

## 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

**Marco F. Ferrão\***

Departamento de Química e Física, Universidade de Santa Cruz do Sul, CP 188, 96815-900 Santa Cruz do Sul - RS, Brasil

**Cesar Mello**

Instituto de Química, Universidade de Franca, CP 32, 14404-600 Franca - SP, Brasil

**Alessandra Borin, Danilo A. Maretto e Ronei J. Poppi**

Instituto de Química, Universidade Estadual de Campinas, CP 6154, 13084-971 Campinas - SP, Brasil

Recebido em 28/4/06; aceito em 29/8/06; publicado na web em 27/4/07

LS-SVM: A NEW CHEMOMETRIC TOOL FOR MULTIVARIATE REGRESSION. COMPARISON OF LS-SVM AND PLS REGRESSION FOR DETERMINATION OF COMMON ADULTERANTS IN POWDERED MILK BY NIR SPECTROSCOPY. Least-squares support vector machines (LS-SVM) were used as an alternative multivariate calibration method for the simultaneous quantification of some common adulterants found in powdered milk samples, using near-infrared spectroscopy. Excellent models were built using LS-SVM for determining  $R^2$ , RMSECV and RMSEP values. LS-SVMs show superior performance for quantifying starch, whey and sucrose in powdered milk samples in relation to PLSR. This study shows that it is possible to determine precisely the amount of one and two common adulterants simultaneously in powdered milk samples using LS-SVM and NIR spectra.

Keywords: support vector machines; multivariate regression; powdered milk

### INTRODUÇÃO

Recentemente, as redes neurais empregando perceptrons de múltiplas camadas e as redes neurais de função base radial vêm sendo amplamente utilizadas em diversos campos da ciência, como por ex.: na teoria de controle, no processamento de sinais e em problemas de modelagem linear e não linear<sup>1,2</sup>. Neste contexto, uma promissora ferramenta denominada máquina de vetor de suporte (SVM)<sup>1-4</sup> tem sido utilizada em problemas de classificação e de modelamento. Este algoritmo, baseado na aprendizagem de máquinas de vetores de suporte, foi recentemente introduzido e aplicado em problemas já conhecidos nas áreas da química<sup>5,9</sup>, farmacêutica<sup>10</sup>, microbiológica<sup>11</sup> e mais amplamente na área médica<sup>12-17</sup>.

Entretanto, na área da quimiometria, à qual se referem os problemas de quantificação (regressão multivariada), as aplicações encontradas na literatura ainda são escassas como a exemplo das aplicações das máquinas de vetor de suporte com mínimos quadrados (LS-SVM) com o emprego de espectros Raman para monitorar reações de copolimerização e de espectros NIR na quantificação de misturas ternárias realizadas por Thissen *et al.*<sup>18,19</sup> na comparação do desempenho da LS-SVM com métodos clássicos de regressão publicado por Chauchard *et al.*<sup>20</sup> visando a previsão da acidez total em uvas, e ainda, do trabalho de Cogdill *et al.*<sup>21</sup> que emprega espectroscopia NIR para estimar propriedades da madeira.

Neste sentido, o presente artigo vem apresentar à comunidade científica os principais fundamentos que tornam as máquinas de vetor de suporte excelentes alternativas aos já consagrados métodos de regressão multivariada, destacando-se sua capacidade de generalização e elevada robustez. Como aplicação desta nova ferramenta quimiométrica são apresentados e discutidos os resultados empregando espectros NIR de amostras de leite em pó puro e com adição de amido, soro de leite e/ou sacarose.

### MÁQUINAS DE VETORES DE SUPORTE (SVM)

As máquinas de vetores de suporte (SVM – “Support Vector Machines”) usadas para regressão são uma generalização do algoritmo “Generalized Portrait” desenvolvido, na Rússia, por Vapnik, Lerner e Chervonenkis na década de 60<sup>22</sup>. Uma abordagem similar, empregando programação linear ao invés de programação quadrática, foi desenvolvida na mesma época nos Estados Unidos<sup>23</sup>. A estratégia de aprendizado baseada em vetores de suporte está fundamentada na teoria de aprendizado estatístico, ou teoria VC (Vapnik-Chervonenkis), que vem sendo desenvolvida ao longo das últimas três décadas visando a proposição de técnicas de aprendizado de máquina que buscam maximizar a capacidade de generalização<sup>24</sup>. Entretanto, a forma atual das máquinas suportadas em vetores foi desenvolvida por Vapnik e colaboradores, nos laboratórios da AT&T Bell, no final da década de 90 com ênfase maior nas aplicações práticas<sup>1,25</sup>.

O problema de modelagem de dados experimentais é pertinente a muitas aplicações da Química e Ciências afins. Na modelagem de dados empíricos, é usado um processo de indução para construir um modelo matemático capaz de expressar as relações de entrada-saída, a partir do qual são deduzidas respostas ainda não observadas. Basicamente, a quantidade e qualidade dos dados disponíveis governam o desempenho deste modelo empírico. Por tratar-se de uma técnica de aprendizado baseada em dados experimentais, o tamanho do conjunto de dados fornece tratabilidade computacional, mas conduz a uma amostragem esparsa do espaço de entrada<sup>26</sup>. Conseqüentemente, o problema de aprendizado tende a ser mal condicionado no sentido de Hadamard, ou seja, não há dependência contínua dos dados e o processo de indução de modelos não possui solução única<sup>27</sup>. Quando temos uma multiplicidade de soluções candidatas igualmente admissíveis, a capacidade de generalização dos modelos resultantes passa a representar um critério de qualidade capaz de atenuar o efeito do mal condicionamento.

\*e-mail: ferrao@unisc.br

No entanto, 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>28,29</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>26</sup>.

Deste contexto origina-se a principal motivação para a proposição de máquinas de vetores-suporte (SVMs), as quais apresentam como principais vantagens em suas aplicações<sup>30</sup>: elevada capacidade de generalização – é uma medida da eficiência na previsão de dados que não pertençam ao conjunto utilizado para o treinamento. Assim sendo, evita-se o “overfitting”, situação na qual se obtêm baixos erros no conjunto de treinamento e erros elevados no conjunto de previsão; robustez em grandes dimensões – possibilita a aplicação de SVMs em vetores de características de grandes dimensões, por ex., imagens. A maioria das técnicas de aprendizado de máquina apresentam dificuldade no tratamento de dados de alta dimensionalidade; convexidade da função objetivo – a aplicação das SVMs implica na otimização de uma função quadrática, que possui apenas um mínimo. Esta é uma vantagem sobre as redes neurais, onde a função objetivo possui vários mínimos locais; teoria bem definida – a base teórica das SVMs, a Teoria de Aprendizagem Estatístico, é bem estabelecida dentro da Matemática e da Estatística.

As SVMs possuem diferentes núcleos que caracterizam seu modo de reconhecimento dos padrões; dentre os mais utilizados podemos citar: Linear, 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 RBF) e Sigmoidal (permite que a SVM tenha comportamento semelhante ao de uma rede MLP). A correta definição do núcleo e de seus respectivos parâmetros possui forte influência nos resultados obtidos por uma SVM<sup>31</sup>.

Freqüentemente, as SVMs são empregadas como ferramentas de classificação binária, podendo ser facilmente adaptadas aos problemas de regressão multivariada. Os princípios da aplicação das SVMs para classificação e regressão linear foram recentemente apresentados na literatura<sup>32,33</sup>, bem como sua extensão aos casos não-lineares.

### Teoria das LS-SVM empregadas para regressão multivariada

A implementação computacional das SVM não é muito simples e o tempo computacional gasto pode ser elevado em alguns casos. Todavia, recentemente foi proposto por Suykens e colaboradores<sup>33</sup> uma extensão da teoria das SVM utilizando o método dos mínimos quadrados, que ficou conhecida como LS-SVM.

As máquinas de vetores de suporte (SVM) em sua presente forma foram desenvolvidas nos laboratórios da AT&T na década de 90<sup>34</sup>. Assim, as SVM são técnicas quimiométricas relativamente novas aplicadas em calibração multivariada e em problemas de reconhecimento de padrões e classificação de dados<sup>35</sup>. As máquinas de vetores de suporte com mínimos quadrados<sup>36</sup> são capazes de tratar de modo relativamente rápido problemas de calibração multivariada lineares ou não lineares. Uma estimativa linear é feita em um núcleo linear ( $y = w^T \phi(x) + b$ ). De modo semelhante ao que ocorre com as SVM nas máquinas de vetores de suporte com mínimos quadrados (LS-SVM), é necessário minimizar uma fun-

ção custo (C) contendo uma penalização dos erros de regressão, conforme a Equação 1:

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

de modo que

$$y_i = w^T \phi(x_i) + b + e_i \quad (2)$$

para todos os valores de  $i = 1, \dots, N$ ; em que  $\phi$  denota a função de linearização dos dados.

A primeira parte desta função custo é um decaimento utilizado para regularizar os pesos e penalizar pesos elevados. Devido a esta regularização, os pesos convergem para valores menores. Este procedimento é necessário pois pesos elevados causam excessiva variância, deteriorando a capacidade de generalização das LS-SVM. A segunda parte da Equação 1 representa os erros de regressão para todos os dados de treinamento. O parâmetro  $\gamma$ , o qual deve ser otimizado pelo usuário, fornece o peso relativo da segunda parte da Equação 1, quando comparado com a primeira parte da mesma. A restrição imposta pela Equação 2 fornece a definição do erro de regressão.

Analisando-se a Equação 1 e sua restrição dada pela Equação 2 é possível concluir que temos um típico problema de otimização convexa<sup>33</sup>, que pode ser resolvido usando-se o método dos multiplicadores de Lagrange<sup>37</sup>:

$$L = \frac{1}{2} \|w\|^2 + \gamma \sum_{i=1}^N e_i^2 - \sum_{i=1}^N \alpha_i \{w^T \phi(x_i) + b + e_i - y_i\} \quad (3)$$

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

O ótimo 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:

$$\frac{\partial L(w, b, e, \alpha)}{\partial w} = w - \sum_{i=1}^N \alpha_i \phi(x_i) = 0 \therefore w = \sum_{i=1}^N \alpha_i \phi(x_i) \quad (4)$$

$$\frac{\partial L(w, b, e, \alpha)}{\partial e} = \sum_{i=1}^N \gamma e - \alpha = 0 \therefore \alpha = \gamma e \quad (5)$$

então:

$$w = \sum_{i=1}^N \alpha_i \phi(x_i) = \sum_{i=1}^N \gamma e_i \phi(x_i) \quad (6)$$

na qual um núcleo positivo de transformação é usado conforme a Equação 7:

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j) \quad (7)$$

Um importante resultado desta abordagem é que os pesos ( $w$ ) podem ser escritos como uma combinação linear dos multiplicadores de Lagrange correspondentes aos dados de treinamento ( $x_i$ ).

Colocando-se o resultado da Equação 6 na reta de regressão original ( $y = w^T \phi(x) + b$ ), o seguinte resultado é obtido:

$$y = \sum_{i=1}^N \alpha_i \phi(x_i)^T \phi(x) + b = \sum_{i=1}^N \alpha_i \langle \phi(x_i), \phi(x) \rangle + b \quad (8)$$

Para um ponto  $y_j$  ser avaliado temos:

$$y_j = \sum_{i=1}^N \alpha_i \phi(x_i)^T \phi(x_j) + b = \sum_{i=1}^N \alpha_i \langle \phi(x_i), \phi(x_j) \rangle + b \quad (9)$$

Os vetores  $\alpha$  são obtidos resolvendo-se o sistema linear de equações como segue:

$$\mathbf{M} \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix} \quad (10)$$

na qual  $\mathbf{M}$  é uma matriz quadrada dada por:

$$\mathbf{M} = \begin{bmatrix} \mathbf{K} + \frac{\mathbf{I}}{\gamma} & \mathbf{1}_N \\ \mathbf{1}_N^T & 0 \end{bmatrix} \quad (11)$$

onde  $\mathbf{K}$  é a matriz núcleo da transformação com  $ij$ -ésimo elemento  $\mathbf{K} = (x_i, x_j) = \phi(x_i)^T \phi(x_j)$  e  $\mathbf{I}$  é a matriz identidade  $N \times N$ ,  $\mathbf{1}_N = [1 \ 1 \ \dots \ 1]_r$ . Assim a solução é dada por:

$$\begin{bmatrix} \alpha \\ b \end{bmatrix} = \mathbf{M}^{-1} \begin{bmatrix} y \\ 0 \end{bmatrix} \quad (12)$$

Como pode ser observado nas Equações 11 e 12, usualmente a maioria dos multiplicadores de Lagrange (os vetores de suporte) não possuem valor zero, indicando que a maioria dos dados de treinamento contribuem para a solução. Assim como ocorre com as SVM padrões, as soluções das LS-SVM não são usualmente esparsas. Entretanto, como descrito por Suykens et al.<sup>33</sup>, uma solução esparsa pode ser facilmente obtida através de técnicas de redução como “pruning”, por ex. Dependendo do número de dados de treinamento, uma solução exata pode ser obtida utilizando-se métodos de solução iterativos, como por ex, o método dos gradientes conjugados, no caso do conjunto de dados ser de alta dimensionalidade. Em ambos os casos, métodos numéricos eficientes devem ser usados.

Em aplicações envolvendo regressão não linear é suficiente trocar o produto interno  $\langle \phi(x_i), \phi(x_j) \rangle$  da Equação 9 por uma função núcleo e o  $ij$ -ésimo elemento da matriz  $K$  igual a  $K_{ij} = \phi(x_i)^T \phi(x_j)$ . Se esta função núcleo satisfizer as condições do teorema de Mercer<sup>38</sup>, a função núcleo determina implicitamente tanto a transformação linear,  $x \rightarrow \phi(x)$ , quanto o correspondente produto interno  $\phi(x_i)^T \phi(x_j)$ . Este fato leva a seguinte função de regressão não-linear:

$$y = \sum_{i=1}^N \alpha_i K(x_i, x) + b \quad (13)$$

Para se avaliar um ponto  $x_j$  faz-se:

$$y_j = \sum_{i=1}^N \alpha_i K(x_i, x_j) + b \quad (14)$$

A obtenção da função núcleo é trabalhosa e dependerá de cada caso. Entretanto, a função núcleo mais usada é  $\exp(-\|x_i - x_j\|^2 / \sigma^2)$ , uma simples função gaussiana, e funções polinomiais  $\langle x_i, x_j \rangle^d$ , em que  $\sigma^2$  é a largura da função gaussiana e  $d$  é o grau do polinômio, o

qual deve ser otimizado pelo usuário, usado para obter os vetores de suporte<sup>39</sup>. Isto é feito obtendo-se o mínimo de uma superfície de erros em função de  $\sigma^2$  e  $\gamma$ , conforme exemplificado na Figura 1. Ressaltamos que os valores de  $\sigma$  e  $d$  devem ser ajustados em conjunto com a constante de regularização  $\gamma$ , para se obter um modelo com boa capacidade de generalização.

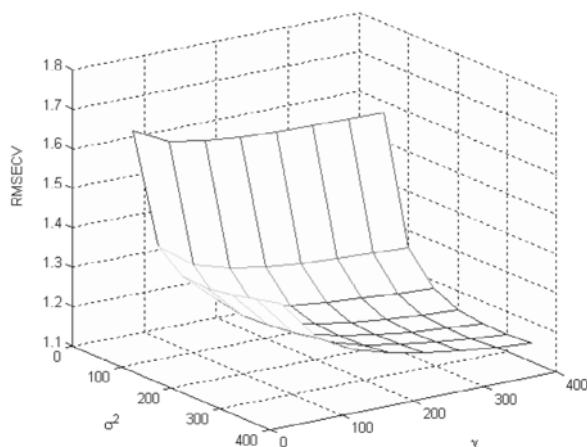


Figura 1. Exemplo da superfície de otimização dos parâmetros  $\sigma^2$  e  $\gamma$  em função do erro (RMSECV)

Em uma abordagem alternativa podemos descrever que as LS-SVM são análogas às redes neurais de avanço, no sentido em que processam a informação de modo sequencial através de camadas e podemos usar funções de base análogas às funções de transferência usadas nas redes neurais. Além disso, existem alguns textos que colocam as LS-SVM como um capítulo das redes neurais. Assim sendo a Figura 2 representa uma LS-SVM como uma rede neural:

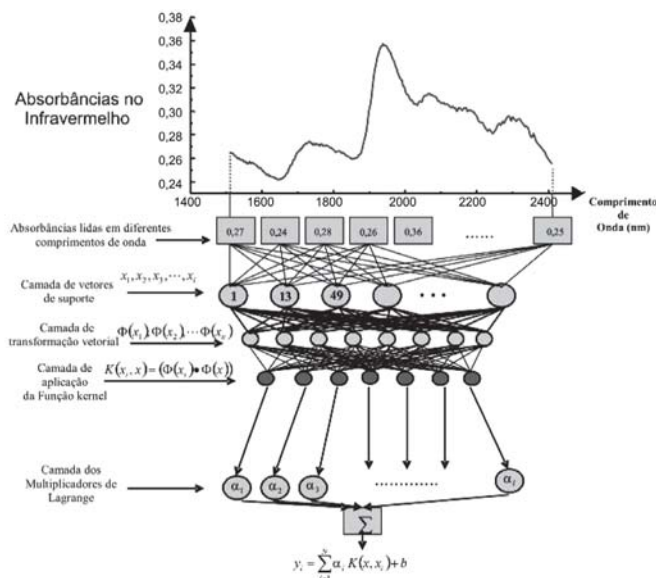


Figura 2. Representação em forma de rede neural para as LS-SVM

Na Figura 2, a informação que entra na rede neural são as absorvâncias lidas em diferentes comprimentos de onda do espectro no infravermelho sendo em seguida processadas pelos vetores suporte, obtidos pelo método de Lagrange. Na seqüência, se necessário, as funções são linearizadas e depois aplicam-se a função núcleo e os pesos, que nada mais são que os multiplicadores de

Lagrange. Finalmente, temos o somatório e a função  $y$  procurada, sendo então comparada com o valor real  $\hat{y}$ . Na grande maioria dos casos é necessário otimizar os valores de  $\sigma^2$  e  $\gamma$  e novamente propagar o sinal pela rede neural até que se obtenha a convergência desejada para os dados de treinamento. Neste caso, realmente as LS-SVM são semelhantes às redes neurais de avanço e às redes com regularização.

Apesar de ambas representações estarem corretas, acreditamos que, como uma primeira abordagem, a apresentação das LS-SVM como redes neurais seja mais simples, uma vez que as redes neurais já são bem conhecidas e usadas em nosso meio acadêmico.

## DETERMINAÇÃO DE AMIDO, SORO E SACAROSE ADICIONADA EM LEITE EM PÓ

### Fraudes em leite em pó

A fiscalização da autenticidade dos alimentos é muito importante, seja por razões econômicas ou de saúde pública. No Brasil, até poucos anos atrás a fraude mais comum era em leite fluido<sup>40</sup>. As contaminações no leite fluido ocorrem adicionando-se água, neutralizadores para mascarar acidez, sal ou açúcar para mascarar água extra ou teor de sólidos elevados, soro de queijo entre outros<sup>41</sup>.

O leite em pó, além do lugar de destaque na alimentação da população, possui grande importância do ponto de vista econômico. Desde 2003 o perfil de exportação mudou, tendo sido exportado em 2004 mais de 55 t de leite em pó, em comparação com o ano de 2000, que foi de menos de 5 t, o que repercute em maior exigência da qualidade do produto, por ex. em análises rápidas em portos e aeroportos, entre outros<sup>42</sup>. Além do grande consumo de leite em pó no comércio varejista, o leite em pó é utilizado largamente na merenda escolar.

A qualidade do leite em pó é regulamentada pelo Regulamento técnico de identidade e qualidade de leite em pó<sup>43</sup> e as fraudes mais frequentes ocorrem com a adição de soro de leite, maltose, sacarose, amido e até soda cáustica. Entre as fraudes mais comuns praticadas no Brasil, está a adição do soro de leite, que é 90% mais barato que o leite e, geralmente, é importado da Argentina, Nova Zelândia e Austrália. Outros contaminantes que têm sido encontrados são amido, sacarose e, com menor frequência, maltodextrina e soda cáustica. A faixa de fraude mais comum é de 20 a 25%, pois até esta o consumidor não percebe diferenças no sabor, mas a fraude pode chegar em até 60%.

### A espectroscopia por reflectância difusa no infravermelho próximo em análises de alimentos

A espectroscopia no infravermelho já foi aplicada em leite para determinação de caseína<sup>44</sup>, proteína e lactose<sup>45</sup>, bem como para classificação<sup>46</sup> de leite e em casos de adulteração com gordura<sup>47</sup>, e com soro em leite em pó desnatado.

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 15 a 90 s), não destrutiva e geralmente sem necessidade de preparo das amostras. Os espectros de NIR de alimentos consistem de amplas bandas correspondendo às sobreposições e combinações de modos vibracionais envolvendo ligações químicas, como C-H, O-H e N-H<sup>48-54</sup>.

A complexidade dos espectros de reflectância difusa, bem como a classificação e quantificação requerem o emprego de ferramentas quimiométricas, que basicamente desenvolvem modelos matemáticos a partir desses sinais instrumentais.

### Preparação das amostras

Amostras de leite em pó tipo a granel Discamp foram contaminadas com soro de leite - fornecido pela Tangará, amido (Synth) e sacarose (Synth), separadamente, em níveis de concentração que foram de 1 a 50% (m/m). Para as misturas ternárias foi realizado um planejamento com 6 níveis de concentração (4, 8, 12, 15, 20 e 25%), conforme Tabela 1, com o propósito de ter amostras com concentrações igualmente espaçadas em toda a faixa estudada para cada um dos contaminantes.

**Tabela 1.** Planejamento de misturas para a adição de dois adulterantes no leite em pó

mistura	leite em pó (%)	adulterante 1 (%)	adulterante 2 (%)
01	92	4	4
02	88	4	8
03	84	4	12
04	81	4	15
05	76	4	20
06	71	4	25
07	88	8	4
08	84	8	8
09	80	8	12
10	77	8	15
11	72	8	20
12	67	8	25
13	84	12	4
14	80	12	8
15	76	12	12
16	73	12	15
17	68	12	20
18	63	12	25
19	81	15	4
20	77	15	8
21	73	15	12
22	70	15	15
23	65	15	20
24	60	15	25
25	76	20	4
26	72	20	8
27	68	20	12
28	65	20	15
29	60	20	20
30	55	20	25
31	71	25	4
32	67	25	8
33	63	25	12
34	60	25	15
35	55	25	20
36	50	25	25
37	92	4	4

### Instrumentação

As amostras foram homogeneizadas com auxílio de vórtex e os espectros obtidos no Espectrofotômetro UV-VIS Cary 5G, em triplicata, na região de 810 m a 2400 nm.

Um total de 7 conjuntos de espectros foram empregados conforme os limites descritos na Tabela 2, incluindo amostras de leite em pó puro.



**Tabela 2.** Descrição dos conjuntos de espectros empregados

conjunto	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

### Modelagem dos dados

Os modelos multivariados foram realizados em um microcomputador com processador Intel Pentium 4 3.00 Ghz e 1 Gbytes de memória RAM, empregando os pacotes PLS-Toolbox<sup>35</sup> e LS-SVM<sup>33</sup> para MATLAB. Aos conjuntos de dados espectrais foi empregada a correção do espalhamento multiplicativo de sinal (MSC).

Para avaliar o desempenho dos modelos de calibração foram calculados os erros de validação cruzada RMSECV (“Root Mean Square Error of Cross-validation”) e de previsão RMSEP (“Root Mean Square Error of Prediction”), conforme a Equação 15.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (15)$$

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

Para os modelos de regressão empregando o método dos mínimos quadrados parciais (PLS), o número de variáveis latentes foi definido através de validação cruzada empregando as amostras do conjunto de calibração descritas na Tabela 3.

**Tabela 3.** Número de espectros empregados na calibração e previsão

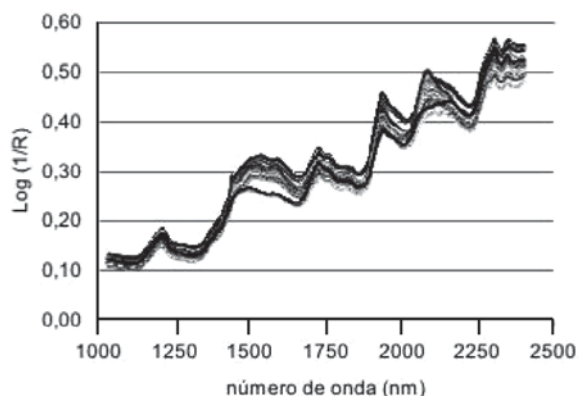
conjunto	espectros para calibração	espectros para previsão
1	30	15
2	30	15
3	30	15
4	20	7
5	20	7
6	20	7
7*	6	6
total	156	72

\* leite em pó puro

### RESULTADOS E DISCUSSÃO

Com o objetivo de corrigir o desvio da linha de base foi aplicada a todos os espectros obtidos a correção do espalhamento multiplicativo (MSC). Os espectros NIR compreendidos entre 1027,5 – 2400 nm de todas as amostras de leite em pó puro e leite adulterado empregadas neste trabalho são apresentados na Figura 3.

Na construção do modelo através das máquinas de vetor suporte por mínimos quadrados é necessário definir os dois parâmetros de entrada (gama e sigma), além das matrizes de espectros e dos vetores das propriedades de interesse para ambos os conjuntos:

**Figura 3.** Conjunto de espectros NIR das amostras de leite em pó puras e adulteradas após tratamento com MSC

calibração e previsão. A Figura 1 ilustra uma típica superfície para otimização do modelo de calibração multivariada empregando LS-SVM. Os parâmetros  $\gamma$  (gama) e  $\sigma^2$  (sigma quadrático) foram combinados, a fim de se verificar a tendência do erro de validação cruzada (RMSECV). Este procedimento é semelhante à escolha do número de variáveis latentes para os modelos de regressão PLS, porém neste caso temos um problema em duas dimensões. Para simplificar o procedimento e a 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 a Equação 16.

$$RMSECV_{(m\u00e9dio)} = \frac{(RMSECV_{(amido)} + RMSECV_{(soro)} + RMSECV_{(sacarose)})}{3} \quad (16)$$

Afim de estudar a dependência do modelo LS-SVM, aqui estudado, com os valores de  $\gamma$  e  $\sigma^2$  foram comparados os valores de  $RMSECV_{m\u00e9dio}$ ,  $RMSEP_{m\u00e9dio}$  e  $R^2$  de calibração para as diferentes combinações destes. De forma semelhante ao apresentado na Figura 1, pode-se observar na Tabela 4 com mais detalhes que, à medida que aumentamos os valores de ambos os parâmetros, os valores de  $RMSECV_{m\u00e9dio}$  e  $RMSEP_{m\u00e9dio}$  decrescem inicialmente de forma intensa, até que por volta de 200 a 300 os valores de  $RMSECV_{m\u00e9dio}$  e  $RMSEP_{m\u00e9dio}$  passam a não apresentar mais uma sensível redução. Por outro lado, quando fixamos o valor de  $\gamma = 200$  e incrementamos  $\sigma^2$ , os valores de RMSE tendem a aumentar. Se fixamos  $\sigma^2 = 200$  e aumentamos  $\gamma$  os valores de RMSEP, apesar de decrescerem, tornam-se mais distantes e com isso aumenta o risco de “overfitting”. Também foi possível observar que se continuarmos incrementando o valor de  $\gamma$  para o infinito, o  $RMSECV_{m\u00e9dio}$  tende a um mínimo, entretanto certamente resultará em um modelo sobredimensionado com maior tendência a erros de previsão das amostras externas.

Para selecionar o número adequado de variáveis latentes nos modelos PLS também foi empregado o RMSECV, cujos resultados para cada adulterante são apresentados na Figura 4. Considerando a quantificação do amido, pode-se observar que acima de 9 variáveis latentes o RMSECV não decresce mais significativamente. Também fica claro que o número mais adequado de VL é diferente para cada caso em estudo, sendo que os mais adequados são apresentados na Tabela 5.

Ao compararmos os resultados para cada um dos analitos, 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 3 analitos, apresentando valores de RMSECV, RMSEC e RMSEP significativamente melhores.

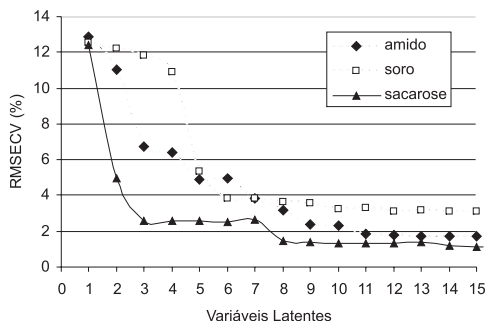
**Tabela 4.** Comparação dos resultados de diferentes combinações dos parâmetros  $\gamma$  e  $\sigma^2$  para os modelos LS-SVM

$\gamma$	$\sigma^2$	RMSECV <sup>a</sup>	R <sup>2</sup>			RMSEP			
			amido	soro	sacarose	amido	soro	sacarose	média
50	50	1,651	0,9997	0,9995	0,9998	1,435	1,445	1,148	1,342
100	100	1,321	0,9997	0,9994	0,9998	1,179	1,192	0,726	1,032
150	150	1,228	0,9997	0,9994	0,9998	1,112	1,112	0,625	0,950
200	200	1,189	0,9997	0,9993	0,9998	1,083	1,092	0,595	0,923
250	250	1,169	0,9997	0,9993	0,9998	1,067	1,095	0,585	0,916
300	300	1,160	0,9997	0,9993	0,9998	1,056	1,105	0,583	0,915
200	300	1,189	0,9995	0,9989	0,9997	1,112	1,133	0,618	0,954
200	350	1,195	0,9994	0,9987	0,9997	1,124	1,163	0,633	0,973
300	200	1,166	0,9998	0,9996	0,9999	1,038	1,084	0,569	0,897
400	200	1,153	0,9998	0,9997	0,9999	1,013	1,082	0,556	0,884

<sup>a</sup> média dos valores entre amido, soro e sacarose

**Tabela 5.** Comparação do desempenho entre os modelos de calibração PLS e LS-SVM para a quantificação de amido, soro e sacarose em amostras de leite em pó

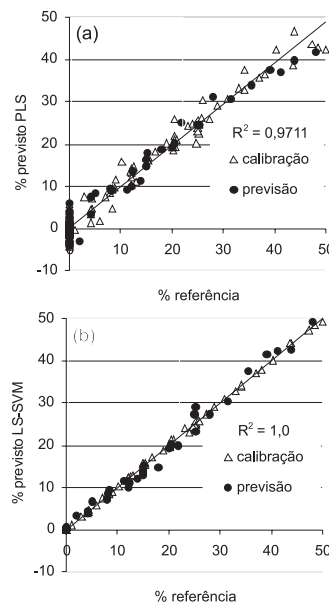
	amido		soro		sacarose	
	PLS	LS-SVM	PLS	LS-SVM	PLS	LS-SVM
R <sup>2</sup>	0,9711	0,9998	0,9174	0,9996	0,9599	0,9999
RMSECV (%)	2,3889	1,0330	3,8426	1,0840	2,5778	0,5690
RMSEP (%)	2,3330	1,0381	3,1514	1,0839	2,8045	0,5697
LV	9	-	6	-	3	-
$\gamma$	-	300	-	300	-	300
$\sigma^2$	-	200	-	200	-	200



**Figura 4.** Valores de RMSECV em função do número de variáveis latentes para os modelos de regressão empregando PLS

Nas Figuras 5 a 7 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. Embora os coeficientes de correlação para os três analitos sejam adequados para ambas as técnicas de regressão, pode-se observar que os modelos empregando o PLS apresentam maior dispersão, tanto para as amostras de calibração quanto de previsão, 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 analitos nas amostras em estudo. Já os modelos empregando LS-SVM sempre foram superiores na capacidade preditiva do ponto zero, em função da sua maior capacidade de generalização.

Afim 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 se encontram nas amostras estudadas, são apresentados na Tabela 6, comparativamente, os valores de previsão para as



**Figura 5.** Gráficos de correlação para calibração e previsão do % de amido adicionado como adulterante: (a) modelo PLS empregando 9 variáveis latentes; (b) modelo LS-SVM

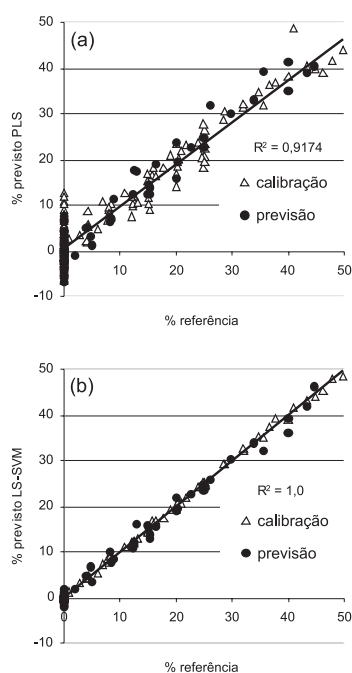
6 amostras de leite em pó puro do conjunto de validação.

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 0,3577% para o amido na amostra 2, enquanto que os modelos PLS além de preverem valores negativos significativos chegam a indicar 8,0796% de sacarose na amostra 4 que, tecnicamente, não apresenta adição do mesmo.

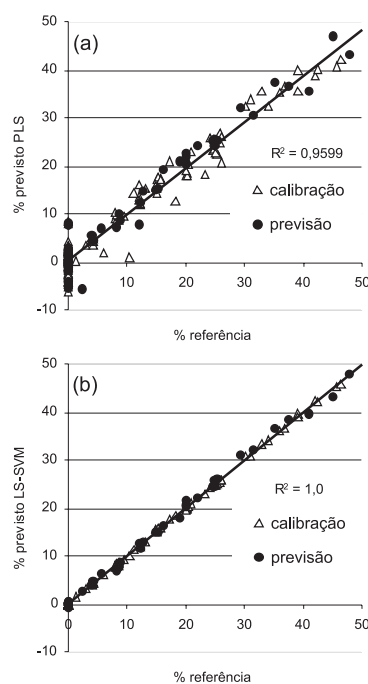
**Tabela 6.** Erro absoluto para a previsão das amostras de leite puro para os modelos de regressão multivariada PLS e LS-SVM

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	2,3226	0,0829	-5,3576	-0,2594	2,0620	0,0082
2	-0,0658	0,3577	-0,5010	-0,2548	1,7390	0,0097
3	1,5587	0,2940	-3,7136	-0,0733	1,7118	-0,0078
4	-4,2903	0,1257	-0,6940	0,0105	8,0796	0,0675
5	-3,4826	-0,0641	-3,3753	0,0533	7,9007	0,0172
6	-3,6347	0,0755	-1,9933	0,0062	7,8189	0,0189
<b>Média<sup>b</sup></b>	<b>2,5591</b>	<b>0,1667</b>	<b>2,6058</b>	<b>0,1096</b>	<b>4,8853</b>	<b>0,0216</b>

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



**Figura 6.** Gráficos de correlação para calibração e previsão do % de soro adicionado como adulterante: (a) modelo PLS empregando 6 variáveis latentes; (b) modelo LS-SVM



**Figura 7.** Gráficos de correlação para calibração e previsão da % de sacarose adicionada como adulterante: (a) modelo PLS empregando 3 variáveis latentes; (b) modelo LS-SVM

## CONCLUSÕES

Neste artigo apresentamos os fundamentos das máquinas de vetor suporte por quadrados mínimos (LS-SVM) e propomos sua aplicação na quantificação de amido, soro e/ou sacarose em amostras de leite em pó. Para efeito de comparação utilizamos o clássico método de regressão por mínimos quadrados parciais (PLSR). Os resultados demonstraram o excelente desempenho dos modelos de regressão multivariada desenvolvidos com LS-SVM. Dentre as principais vantagens das LS-SVM podemos verificar sua capacidade de inferir sobre a ausência de um ou mais dos adulterantes avaliados, enquanto que todos os modelos PLSR apresentaram falso positivo.

Também foram demonstradas a grande capacidade de generalização e a flexibilidade das LS-SVM, combinando diferentes conjuntos de espectros na construção de um único modelo de calibração para cada analito considerado.

Finalmente, a utilização da LS-SVM na construção de modelos de regressão multivariados mostrou-se uma técnica bastante promissora para ser empregada na avaliação da qualidade de produtos alimentícios através de medidas indiretas de análise, como por ex., espectros de reflexão no infravermelho próximo (NIR).

## AGRADECIMENTOS

Ao CNPq pelas bolsas concedidas.

## REFERÊNCIAS

- Vapnik, V. Em *Nonlinear Modeling: Advanced Black-Box Techniques*; Suykens, J. A. K.; Vandewalle, J., eds.; Kluwer Academic Publishers: Boston, 1998, p. 55-85.
- Suykens, J. A. K.; *Eur. J. Control* **2001**, *7*, 311.
- Burges, C. J. C.; *Data Min. Knowl. Discov.* **1998**, *2*, 121.
- Smola, A. J.; Schölkopf, B.; *Stat. Comput.* **2004**, *14*, 199.
- Belousov, A. I.; Verzakov, S. A.; von Frese, J.; *J. Chemom.* **2002**, *16*, 482.
- Zomer, S.; Brereton, R. G.; Carter, J. F.; Eckers, C.; *Analyst* **2004**, *129*, 175.
- Pierna, J. A. F.; Volery, P.; Besson, R.; Baeten, V.; Dardenne, P.; *J. Agric. Food Chem.* **2005**, *53*, 6581.
- Brudzewski, K.; Kesik, A.; Kolodziejczyk, K.; Zborowska, U.; Ulaczyk, J.; *Fuel* **2006**, *85*, 553.
- Brudzewski, K.; Osowski, S.; Markiewicz, T.; Ulaczyk, J.; *Sens. Actuators, B* **2006**, *113*, 135.
- Burbidge, R.; Trotter, M.; Buxton, B.; Holden, S.; *Comput. Chem.* **2001**, *26*, 5.
- Harz, M.; Rosch, P.; Peschke, K. D.; Ronneberger, O.; Burkhardt, H.; Popp, J.; *Analyst* **2005**, *130*, 1543.

12. Peng, S. H.; Xu, Q. H.; Ling, X. B.; Peng, X. N.; Du, W.; Chen, L. B.; *FEBS Lett.* **2003**, 555, 358.
13. Lu, C.; van Gestel, T.; Suykens, J. A. K.; van Huffel, S.; Vergote, I.; Timmerman, D.; *Artif. Intell. Med.* **2003**, 28, 281.
14. Devos, A.; Lukas, L.; Suykens, J. A. K.; Vanhamme, L.; Tate, A. R.; Howe, F. A.; Majos, C.; Moreno-Torres, A.; van der Graaf, M.; Arus, C.; van Huffel, S.; *J. Magn. Reson.* **2004**, 170, 164.
15. Martin, T. C.; Moecks, J.; Belousov, A.; Cawthraw, S.; Dolenko, B.; Eiden, M.; von Frese, J.; Köhler, W.; Schmitt, J.; Somorjai, R.; Udelhoven, T.; Verzakov, S.; Petrich, W.; *Analyst* **2004**, 129, 897.
16. Acir, N.; *Neural Comput. Appl.* **2005**, 14, 299.
17. Majewski, P.; Jedruch, W.; *Lect. Notes Artif. Intell.* **2005**, 3533, 400.
18. Thissen, U.; Üstün, B.; Melssen, W. J.; Buydens, L. M. C.; *Anal. Chem.* **2004**, 76, 3099.
19. Thissen, U.; Pepers, M.; Üstün, B.; Melssen, W. J.; Buydens, L. M. C.; *Chemom. Intell. Lab. Syst.* **2004**, 73, 169.
20. Chauchard, F.; Cogdill, R.; Roussel, S.; Roger, J. M.; Bellon-Maurel, V.; *Chemom. Intell. Lab. Syst.* **2004**, 71, 141.
21. Cogdill, R. P.; Schimleck, L. R.; Jones, P. D.; Peter, G. F.; Daniels, R. F.; Clark, A.; *J. Near Infrared Spectrosc.* **2004**, 12, 263.
22. Vapnik, V.; Lerner, A.; *Automat. Remote Control* **1963**, 24, 774.
23. Mangasarian, O. L.; *Oper. Res.* **1965**, 13, 444.
24. Vapnik, V. N.; *The Nature of Statistical Learning Theory*, Springer: New York, 2<sup>nd</sup> ed., 2000.
25. Vapnik, V.; Golowich, S.; Smola, A. Em *Neural Information Processing Systems*; Mozer, M.; Jordan, M.; Petsche, T., eds.; MIT Press: Cambridge, 1997, vol. 9.
26. Lima, C. A. M.; *Tese de Doutorado*, Universidade Estadual de Campinas, Brasil, 2004.
27. Poggio, T.; Torre, V.; Koch, C.; *Nature* **1985**, 317, 314.
28. Hornik, K.; Stinchcombe, M.; White, H.; *Neural Netw.* **1989**, 2, 359.
29. Cerqueira, E. O.; Andrade, J. C.; Poppi, R. J.; Mello, C.; *Quim. Nova* **2001**, 24, 864.
30. Smola, A. J.; Barlett, P.; Schölkopf, B.; Schuurmans, D. Em *Advances in Large Margin Classifiers*, Barlett, P.; Schuurmans, D.; Schölkopf, B., eds.; MIT Press: Cambridge 2000, cap 1, p. 1-28.
31. Giampaolo, L. L.; Netto, A. V.; de Carvalho, A. P. L. F.; de Oliveira, M. C. F.; *Sba, Soc. Bras. Autom.* **2005**, 16, 146.
32. Schölkopf, B.; Burges, C. J. C.; Smola, A. J.; *Advances in Kernel Methods: Support Vector Learning*, MIT Press: Cambridge, 1999.
33. Suykens, J. A. K.; van Gestel, T.; de Brabanter, J.; de Moor, B.; Vandewalle, J.; *Least-Squares Support Vector Machines*, World Scientific: Singapore, 2002.
34. Vapnik, V.; *Statistical Learning Theory*, John Wiley & Sons: New York, 1998.
35. Thissen, U.; van Brakel, R.; de Weijer, A. P.; Melssen, W. J.; Buydens, L. M. C.; *Chemom. Intell. Lab. Syst.* **2003**, 69, 35.
36. Suykens, J. A. K.; Vandewalle, J.; *Neural Proc. Lett.* **1999**, 9, 293.
37. Arfken, G. B.; Weber, H. J.; *Mathematical Methods for Physicists*, 4<sup>th</sup> ed., Academic Press: New York, 1995, p. 973.
38. Mercer, J.; *Philos. Trans. Roy. Soc. London Ser. A* **1909**, 209, 415.
39. Cogdill, R. P.; Dardenne, P.; *J. Near Infrared Spectrosc.* **2004**, 12, 93.
40. Silva, H. F.; *Físico-química do leite e derivados: métodos analíticos*, Ed. Gemacom: Juiz de Fora, 1997.
41. Harding, F.; *Adulteration of milk - Milk Quality*, Chapman e Hall Food Science Book: New York, 1999.
42. <http://www.cnpq.br/producao/06mercado/tabela06.20.php>, acessada em Dezembro 2005.
43. <http://extranet.agricultura.gov.br/consultas/legis/do/consultaLei?op=viewTextual&codigo=5493>, acessada em Março 2006.
44. Sorensen, L. K.; Lund, M.; Juul, B.; *J. Dairy Res.* **2003**, 70, 445.
45. Sasic, S.; Ozaki, Y.; *Anal. Chem.* **2001**, 73, 64.
46. Downey, G.; Robert, P.; Bertrand, D.; Kelly, P. M.; *Appl. Spectrosc.* **1990**, 44, 150.
47. Sato, T.; Kawano, S.; *J. Dairy Sci.* **1990**, 73, 3408.
48. Osborne, B. G. Em *Near-infrared spectroscopy in food analysis*; Encyclopedia of Analytical Chemistry, Meyers, R. A., ed.; Wiley: Chichester, 1999.
49. Parreira, T. F.; Ferreira, M. M. C.; Sales, H. J. S.; de Almeida, W. B.; *Appl. Spectrosc.* **2002**, 56, 1607.
50. Pasquini, C.; *J. Braz. Chem. Soc.* **2003**, 14, 198.
51. Rodrigues, J. C.; Nascimento, A. C.; Alves, A.; Osório, N. M.; Pires, A. S.; Gusmão, J. H.; da Fonseca, M. M. R.; Ferreira-Dias, S.; *Anal. Chim. Acta* **2005**, 544, 213.
52. Morgano, M. A.; Faria, C. G.; Ferrão, M. F.; Bragagnolo, N.; Ferreira, M. M. C.; *Ciênc. Tecnol. Aliment.* **2005**, 25, 25.
53. Liu, Y.; Ying, Y.; Yu, H.; Fu, X.; *J. Agric. Food Chem.* **2006**, 54, 2810.
54. Kovalenko, I. V.; Rippke, G. R.; Hurburgh, C. R.; *J. Am. Oil Chem. Soc.* **2006**, 83, 421.
55. 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.