada leite, ou pelo menos para cada grupo, se for empregado um modelo de calibração linear. Seria necessário um modelo de quantificação para cada tipo (ou grupo) de amostras de leite.

Na Figura 11 são apresentados os espectros destas amostras de leite com e sem pré-tratamento por MSC. Pode-se verificar que, entre 1560 a 2000nm encontra-se o 1º sobretom e, entre 1030 a 1330nm, o 2º sobretom do estiramento O-H de ligações intramoleculares de açúcares, como a lactose e a galactose. Os estiramentos das ligações C-O-O da maioria dos aminoácidos (proteínas) aparece em 2330 a 2420nm (1º sobretom), 1550 a 1610nm (2º sobretom) e 1160 a 1210 nm (3º sobretom). Entre 2220 a 2380nm, observa-se o 3º sobretom do estiramento C-O de ácidos graxos de cadeias longas (gordura)<sup>31,32</sup>.

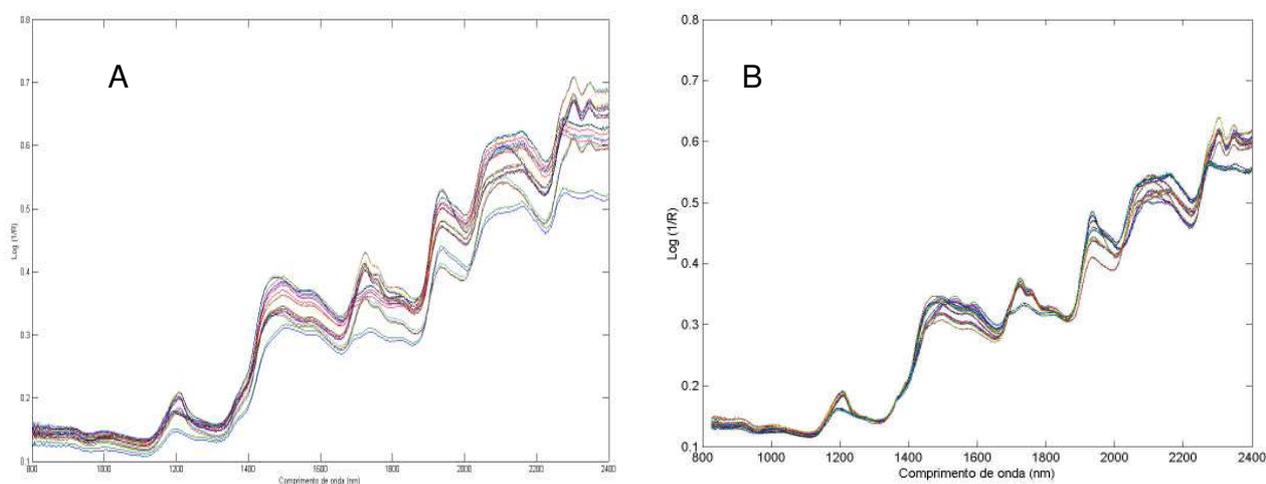


Figura 11. Espectros no Infravermelho Próximo (NIR) amostras das diferentes marcas de leite: A- sem o pré-tratamento com MSC; B- com pré-tratamento MSC.

Na Figura 12, é mostrado o resultado da Análise de Componentes Principais, onde o modelo foi escolhido com quatro componentes principais (97,65% de variância dos dados). Com os escores dos dois primeiros componentes principais, foi possível visualizar a separação entre os leites de **formulação infantil** (grupos 1 e 2), leites **inteiros** (grupos 5, 6 e 7 ) e leites **desnatados** (grupos 3 e 4). Dentre os grupos dos leites inteiros, o grupo 6 se refere às amostras de leite integral degradado.

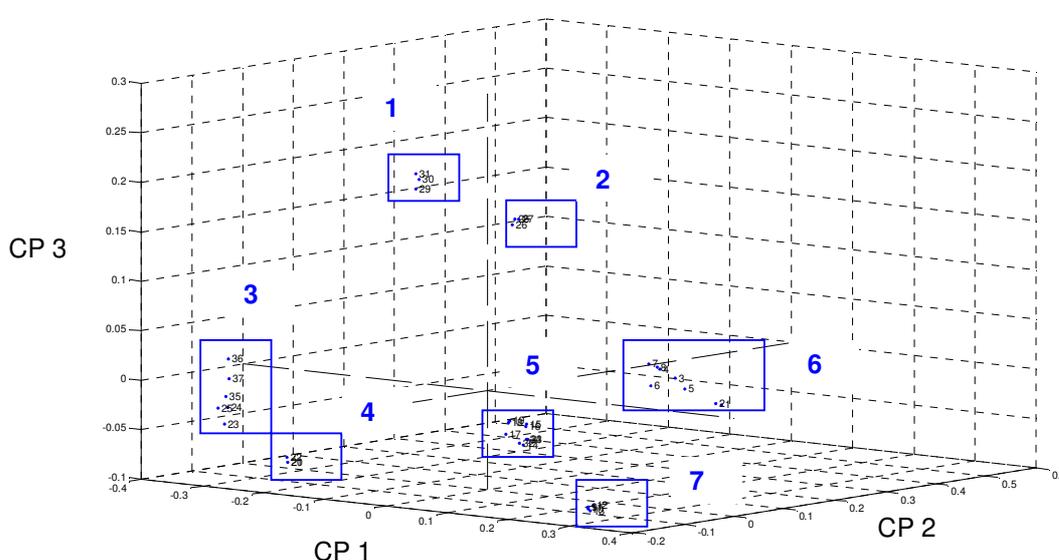


Figura 12. Gráfico de escores gerado pela Análise de Componentes Principais (PCA) com a separação dos diferentes tipos de leite em pó.

Grupo 1: Formulação infantil D; Grupo 2: Formulação infantil F; Grupo 3: Desnatado C e Desnatado D; Grupo 4: Desnatado B; Grupo 5: Integral A, Integral B e Integral G; Grupo 6: Integral degradado; Grupo 7: Integral E.

### 2.3.2. Análise qualitativa de leite em pó com diferentes adulterantes

Na Figura 13, estão apresentados os espectros na região do infravermelho próximo do leite em pó “E” e dos contaminantes em pó usados na adulteração artificial: amido, soro de leite e sacarose. É possível observar que o espectro do soro é bem semelhante ao espectro do leite, indicando que esta composição

química parecida pode gerar dificuldade para uma modelagem de quantificação e identificação da presença do soro no leite em pó.

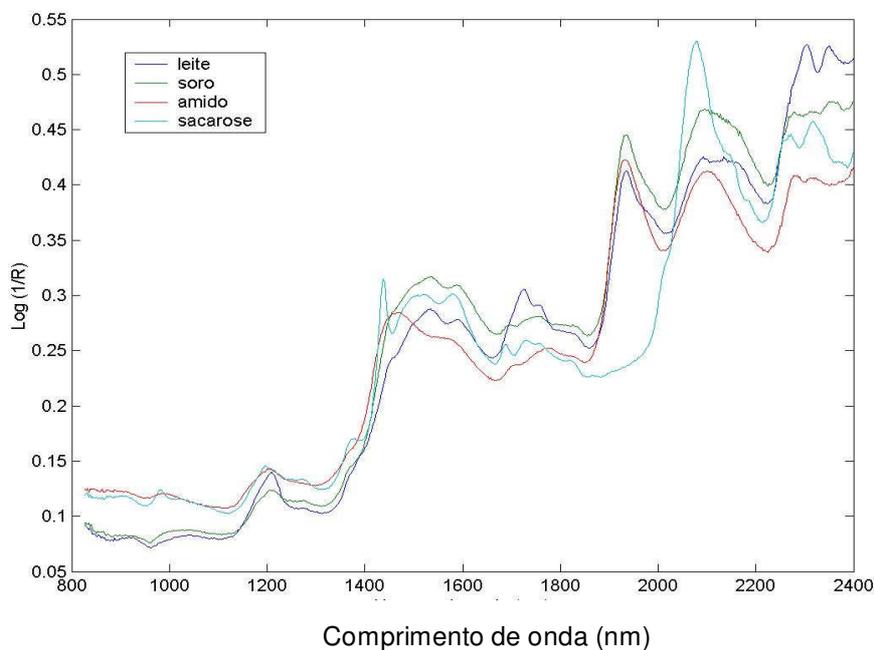


Figura 13. Espectro do leite em pó usado para adulteração e espectros dos contaminantes em pó: amido, soro de leite e sacarose.

Na Figura 14, estão apresentados os espectros das amostras do conjunto do leite “E” (incluindo as amostras degradadas) e os conjuntos das amostras com os diferentes adulterantes e um conjunto com amostras de leite com presença dos três adulterantes, porém que não participaram dos modelos de quantificação.

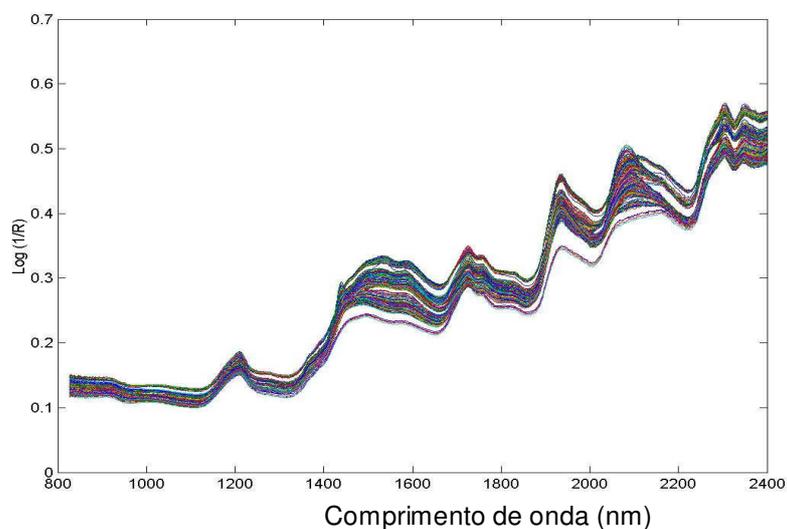


Figura 14. Espectros NIR das 306 amostras de leite em pó puras e adulteradas.

Na Figura 15 está apresentado os escores dos dois primeiros componentes onde é possível visualizar a separação em classes das amostras de leite em pó integral sem adulterantes (grupo A) das amostras de leite em pó com diferentes adulterantes isolados ou misturas deles (amido, soro, sacarose, amido e soro, amido e sacarose e soro e sacarose). As amostras do leite em pó integral, sem adulterantes, porém degradadas, se aproximam do grupo E, que é o leite em pó adulterado com sacarose como indica a seta na Figura 8. É através desta separação clara das amostras do leite E dos outros grupos (de amostras de leite E adulteradas) que admitimos que estas amostras não possuem nenhum destes adulterantes, pois caso contrário estaria próxima a um dos agrupamentos referentes aos leites adulterados. Assim, foi usado o termo “amostras puras” para as amostras do leite E.

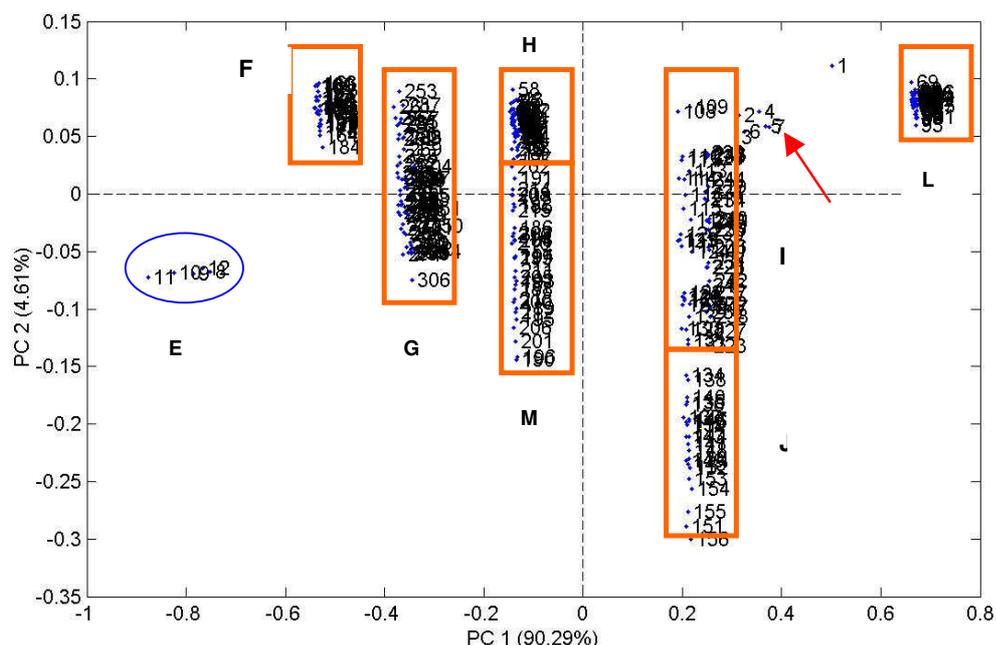


Figura 15. Gráfico dos escores do conjunto das 306 amostras constituído de amostras de leite em pó sem adulteração e adulterado com diferentes adulterantes.

Seta: Amostras de 1 a 7 leite E degradado. E: leite E sem adulterante; F: leite E adulterado com amido e soro; G: leite E adulterado com amido, soro e sacarose; H: leite E adulterado com amido e sacarose; I: leite E adulterado com sacarose; J: leite E adulterado com soro e sacarose; L: leite E adulterado com soro e M: leite E adulterado com amido.

### 2.3.3. Modelos para quantificação simultânea de adulterantes para conjunto de amostras contendo um e dois adulterantes

Para a escolha do número de variáveis latentes, para o modelo PLS, foram avaliados os valores do erro médio quadrático da validação cruzada para cada variável latente. Para simplificar o procedimento e construção de um modelo único para a previsão simultânea de amido, soro e sacarose nas amostras de leite em pó, consideramos a resposta média do RMSECV para os três analitos conforme equação 19.

$$RMSECV_{(m\u00e9dio)} = \frac{(RMSECV_{(amido)} + RMSECV_{(soro)} + RMSECV_{(sacarose)})}{3} \quad (\text{Eq.19})$$

Na Figura 16 \u00e9 apresentado os valores de RMSECV partindo de quatro at\u00e9 16 vari\u00e1veis latentes, para os tr\u00eas adulterantes. Os valores de RMSECV foram relativamente altos das tr\u00eas primeiras vari\u00e1veis por isso n\u00e3o foram colocadas no gr\u00e1fico. Dessa forma, para o desenvolvimento do PLS foram encontrados 10, 8 e 8 vari\u00e1veis latentes para amido, soro e sacarose, respectivamente.

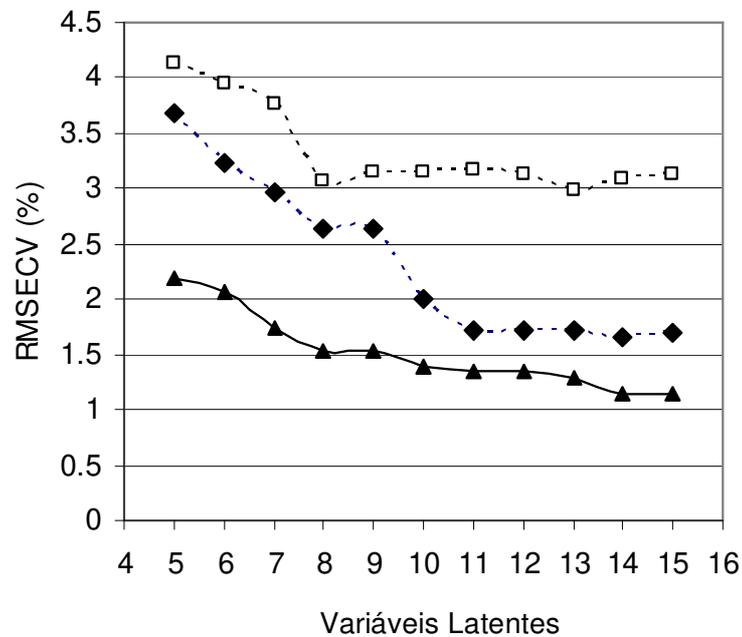


Figura 16. Gr\u00e1fico com valores de RMSECV em fun\u00e7\u00e3o do n\u00famero de vari\u00e1veis latentes. (□ - soro; ◆ - amido; ▲ - sacarose).

Para o desenvolvimento do modelo LS-SVM, foram realizadas v\u00e1rias otimiza\u00e7\u00f5es entre os par\u00e2metros  $\gamma$  (gama) e  $\sigma^2$  (sigma quadrado) at\u00e9 se encontrar um m\u00ednimo do RMSEV que n\u00e3o apresentassem mais uma diminui\u00e7\u00e3o significativa do seu valor com o aumento dos par\u00e2metros  $\gamma$  e  $\sigma^2$ . Atrav\u00e9s da superf\u00edcie apresentada na Figura 17 podemos observar o m\u00ednimo do valor de RMSECV para

os parâmetros  $\gamma$  de 1000 e  $\sigma^2$  de 300, os quais foram usados no desenvolvimento do modelo LS-SVM.

Uma discussão mais detalhada sobre a escolha dos valores dos parâmetros  $\gamma$  e  $\sigma^2$  é apresentada no artigo: Borin A, Maretto D A, Ferrão M F, Mello C, Poppi R J, Least-Squares Support Vector Machines and Near Infrared Spectroscopy for Quantification of Common Adulterants in Powdered Milk, *Analytica Chimica Acta*, 579, 2006, 25.

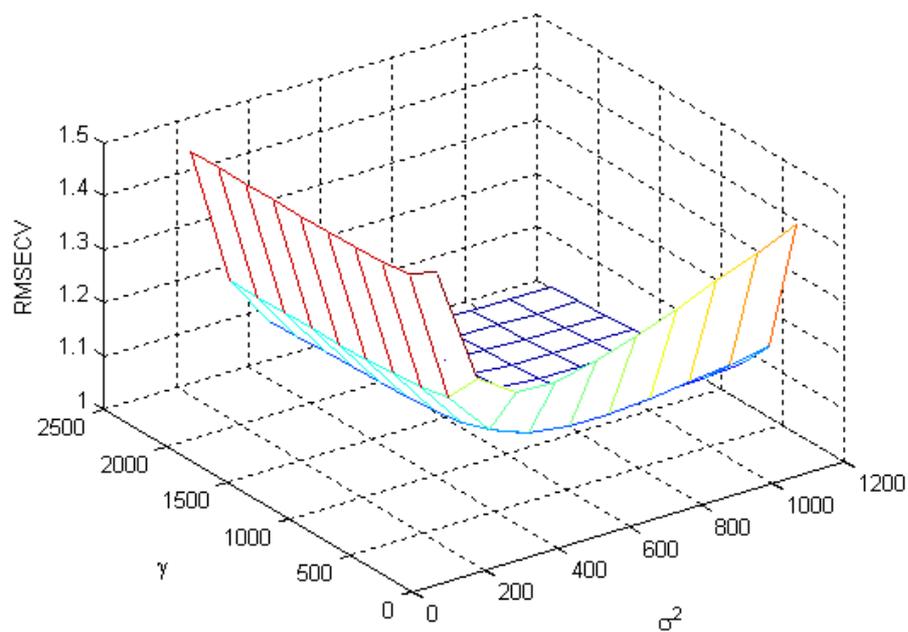


Figura 17. Superfície de resposta para a otimização dos parâmetros  $\gamma$  e  $\sigma^2$  a serem aplicados no modelo LS-SVM.

Escolhidos os números de variáveis latentes para o PLS e os parâmetros  $\gamma$  e  $\sigma^2$  para a LS-SVM, foram desenvolvidos os modelos PLS e LS-SVM. Na Tabela 8, são comparados os resultados de ambos os modelos em termos dos erros.

a

b

Tabela 8. Comparação em termos de erros, entre os resultados de PLS e LS-SVM.

	Amido		Soro		Sacarose	
	PLS	LS-SVM	PLS	LS-SVM	PLS	LS-SVM
$R^2$	0,98	1,00	0,94	1,00	0,99	1,00
RMSECV (%)	2,00	1,02	3,06	1,47	1,53	0,91
RMSEP (%)	1,96	1,01	3,17	1,12	1,38	0,56
LV	10	-	8	-	8	-
$\gamma$	-	1000	-	1000	-	1000
$\Sigma^2$	-	300	-	300	-	300

Na Figura 18, estão apresentados os gráficos dos valores previstos, tanto para o conjunto de calibração quanto para o conjunto de previsão, e as retas de regressão ajustadas para o adulterante amido.

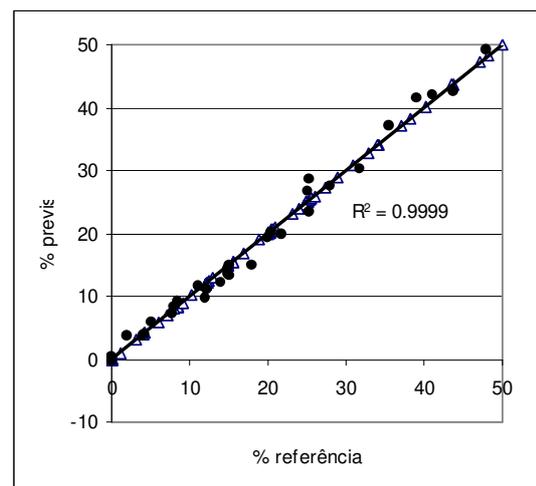
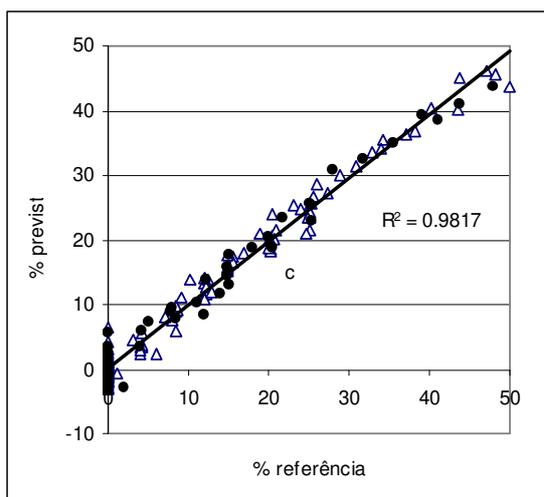


Figura 18. Gráficos de correlação para a calibração e previsão da % de amido adicionado como adulterante (a) do modelo PLS, empregando 10 variáveis latentes. (b) Modelo do LS-SVM ( $\Delta$ - amostras de calibração;  $\bullet$  amostras de previsão).

Os modelos PLS e LS-SVM apresentaram uma boa correlação entre os valores de referência e os valores previstos, com coeficiente de correlação ( $R^2$ ) de 0,98 para o PLS e próximo de 1,00 para LS-SVM. Porém, é possível observar que os dados na reta de regressão do modelo PLS (Figura 18a) estão um pouco mais dispersos pela reta em relação aos dados da reta de regressão do modelo da LS-SVM (Figura 18b), onde os dados aparecem bem distribuídos e próximos da reta de correlação. Outro ponto importante a se destacar é que, próximo do ponto zero da reta referente ao modelo do PLS, as amostras puras, ou seja, sem adição de amido, aparecem com valores diferentes de zero (ou valores negativos ou maiores que zero) na previsão do PLS. Já, no modelo da LS-SVM, as amostras puras são previstas com concentrações próximas do zero, sem indicação de um resultado falso positivo de fraude.

E essa observação realizada para o adulterante amido se repetem para os outros dois adulterantes. Na Figura 19, estão apresentados os valores previstos contra os de referência e as retas de regressão para o adulterante sacarose, o qual com o modelo PLS apresenta, entre os três adulterantes, o melhor coeficiente de correlação para o modelo PLS de  $R^2$  de 0,99. Finalmente, na Figura 20, estão apresentadas as retas de regressão para o adulterante soro, para o modelo PLS e para o modelo LS-SVM, respectivamente.

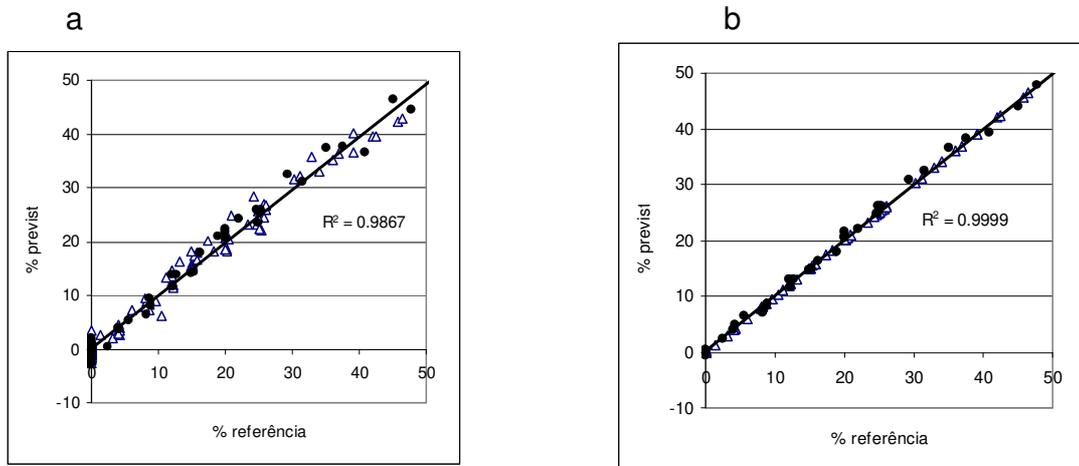


Figura 19. Gráficos de correlação para a calibração e previsão da % de sacarose adicionado como adulterante (a) do modelo PLS, empregando oito variáveis latentes. (b) Modelo do LS-SVM ( $\Delta$ - amostras de calibração;  $\bullet$  amostras de previsão).

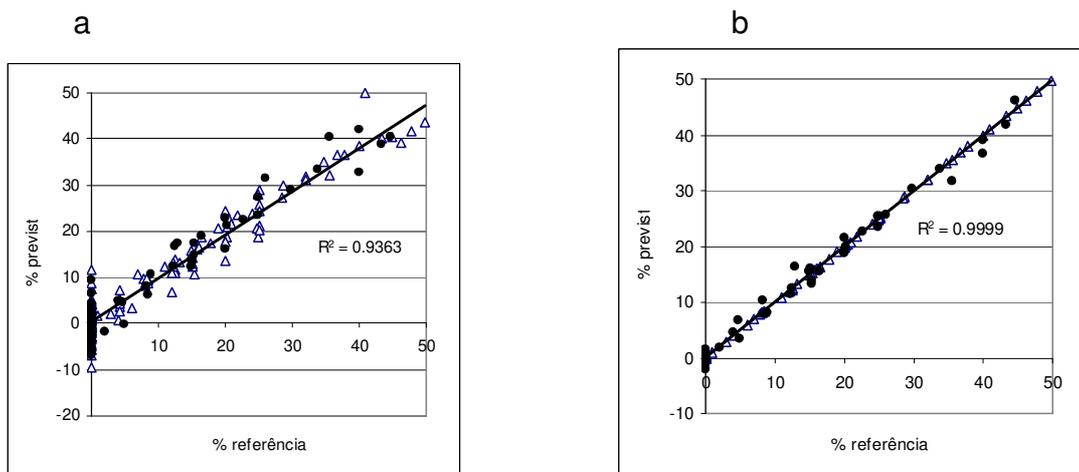


Figura 20. Gráficos de correlação para a calibração e previsão da % de soro adicionado como adulterante (a) do modelo PLS empregando oito variáveis latentes. (b) Modelo do LS-SVM ( $\Delta$ - amostras de calibração;  $\bullet$  amostras de previsão).

Observando o perfil das retas com os valores previstos e de referência dos conjuntos de calibração e previsão para os três adulterantes, nota-se que, empregando o PLS, o adulterante soro (Figura 20a) é o que possui o mais baixo

coeficiente de correlação dos dados,  $R^2$  de 0,94, e as amostras também estão mal distribuídas e dispersas ao longo da reta. Esta dificuldade de para o adulterante soro se deve pela composição parecida do soro com o leite como observado na Figura 13 referente ao espectro do leite e do soro apresentados separadamente. A observação oposta ocorre com o modelo LS-SVM para o adulterante soro, com as amostras próximas e bem distribuídas ao longo da reta de correlação,  $R^2$  de 1,00 e não apresentando amostras com valor de referência de zero% (ausência do soro) sendo previstas com valores negativos ou maiores que zero% de adulterante.

#### 2.3.4. Modelos para quantificação simultânea de adulterantes para o conjunto de amostras “puras” (sem adulterante) e amostras contendo um e dois adulterantes

Nestes modelos de PLS e LS-SVM, foram usados todos os conjuntos (experimentos 1 a 7 da Tabela 5). Para o desenvolvimento dos modelos de calibração por PLS, foram encontrados 9, 6 e 3 variáveis latentes para amido, soro e sacarose, respectivamente, como é apresentado na Figura 21.

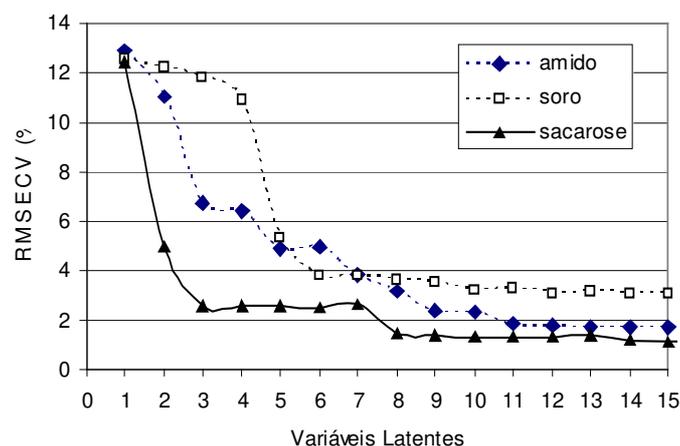


Figura 21. Gráfico com valores de RMSECV em função do número de variáveis latentes.

Na Figura 22 é apresentada a superfície de resposta para a otimização dos parâmetros  $\gamma$  e  $\sigma^2$  até se encontrar um mínimo do RMSECV onde a diminuição

deste valor não seja mais significativa. Para este modelo LS-SVM, foram escolhidos  $\gamma$  de 300 e  $\sigma^2$  de 200.

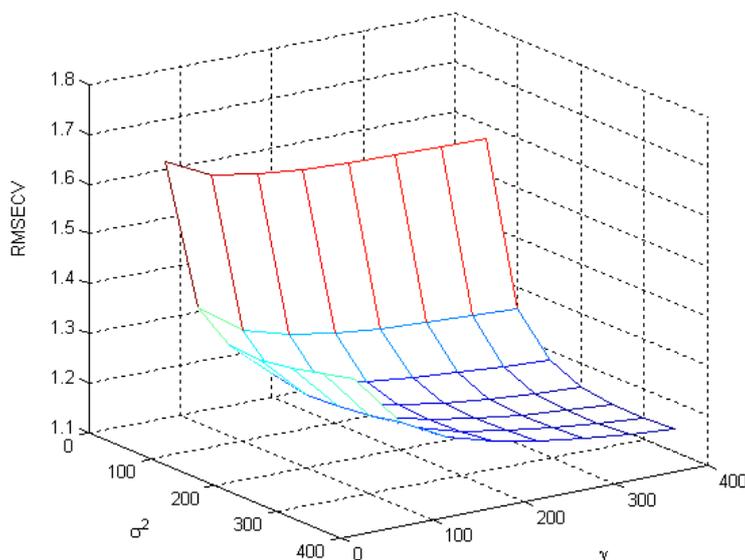


Figura 22. Superfície de resposta da otimização dos parâmetros  $\gamma$  e  $\sigma^2$  para o modelo LS-SVM.

É interessante observar que, com a inclusão do conjunto com amostras de leite puras (sem adulterantes), a quantidade de variáveis latentes foi menor para o modelo PLS (9, 6 e 3 variáveis latentes para amido, soro e sacarose, respectivamente), em comparação ao PLS do conjunto anterior (10, 8 e 8 variáveis latentes para amido, soro e sacarose, respectivamente). A mesma observação é feita para a escolha dos parâmetros  $\gamma$  e  $\sigma^2$  para ambos os conjuntos. Neste modelo com a adição do conjunto de amostras puras foram escolhidos  $\gamma$  de 300 e  $\sigma^2$  de 200. Em contrapartida com o conjunto anterior foram escolhidos  $\gamma$  de 1000 e  $\sigma^2$  de 300. Supõe-se que a introdução dos espectros das amostras puras com medidas no vetor  $Y$  de zeros melhora o desenvolvimento dos modelos. Como pode ser visto na Tabela 9, ao compararmos os resultados para cada um dos adulterantes, percebe-se que, embora tanto os modelos PLS quanto os LS-SVM apresentem bons coeficientes de correlação para a calibração ( $R^2$ ), o modelo

empregando LS-SVM é superior para quantificação de qualquer um dos três adulterantes, apresentando valores de RMSECV e RMSEP significativamente melhores.

Tabela 9. Comparação dos resultados entre os modelos PLS e LS-SVM com amostras sem adulterantes, com amostras com um e com dois adulterantes.

	Amido		Soro		sacarose	
	PLS	LS-SVM	PLS	LS-SVM	PLS	LS-SVM
<b>R<sup>2</sup></b>	0,97	1,00	0,92	1,00	0,96	1,00
<b>RMSECV (%)</b>	2,39	1,03	3,84	1,08	2,58	0,57
<b>RMSEP (%)</b>	2,33	1,04	3,15	1,08	2,80	0,57
<b>LV</b>	9	-	6	-	3	-
<b>γ</b>	-	300	-	300	-	300
<b>σ<sup>2</sup></b>	-	200	-	200	-	200

Nas Figuras 23 a 25, são apresentados os gráficos de correlação entre os valores de referência e os valores previstos pelos modelos PLS e LS-SVM para a quantificação de amido, soro e sacarose em leite em pó, respectivamente.

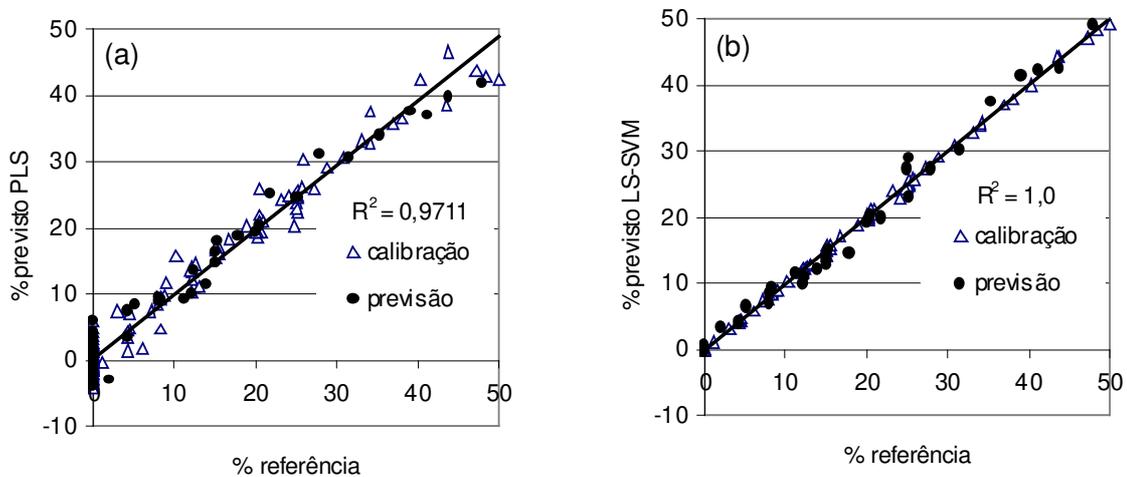


Figura 23. Gráficos de correlação para a calibração e previsão da % de amido adicionado como adulterante do modelo PLS, empregando nove variáveis latentes. (b) Modelo do LS-SVM.

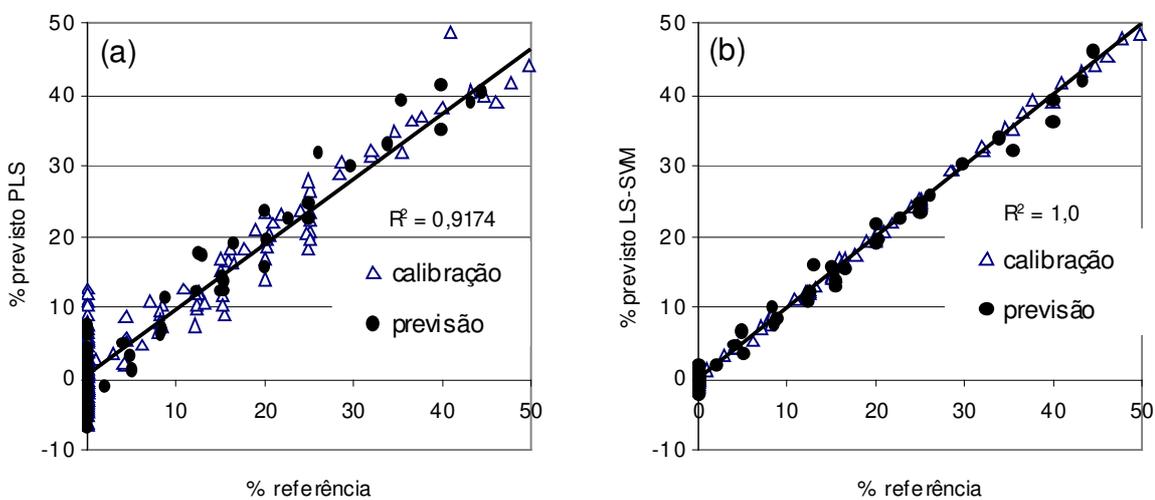


Figura 24. Gráficos de correlação para a calibração e previsão da % de soro adicionado como adulterante: (a) modelo PLS, empregando seis variáveis latentes; (b) modelo LS-SVM.

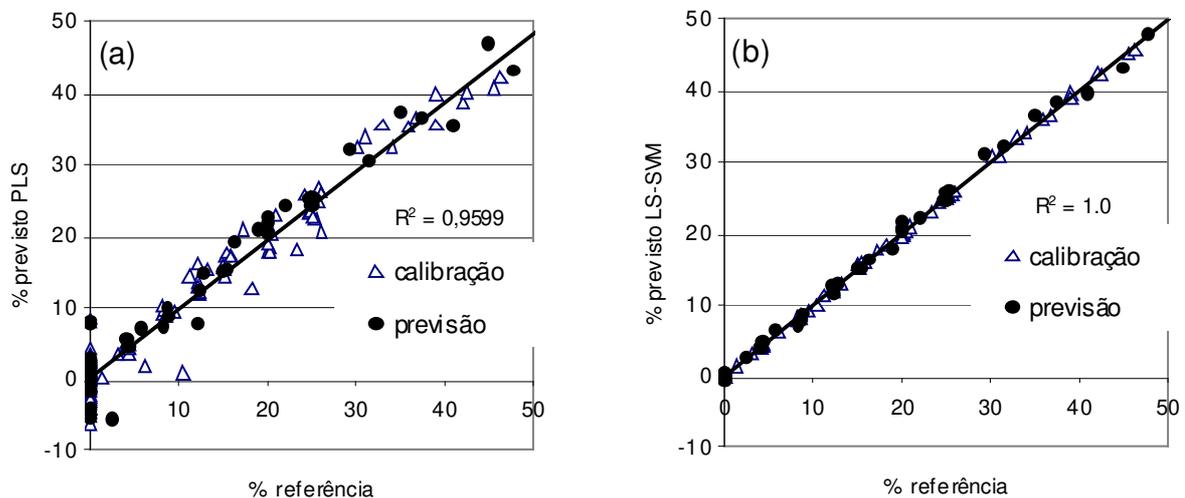


Figura 25. Gráficos de correlação para a calibração e previsão da % de sacarose adicionada como adulterante: (a) modelo PLS, empregando três variáveis latentes; (b) modelo LS-SVM.

Embora os coeficientes de correlação para os três adulterantes sejam adequados para ambas as técnicas de regressão, pode-se observar que os modelos empregando o PLS apresentam uma maior dispersão, tanto para as amostras de calibração, quanto de previsão, do que os análogos empregando LS-SVM. A grande dificuldade dos modelos PLS apresentados é a incapacidade destes em prever a ausência de qualquer um dos adulterantes nas amostras em estudo. Já os modelos empregando LS-SVM sempre foram superiores na capacidade preditiva da ausência do adulterante, em função da sua maior capacidade de generalização. A fim de demonstrar a capacidade do modelo LS-SVM em indicar, não só quem possivelmente esteja adulterando as amostras de leite em pó, mas também que nenhum dos três potenciais adulterantes encontram-se nas amostras estudadas, são apresentados na Tabela 10, comparativamente, os valores de previsão para as seis amostras de leite em pó puro do conjunto de validação.

Tabela 10. Erro absoluto para a previsão das amostras de leite puro para os modelos de regressão multivariada PLS e LS-SVM.

Amostra	amido (%)		soro (%)		sacarose (%)	
	PLS (9) <sup>a</sup>	LS-SVM	PLS (6) <sup>a</sup>	LS-SVM	PLS (3) <sup>a</sup>	LS-SVM
1	2,32	0,08	-5,36	-0,26	2,06	0,01
2	-0,06	0,36	-0,50	-0,25	1,74	0,01
3	1,56	0,29	-3,71	-0,07	1,71	-0,01
4	-4,29	0,13	-0,69	0,01	8,08	0,07
5	-3,48	-0,06	-3,37	0,05	7,90	0,02
6	-3,63	0,07	-1,99	0,01	7,82	0,02
<b>Média<sup>b</sup></b>	2,56	0,17	2,60	0,11	4,88	0,02

<sup>a</sup> número de variáveis latentes empregadas no PLS; <sup>b</sup> média dos módulos dos erros absolutos.

Verifica-se que o modelo empregando LS-SVM é capaz de indicar a ausência dos referidos adulterantes, uma vez que o maior valor encontrado foi de 0,36 % para o amido na amostra dois, enquanto que os modelos PLS, além de preverem valores negativos significativos, chegam a indicar 8,08 % de sacarose na amostra quatro que tecnicamente não apresenta adição do mesmo. Discussões mais detalhadas sobre esta aplicação se encontram nos seguintes artigos: **1)** 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*, 30 (4), **2007**, 852 . **2)** Borin A, Maretto D A, Ferrão M F, Mello C, Poppi R J, Least-Squares Support Vector Machines and Near

Infrared Spectroscopy for Quantification of Common Adulterants in Powdered Milk, *Analytica Chimica Acta*, 579, **2006**, 25.

## 2.4. Referências Bibliográficas

<sup>1</sup> Silva H F, *Físico-química do leite e derivados: métodos analíticos*, Ed. Gemacom, Juiz de Fora, **1997**.

<sup>2</sup> Harding F, *Adulteration of milk - Milk Quality*, Chapman e Hall Food Science Book, New York, **1999**.

<sup>3</sup> Jha S N, Matsuoka T, Detection of adulterants in milk using near infrared spectroscopy, *Journal of Food Science Technology*, 41, **2004**, 313-316.

<sup>4</sup> Farina E M M Q, Gutman G E, Lavarello P J, Nunes R, Reardn T, Private and public milk standards in Argentina and Brazil, *Food Policy*, 30, **2005**, 302-315.

<sup>5</sup> Embrapa Gado de Leite - Leite em Números – Mercado, Tabela 06.20 – Exportação brasileira de produtos lácteos – 1998/2005 – Embrapa, Brazil. Disponível em: <<http://www.cnp.gl.embrapa.br/producao/06mercado/tabela006.20.php>>. Acesso em: 05 de maio **2005**

<sup>6</sup> Departamento da Polícia Federal, Operação “Soro” em Recife/PE. Disponível em: <<http://dpf.gov.br/DCS/noticias/2004/março/050322004>>. Acesso em: 30 de Março **2004**.

<sup>7</sup> Nota à imprensa, Ministério Público denuncia quadrilha por fraude de leite, Disponível em: <<http://www.prpe.mpf.gov.br/internet/content/view/full/1869>> .Acesso em: 30 de Março **2004**.

<sup>8</sup> Rosa R, Santos R, Fraude no Leite é investigada, *Jornal do Brasil*, Disponível em: <[http://www2.rio.rj.gov.br/cgm/publicacoes/fraudes\\_corupcao](http://www2.rio.rj.gov.br/cgm/publicacoes/fraudes_corupcao)>. Acesso em: 19 de abril **2002**.

<sup>9</sup> BRASIL. Ministério da Agricultura, Pecuária e Abastecimento, Portaria n.124, de 23 de Setembro de **1991**. SISLEGIS – Sistema de Legislação Agrícola Federal, Disponível em: <<http://extranet.agricultura.gov.br/consultasislegis/do/consultaLei?op=viewTextual&codigo=5493>>. Acesso em: 05 de maio, **2005**

<sup>10</sup> Veloso A C A, Teixeira N, Ferreira, I M P L V O, Ferreira M A, Metodologias de doseamento das proteínas do leite, *Boletim de Biotecnologia*, **2001**, 17-31.

<sup>11</sup> Veloso A C A, Teixeira N, Ferreira, I M P L V O, Ferreira M A, Detecção de adulterações em produtos alimentares contendo leite e/ ou proteínas lácteas, *Química Nova*, 25(4), **2002**, 609 -615.

<sup>12</sup> Bobbio F O, Bobbio P A, *Introdução à Química dos Alimentos*, Editora UNICAMP, Campinas, **1985**.

<sup>13</sup> BRASIL. Ministério da Agricultura, do Abastecimento e da Reforma agrária. Portaria n.124, de 23 de Setembro de 1991. Disponível em: <<http://www.agricultura.gov.br>>. Acesso em: 05 de maio **2005**

<sup>14</sup> López- Fandiño R, Corzo N, Villamiel M, Delgado T, Olano A, Ramos M, Assessment of quality of commercial UHT milks by chromatographic and electrophoretic methods, *Journal of Food Protection*, 56, **1993**, 263-269.

<sup>15</sup> López-Fandiño R, Olano A, Corzo N, Ramos M, Proteolysis during storage of UHT milk: differences between whole and skim milk, *Journal of Dairy Research*, 60, **1993**, 339-347.

<sup>16</sup> da Silva P H F, *Leite UHT : fatores determinantes para sedimentação e gelificação*, EPAMIG, Juiz de Fora, **2004**.

- 
- <sup>17</sup> Recio I, López-Fandino R, Olano A, Olieman C, Ramos M, Study of the formation of caseinomacropptides in stored ultra-high-temperature-treated milk by capillary electrophoresis, *Journal of Agricultural and Food Chemistry*, 44, **1996**, 3845-3848.
- <sup>18</sup> Cáceres A, Cárdenas S, Gallego M, Valcárcel M, A continuous spectrophotometric system for the discrimination/ determination of monosaccharides and oligosaccharides in foods, *Analytica Chimica Acta*, 404, **2000**, 121-129.
- <sup>19</sup> Fertig C C, Podczeck F, Jee R D, Smith M R, Feasibility study for the rapid determination of the amylose content in starch by near-infrared spectroscopy, *European Journal of Pharmaceutical Sciences*, 21, **2004**, 155-159.
- <sup>20</sup> Skoog D A, Holler F J, Nieman T A, *Princípios de Análise Instrumental*, 5<sup>a</sup> Ed, Ed. Bookman, São Paulo, **2002**
- <sup>21</sup> Lee K A, Review of applications of near infrared spectroscopy to food analysis, *The NIR Spectrum*, 2(2), **2004**, 11- 16.
- <sup>22</sup> Osborne B G, *Near-infrared spectroscopy in food analysis*, Encyclopedia of Analytical Chemistry, Wiley, Chichester, **1999**.
- <sup>23</sup> Rodriguez-Otero J L, Hermida M, Centeno J, Analysis of Dairy products by near-infrared spectroscopy: A Review, *Journal of Agricultural and Food Chemistry*, 45(8), **1997**, 2815-2819.
- <sup>24</sup> Holroyd S, Meagher L, Illingworth F, At-line near infrared spectroscopy for the measurement of solid fat content in the dairy industry, *NIR News*, 18(1), **2007**, 6-8.
- <sup>25</sup> Sasic S, Ozaki Y, Short-wave near-infrared spectroscopy of biological fluids.1. Quantitative analysis of fat, protein, and lactose in raw milk by partial least-squares regression and band assignment, *Analytical Chemistry*, 73, **2001**, 64 -71.
- <sup>26</sup> Jha S N, Matsuoka T, Detection of adulterants in milk using near infrared spectroscopy, *Journal Food Science and Technology*, 41(3), **2004**, 313-316.
- <sup>27</sup> Ferrão M F, Aplicações de técnicas Espectroscópicas de reflexão o infravermelho no controle de qualidade de farinha de trigo, Tese de Doutorado. Instituto de Química. UNICAMP **2000**.
- <sup>28</sup> Maretto D A, Comparação entre máquinas de vetores de suporte por mínimos quadrados (LS-SVM) e métodos lineares para transferência de calibração, Dissertação de Mestrado, Instituto de Química, UNICAMP, **2007**.
- <sup>29</sup> 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**.
- <sup>30</sup> Pelckmans K, Suykens J A K, Van Gestel T, de Brabanter J, Lukas L, Hamers B, de Moor B, Vandewalle J, *LS-SVMlab Toolbox User's Guide version 1.5*, Department of Electrical Engineering, Katholieke Universiteit Leuven, **2003**.
- <sup>31</sup> Murray I, Williams PC, *Chemical Principles of Near-infrared technology* in Williaams P, Norris K, *Near-infrared technology in the agricultural and food industries*, American Association of Cereal Chemists, Minesota, **1990**.
- <sup>32</sup> Frankhulzen R, *NIR Analysis of Dairy Products*, in Burns DA, Ciurczak EW, *Handbook of Near-infrared Analysis*, 2<sup>a</sup> Ed. Marcel Dekker , New York, **2001**.

# Capítulo 3

**APLICAÇÃO 2 – QUANTIFICAÇÃO DE LACTOBACILLUS  
EM LEITE FERMENTADO POR IMAGEM ESCANEADA E  
MÁQUINAS DE VETORES DE SUPORTE POR MÍNIMOS  
QUADRADOS (LS-SVM).**

### **3.1. Introdução**

#### **3.1.1. Os Lactobacillus**

Os Lactobacillus pertencem a um gênero de microrganismos com cerca de 56 espécies e fazem parte do grupo das bactérias lácticas, que também inclui os *Leuconostoccus*, *Pediococcus*, *Streptococcus*, *Lactococcus* e os *Enterococcus*<sup>1</sup>. Eles são microrganismos relacionados com a produção de alimentos de alta e média acidez, da qual podem participar como coadjuvantes da fabricação (leites fermentados, iogurtes, queijos, salames, presuntos, chucrute e outros vegetais fermentados) ou como deteriorantes, respondendo por perdas significativas na indústria de sucos, refrigerantes, maioneses etc<sup>2</sup>.

Os leites fermentados são basicamente constituídos somente de leite e fermentos. De acordo com a Federação Internacional de Laticínios (IDF), leite fermentado é um produto elaborado a partir de leite, desnatado ou não, com culturas lácticas específicas, mantidas vivas até a aquisição do consumidor, e não apresentando patógenos. As culturas utilizadas incluem várias bactérias, dentre elas: *Lactobacillus*, *Streptococcus*, *Propionibacterium*, *Leuconostoc*, *Bifidobacterium* e *Acetobacter*<sup>3</sup>. O termo bebida láctea fermentada, segundo o regulamento técnico de identidade e qualidade de bebidas lácteas do Ministério da Agricultura, é o produto lácteo (mistura de leite e soro de leite) fermentado mediante a ação de cultivo de microrganismos específicos e/ou adicionados de leite(s) fermentados e que não pode ser submetido a tratamento térmico após a fermentação. A contagem total de bactérias lácticas viáveis deve ser no mínimo de  $10^6$  ufc/g (unidades formadoras de colônias por grama), no produto final, para o(s) cultivo(s) láctico (s) específico (s) empregado(s), durante todo o prazo de validade<sup>4</sup>.

#### **3.1.2. Métodos padrões de Contagem de Bactérias**

Em microbiologia, particularmente na bacteriologia, o crescimento microbiológico é utilizado como referência à grandeza da população total e pode ser determinado por numerosas técnicas<sup>5</sup>. Porém, o método mais comumente

utilizado para contagem do número de células numa população é a contagem de células viáveis, também chamada de contagem em placa. Trata-se de um método indireto que se baseia no princípio de que cada célula viável, quando presente num meio sólido, pode se multiplicar repetidas vezes e originar uma colônia visível a olho nú<sup>6</sup>.

### **3.1.3. Interface entre Química Analítica e Microbiologia**

A interface entre Química Analítica e Microbiologia vem se ampliando à medida que a instrumentação analítica se aprimora, bem como o tratamento dos muitos dados gerados. Entre as técnicas modernas<sup>7</sup> para caracterização e detecção de microorganismos existem a cromatografia gasosa, eletroforese, espectrometria de massas, fluorescência, ressonância magnética nuclear, espectroscopia no infravermelho<sup>8</sup> e Raman<sup>9</sup>.

Dentre estas técnicas inovadoras, há a análise de imagens multivariadas<sup>10</sup>. O conceito de imagem multivariada está ligado à presença de vários canais (respostas) para o valor de intensidade do pixel. Seu desenvolvimento surgiu com o uso de imagens coloridas na microscopia ótica, e continuou com a introdução de imagens espectrais nas várias áreas da ciência. O caso mais simples de imagem multivariada são as imagens coloridas e decompostas nas cores primárias RGB (Red, Green, Blue)<sup>11</sup>.

Dentre as diferentes maneiras de aquisição das imagens: projeção, varredura, tomografia entre outros, há o escaneamento do objeto a ser estudado<sup>12</sup>. Entre as vantagens de se obter imagens através de scanner estão: o baixo custo e a multifuncionalidade do instrumento, a não necessidade de controle de iluminação e, depois de adquirida a imagem, esta pode ser armazenada para posterior interpretação<sup>13</sup>.

Em relação a trabalhos empregando imagens, é muito comum a análise de imagens para classificação e/ou reconhecimento de padrões. Em 2004, Antonelli *et al.*<sup>14</sup> empregaram um histograma, constituído das frequências dos parâmetros das cores e também dos escores obtidos pela Análise de Componentes Principais (PCA). Este histograma foi usado para a classificação da qualidade de molhos de

manjeriço, pela imagem das cores destes, a partir de uma câmera digital. Porém, o uso dos parâmetros das cores, que forma o histograma, possui poucos trabalhos com foco para quantificação. Em 2002, Gabrielson *et al.*<sup>13</sup> utilizaram scanner com objetivo de quantificação. Porém, neste caso não foi usado histograma de cores, mas sim o relacionamento das cores produzidas pelos corantes com a medida obtida em determinados comprimentos de onda no espectrofotômetro.

Pode-se encontrar a patente US2005053266-A1<sup>15</sup>, que emprega a leitora 3M para Placas Petrifilm para contagem de Aeróbios, Coliformes ou *E.coli*/coliformes. Esta leitura fica condicionada ao uso das placas Petrifilm, específicas para os determinados organismos, acima citados. A patente US2006176400-A1<sup>16</sup> descreve uma forma de processamento de imagem, onde se gera o histograma para contar o número de pixels para cada cor pré-determinada na imagem. Neste método o processamento da imagem também emprega o uso de histograma, porém, não foi usada para a quantificação de microrganismos, como é a proposta neste trabalho, no caso quantificação de *Lactobacillus*.

#### **3.1.4. Imagem colorida RGB**

A cor é uma propriedade de grande importância nas indústrias alimentícias<sup>17,18</sup>, textil, fotografia, e em muitas outras aplicações. As cores podem ser simuladas através da mistura de três cores, chamadas primárias: vermelha, verde e azul (das siglas em inglês referentes a estas cores: Red (R), Green (G) e Blue (B)).

Uma imagem colorida RGB é um tensor  $M \times N \times 3$  de pixels de cor, onde cada cor é um triplete correspondente aos componentes vermelho, verde e azul da imagem RGB. Na Figura 26 pode-se observar uma representação de um tensor com três matrizes correspondendo as matrizes de valores de R, de G e de B.

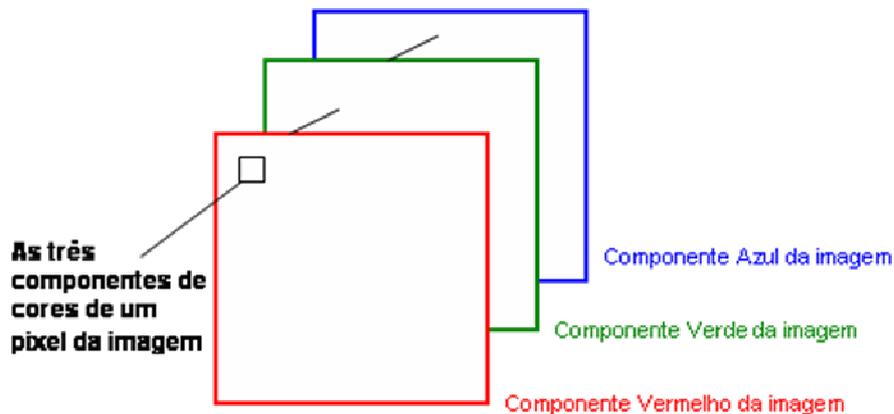


Figura 26. Representação de um tensor com matrizes de valores de R, de G e de B.

O número de bits usados para representar os valores do pixel dos componentes da imagem determina a resolução da imagem RGB. Por exemplo, se cada componente da imagem é uma imagem de 8 bits, ou seja, valores de pixel de  $[0, 255]$ , o valor da resolução da imagem RGB é de 24 bits. Neste caso, o número de cores possíveis é  $(2^b)^3$ , onde  $b$  é o número de bits em cada componente da imagem, ou seja, aproximadamente 16,8 milhões de combinações possíveis de cor. Na análise de imagens, geralmente, é realizada a decomposição da imagem colorida em imagens de cores primárias, gerando três imagens que serão processadas como uma imagem multivariada<sup>19</sup>.

### 3.1.5. Convertendo cores de RGB para HSI

As cores formadas na composição da base RGB também podem ser expressas em outros parâmetros de composição de cores, que vão representar outras características das cores. Estes parâmetros são o contraste (do inglês Hue, H), a saturação (Saturation, S) e a intensidade (Intensity, I).

As equações para conversão de valores de RGB para valores de HSI são dadas por equações a seguir. O componente de contraste (H) é dado usando a seguinte equação:

$$H = \begin{cases} \theta & \text{se } B \leq G \\ 360 - \theta & \text{se } B > G \end{cases} \quad (\text{Eq.20})$$

onde:

$$\theta = \cos^{-1} \left\{ \frac{\frac{1}{2}[(R-G) + (R-B)]}{\left[ \frac{(R-G)^2 + (R-B)(G-B)}{2} \right]^{1/2}} \right\} \quad (\text{Eq. 21})$$

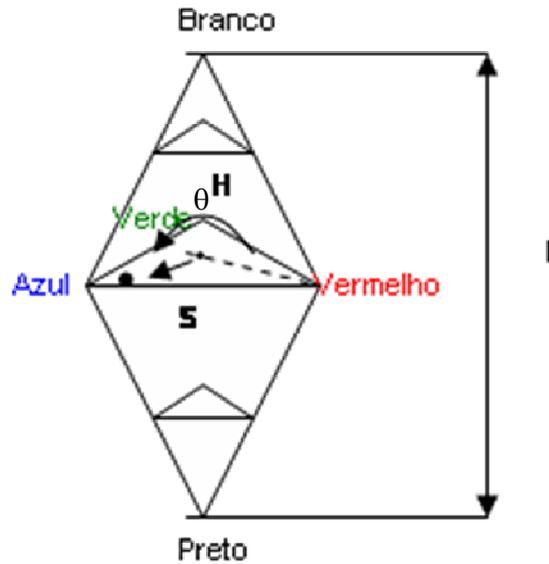
O componente da saturação (S) é dado por

$$S = 1 - \frac{3}{(R+G+B)} [\min(R, G, B)] \quad (\text{Eq. 22})$$

Finalmente, o componente da intensidade (I) é dado por

$$I = \frac{1}{3}(R+G+B) \quad (\text{Eq. 23})$$

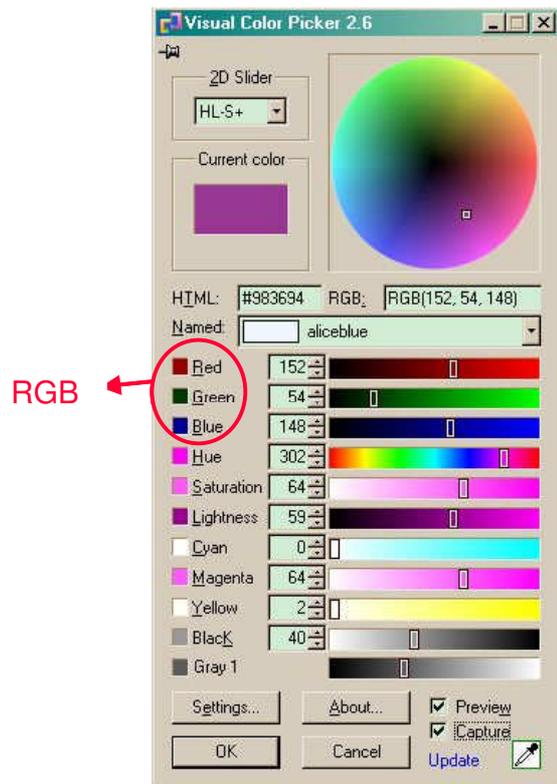
Uma representação gráfica das cores em valores de HSI a partir de RGB pode ser observada na figura 27.



**Figura 27.** Representação gráfica das cores em valores de HSI, a partir de RGB. A bolinha preta (•) na figura é um exemplo de cor.

O valor do contraste H é dado em grau ( $^{\circ}$ ) dependendo em que ângulo ( $\theta$ ) o contraste da cor está em relação a cor de partida, que é o vermelho como pode ser observado na Figura 27. A intensidade (I) varia no eixo vertical tendo como extremidades as cores branco e preto. Já, a saturação depende da mistura entre RGB partindo do ponto central da mistura destas.

Observa-se na Figura 28 um exemplo de uma cor com os valores de RGB e HSI.



**Figura 28.** Valores RGB indicadas pela flecha, referente a cor indicada neste exemplo.

Para a formação do histograma de cores das imagens usadas neste trabalho, foram utilizadas todas as formas de combinação para a formação das cores, ou seja, as cores na base RGB e na base HSI, porque cada uma delas é uma base importante na definição específica das cores.

## 3.2. Parte Experimental

### 3.2.1. Preparação da amostra e diluições seriadas

O leite fermentado usado neste experimento é constituído de leite desnatado, açúcar e culturas lácteas como indicado pelo fabricante. Todo o procedimento de plaqueamento para a contagem foi realizado em capela de fluxo laminar, com

presença do bico de Bunsen, com as vidrarias e meios previamente esterilizados em autoclave a 120°C por 20 minutos.

Quatro soluções estoque foram preparadas transferindo-se 1,0 mL do leite fermentado (de quatro embalagens diferentes do mesmo lote) em 9,0 mL de água esterilizada. De cada solução estoque foram realizadas diluições sucessivas de 1,0 mL para 9,0 mL de água até chegar à diluição de  $10^5$  ufc/mL (unidades formadoras de colônias por mL).

A transferência do volume foi feita por pipetas estéreis e diferentes para cada diluição. Antes da tomada do volume a ser diluído agitou-se vigorosamente o tubo em agitador tipo vórtex<sup>20</sup>.

### **3.2.2. Inoculação**

Preparou-se o meio de cultura onde foi utilizado o ágar Man-Rogosa Sharpe agar (MRS agar -Fluka Biochemika) que é seletivo para *Lactobacillus*. No preparo deste meio, foi adicionado 1,20 g do marcador celular Tetrazolium blue chloride (Fluka, AB), que neste caso teve apenas a função de deixar o meio com cor azul, para maior contraste na obtenção das imagens.

As placas de Petri usadas foram de tamanho e alturas os mais semelhantes possíveis. As placas de Petri contendo 30 mL de meio MRS agar foram inoculadas com 1,0 mL das diluições bacterianas. Depois disto, cada placa de Petri foi coberta com 4,0 mL de meio MRS para criar uma condição microaeróbica. Para a diluição de cada solução estoque, foi realizado este plaqueamento em triplicata, sendo que no total foram incubadas 60 placas.

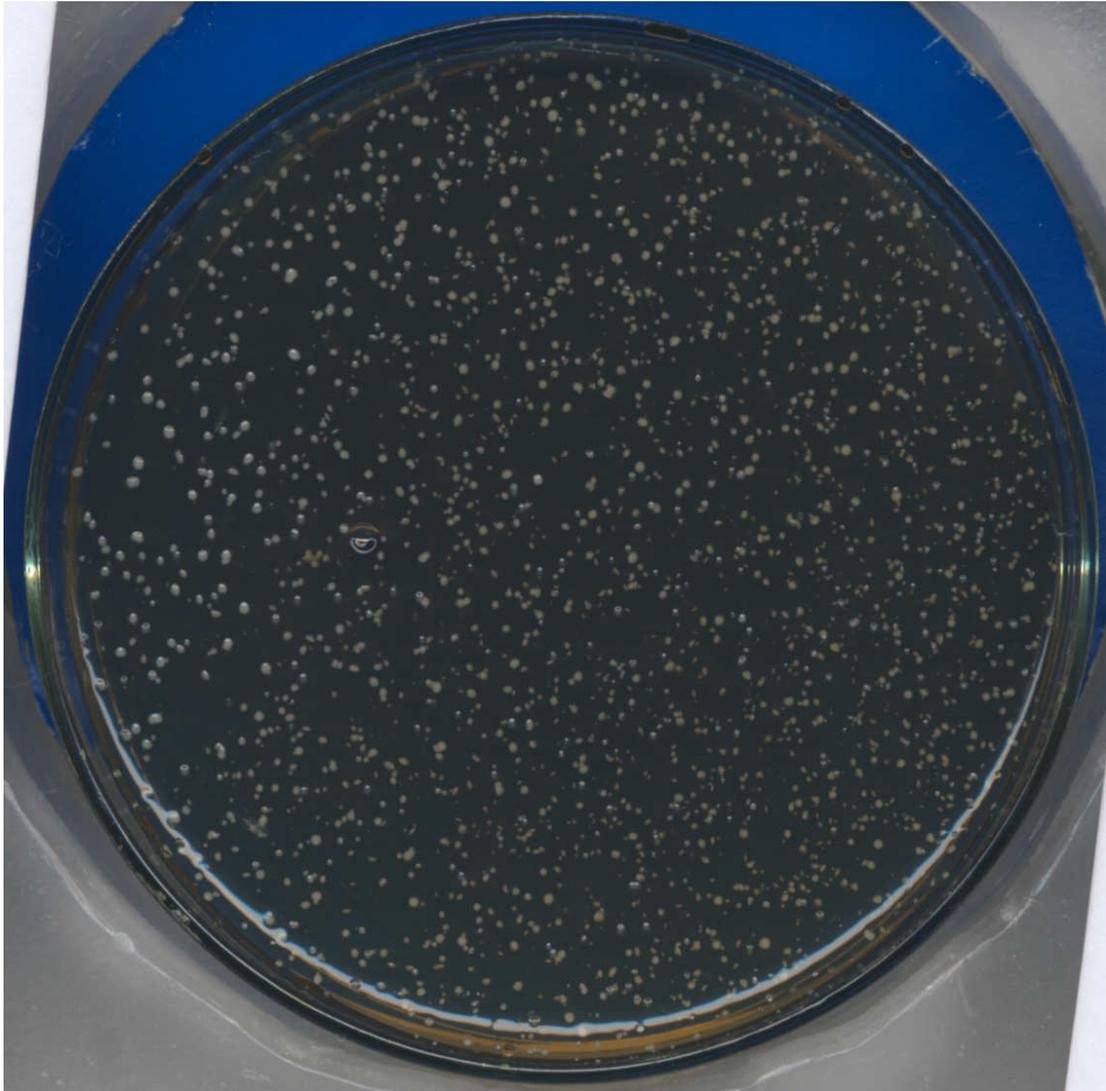
### **3.2.3. Incubação**

Após a inoculação, as placas foram colocadas em estufa microbiológica a 37°C por 48 horas.

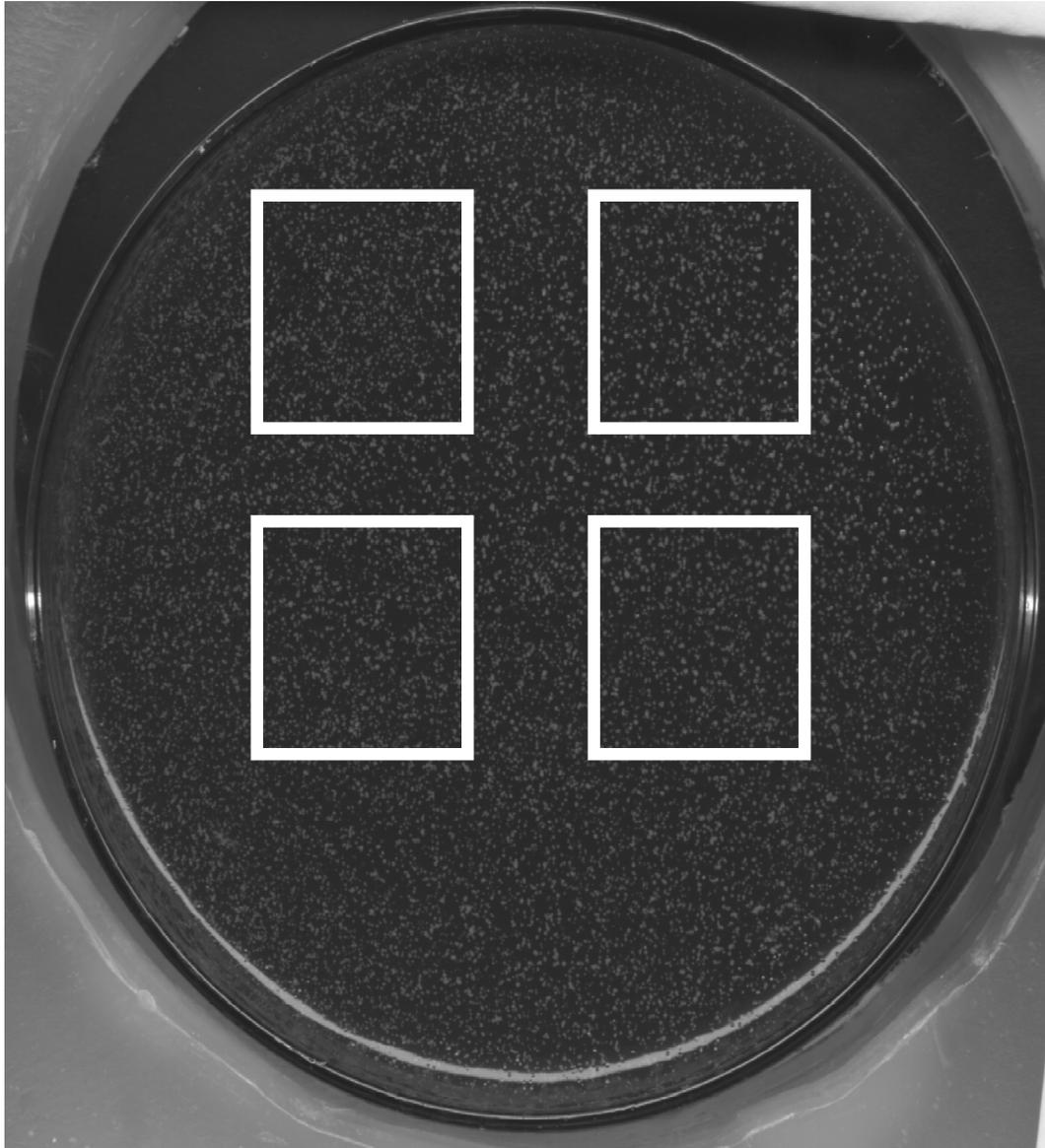
### **3.2.4. Obtenção e tratamento das imagens das placas**

Depois das 48 horas de crescimento, foram obtidas as imagens das placas. Estas imagens foram obtidas no scanner Canon, modelo CanoScan D646U, que é

um scanner de mesa comum. Para a aquisição de imagens das placas de Petri, foi usado um suporte de parafina (modelado para as placas) e colocado sobre o scanner. Para um melhor contraste, foi usada uma cartolina azul entre a placa e este suporte de parafina. Nas Figuras 29 e 30, pode-se observar as imagens das placas bacterianas nas diluições  $10^4$  e  $10^3$  ufc/ml, respectivamente. As placas foram escaneadas com a placa aberta (sem a tampa). Na Figura 30, referente à placa com diluição  $10^3$  ufc/ml, pode-se observar um maior número de unidades formadoras de colônias, na qual é praticamente impossível se realizar a contagem visual. Em ambas as diluições, o crescimento das unidades formadoras de colônias foi homogêneo, ou melhor, bem distribuído em toda a placa.



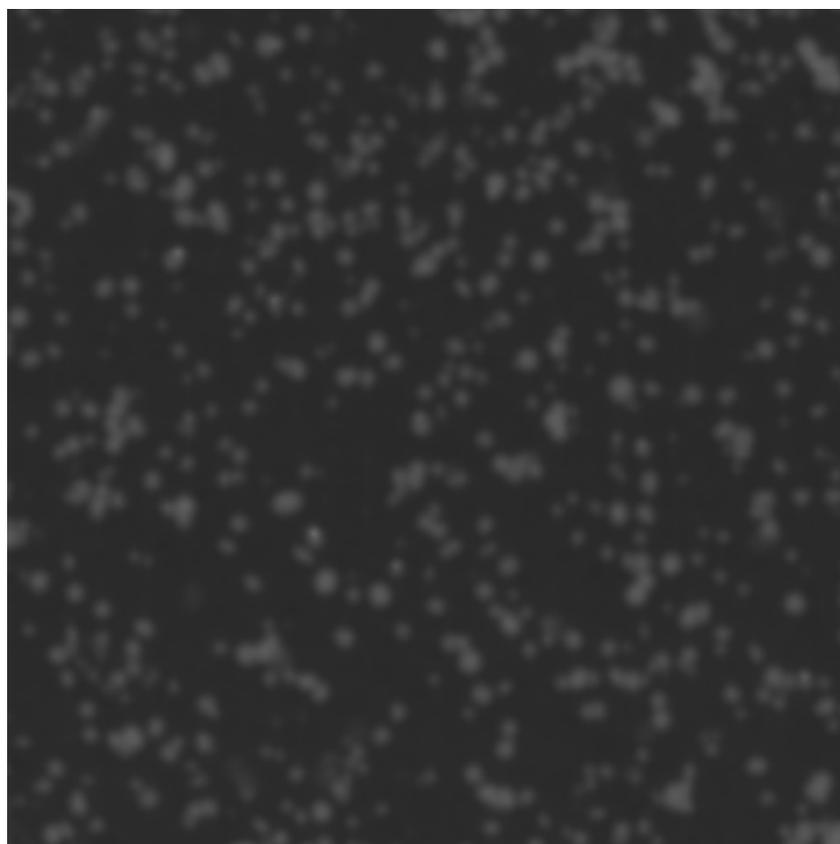
**Figura 29.** Imagem de Placa de crescimento bacteriano na diluição de  $10^4$  ufc/ mL.



**Figura 30.** Imagem da placa de crescimento bacteriano na diluição de  $10^3$  ufc/ mL com a indicação das áreas escolhidas (quadrado) em cada quadrante da placa. Esta indicação dos quadrados é apenas uma representação: está fora da escala real.

Todas as imagens foram obtidas com resolução de 24 bits RGB (16.8 milhões de cores) e 300 dpi (dpi – do inglês, pontos por polegada). Estes parâmetros foram fixos para obtenção das imagens no scanner. Estas imagens foram salvas no formato tiff (Target Image File Format). A partir de cada imagem, foram selecionadas regiões quadradas de 300 x 300 pixels em cada quadrado de cada quadrante utilizando o editor de imagens GNU Image Manipulation Program (GIMP).

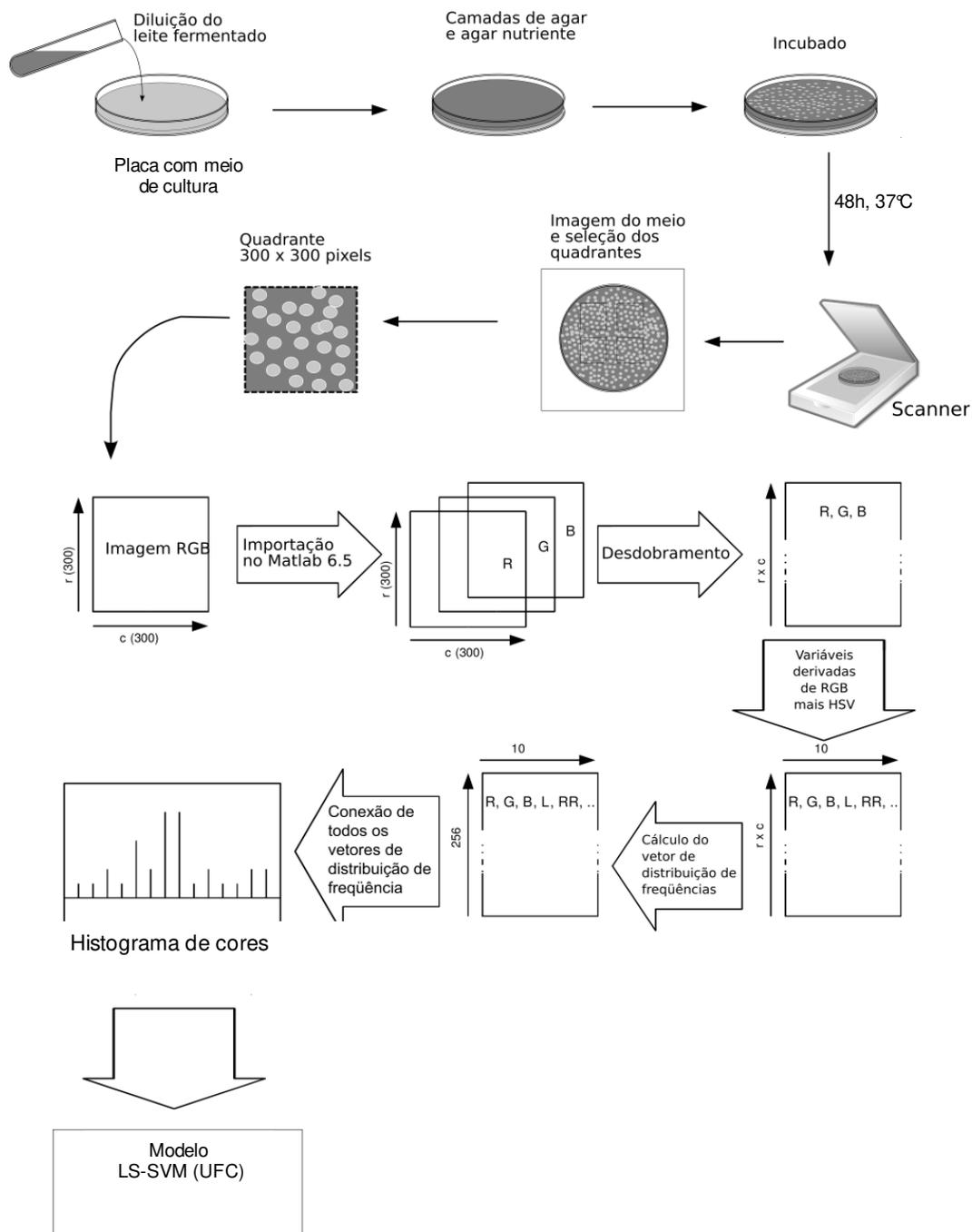
Na figura 31 é apresentada a imagem de um quadrado de cada quadrante escolhido da placa de crescimento bacteriano na diluição de  $10^3$  ufc/ml.



**Figura 31.** Imagem do quadrado de um dos quadrantes da placa de crescimento na diluição  $10^3$  ufc/ mL.

No GIMP, as imagens dos quadrantes foram salvas no formato Tiff, sem a camada de transparência, que também são conjunto de informações sobre as

cores, porém não utilizadas neste trabalho. Fez-se a importação das imagens no programa Matlab®, versão 6.5, gerando um tensor tridimensional {300x300x3} (canais vermelho (R), verde (G) e azul (B)). O tensor foi desdobrado numa matriz de três colunas (R, G e B) e 90000 linhas (300x300). As três variáveis R, G e B podem assumir valores inteiros entre 0 e 255. A partir desta matriz, foram criadas novas variáveis, usando operações algébricas descritas por Antonelli<sup>14</sup>. Primeiramente, foi calculada a luminosidade (L), que é a soma de R+G+B. Calculou-se ainda os valores relativos das três cores – vermelho relativo (RR), verde relativo (RG), e azul relativo (RB) – que são as razões entre os valores de R, G, B e sua soma (L). Na seqüência, foram determinados o contraste (H), a saturação (S) e a intensidade (I) transformando a imagem no padrão HSI. Calculou-se o vetor de distribuição de freqüências de valores 0 a 255, para cada uma das variáveis descritas anteriormente: R, G, B, L, RR, RG, RB, H, S, I. Os vetores resultantes para cada variável foram unidos, formando um vetor único com comprimento de 2560 (256x10) onde foram graficados os histogramas, como apresentado no esquema da Figura 32.



**Figura 32.** Esquema da obtenção do histograma de cores a partir das imagens escaneadas das placas microbiológicas.

### 3.2.5. Escolha das amostras para calibração e previsão

A matriz, referente aos histogramas dos quadrantes, foi ordenada em ordem crescente de unidades formadoras de colônias (ufc) e a escolha das

amostras para o conjunto de previsão foi feita pelo critério de selecionar uma amostra no intervalo de outras três (ou seja, de três em três amostras) e que também não estivesse nos extremos das faixas do número de unidades formadoras de colônias (ufc).

#### **3.2.6. Modelagem dos dados**

Conforme item 2.2.6.

#### **3.2.7. Avaliação do desempenho dos modelos**

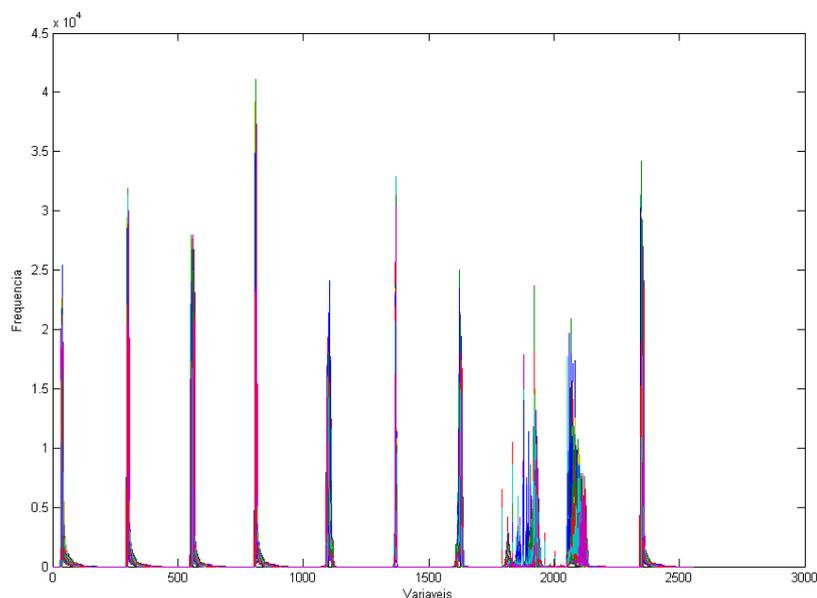
Conforme item 2.2.7.

### **3.3. Resultados e Discussões**

#### **3.3. 1. Histograma e Contagem das colônias na Placa de crescimento**

Para o desenvolvimento dos modelos foram escolhidas as placas das diluições  $10^3$  e  $10^4$  ufc/mL. Assim, das 60 placas iniciais (referente às cinco diluições), foram obtidas as imagens de 24 placas (referentes às diluições  $10^3$  e  $10^4$  ufc/mL). E destas, obtiveram-se 96 quadrantes.

Porém, destes 96 quadrantes, sete foram descartados devido à presença de bolhas ou outras falhas no meio de cultura, conseqüentemente afetando as imagens, restando as imagens dos 89 quadrantes cujos histogramas são apresentados na Figura 33.

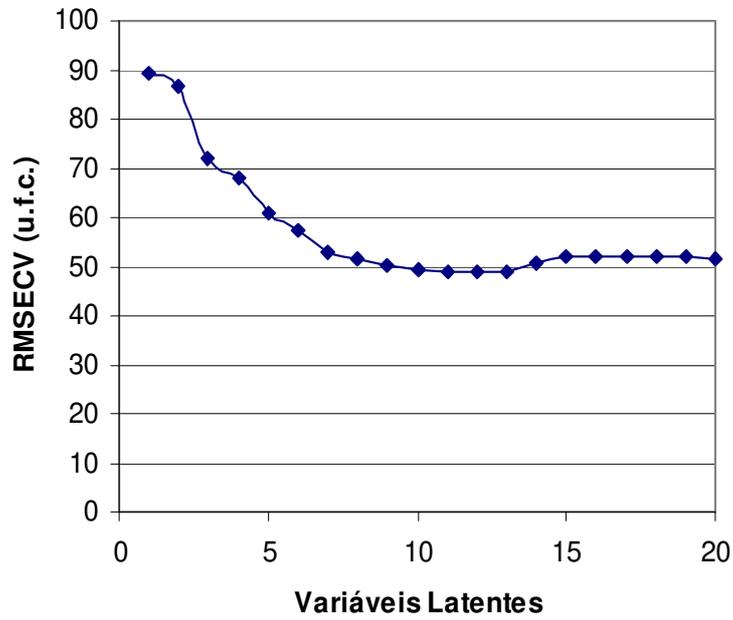


**Figura 33.** Histograma de cores.

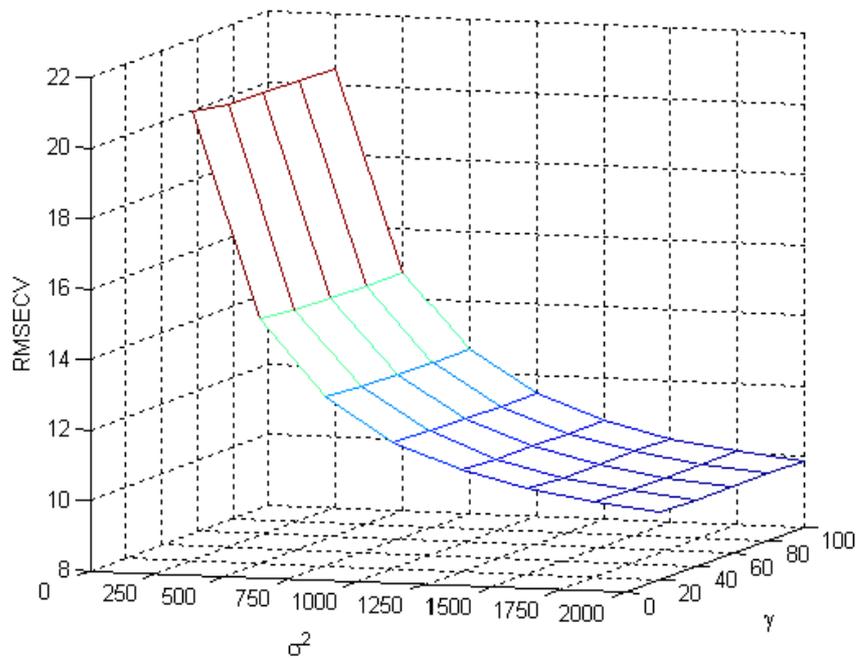
Foi feita a contagem das unidades formadoras de colônias (ufc) da imagem de cada quadrante, ou seja, na própria tela do computador. Os valores de contagem ficaram na faixa de 28 a 366 ufc. A porcentagem do desvio para a contagem das unidades formadoras de colônias (ufc) dos 45 quadrantes da diluição  $10^3$  ufc/mL foi de 10,84% e o desvio para os 44 quadrantes da diluição  $10^4$  ufc/mL foi de 20,38%.

### 3.3.2. Resultados dos Modelos PLS e LS-SVM

Para o desenvolvimento dos modelos PLS e LS-SVM, foi usado o conjunto de calibração com 65 imagens dos quadrados, e, para a previsão, 24 imagens dos quadrados. Na Figura 34, são apresentados os valores de RMSECV contra o número de variáveis latentes do PLS e, na Figura 35, a superfície com a otimização dos parâmetros da LS-SVM.



**Figura 34.** Valores de RMSECV contra o número de variáveis latentes.



**Figura 35.** Superfície de otimização dos parâmetros para modelagem LS-SVM.

Pela Figura 34, observa-se que, para desenvolver um modelo com o erro médio quadrático da validação cruzada (RMSECV) menor possível, há a necessidade do uso de um número consideravelmente grande, a partir de sete variáveis latentes, o que já é um indício que o modelo linear não está sendo adequado para ajustar os valores de  $\mathbf{X}$  (no caso, a matriz dos histogramas de cores) com os valores de  $\mathbf{Y}$  (no caso, os valores de unidades formadoras de colônias). Para o modelo do PLS foram empregadas sete variáveis latentes.

Na figura 35 é apresentada a superfície entre os parâmetros  $\sigma^2$  e  $\gamma$  em relação ao RMSECV. Foram realizadas várias combinações testando valores de  $\sigma^2$  e  $\gamma$  até se obter esta superfície de resposta apresentando um mínimo de RMSECV em função destes dois parâmetros. Nestas condições, foram escolhidos para o modelo LS-SVM um  $\gamma$  de 20 e um  $\sigma^2$  de 1750.

Por fim, os resultados obtidos para o modelo PLS e para o LS-SVM estão apresentados na Tabela 11.

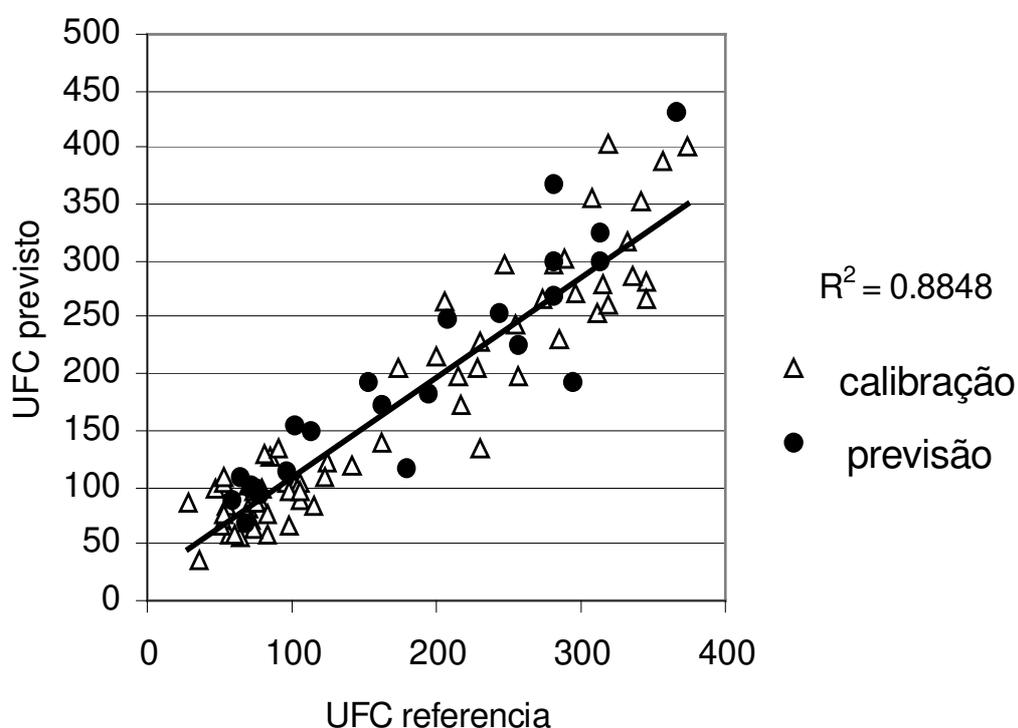
**Tabela 11.** Parâmetros para os modelos de PLS e LS-SVM usando os quadrantes.

	PLS	LS-SVM
<b>R<sup>2</sup></b>	0,8848	0,9982
<b>RMSECV (ufc)</b>	52,84	10,13
<b>RMSEC (ufc)</b>	36,33	6,83
<b>RMSEP (ufc)</b>	41,56	22,16
<b>Variáveis Latentes</b>	7	-
<b><math>\gamma</math></b>	-	20
<b><math>\sigma^2</math></b>	-	1750

Como pode se observar na Tabela 8, os valores do coeficiente de correlação dos dados ( $R^2$ ), os erros médios quadráticos da validação cruzada, da

calibração e da previsão (RMSECV, RMSEC e RMSEP, respectivamente) são melhores para o modelo LS-SVM, em relação ao modelo PLS.

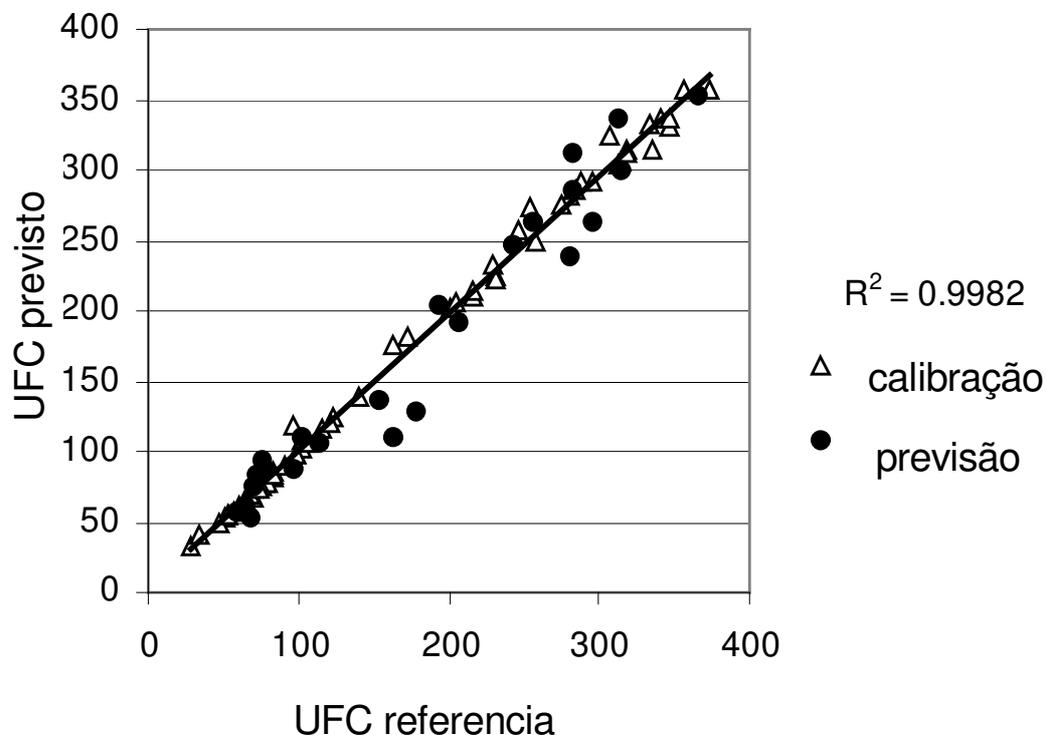
Na Figura 36, apresenta-se o gráfico com os valores de referência versus os valores previstos dos conjuntos de calibração e de previsão para o modelo PLS. Na Figura 37, são apresentados os valores de referência versus os valores previstos, também para os conjuntos de calibração e previsão pelo modelo LS-SVM.



**Figura 36.** A. Gráfico dos valores de ufc de referência versus valores previstos pelo modelo PLS.

Na Figura 36, com o gráfico dos valores de referência versus os de previsão para os conjuntos de calibração e previsão. Nota-se que as amostras não estão muito próximas da reta de correlação, ou seja, estão bem dispersas. Algumas amostras estão até significativamente distantes da reta, apresentando um coeficiente de correlação ( $R^2$ ) de 0,88. Já com o modelo LS-SVM, as amostras se

distribuem melhor e mais próximas ao longo da reta de correlação, com  $R^2$  aproximado de 1,00, como mostra a Figura 37.



**Figura 37.** A. Gráfico dos valores de ufc de referência versus valores previstos pelo modelo LS-SVM.

Uma hipótese para inadequação do modelo linear PLS é que o número de colônias na diluição  $10^3$  ufc/mL é maior e o tamanho delas menor. O efeito oposto ocorre na diluição  $10^4$  ufc/mL, ou seja, têm-se menos colônias, mas com tamanhos maiores. Numa mesma diluição a diferença entre os tamanhos das colônias não é significativa. Isso é natural ocorrer em microbiologia, mesmo considerando que as condições de incubação e crescimento são constantes (igual quantidade de material nutriente, temperatura, tempo de crescimento etc).

Por fim, a proposta apresentada neste trabalho tenta minimizar a limitação da contagem visual humana, que utiliza um menor número de diluições para o plaqueamento, cuja imagem é obtida por scanner (um instrumento acessível e

dispensa controle de iluminação) e depois de adquirida a imagem, esta pode ser armazenada para posterior contagem e interpretação.

### **3.3.3. Conversão de ufc por quadrado de cada quadrante para ufc/ ml por placa**

É usual em microbiologia a contagem de uma área menor dentro de uma placa e fazer a conversão proporcional para a área da placa inteira. A área do quadrado de cada quadrante é de 90000 pixels<sup>2</sup> e a área da placa, seguindo uma média dos diâmetros das placas usadas, é obtida multiplicando a área de cada quadrado por 36. Ou seja, os valores de ufc contados em cada quadrado foram multiplicados por 36 para se obter o número de ufc para a placa inteira.

Esta proporcionalidade calculada matematicamente pode ser considerada válida para as diluições utilizadas neste trabalho, pois como pôde se observar nas Figuras 29 e 30 a distribuição das colônias é homogênea pela placa. Depois de se obter os resultados de unidades formadoras de colônia proporcional a área da placa (número de ufc) calcula-se ufc/ml, seguindo a Equação 24.

$$\text{Número de ufc} * \text{fator de diluição} = \text{ufc/ mL}^{21} \quad (\text{Eq. 24})$$

Os resultados dos cálculos para a unidade de ufc/ mL estão apresentados na Tabela 12.

**Tabela 12.** Valores do conjunto de referência e as previsões para os modelo LS-SVM e PLS.

Referência (ufc/ mL) x10 <sup>6</sup>	Previsto PLS (ufc / mL) x10 <sup>6</sup>	Previsto LS-SVM (ufc/ mL) x10 <sup>6</sup>
5,51	6,93	4,91
5,87	6,20	3,94
6,44	4,17	4,65
6,98	6,58	7,34
7,45	8,91	6,92
8,75	9,11	8,90
9,22	8,14	9,49
10,10	13,30	8,60
10,20	9,71	10,30
10,20	10,80	11,30
10,20	15,50	12,70
10,60	6,92	9,46
11,30	10,70	10,80
11,30	11,70	12,10
20,90	31,90	20,70
23,40	39,40	22,40
24,50	24,90	18,90
25,60	37,00	27,10
25,90	35,50	30,20
27,40	36,00	33,90
28,10	33,40	31,90
34,90	40,70	31,40
36,70	55,80	39,40
41,00	53,50	38,40
Média dos erros relativos (%)	21,97	11,39

Na Tabela 12, são apresentados os valores de ufc/mL para a placa inteira, os quais usando o PLS, apresenta uma média dos erros relativos das amostras de 21.97% . Para o modelo LS-SVM a média dos erros relativos é de 11.39%.

Os resultados desta aplicação são apresentados no artigo: · Borin A, Ferrão M F, Mello C, Cordi L, Pataca L M, Poppi R J, Quantification of Lactobacillus in fermented milk by multivariate image analysis and least squares support vector machines, *Analytical and Bioanalytical Chemistry*, 387, **2007**, 1105.

### 3.4. Referências Bibliográficas

- 
- <sup>1</sup> Pelczar M, Reid R, Chan E C S, *Microbiologia*, vol.2, McGraw-Hill, São Paulo, **1981**.
  - <sup>2</sup> Curk M C, Peladan F, Hubert J C, Fourier transform infrared (FTIR) spectroscopy for identifying Lactobacillus species, *FEMS Microbiology Letters*, 123, **1994**, 241-248.
  - <sup>3</sup> IDF Bulletin of the international Dairy Federation, 117, **1983**, 1– 4.
  - <sup>4</sup> BRASIL. Ministério da Agricultura. Regulamento técnico de identidade e qualidade de bebidas lácteas. Disponível em: <<http://www.agricultura.gov.br>>. Acesso em: 05 de maio **2005**.
  - <sup>5</sup> Silva F<sup>o</sup> G N, Oliveira V L, *Microbiologia – Manual de Aulas Práticas*, ED UFSC, Florianópolis, **2004**.
  - <sup>6</sup> Marshall R T, Adams D M, Morgan D R, Olsen N F, White C H, *Mirobiological methods for dairy products*; Richardson G H, *Standard methods for the examination of Dairy products*, 15<sup>a</sup>Ed., **1985**.
  - <sup>7</sup> Nelson W H, *Modern Techniques for Rapid Microbiological Analysis*, VCH Publishers, New York, **1991**.
  - <sup>8</sup> Whittaker P, Mossoba M M, Al-khaldi S, Fry F S, Dunkel V C, Tall B D, Yurawecz M P, *Journal of Microbiological Methods*, 55, **2003**, 709-716.
  - <sup>9</sup> Lopez-Diez E C, Goodacre R, Characterisation of microorganisms using UV. resonance Raman spectroscopy and chemometrics, *Analytical Chemistry*, **2004**, 76(3), 585-589.
  - <sup>10</sup> Geladi P, Grahn H, *Multivariate Image Analysis*. First Edition. New York: Ltd, John Wiley & Sons, **1996**.
  - <sup>11</sup> Gonzalez R C, Woods R E, Eddins S L, *Digital image processing using Matlab* Pearson Prentice Hall, London, **2004**.
  - <sup>12</sup> Byrne L, Barker J, Pennarun-Thomas G, Diamond D, Digital imaging as a detector for generic analytical measurements, *Trends in Analytical Chemistry*, 19 (8), **2000**, 517- 522.
  - <sup>13</sup> Gabrielson J, Hart M, Jarelov A, Kuhn I, McKenzie D, Mollby R, Evaluation of redox indicators and the use of digital scanners and spectrophotometer for

---

quantification of microbial growth in microplates, *Journal of Microbiological Methods*, 50, **2002**, 63-73.

<sup>14</sup> Antonelli A, Cocchi M, Fava P, Foca G, Franchini GC, Manzini D, Ulrici A, *Analytica Chimica Acta*, 515, **2004**, 3-13.

<sup>15</sup> Plumb M R, Binsfeld C A, Potter D T, Methods for detection and enumeration of biological agents (especially bacterial colonies) by analysis of image of bacterial growth plates, *US Patent n<sup>o</sup> US2005053266-A1*, **2005**.

<sup>16</sup> Shimizu M, Image processing method involves determining local peak in histogram generated by counting number of pixels in image, and specifying pixel in particular image based on local peak color value, *US Patent n<sup>o</sup> US2006176400-A1*, **2006**.

<sup>17</sup> Yu H, MacGregor F, Multivariate image analysis and regression for prediction of coating content and distribution in the production of snack foods, *Chemometrics and Intelligent Laboratory systems*, 67, **2003**, 125-144.

<sup>18</sup> Basset O, Buquet B, Abouelkaram S, Delachartre P, Culioli J, Application of texture image analysis for the classification of bovine meat, *Food Chemistry*, 69, **2000**, 437-445.

<sup>19</sup> Pataca L C M, Análises de Mel e Própolis Utilizando Métodos Quimiométricos de Classificação e Calibração, Tese de Doutorado, Instituto de Química, UNICAMP, **2006**.

<sup>20</sup> Silva N, Junqueira V C A, Silveira N F A, *Manual de métodos de análise microbiológica de alimentos*, 2<sup>a</sup> Ed., Ed. Varela, São Paulo, **2001**.

<sup>21</sup> Madigan M T, Martenko J M, Parker J, *Brock Biology of Microorganisms*, 9th Ed. Prentice Hall, New Jersey, **2000**.

# Capítulo 4

***APLICAÇÃO 3 – ANÁLISE EXPLORATÓRIA DE DIFERENTES TIPOS DE LEITE E DETERMINAÇÃO DO ÍNDICE DE GORDURA EM LEITE HUMANO POR REFLECTÂNCIA TOTAL ATENUADA NO INFRAVERMELHO MÉDIO (FTIR) E MÁQUINAS DE VETORES DE SUORTE POR MÍNIMOS QUADRADOS (LS-SVM).***

## 4.1. Introdução

### 4.1.1. Leite Humano

O aleitamento materno é considerado um dos pilares fundamentais para apoio, promoção e proteção da saúde das crianças em todo mundo. Porém, por questões clínicas do recém-nascido ou maternas, o bebê pode ficar impossibilitado de sugar ao seio após o parto e tem que se alimentar por outros métodos. Por estes motivos, muitos lactentes são alimentados com leites obtidos em Bancos de Leite Humano (BLH), produto de ações voluntárias de mulheres que têm produção de leite excedente<sup>1</sup>.

Basicamente, o leite materno<sup>2</sup> é uma mistura homogênea, rica em gordura, minerais, vitaminas, enzimas e imunoglobulinas, que suprem as necessidades nutritivas, metabólicas e digestivas do recém-nascido (RN), além de fornecer hormônios, aminoácidos essenciais e fatores imunológicos, tornando-se a melhor escolha para a nutrição dos RN, especialmente os prematuros e aqueles com doenças graves<sup>3</sup>.

A composição do leite humano varia no decorrer do ciclo da lactação: colostro (até sete dias após o parto), transição (entre o 7º e 15º dia pós-parto) e maduro (a partir do 15º dia). Varia também entre mulheres, numa mesma mulher e até em uma mesma mamada (início e fim)<sup>4</sup>. Dentro desta composição do leite humano, a gordura é o seu constituinte de maior oscilação, sendo influenciada por diversos fatores, entre eles: o intervalo entre as mamadas, a paridade (número de filhos gestados), idade da lactante, nível sócio-econômico, o tipo de ordenha, congelamento e degelo repetitivos, estado nutricional da nutriz, entre outros<sup>1</sup>. Aproximadamente 50% das calorias no leite humano são fornecidos aos recém-nascidos como gordura. A avaliação do teor energético do leite humano imediatamente antes da administração ao RN é importante para evitar o abandono do uso do leite humano em favor do uso de fórmulas lácteas sem a tentativa de melhorar seu conteúdo calórico-nutricional. Os compostos pertencentes ao grupo das gorduras do leite de vaca e derivados são caracterizados pela composição de:

30 a 40% de ácido oléico; 20 a 30% de ácido palmítico e 10 a 15% de ácido esteárico; é também o único grupo de gorduras que contêm ácido butírico, um ácido graxo de cadeia curta, que pode chegar até 15%<sup>5</sup>.

Na Tabela 13, é mostrada a média de alguns macronutrientes presentes no leite humano nas diferentes fases, do leite de vaca e da formulação infantil.

**Tabela 13.** Média de alguns macronutrientes na composição do leite humano<sup>6</sup>.

Nutriente	Colostro	Leite Humano Transição	Leite Humano maduro	Leite de vaca	Formulação infantil
Água (g)	88,2	87,4	<b>87,1</b>	87,8	87,9
Proteína (g)	2,0	1,5	<b>1,3</b>	3,2	1,4
Gordura (g)	2,6	3,7	<b>4,1</b>	3,9	3,6
Carboidrato (g)	6,6	6,9	<b>7,2</b>	4,8	7,1
Energia (Kcal)	56	67	<b>69</b>	66	64
Nitrogênio Total (g)	0,31	0,23	<b>0,20</b>	0,50	0,22
Ácidos graxos saturados (g)	1,1	1,5	<b>1,8</b>	2,4	1,4
Ácidos graxos mono-insaturados (g)	1,1	1,5	<b>1,6</b>	1,1	1,3
Ácidos graxos poli-insaturados (g)	0,3	0,5	<b>0,5</b>	0,1	0,6
Colesterol (mg)	21	24	<b>16</b>	14	6

Do ponto de vista físico, coexistem no leite três estados: **emulsão**- de matéria graxa sob forma celular; **suspensão** - de caseína ligada aos sais minerais; e **solução** ou fase hídrica - que são as substâncias dissolvidas na água.

A fração emulsão congrega os constituintes lipossolúveis – gordura, óleos, vitaminas, pigmentos e alguns ácidos graxos livres. Praticamente todos os constituintes lipossolúveis, ou sua grande maioria, estão presentes na forma de

glóbulos, envoltos por uma membrana fosfolipoprotéica. Essa membrana é a mesma da célula alveolar da glândula mamária, e é responsável por dar estabilidade à emulsão<sup>7</sup>.

A fração suspensão é constituída de micelas de caseína, formadas por subfrações, como a  $\kappa$ -caseína,  $\gamma$ -caseína,  $\alpha$ -caseína, dentre outras. O sistema caseína forma uma suspensão coloidal do tipo gel, cuja estabilidade é conferida pela fração  $\kappa$ -caseína que envolve a micela. A quase totalidade do cálcio e do fósforo presentes no leite humano encontra-se associado às micelas, quimicamente ligadas às frações que as integram.

A fração solução reúne a água, principal constituinte do leite humano, que apresenta concentração de 87% v/v, bem como os demais hidrossolúveis, a exemplo das proteínas do soro, sais minerais, carboidratos e a maior parte dos imunobiológicos presentes no leite humano<sup>7</sup>.

Essas três frações apresentam uma relação de proporcionalidade entre si, decorrente do próprio movimento de síntese do leite humano. Dessa maneira, a variação na concentração de um dos constituintes do leite sempre acarreta alteração nos demais, podendo essa relação de proporcionalidade se dar de forma direta ou indireta, dependendo dos constituintes considerados. Os constituintes lipossolúveis, que integram a fração emulsão, por exemplo, tendem a se relacionar de forma inversamente proporcional com as proteínas do soro do leite ou proteínas solúveis, principais representantes dos imunobiológicos. Tal tendência permite afirmar que quanto maior o conteúdo de gordura maior será o aporte energético e menor será a concentração de imunobiológicos<sup>7</sup>.

A fração emulsão congrega os componentes de menor densidade, resultando em valor médio situado na ordem de  $0,9 \text{ g/cm}^3$ . Por essa razão, ao se submeter o leite à centrifugação, a fração emulsão tende a ascender no tubo e separar-se dos demais constituintes. Contudo, ao sofrer a ação da força centrífuga, a fração emulsão arrasta consigo as micelas de caseína, formando um aglomerado denominado creme, que se separa do soro do leite ou fração hidrossolúvel. Estudos acerca da proporcionalidade dos constituintes do leite

humano permitiram o estabelecimento da relação matemática entre creme, soro, gordura e conteúdo energético<sup>7</sup>.

#### **4.1.2. A importância do Controle de Qualidade nos Bancos de Leite Humano (BLHs) para Saúde Pública**

A Rede Nacional de Bancos de Leite Humano do Brasil (RNBLH) é a maior rede de bancos de leite humano do mundo, além de ser referência internacional para outros bancos de leite humanos. Atualmente, conta com mais de 150 unidades espalhadas por todo o país e uma de suas prioridades é de atender às mães e os prematuros e recém-nascidos de baixo peso internados em unidades hospitalares.

Em vista disso, o controle de qualidade é muito importante. A qualidade tem um significado amplo, incluindo as características responsáveis pelo valor biológico do produto e a segurança para o consumidor, sob o ponto de vista de saúde pública. A qualidade do leite humano ordenhado pode ser definida como uma grandeza que resulta da avaliação conjunta de uma série de parâmetros, que incluem as características nutricionais, imunológicas, químicas e microbiológicas. Entre as análises físico-químicas realizadas para a liberação do consumo do leite humano estão: cor, flavor (sabor e odor), acidez e o índice de gordura.

Porém os bancos de leite humano enfrentam hoje no Brasil o sério problema de certificar a qualidade de seus produtos, tanto a nível sanitário como físico-químico/ nutricional. Isto porque as metodologias são caras e/ou de difícil acesso. Há a necessidade de se desenvolver e adaptar técnicas laboratoriais de fácil execução, de baixo custo, que sejam precisas e sensíveis o suficiente para serem empregadas no controle de qualidade de rotina dos BLHs<sup>8</sup>.

Enfim, os BLHs representam no Brasil um valioso instrumento na área de Saúde Pública, para atender aos lactentes clinicamente impossibilitados de serem amamentados por suas mães. Através da publicação da Portaria nº 322/88 do Ministério da Saúde e da criação da Comissão Central de Bancos de Leite, muitos investimentos e pesquisas têm sido realizados, no intuito de desenvolver

metodologias de controle adaptadas às típicas necessidades nacionais, suficientemente seguras e sensíveis para uso rotineiro em BLHs.

#### **4.1.3. A Espectroscopia no Infravermelho Médio (FTIR) por Reflectância Total Atenuada (ATR)**

Métodos analíticos baseados na espectroscopia no infravermelho médio (MIR) são amplamente utilizados para determinar gordura em produtos lácteos<sup>9,10,11</sup>, água<sup>12</sup> e na avaliação do tempo de prateleira (shelf-life) do produto<sup>13</sup>. Porém para determinação de gordura do leite materno existem na literatura apenas trabalhos empregando outras técnicas como cromatografia gasosa<sup>14</sup> e espectrometria de massas<sup>15</sup>.

A Espectroscopia no Infravermelho Médio, de 4000 a 200  $\text{cm}^{-1}$ , pode ser considerada a região mais amplamente utilizada na identificação e quantificação de espécies químicas, e dentro desta, a região de “impressão digital”, de 1200 a 600  $\text{cm}^{-1}$ , é mais útil já que pequenas diferenças na estrutura e composição produzem variações significativas nos picos de absorção nesta região.

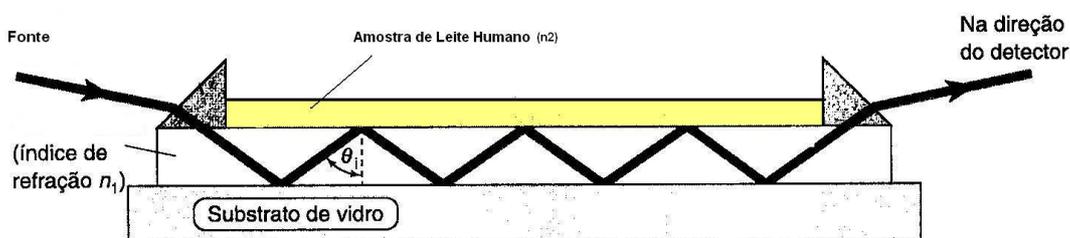
As vantagens da espectroscopia no infravermelho como praticidade, caráter não destrutivo, não necessidade de preparo da amostra, fazem que esta técnica seja empregada na agricultura, na indústria de alimentos<sup>16</sup>, farmacêutica<sup>17</sup> e nos mais diversos laboratórios de controle de qualidade.

Além disso, a espectroscopia no infravermelho médio tem alta versatilidade devido a diferentes acessórios, como reflectância total atenuada (ATR-attenuated total reflection). A reflectância total atenuada (ATR) é uma técnica de amostragem rápida que requer uma mínima preparação da amostra e, principalmente, que apresenta fácil obtenção de espectros de materiais espessos e fortemente absorventes, difíceis de serem analisados por espectroscopia por transmissão<sup>18</sup>.

Tecnicamente, a reflectância total atenuada se caracteriza pelas múltiplas reflexões da radiação infravermelha que ocorrem no interior de cristais, de materiais com alto índice de refração (como por exemplo, ZnSe), interagindo apenas com a amostra que estiver superficialmente no cristal. Este cristal deve ser

composto de um material com alto índice de refração, para que somente uma pequena parte do feixe de radiação incidente seja refletido ao atingir o cristal.

Após o feixe incidente sofrer a difração penetrando no meio  $n_2$ , a geometria do cristal deve promover a reflexão e para tanto  $\Theta_i < \text{sen}^{-1}(n_2/n_1)$ . Esta reflexão interna cria uma onda evanescente a qual se estende na amostra em contato com o cristal. Na região do infravermelho onde a amostra absorve energia, a onda evanescente será atenuada. Na figura 38 é apresentada uma representação da obtenção de espectro usando o acessório de ATR.



**Figura 38.** Representação do acessório de Reflexão total atenuada para a obtenção do espectro de amostra de leite humano.

O fenômeno da reflexão interna é observado sob certas condições. Quando a radiação entra em um material feito com índice de refração alto em relação ao meio externo (cristal de ATR) a radiação será refletida totalmente internamente<sup>19</sup>.

O ângulo acima do qual esta onda refratada deixa de ser real é denominado de ângulo crítico ( $\theta_c$ ) e é expresso conforme a equação

$$\theta_c = \text{sen}^{-1}(n_2 / n_1) \quad (\text{Eq. 25})$$

Nestas condições, o feixe é totalmente refletido na interface  $n_2 / n_1$ . As propriedades do campo evanescente, onde a propagação da radiação infravermelha no meio 1 (com índice de refração  $n_1$ ), sofre reflexão interna na interface com o meio 2 (com índice de refração  $n_2$ ), quando o ângulo de incidência

excede o ângulo crítico. A profundidade de penetração da radiação no segundo meio pode ser calculada pela equação 26.

$$d_p = \frac{\lambda}{2\pi n_1 \left( \sin^2 \theta_i - \left( \frac{n_2}{n_1} \right)^2 \right)^{1/2}} \quad (\text{Eq. 26})$$

onde  $d_p$  é a profundidade de penetração da radiação;  $\theta_i$  é o ângulo de incidência de refração do elemento de ATR;  $\lambda$  é o comprimento de onda da radiação;  $n_2$  é o índice de refração da amostra e  $n_1$  é o índice de refração do cristal de ATR.

Assim, a profundidade de penetração, depende do comprimento de onda, do índice de refração dos dois materiais e do ângulo do feixe em relação a interface. A radiação que atinge a amostra penetra poucos microns de profundidade. Desta forma, qualquer material que esteja em contato com o cristal de ATR pode absorver a radiação incidente atenuando sua intensidade, dando origem ao espectro infravermelho. Em função disso, a composição da superfície pode ser estudada e o espectro torna-se independente da espessura da amostra.

Contudo, a profundidade de penetração é proporcional ao comprimento de onda e, portanto, é observado um aumento constante na profundidade de penetração na amostra, com a varredura do espectro de comprimento de onda maiores para menores. Como consequência, as intensidades relativas podem ser diferentes das observadas pelo emprego da espectroscopia por transmissão. A profundidade de penetração também varia com o índice de refração da amostra. Esse efeito é menos pronunciado quando o índice de refração do cristal de ATR é maior. Um dos materiais mais comumente usados para o cristal de ATR é o ZnSe, cujo índice de refração é de 2,4 e sua faixa de utilização é de 20.000 até 650  $\text{cm}^{-1}$ , sendo recomendado para medidas em soluções aquosas por ser insolúvel em água 20.

#### **4.1.4. Comitê de Ética de Pesquisa**

Quando um projeto de pesquisa trabalha com amostras que envolvam direta ou indiretamente seres humanos faz-se necessário o projeto ser aprovado por comitês de ética das instituições relacionadas ao trabalho. Esta parte do projeto de análises de leite humano foi aprovada pelo comitê de Ética de Pesquisa da Faculdade de Ciências Médicas da Unicamp (projeto nº432/2004) e pelo Comitê de ética da Maternidade Municipal de Campinas.

## **4.2. Parte Experimental**

### **4.2.1. Amostras para a separação de tipos de leite fluido por Análises de Componentes Principais (PCA)**

Para um estudo inicial da similaridade entre as amostras de leite, foram obtidos espectros por reflectância atenuada total (ATR) no infravermelho médio de três amostras de cada tipo de leite fluido (ou seja, de três embalagens diferentes), em triplicada de: leite integral pasteurizado e light pasteurizado, leite caprino integral e light, Leite de soja integral e light, Leite de Côco Integral e Light, Leite Infantil e leite cru de vaca (diretamente do produtor).

### **4.2.2. Amostras e homogeneização**

Após a aprovação do Comitê de Ética da Unicamp e do Comitê de Ética da Maternidade Municipal de Campinas, foram coletadas 62 amostras de 10 mL de leite humano **maduro pasteurizado** na Maternidade Municipal de Campinas. As voluntárias que doaram as amostras assinaram um termo de consentimento e responderam um questionário com algumas informações pessoais (Anexo A). As coletas das amostras foram agendadas nos dias que estas eram descongeladas para o fracionamento e distribuição para os bebês. Em média, eram coletadas de quatro a seis amostras de cada vez. O transporte das amostras da Maternidade ao laboratório foi realizado, em média, de 20 a 25 minutos. Neste percurso as

amostras foram condicionadas em caixa térmica com gelok para mantê-las resfriadas. Assim que chegavam ao laboratório, as amostras foram analisadas imediatamente, depois de permanecerem 15 minutos em banho-maria a 40°C e 15 segundos no vórtex, para serem bem homogêneas. Em seguida foram obtidos os espectros na região do infravermelho por reflectância total atenuada (ATR).

Todas as amostras foram rotuladas com as seguintes informações:

- a) Número da amostra;
- b) Data da Pasteurização;
- c) Índice de Gordura obtido pelo método padrão (Kcal/ L)

As 62 amostras foram fornecidas por 41 doadoras. As respostas do questionário com informações pessoais destas doadoras estão apresentadas no Anexo B. Neste grupo de 41 doadoras, há três lactantes fumantes e algumas que usam frequentemente algum tipo de medicação.

#### **4.2.3. Técnica do Crematócrito - Método de referência para índice de gordura em leite humano**

A medida de gordura foi fornecida pelo banco de leite da maternidade. Para um melhor desenvolvimento da rotina do banco, esta medida de gordura é realizada antes da pasteurização, lembrando que o leite humano pasteurizado (LHP), antes de ser administrado ao recém-nascido (RN), passa por um novo congelamento, sendo depois descongelado e aquecido. Durante todo este processo, são verificadas perdas energéticas, principalmente no teor de gordura, pois o congelamento e descongelamento acarretam rompimento das membranas dos glóbulos de gordura, propiciando sua coalescência e facilitando sua aderência às paredes dos frascos armazenadores<sup>21</sup>.

A metodologia do crematócrito, adaptada pelo Instituto Fernandes Figueira da Fundação Oswaldo Cruz, para uso corrente em Bancos de Leite, tem sido utilizada recentemente como critério classificatório do conteúdo gorduroso, e conseqüentemente energético, do leite humano coletado<sup>22</sup>.

A técnica do crematócrito<sup>23</sup>, originalmente descrita por Lucas e colaboradores em 1978, foi adaptada e testada com êxito na rotina operacional dos BLHs no país. O crematócrito é semelhante ao micro-hematócrito, devendo-se utilizar leite em lugar do sangue. Após a centrifugação dos capilares por 15 minutos, ocorre a separação do creme e soro do leite, o creme ocupa a parte posterior do capilar e corresponde à fração de coloração mais densa. O soro, de aspecto “mais ralo”, fica abaixo do creme. Com auxílio de uma régua milimetrada deve-se medir o comprimento da coluna de creme (mm) e da coluna total do produto (coluna de creme + coluna de soro, expressos em mm). De posse destes valores, empregando as equações abaixo, obtém-se:

$$\text{Teor de creme} = \frac{\text{Coluna de Creme (em mm)} \times 100}{(\%) \text{ Coluna total de produto (em mm)}} \quad (\text{Eq.27})$$

$$\text{Teor de gordura} = \frac{\text{Teor de creme (\%)} - 0,59}{(\%) 1,46} \quad (\text{Eq. 28})$$

$$\text{Energia (Kcal/litro)} = 66,8 \times \text{Teor de creme (\%)} + 290 \quad (\text{Eq.29})$$

As amostras coletadas variaram de **384,85 a 766,95 Kcal/L**.

Num estudo com 2568 amostras de leite materno, observou-se que os teores calóricos com valores de 464 a 700 Kcal/L vieram de amostras coletadas no início da mamada e de 720 a 1100 Kcal do final da mamada<sup>24</sup>.

#### 4.2.4. Instrumental

Os espectros foram obtidos em um Espectrômetro FT-IR MB100 Bomen, com detector de sulfato de triglicina deuterada (DTGS), empregando o acessório de Reflectância Total Atenuada (ATR), na região de 980 a 3000 cm<sup>-1</sup>. As medidas foram obtidas em triplicata, com 64 varreduras, sensibilidade média (nível C) e

resolução de  $4\text{ cm}^{-1}$ . O espectro de referência foi feito com o cristal vazio (ar) de ZnSe do acessório de ATR horizontal. Em cada replicata ou troca de amostra, o acessório de ATR foi limpo com solução de detergente 5%, depois com etanol, metanol e em seguida seco para colocar a amostra seguinte<sup>25</sup>.

#### **4.2.5. Pré-tratamento dos espectros**

Devido às variações nos espectros, principalmente pelo espalhamento de luz nos glóbulos de gordura, foi feita a correção do espalhamento multiplicativo de sinal (MSC) no conjunto dos espectros de leite humano.

#### **4.2.6. Escolha das amostras para calibração e previsão**

Neste caso a escolha das amostras para a calibração e previsão foi pelo algoritmo de Kennard–Stone (kenston) do pacote PLS-Toolbox versão 3.5<sup>26</sup>. Definido o número de amostras para calibração, este algoritmo inicia com a seleção de dois espectros: aquele mais próximo do ponto central e o outro que possui a maior distância euclidiana em relação ao primeiro. Depois a outra amostra escolhida é a que possui a maior distância em relação a esta última e assim sucessivamente até chegar no número escolhido de amostras para calibração. O restante das amostras será o conjunto de previsão.

#### **4.2.7. Modelagem dos dados**

Conforme item 2.2.6.

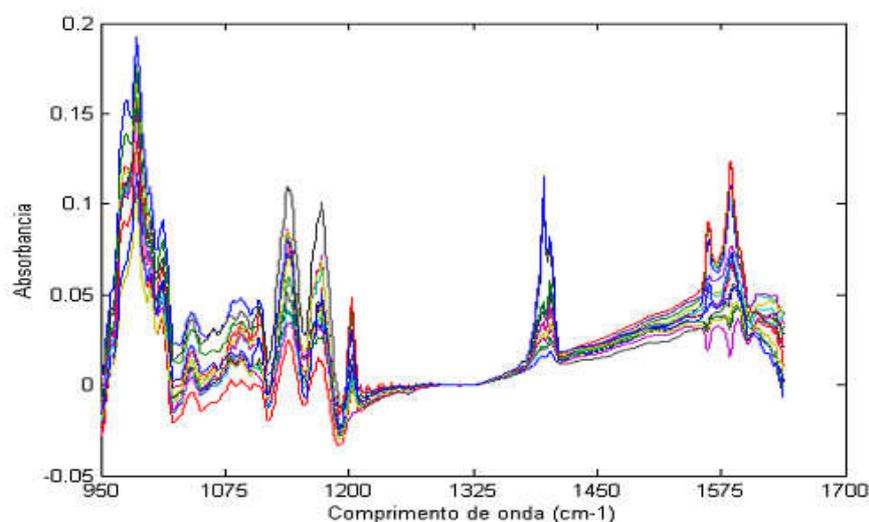
#### **4.2.8. Avaliação do desempenho dos modelos**

Conforme item 2.2.7.

### 4.3. Resultados e Discussão

#### 4.3.1. Separação de tipos de leite por Análise de Componentes Principais (PCA)

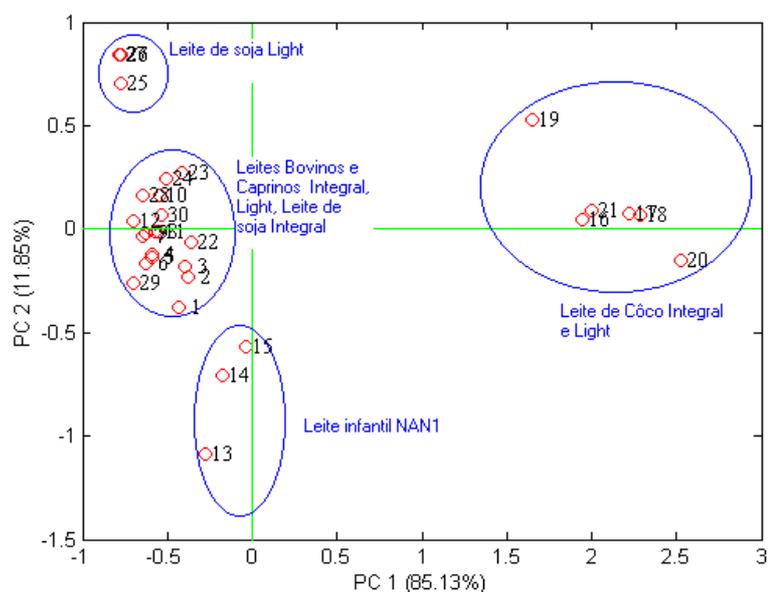
Os espectros obtidos das amostras de diferentes tipos de leite foram corrigidos pela correção do espalhamento multiplicativo (MSC) e são apresentados na figura abaixo.



**Figura 39.** Espectros das 30 amostras de leites com pré-tratamento MSC.

A Figura 39 mostra os espectros dos leites. Os picos na região entre 1500-1700  $\text{cm}^{-1}$  são devidos as bandas de proteína das amostras de leite, enquanto a faixa de 1200 a 1450  $\text{cm}^{-1}$  refere-se aos grupos carboxílicos das proteínas. Entre 1520 a 1560  $\text{cm}^{-1}$  estão as bandas de amida das proteínas. As características vibracionais de mono e polissacarídeos encontram-se geralmente na região de 900 a 1200  $\text{cm}^{-1}$  devido às absorções OH <sup>27</sup> e estiramentos C-O dos açúcares<sup>28, 29</sup>.

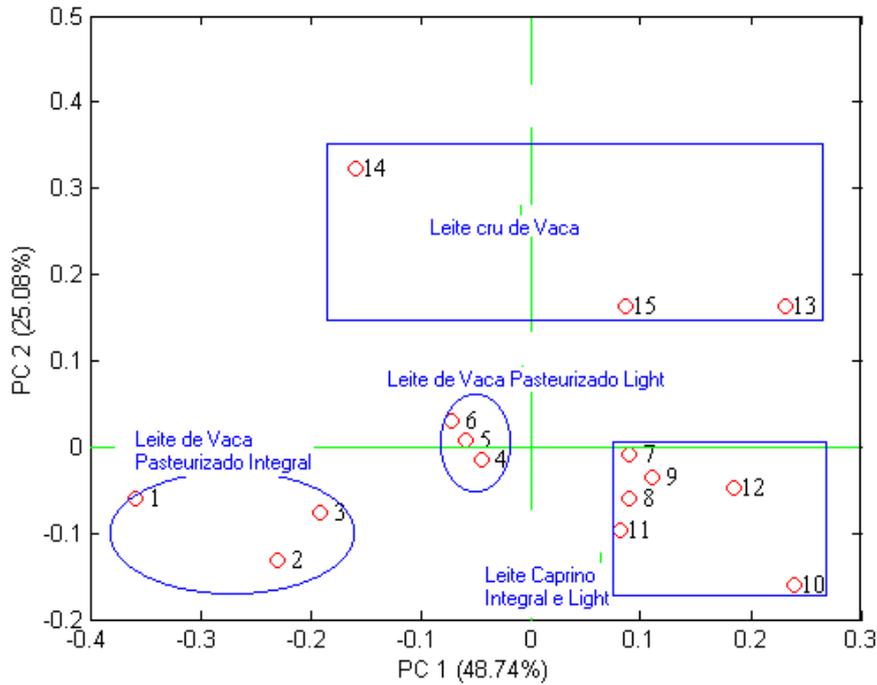
A partir dos espectros das 30 amostras, utilizando Análise de Componentes Principais, observou-se a separação de quatro grupos, quando os escores do 1º componente são graficados contra os do 2º componente principal (96,08% de variância dos dados). Na Figura 40, pode-se observar um grupo com as amostras de leite de soja light, outro com leite infantil, um terceiro grupo com as amostras de leite de coco integral e desnatado, e o último grupo, com a maior parte das amostras (leite de vaca cru, leite de vaca pasteurizado integral e light, leite de cabra pasteurizado integral e light e leite de soja integral) como se pode observar na Figura 40.



**Figura 40.** Gráfico da separação dos grupos das 30 amostras de diferentes tipos de leites por Análise de Componentes Principais.

Realizando a PCA das 15 amostras de leite de origem animal (vaca e cabra) e novamente graficando o 1º componente principal contra o 2º componente (73,82% da variância dos dados), se observa na Figura 41 a separação do leite de

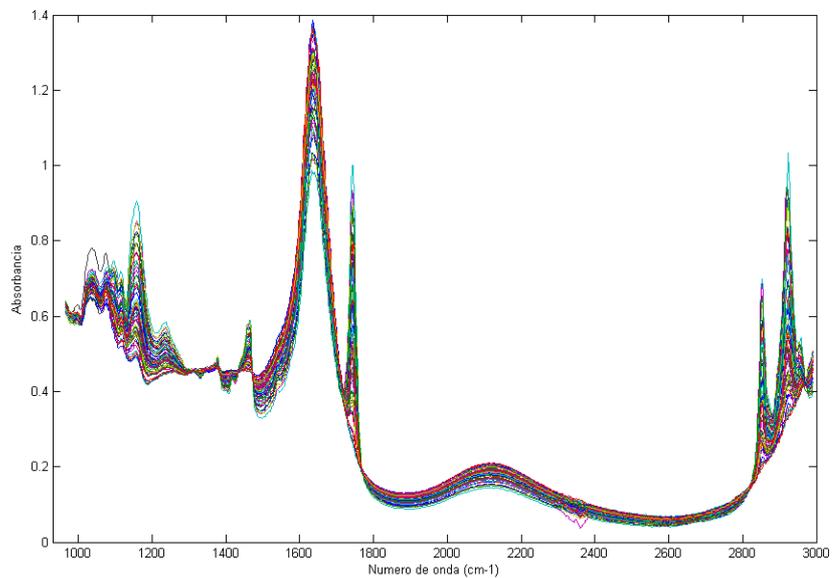
cabra das amostras de leite de vaca, e também a separação do grupo das amostras de leite de vaca cru, em relação às amostras pasteurizadas. Dentro do grupo do leite de vaca pasteurizada, ocorre um agrupamento das amostras integrais e outro das amostras light.



**Figura 41.** Análise de componentes principais (PCA) de 15 amostras de leites de origem animal.

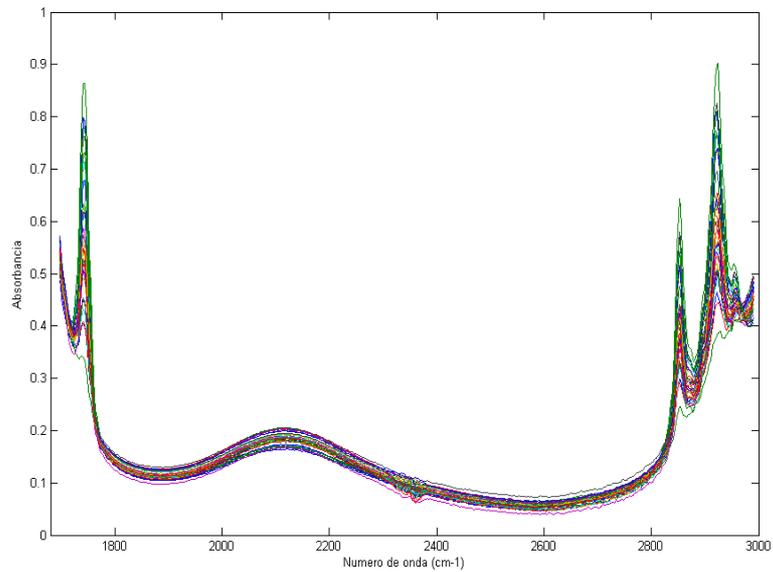
#### 4.3.2. Determinação do Índice de gordura no leite humano

Na Figura 42, são apresentados os espectros na região de 980 a 3000  $\text{cm}^{-1}$  das 62 amostras de leite humano.



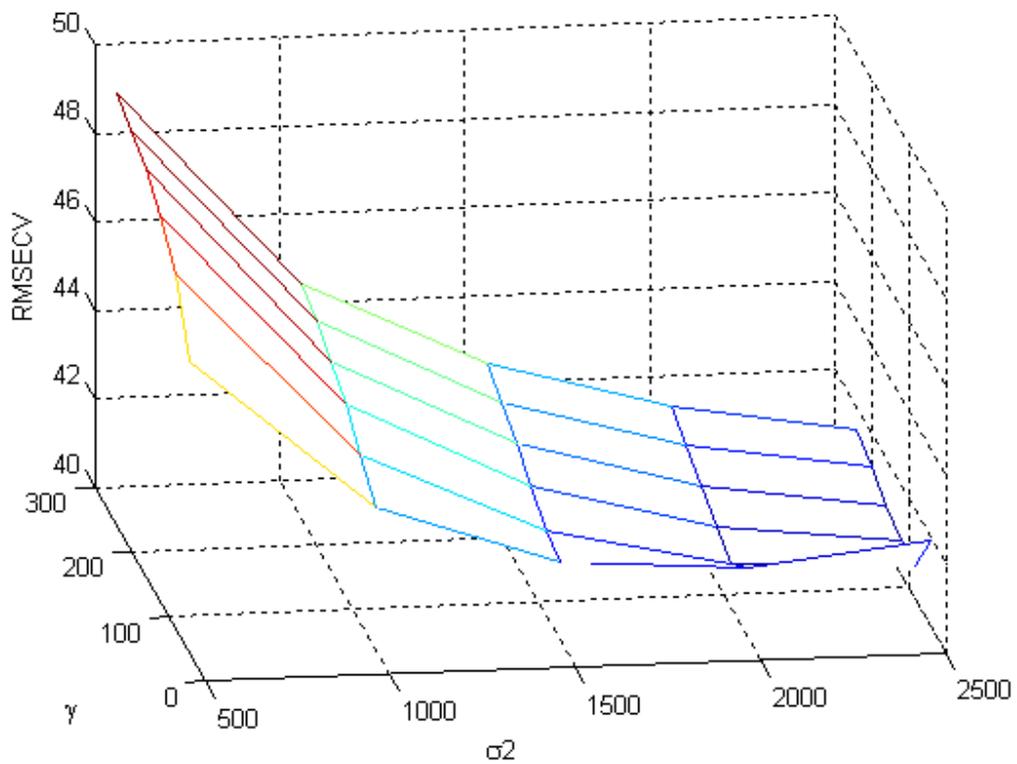
**Figura 42.** Espectros das 62 amostras de leite humano na região no infravermelho de 980 a 3000  $\text{cm}^{-1}$  com pré-tratamento MSC.

Basicamente, a estrutura química das moléculas de gordura pode ser dividida em duas porções: (A) o glicerol conectado por um éster e (B) cadeias alquil de ácidos graxos<sup>17</sup>. Os grupos carbono-hidrogênio (C-H) da gordura absorvem na região de 2850.0 – 2950.0  $\text{cm}^{-1}$  e os grupos carbonilas (C=O) da gordura do leite absorvem na faixa de 1730.0- 1760.0  $\text{cm}^{-1}$ . Dessa maneira, o modelo de gordura foi desenvolvido de 1700 a 300  $\text{cm}^{-1}$  como pode ser observado na Figura 43. Para o desenvolvimento dos modelos das 62 amostras iniciais restaram 38 amostras, depois da retirada de amostras anômalas, pelo gráfico de resíduos e leverage do PLS, e principalmente de amostras que não apresentaram bandas nas regiões esperadas para gordura. A partir das 38 amostras, foram escolhidas 31 amostras para o conjunto de calibração e sete para previsão.



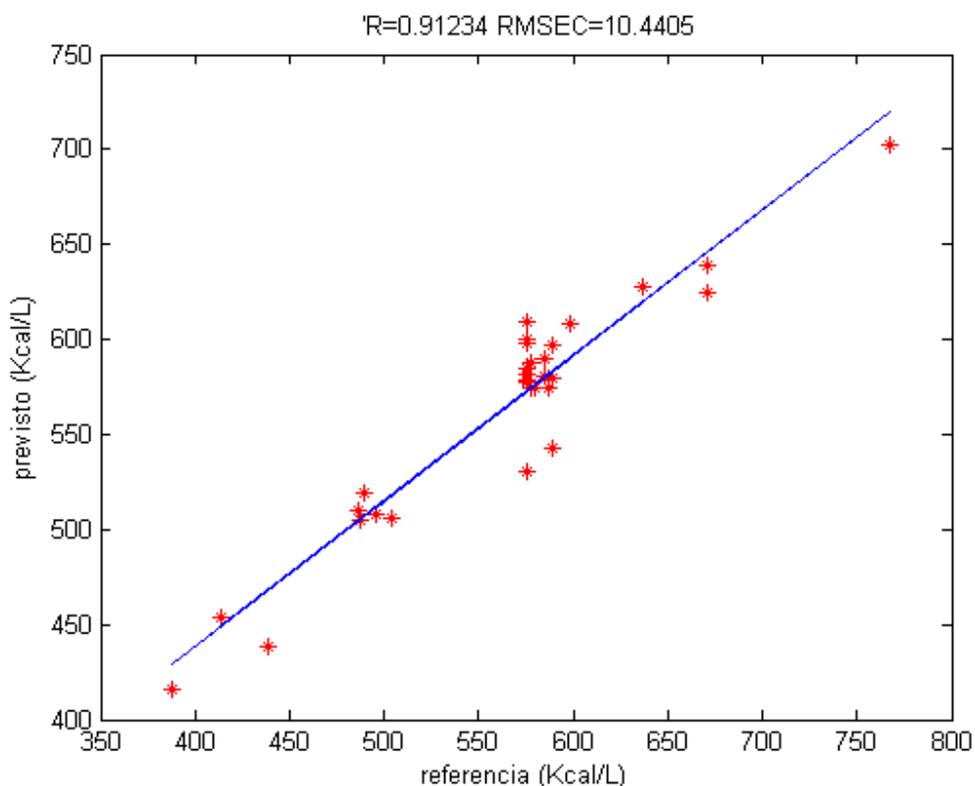
**Figura 43.** Espectros das 38 amostras de leite humano na região no infravermelho de 1700 a 3000  $\text{cm}^{-1}$ .

Para a LS-SVM, também foi realizada a otimização dos parâmetros  $\gamma$  e  $\sigma^2$  até atingir um mínimo da superfície (menor RMSECV) a partir do qual a diminuição do valor do RMSECV não seja mais significativa. Como apresentado na Figura 44 os parâmetros usados no modelo LS-SVM foram  $\gamma$  de 120 e  $\sigma^2$  de 2500.



**Figura 44.** Superfície de otimização da LS-SVM para o conjunto de calibração para determinação do índice de gordura em leite humano.

Na Figura 45, observa-se a previsão das amostras do próprio conjunto de calibração.



**Figura 45.** Conjunto de calibração do modelo LS-SVM para determinação do índice de gordura em leite humano.

Na Figura 45, observa-se uma distribuição desigual das amostras ao longo da reta de correlação dos dados, que tem  $R^2$  de 0,91. Na Tabela 14, são apresentados os resultados para os modelos PLS e LS-SVM para o conjunto de previsão. Os valores do coeficiente de correlação dos valores previstos versus de referência para o conjunto de previsão ( $R^2_{\text{prev}}$ ), e o RMSECV, RMSEC e RMSEP apresentados pelo modelo LS-SVM são melhores que os do modelo PLS, porém ainda não adequados para ser implantando, por exemplo, no controle de qualidade de um banco de leite humano.

**Tabela 14.** Resultados dos modelos PLS e LS-SVM para determinação do índice de gordura em leite humano.

<b>Gordura</b>	<b>PLS</b>	<b>LS-SVM</b>
<b>R<sup>2</sup> prev</b>	0,05	0,78
<b>RMSEC (Kcal/L)</b>	13,00	10,44
<b>RMSECV (Kcal/L)</b>	102,80	24,73
<b>RMSEP (Kcal/L)</b>	102,91	40,76
<b>VL</b>	10	
<b>gama</b>		120
<b>sigma</b>		2500

Uma das justificativas para explicar o mau desempenho do PLS pode ser a ocorrência de desvios da linearidade devido as amostras que, além de serem fornecidas por doadoras diferentes, numa mesma doadora este tipo de amostra apresenta diferenças. Outra justificativa pode ser devido a perdas de gordura das amostras durante os processos de congelamento/ descongelamento que foram realizadas entre as medidas de crematócrito até as obtenções dos espectros.

#### 4.4. Referências Bibliográficas

<sup>1</sup> Gomes A M C, Kuzuhara J S W, Mattar M J G, Makssoudian P, Cipriano M C P, Galisa M C, Análise do leite humano impróprio para o consumo: aspectos físico-químicos e microbiológicos, Caderno de Resumos. II Congresso Nacional de Banco de Leite Humano, **2000**.

<sup>2</sup> Pequeno manual de Amamentação. Disponível em: <<http://www.aleitamento.org.br/manual/extraire.htm>>. Acesso em: 10 de Abril **2004**.

<sup>3</sup> Oliveira J E D, Marchini J S, *Ciências Nutricionais*, Ed. Sarvier, São Paulo, **1998**, 222- 224.

<sup>4</sup> Hibberd C M, Brooke O G, Carter N D, Haug M, Harzer G, Variation in the composition of breast milk during the first 5 weeks of lactation: implications for the feeding of preterm infants, *Archives of Disease in Childhood*, 57, **1982**, 658-662.

<sup>5</sup> Bobbio F O, Bobbio P A, *Introdução à Química dos Alimentos*, Editora UNICAMP, Campinas, **1985**.

- 
- <sup>6</sup> Emmett P M, Rogers I S, Properties of human milk and their relationship with maternal nutrition, *Early Human Development*, 49, **1997**, S7 - S28.
- <sup>7</sup> de Almeida J A G, Guimarães V, Novak F R, Leite Humano Ordenhado: Determinação do Crematócrito, Normas técnicas para bancos de leite humano, IFF/NT- 30.05, Centro de Referência Nacional para Bancos de Leite Humano, Instituto Fernandes Figueira, Fundação Oswaldo Cruz, Ministério da Saúde, Fevereiro **2005**.
- <sup>8</sup> Giugliani E R J, Rede nacional de bancos de leite humano do Brasil: tecnologia para exportar, *Jornal de Pediatria*, 78 (3), **2002**, 183.
- <sup>9</sup> Moros J, Iñon F A, Garrigues S, de la Guardia M, Determination of the energetic value of fruit and milk-based beverages through partial-least-squares attenuated total reflectance-Fourier transform infrared spectrometry, *Analytica Chimica Acta*, 538, **2005**, 181-193.
- <sup>10</sup> Cocciardi R A, Ismail A A, Van de Voort F R, Sedman J, A rapid Fourier transform infrared quality control method for the determination of lactose, sucrose, fat and total solids in chocolate milk, *Milchwissenschaft*, 57 (2), **2002**, 690-693.
- <sup>11</sup> Stegeman G A, Baer R J, Schingoethe D J, Casper D P, Influence of milk fat higher in unsaturated fatty acids on the accuracy of milk fat analyses by the mid-infrared spectroscopic method, *Journal of Food Protection*, 54, **1991**, 890-893.
- <sup>12</sup> Hop E, Luinge H J, Van Hemert H, Quantitative analysis of water in milk by FT-IR spectrometry, *Applied Spectroscopy*, 47, **1993**, 1180- 1182.
- <sup>13</sup> Lynch J M, Barrano D M, Evaluation of commercially available milk powders for calibration of mid-infrared analysers, *Journal of AOAC International*, **1995**, 78, 1219 - 1224
- <sup>14</sup> López-López A, Castellote-Bargalló A I, López-Sabater M C, Comparison of two direct methods for the determination of fatty acids in human milk, *Chromatographia*, 54 (11/12) **2001**, 743 -747.
- <sup>15</sup> Charlwood J, Tolson D, Dwek M, Camilleri P, A detailed analysis of neutral and acidic carbohydrates in human milk, *Analytical Biochemistry*, 273, **1999**, 261-277.
- <sup>16</sup> Van de Voort F R, Fourier transform infrared spectroscopy applied to food analysis, *Food Research International*, 25, **1992**, 397-403.
- <sup>17</sup> Lewis E N, Carrol J E, Clarke F, A near infrared view of pharmaceutical formulation analysis, *NIR news*, 12, **2001**, 16-18.
- <sup>18</sup> Lisheng X, Schlup J R, Application of near-infrared attenuated total reflectance spectroscopy for monitoring epoxy resin/amine cure reactions, *Applied Spectroscopy*, 50(1), **1996**, 109-114.
- <sup>19</sup> Harrick N J, Multiple reflection cells for internal reflection spectrometry, *Analytical Chemistry*, 36, **1964**, 188-191.
- <sup>20</sup> Ferrão M F, Aplicações de técnicas Espectroscópicas de reflexão o infravermelho no controle de qualidade de farinha de trigo, Tese de Doutorado. Instituto de Química. UNICAMP **2000**.
- <sup>21</sup> Vieira A A, Moreira M E L, Rocha A D, Pimenta H P, Lucena S L, *Jornal de Pediatria*, 80 (6), **2004**, 490 -494.

- 
- <sup>22</sup> Bortolozo E A F Q, Cândido L M B, O uso do crematócrito no controle de qualidade em bancos de Leite Humano, Caderno de Resumos. III Congresso Brasileiro de Bancos de Leite Humano, **2002**.
- <sup>23</sup> Manual Técnico do Leite Humano. Disponível em: <[www.fiocruz.br/manual](http://www.fiocruz.br/manual)>. Acesso em: 10 de Abril 2004.
- <sup>24</sup> Afonso F M, Laxes S M C, Lanzillotti H S, Lessa M C, Gomes V A, Martins R C V, Análise da variabilidade energética do leite humano ordenhado no hospital maternidade Camela Dutra- RJ, Caderno de Resumos. III Congresso Nacional de Banco de leite humano, **2002**.
- <sup>25</sup> Iñon F A, Garrigues S, de la Guardiã M, Nutritional parameters of commercially available milk samples by FTIR and chemometric techniques, *Analytica Chimica Acta*, 513, **2004**, 401- 412.
- <sup>26</sup> 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**.
- <sup>27</sup> Bellanato J, Juárez M, Ramos M, Aplicación de la espectroscopia infrarroja a control de calidad de la leche, *Revista de Agroquímica e Tecnologia Alimentar*, **1978**, 18, 353 – 363.
- <sup>28</sup> Sivakesava S, Irudayaraj J, Rapid determination of tetracycline in milk by FT-MIR and FT-NIR Spectroscopy, *Journal of Dairy Science*, 85, **2002**, 487-493.
- <sup>29</sup> Cocciardi R A, Ismail A A, Van de Voort F F, Sedman J, A rapid Fourier transform infrared quality control method for the determination of lactose, sucrose, fat and total solids in chocolate milk, *Milchwissenschaft*, **2002**, 57, 90 – 93.

# Conclusões

## Conclusões

A partir dos resultados obtidos neste trabalho, pode-se concluir que as Máquinas de vetores de suporte por mínimos quadrados (LS-SVM), apesar de recentes na área de regressão de dados analíticos, são promissoras e, em todas as aplicações realizadas, apresentaram significativamente melhor desempenho em relação ao método de regressão por mínimos quadrados parciais (PLS).

A determinação simultânea dos adulterantes em leite em pó, usando a Espectroscopia no Infravermelho Próximo por reflectância difusa, foi muito bem sucedida com o emprego da LS-SVM. Foram obtidos erros de previsão baixos (da ordem de no máximo 1,10% para LS-SVM) e, principalmente, conseguindo amostras com ausência de um ou todos adulterantes (ausência de adulterantes) no caso da amostra pura, que o PLS indica como falso positivo de fraude, como por exemplo, determinando 8,00% de sacarose numa amostra pura.

A proposta do uso de imagem para quantificação tem por objetivo contribuir para os poucos estudos que existem na área de química para o uso de imagem em quantificação, o qual é usada preferencialmente para classificação. O uso da imagem escaneada teve por objetivo a proposta de obtenção das imagens da maneira mais simples possível. Mais uma vez o desempenho da LS-SVM em relação ao PLS foi melhor também para a quantificação de *Lactobacillus* com erros na ordem de no máximo 11,00 ufc.

Na determinação de índice de gordura em leite humano, o desafio era a quantificação deste parâmetro muito importante no controle de qualidade de bancos de leite humano, desenvolvendo um modelo de a partir de amostras de

doadoras diferentes, ou seja, sujeito a uma alta variação da matriz. Empregando a técnica de Espectroscopia no infravermelho médio por reflectância total atenuada observam-se razoáveis resultados, sujeitos a algumas sugestões como o uso de ATR com controle de temperatura e medidas do índice de gordura paralelamente com a obtenção dos espectros, assim evitando perdas durante processos de congelamento/descongelamento das amostras. Porém, mais uma vez foi verificado o melhor desempenho da LS-SVM em relação ao PLS, especialmente para conjuntos de dados que não se ajustam adequadamente a um perfil linear, seja devido a complexidade da amostra seja a respostas dos métodos de referência não se ajustarem linearmente a medidas instrumentais.

# Perspectivas

## Perspectivas

Dentre as perspectivas futuras para o uso das máquinas de vetores de suporte por mínimos quadrados estão: o emprego de outras funções núcleo para o desenvolvimento de modelos e o desenvolvimento de uma rotina que possibilite uma escolha sistemática dos parâmetros (referentes a cada função núcleo), por exemplo, os parâmetros  $\gamma$  e  $\sigma^2$  da função núcleo RBF.

A proposta para um trabalho futuro com determinação de adulterantes em leite em pó, é a construção de um modelo único com diferentes tipos de leite em pó, também adulterados com amido, soro de leite e sacarose.

No caso do uso de imagem para quantificação, além de poder ser empregada em outras áreas da microbiologia (alimentos, clínica) pode se estender para outras aplicações, que além do histograma de cores, pode usar outras propriedades das imagens, como por exemplo, imagens em escala de cinza.

Em relação ao modelo desenvolvido para determinação de gordura em leite humano, para um melhor desempenho da LS-SVM sugere-se: o emprego de uma população maior de amostras; o emprego de acessório de ATR com controle de temperatura; obtenção simultânea dos valores de gordura com a aquisição dos espectros (para evitar possíveis perdas do teor de gordura nos processos de congelamento/descongelamento). Além disso, para bancos de leite humano em que a aquisição de um espectrofotômetro no Infravermelho seja acessível, inclui-se a perspectiva de que, além do parâmetro de gordura, sejam quantificados outros parâmetros como proteínas, acidez e açúcares a partir de um mesmo espectro no infravermelho médio.

# Anexos

## Anexos

### Anexo A - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

#### DADOS SOBRE A PESQUISA

Pesquisadora: Alessandra Borin

Cargo/Função: Aluna de doutorado em Química na UNICAMP

Avaliação do Risco da pesquisa: **SEM RISCO** para doadoras (nutrizes)

#### TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO A

Meu nome é Alessandra, aluna de doutorado em Química na UNICAMP. Juntamente com meu orientador, prof. Dr. Ronei Jesus Poppi, estamos pesquisando uma nova maneira de analisar a qualidade do leite materno doado ao BLH que é analisado e fornecido para os recém-nascidos.

Para podermos realizar esta pesquisa necessitamos que a senhora \_\_\_\_\_, doadora do BLH, concorde em responder algumas questões sobre seus dados pessoais, da gravidez e em relação ao retiro do leite. Gostaria que também nos informasse sobre hábito de fumar, da ingestão de café, chá preto, chocolate e refrigerantes e uso de medicação. Caso a senhora concorde, utilizaremos **10 mL** do leite ordenhado, para ser doado.

As doadoras não serão identificadas nos resultados do estudo, mas os resultados individuais serão informados para o Banco de leite Humano, o qual pode disponibilizá-los para cada doadora. Não haverá nenhum ressarcimento financeiro as doadoras.

Declaro que, após esclarecida pela pesquisadora e ter entendido o que me foi explicado, consinto em participar do presente Projeto de Pesquisa.

Campinas, \_\_\_\_\_ de \_\_\_\_\_ de 20\_\_\_\_\_.

\_\_\_\_\_  
Assinatura da doadora

\_\_\_\_\_  
Assinatura do pesquisador

\*Para maiores informações, dúvidas e esclarecimentos sobre a pesquisa : Falar com Alessandra Borin , Fone: (19) 37883126 / \* Para Eventuais denúncias: Comitê de Ética em Pesquisa (19) 3788-8936

## PROTOCOLO DE IDENTIFICAÇÃO DO PACIENTE

### **Identificação Paciente**

Nome: \_\_\_\_\_

Endereço: \_\_\_\_\_

\_\_\_\_\_

R.G.: \_\_\_\_\_

Fone: \_\_\_\_\_

Número de caso da pesquisa: \_\_\_\_\_

Número do cadastro da doadora no BLH: \_\_\_\_\_

## PROTOCOLO DE COLETA DE DADOS

1) Identificação do caso da Pesquisa: Grupo/ Número \_\_\_\_\_

Idade: \_\_\_\_\_ Paridade (nº de filhos): \_\_\_\_\_

Tempo de Lactação: \_\_\_\_\_ (dias/ meses)

Cadastrada no BLH desde: \_\_\_\_\_

Número do cadastro: \_\_\_\_\_

Data da coleta da amostra: \_\_\_/\_\_\_/\_\_\_

Local da Coleta:

BLH     Domicílio

2) Condições da coleta:

Manual     Bomba mecânica     Bomba manual

3) Momento da ordenha

Inicial     Final

4) Condições de estocagem

Refrigerado     Congelado

5) Hábitos de Vida

### **FUMO**

Fumante:     Não     Sim

Quanto tempo \_\_\_\_\_ Quantos cigarros/dia: \_\_\_\_\_

No seu lar há pessoas que fumam?

Não     Sim Quem? \_\_\_\_\_

Trabalha Fora? No seu local de trabalho há pessoas que fumam?

Não     Sim Quem? \_\_\_\_\_

De uma maneira geral, convive com pessoas que fumam

Não     Sim Quem? \_\_\_\_\_

### **ALIMENTAÇÃO**

Faz uso de bebida alcoólica: ( ) Não ( ) Sim

Tipo de bebida: \_\_\_\_\_ Dose diária: \_\_\_\_\_

Ingeri:

( ) Café - quantidade diária: \_\_\_\_\_ (xícaras de café)

( ) Chá mate, preto - quantidade diária: \_\_\_\_\_ (xícaras de chá)

( ) Chocolate - quantidade diária: \_\_\_\_\_ (gramas)

( ) Coca-cola ou similares - quantidade diária: \_\_\_\_\_ (copos)

### **MEDICAÇÃO**

Faz uso de medicamento atualmente (freqüente ou esporádico):

( ) Não ( ) Sim

Qual: \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

Dose : \_\_\_\_\_

Quanto tempo: \_\_\_\_\_

**Faz** uso de Amoxicilina? ( ) Não ( ) Sim

Há quanto tempo: \_\_\_\_\_

Observações ou comentários :

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

## Anexo B

Perfil das doadoras de amostras de leite humano																			
Doadoras	Amostra	Idade	Nº de Filhos	Tempo Lactação	Local da coleta	Tipo de coleta	momento ordenha	estocagem	fumante	cig/dia	no lar fumantes	no trabalho fumantes	bebida alcoólica	café	chá mate	chocolate	poca-cola/ similares	medicamento	Dose/ dia
1	B2	27	3	5m	D	M	I	C	S	nr	N	S	N	av	av	av	av	1	1 cp
2	B5*	26	3	*	BLH	M	I	C	S	5	S	S	av	N	N	av	N	N	
3	B1	30	2	2m	D	Bm	I e F	C	S	20	N	N	N	N	N	av	av	1	diário
4	A1 / A20	34	1	2m	D	M	I e F	C	N	N	N	N	N	S	N	N	N	1	1 cada 2 dias
5	A2 / A38	33	2	9m	D	M	I	C	N	N	N	N	N	N	N	N	av	N	N
6	A3 / A22	25	2	1m	D	Bm	I e F	C	N	N	N	N	N	av	av	av	av	N	N
7	A4	22	1	6m	D	M	I e F	C	N	N	N	N	N	S	N	av	av	1	nr
8	A5	23	1	8m	nr	M	I	C	N	N	S	S	N	S	N	N	S	1	1cp/dia
9	A6 / A37	31	1	3m	D	M	I e F	C	N	N	S	N	N	S	N	av	av	3	500mg
10	A7 / A36	24	1	5m	D	Bm	I	C	N	N	N	N	N	N	N	av	S	N	N
11	A8	32	2	4m	D	M	I e F	C	N	N	N	N	N	av	N	N	av	1	1 cp/ dia
12	A9	29	1	2m	D	M e Bm	I	C	N	N	N	N	N	av	N	av	av	1	1 cp/ dia
13	A10 / A23	38	2	17d	D	Bm	I e F	C	N	N	N	N	N	av	N	av	av	2	2 cp/dia
14	A12 / A25	33	1	2m	D	M	I e F	nr	N	N	S	N	N	N	N	N	N	1	1 cp/ dia
15	A13 / A26	25	1	5m	D	M	I	C	N	N	N	N	N	N	N	N	N	1	1 cp/ dia
16	A14 / A17	27	2	4m	D	M	I e F	C	N	N	N	N	N	S	N	N	N	N	N
17	A11 / A24	19	2	2m	D	M	I	C	N	N	S	N	N	N	av	av	av	2	nr
18	A15 / A28	23	2	2m	D	M	I e F	C	N	N	S	N	N	N	N	N	N	1	1 cp/ dia

Legenda: N- Não; S- Sim; av- as vezes; I – inicial; F- final; M – manual; Bm- bomba manual; Bmec- bomba mecânica; Do – domicílio; BLH- banco de leite humano; C – congelado; m- meses; d-dias; nr- não respondeu.  
 Medicamentos: 1- anticoncepcionais (Cerazette, Micronor, Depovera); 2 – Ferro (Sulfato ferroso); 3- Cefalexina;  
 4- Diclofenado; 5 –Lidicina; 6- Vasken.

## Anexo B

Doadoras	Amostra	Idade	Nº de Filhos	Tempo Lactação	Local da coleta	Tipo de coleta	momento ordenha	estocagem	fumante	cig/dia	no lar fumantes	no trabalho fumantes	bebida alcoólica	café	chá mate	chocolate	coca-cola/ similares	medicamento	dose/ dia
19	A16 / A39	27	3	4m	D	M	I e F	C	N	N	S	N	N	N	N	N	av	2	109mg 3x/dia
20	A19 / A27	30	3	2m	D	M	I e F	C	N	N	N	S	S	S	N	av	av	1	1 cp/ dia
21	A18 / A29	30	3	3m	D	M	I	C	N	N	N	N	N	S	av	av	av	N	N
22	A30	31	1	5m	D	Bm	I e F	C	N	N	N	S	N	S	N	N	N	4	8 /8 hs
23	A31 / A44	33	2	20d	D	M	I	C	N	N	N	S	N	S	N	N	N	N	N
24	A32 / A41	19	1	3m	D	Bm e M	I	C	N	N	S	N	av	S	N	N	av	N	N
25	A33 / A40	25	1	11m	D	M	I e F	C	N	N	N	N	N	N	N	av	N	5	1x/ mês
26	A34 / A42	26	1	4m	D	M	I	C	N	N	N	N	N	av	N	av	av	N	N
27	A35	26	1	1m	D	M	I e F	C	N	N	N	N	av	N	N	av	av	N	N
28	A43	35	2	9d	D	M	I e F	C	N	N	S	N	N	N	av	av	av	3 e 4	4 cp/ dia
29	A45	29	1	2m	D	M	F	C	N	N	N	N	N	av	N	N	N	1	1 cp/ dia
30	A46	30	1	3m	D	Bm	I e F	C	N	N	N	N	N	av	N	N	N	1	nr
31	A47 / A52 / A53 / A54	34	1	1m	D	Bm	I e F	C	N	N	N	N	N	N	av	av	av	6	5 mg 2x/dia
32	A48	33	2	4m	D	M	I e F	C	N	N	N	N	av	S	N	av	av	nr	nr
33	A49	30	1	5d	D	Bmec	I	C	N	N	N	N	N	S	N	N	N	1	1 cp/ dia
34	A50	34	2	3m	D	M	I e F	C	N	N	N	N	N	S	N	N	N	1	nr
35	A51	37	1	15d	D	M	I e F	C	N	N	N	S	N	av	N	N	av	2	nr
36	A55	32	1	3m	D	M	F	C	N	N	N	N	N	N	nr	nr	nr	nr	nr
37	A56	20	1	2m	D	M	I e F	C	N	N	S	N	N	N	N	av	N	1 e 2	nr
38	A58	18	1	5m	D	Bm	I	C	N	N	N	N	N	N	N	N	S	N	nr
39	A60	30	2	4m	D	M	F	C	N	N	N	N	N	N	S	S	N	nr	nr
40	A61	16	1	2m	D	M	F	C	N	N	S	N	N	S	N	S	N	nr	nr
41	A64	29	1	5m	D	M	I	C	N	N	N	N	N	N	N	S	N	N	nr

Legenda: N- Não; S- Sim; av- as vezes; I – inicial; F- final; M – manual; Bm- bomba manual; Bmec- bomba mecânica; Do – domicílio; BLH- banco de leite humano; C – congelado; m- meses; d-dias; nr- não respondeu.  
 Medicamentos: 1- anticoncepcionais (Cerazette, Micronor, Deprovera); 2 – Ferro (Sulfato ferroso); 3- Cefalexina;  
 4- Diclofenado; 5 –Licidina; 6- Vasken.