

A ESCOLHA DA FAIXA ESPECTRAL NO USO COMBINADO DE MÉTODOS ESPECTROSCÓPICOS E QUIMIOMÉTRICOS

Flavia C. C. Oliveira, Antônio T. P. C. de Souza, José A. Dias, Silvia C. L. Dias e Joel C. Rubim*

Instituto de Química, Universidade de Brasília, CP 04478, 70910-970 Brasília-DF

Recebido em 6/3/03; aceito em 29/9/03

THE CHOICE OF THE SPECTRAL REGION IN THE USE OF SPECTROSCOPIC AND CHEMOMETRIC METHODS. A method is presented for the choice of spectral regions when absorption measurements are coupled to chemometric tools to perform quantitative analyses. The method is based on the spectral distribution of the relative standard deviation of concentration (s/c). It has been applied to the development of PLS-FTNIR calibration models for the determination of density and MON of gasoline, and ethanol content and density of ethanol fuel. The new method was also compared with the correlation (R^2) method and has proved to generate PLS calibration models that present better accuracy and precision than those based on R^2 .

Keyword: FTNIR; PLS; spectral region selection.

INTRODUÇÃO

O uso da espectroscopia de absorção no infravermelho próximo (NIR, "near infrared"), combinada com o tratamento dos dados por métodos quimiométricos, tornou-se rotineiro, encontrando aplicações na classificação de restos de demolição¹, no monitoramento "on line" de processos de polimerização², na análise de fármacos³⁻⁵, solos^{6,7} e alimentos⁸, na agro-indústria^{8,9}, na indústria têxtil¹⁰, em refinarias^{11,12}, além de várias outras aplicações.

Dentre os métodos quimiométricos que têm sido propostos para a análise espectral de multicomponentes, os mais populares são o PCR ("principal component regression")¹³, o PLS ("partial least-squares")¹⁴ e seus análogos¹⁵⁻¹⁹. Com base em algumas suposições, conseguiu-se demonstrar, teoricamente, que a inclusão de mais regiões espectrais sempre melhora a eficiência de previsão dos modelos baseados em análise multivariada²⁰. Como bem colocado por Jiang *et al.*²¹, a consequência direta deste tipo de suposição sugere que tais métodos podem levar à construção de modelos de calibração usando espectros inteiros, sem a necessidade de seleção de regiões espectrais específicas. No entanto, evidências, tanto teóricas^{22,23} como experimentais^{24,25}, indicam que a escolha das regiões espectrais pode melhorar significativamente a eficiência destas técnicas de calibração.

A escolha da região espectral envolve a escolha de determinados canais espectrais (um comprimento de onda ou um conjunto de comprimentos de onda), que permitem ao modelo de calibração minimizar os erros de previsão. Os benefícios da seleção da região espectral vão desde a estabilidade do modelo de calibração em relação à colinearidade, como também na interpretação das relações entre o modelo e a composição da amostra²¹.

Existem vários procedimentos para a escolha da região espectral. Alguns baseiam-se em critérios objetivos para a avaliação do melhor conjunto de canais espectrais, enquanto outros se baseiam em algoritmos que localizam o conjunto de canais espectrais que podem dar os melhores resultados. Como exemplos de critérios objetivos podemos citar a análise da distribuição espectral da relação sinal-ruído, do determinante da matriz de calibração, bem como a avaliação da raiz quadrada dos erros médios de previsão (RMSEP)²⁶. No

caso de algoritmos para pesquisa e seleção de regiões espectrais, existem várias metodologias^{21,24,27-31}, dentre as quais destacamos os algoritmos genéticos^{30,31} e, mais recentemente, o modelo PLS de janela móvel ("MWPLSR – moving window partial least-squares regression")²¹.

Os manuais de programas de aplicativos para PCR e PLS^{32,33} ou não discutem as implicações da falta de critérios para a escolha da região espectral, ou o fazem de modo superficial. Além disso, em alguns desses pacotes de aplicativos para PLS, a única ferramenta oferecida para a escolha da região espectral é o espectro de correlação (R^2 vs λ em nm). No entanto, o uso desta ferramenta não leva em consideração alguns aspectos importantes como os que discutiremos neste trabalho.

Neste trabalho pretendemos apresentar uma metodologia relativamente simples para a escolha da região espectral a qual considera os seguintes aspectos:

1. Regiões espectrais inúteis – São as regiões do espectro onde variações nas concentrações dos componentes da mistura não causam variações na absorvância. A eliminação dessas regiões reduz o número de dados e, conseqüentemente, a quantidade de memória de computador e o tempo necessário para realizar todos os cálculos de calibração. Além disso, as incertezas devido ao ruído instrumental seriam eliminadas, pois o ruído associado a cada canal espectral não considerado não contribuiria para a incerteza da medida.
2. Não linearidade – Regiões onde a absorvância é muito grande levam a desvios da lei de Beer-Lambert, ou seja, a relação entre a absorvância e a concentração não é mais linear. Por exemplo, a absorvância molar de uma determinada espécie num determinado comprimento de onda depende do índice de refração do meio. Por sua vez, o índice de refração depende das concentrações das espécies absorventes naquele comprimento de onda. Logo, para concentrações elevadas (> 0,01 mol/L), não se observará uma relação linear entre a absorvância e a concentração^{34,35}. Análise de fatores por PLS ou PCA pode corrigir alguns tipos de não linearidade, mas não podem corrigir saturação de absorção quando, por exemplo, se usa água como referência na determinação de espécies em solução aquosa.
3. Relação sinal-ruído – É dada pela razão entre a amplitude média do sinal e o desvio padrão do sinal medido. Existem algumas

*e-mail: jocrubim@unb.br

referências onde este problema é discutido especificamente em relação às medidas de absorvância^{34,35}. De forma geral, pode-se dizer que a precisão de uma medida espectroscópica é limitada pelas incertezas ou ruídos associados ao instrumento utilizado.

Uma medida espectrofotométrica envolve o registro de I_o , a intensidade de luz captada pelo detector na ausência da amostra, ou seja, a referência e a medida de I , a intensidade de luz captada pelo detector na presença da amostra. O espectro de transmissão, $T(\%)$ vs $\lambda(\text{nm})$, será dado pela relação I/I_o em cada comprimento de onda (λ). No entanto, a lei de Beer-Lambert mostra que a absorvância é que apresenta uma relação linear com a concentração. Neste caso, o instrumento também pode fornecer o espectro de absorção, A vs $\lambda(\text{nm})$, registrando os valores de $\log(I_o/I)$ ou $-\log T$ para cada comprimento de onda. Portanto, a incerteza na determinação da concentração através da lei de Beer-Lambert:

$$A = \epsilon bc \quad (1)$$

onde ϵ é a absorvidade molar, b o caminho óptico e c a concentração, será determinada pelas incertezas nas medidas de I_o e I , ou seja, nas medidas de T , uma vez que $T = I/I_o$.

A partir da lei de Beer-Lambert é possível derivar-se uma expressão que relaciona a incerteza relativa de concentração (s_c/c) com o desvio padrão absoluto de transmitância (s_T) como sendo³⁴:

$$\frac{s_c}{c} = \frac{0,434s_T}{-T \log T} \quad (2)$$

A partir da Equação 2 pode-se ver que a incerteza associada à concentração obtida por medidas espectroscópicas varia de uma forma complexa em relação à transmitância, podendo ser mais complexa ainda, uma vez que s_T também pode depender de T ³⁴.

Rothman *et al.*³⁴ mostraram que existem várias fontes instrumentais de erro em medidas espectroscópicas, as quais afetam a precisão de medidas de absorção (transmissão) e, portanto, a determinação de concentrações com base nestas medidas. Dentre elas podemos destacar os ruídos associados ao tipo de detector utilizado, ruído térmico (p.ex. agitação térmica dos transportadores de carga de detectores baseados em semi-condutores, responsável pelo ruído térmico ou “dark current”) e “photon shot noise” (ruído associado à transferência de carga, por exemplo, em junções p-n ou em detectores baseados em tubos fotomultiplicadores), ruídos de baixa frequência (“flicker noise” - são importantes para frequências abaixo de 100 Hz - um exemplo deste tipo de ruído é a variação na intensidade da fonte luminosa³⁴), incertezas associadas ao posicionamento da amostra e variações na temperatura ambiente (ruído ambiental).

De acordo com o trabalho de Rothman *et al.*³⁴, na ausência de ruídos de baixa frequência, ruídos ambientais ou ruídos devido à incerteza no posicionamento da amostra (diferenças de caminho óptico), a incerteza da medida de transmitância está associada ao ruído introduzido pelo sistema de leitura do sinal luminoso. Se ruídos térmicos forem as únicas fontes de ruído, o desvio padrão de transmitância, s_T , será uma constante (comum em detectores para medidas no infravermelho e infravermelho próximo, baseados em semi-condutores). Caso a incerteza da transmitância seja governada pelo “photon shot noise” na medida de I_o (referência) e I (amostra), o desvio padrão absoluto de transmitância será dado por^{34,35}:

$$s_T = k\sqrt{T^2 + T} \quad (3)$$

É possível encontrar na literatura alguns artigos que, em sua busca para a construção de um modelo de calibração, não levam em consideração alguns dos pontos levantados acima, principalmente os pon-

tos 2 e 3. Encontramos, assim, artigos onde somente os parâmetros estatísticos obtidos pelos modelos de calibração são apresentados, sem a apresentação dos espectros³⁶, não permitindo uma avaliação da confiabilidade do modelo obtido em função da faixa espectral escolhida. Encontramos também trabalhos onde a água é usada como referência, resultando em erros de previsão significativamente elevados^{37,38}, bem como trabalhos onde as faixas espectrais escolhidas estão sujeitas a erros significativos nas medidas de absorvância^{39,40}.

Poucos são os trabalhos encontrados na literatura que discutem a questão da escolha da região espectral, considerando aspectos como os três pontos citados acima, principalmente o ponto relativo à incerteza relativa de concentração (ponto 3) e suas implicações nos resultados obtidos. Na literatura encontramos apenas um artigo que fala brevemente sobre o efeito de se trabalhar em regiões onde a absorvância leva a desvios da lei de Beer-Lambert⁴¹.

O objetivo deste trabalho é mostrar como os pontos citados acima, quando não devidamente avaliados, durante a escolha da região espectral, podem comprometer a precisão de modelos de calibração com base em medidas espectroscópicas. Para tanto, faremos uso de um conjunto de resultados obtidos na construção de modelos de calibração PLS-FTNIR para a determinação de MON (número de octanas pelo método do motor) e densidade em gasolina, e teor de etanol e densidade em álcool etílico hidratado combustível (AEHC). Os resultados serão comparados com aqueles obtidos empregando-se o método do espectro de correlação (R^2)⁴² para a seleção da região espectral.

PARTE EXPERIMENTAL

Amostras e procedimentos para obtenção das curvas s_c/c vs A (absorvância) e s_c/c vs cm^{-1}

Para estas medidas foram usadas três amostras de etanol/água, preparadas por pesagem em balança analítica (Mettler Toledo, $\pm 0,0001$ g), misturando-se etanol anidro (Merk) e água destilada, com teores de etanol de 100%, 70% e 50%. Foram empregadas também 4 amostras de gasolina, uma do tipo A e três gasolinas tipo C (fornecidas pelo CEPAT), com teores de etanol de 20%, 30% e 35%.

Para a obtenção das curvas s_c/c vs A foram registrados 20 espectros de absorção de cada amostra. Para cada espectro foram registrados I e I_o em intervalos constantes de tempo (5 min). Para cada conjunto de 20 espectros foram obtidos os valores do desvio padrão absoluto da transmitância, s_T , para as absorções consideradas. Os valores de s_c/c foram obtidos com base nos valores de transmitância média de cada banda de absorção considerada, através da Equação 2.

Os espectros de incerteza relativa de concentração (s_c/c vs cm^{-1}) foram obtidos calculando-se a média de 20 espectros de transmissão e o espectro do desvio padrão absoluto, ou seja, o desvio padrão absoluto em cada número de onda. Conhecendo-se T (a transmitância) para cada número de onda da média dos 20 espectros, e aplicando-se a Equação 2, chega-se facilmente à distribuição espectral da incerteza relativa de concentração.

Amostras de gasolina

Para a construção do modelo de calibração para determinação da densidade e MON (número de octanas pelo método do motor) em gasolinas foram usados 30 padrões de gasolina A e 50 de gasolina C. Os valores de densidade foram determinados pela norma ASTM D 4052 e variam de 0,7413 a 0,7727 g/mL, com um valor médio de 0,7458 e desvio padrão de 0,009 g/mL. Os valores de MON foram obtidos segundo a norma ASTM D 2700 e variam de 81 a 85, com um valor médio de 82,4 e desvio padrão de 1,2.

Amostras de álcool etílico hidratado combustível (AEHC)

Na construção dos modelos de calibração para a determinação de teor de etanol e densidade foram empregadas 120 amostras certificadas de AEHC fornecidas pela Copersucar. Os teores alcoólicos das 120 amostras consideradas variam entre 90,26 e 100% m/m (valor médio 92,85% m/m e desvio padrão de 0,25% m/m) e suas densidades variam de 0,8094 a 0,8129 g/mL com um valor médio de 0,8117 g/mL e desvio padrão de 0,00067 g/mL. Os teores de etanol e densidade foram determinados segundo a norma ASTM D4052.

Obtenção dos espectros FTNIR

Os espectros das amostras utilizadas nos modelos de calibração PLS foram obtidos em triplicada, usando-se um equipamento FTNIR da Bruker (Equinox 55) equipado com acessório de transflctância, que se conecta ao equipamento por meio de fibras ópticas. O caminho óptico utilizado foi o menor disponível, de 1,0 mm (0,5 mm é a distância entre a superfície refletora e a fibra óptica). Cada espectro é o resultado de 16 varreduras usando-se o ar como referência. Em cada caso será citada a resolução espectral utilizada.

RESULTADOS E DISCUSSÃO

A Figura 1 mostra espectros de absorção no infravermelho próximo de amostras de gasolina com diferentes teores de etanol e amostras de etanol com diferentes teores de água. As linhas base dos espectros das amostras de etanol/água foram deslocadas para valores maiores de absorbância para facilitar a visualização dos espectros e ressaltar suas diferenças. A absorbância média próximo a 10000 cm^{-1} é de 0,2 para os dois sistemas. Para ambos os casos, a resolução espectral foi de 8 cm^{-1} . Estas amostras foram escolhidas para a avaliação do comportamento da incerteza relativa de concentração (s_r/c) em função da absorbância (ver Equação 2). As curvas que mostram a dependência de s_r/c em relação à absorbância são apresentadas nas Figuras 2 (a) e (b). A Tabela 1 apresenta as faixas espectrais consideradas, em cada caso, para a realização das medidas de transmitância.

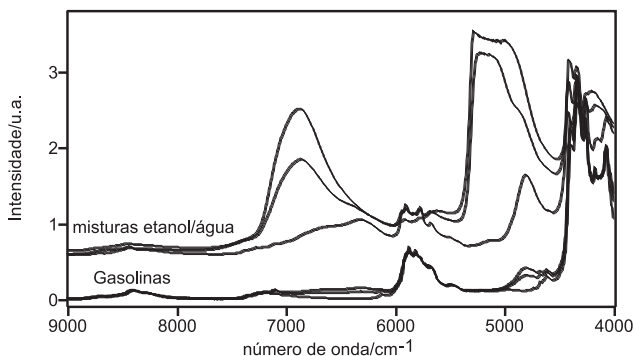


Figura 1. Espectros FTNIR de amostras de gasolina e amostras de misturas etanol/água

As curvas apresentadas na Figura 2 foram obtidas por ajuste dos pontos experimentais usando-se a expressão:

$$y(\%) = \frac{0,434a}{x10^{-x}} \quad (4)$$

onde $10^{-x} = T$ e a é um parâmetro ajustável. O ajuste obtido foi razoável, considerando o fato de o termo Q^2 (ver quadros da Figura 2), que dá a média do quadrado das diferenças entre os valores experi-

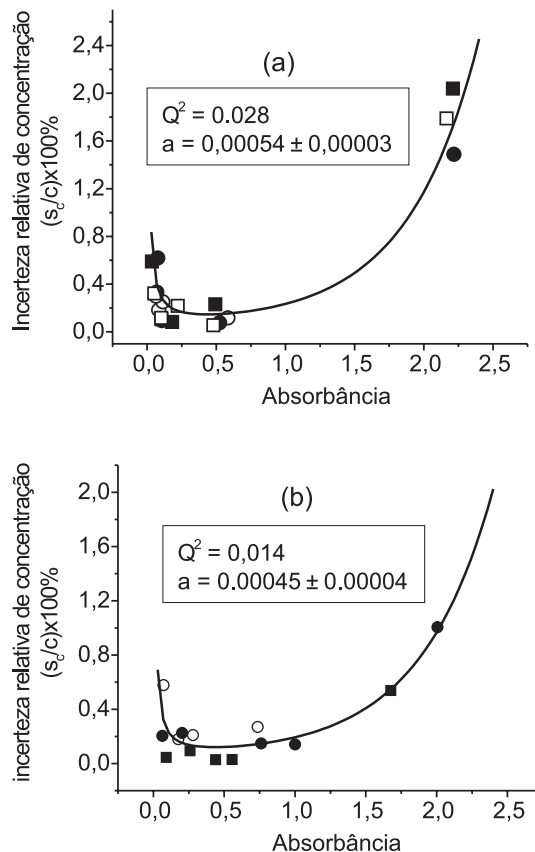


Figura 2. Incerteza relativa de concentração (%) em função da absorbância para (a) gasolina com teores de EtOH (% v/v) de 0 (○), 20 (●), 30 (■), 35 (□) e para (b) misturas etanol/água com teores de EtOH (%v/v) de 100 (■), 70 (○) e 50 (●). As linhas são o resultado de um ajuste dos pontos experimentais à Equação 4

Tabela 1. Faixas espectrais consideradas nas medidas de transmitância para amostras de gasolina e misturas etanol/água

Faixa espectral cm^{-1}	amostra	Atribuição
8012-8784	EtOH/água	3x $\nu^s(\text{CH})$
6032-7418		2x $\nu(\text{OH})$ da água
5378-5992		2x $\nu(\text{CH})$
5002-5477		$\nu(\text{OH}) + \delta^s(\text{OH})$ da água
4548-5050		$\nu(\text{OH}) + \delta(\text{OH})$ do etanol
7862-8847		3x $\nu(\text{CH})$
6536-7521		2x $\nu(\text{CH}) + \delta(\text{CH})$
5362-6233	Gasolina	2x $\nu(\text{CH})$
4737-5135		$\nu(\text{OH}) + \delta(\text{OH})$ do etanol
4135-4500		$\nu(\text{CH}) + \delta(\text{CH})$

^a ν e δ representam modos de estiramento e deformação angular, respectivamente.

mentais e os valores ajustados, ser relativamente próximo de zero. De acordo com a Equação 2, $a = s_r$ e os quadros inseridos na Figura 2 mostram os valores deste parâmetro para os dois sistemas investigados. Os resultados mostram que s_r é da ordem de 5×10^{-4} %, o que é compatível com o tipo de detector utilizado, um semi-condutor, para o qual são os ruídos térmicos que determinam a incerteza na medida da absorbância³⁴.

Os resultados da Figura 2 mostram que os valores de incerteza relativa de concentração, para absorvâncias menores do que a unidade, estão abaixo de 0,2%, voltando a crescer somente para absorvâncias menores que 0,1, chegando a serem inferiores a 1% para absorvâncias da ordem de 0,03. Os valores mínimos de s_c/c foram de 0,15 e 0,12% para gasolina e AEHC, respectivamente. Estes valores foram obtidos para a absorvância de 0,43, nos dois casos. De acordo com Rothman *et al.*³⁴, para o tipo de detector utilizado, a teoria prevê um valor mínimo de s_c/c para absorvância de 0,475. Em seu trabalho, estes autores obtiveram um valor de absorvância experimental de 0,423.

Os resultados da Figura 2 mostram também que a escolha de regiões espectrais onde a absorvância é maior do que a unidade leva a valores de s_c/c elevados. Por exemplo, para absorvância 2,0 os valores de s_c/c podem ser quatro a cinco vezes maiores que para absorvâncias menores que 1,0. Chung *et al.*⁴¹ consideram que bandas com absorvâncias maiores que a unidade não devam ser incluídas devido a efeitos de não linearidade. No entanto, os efeitos de não linearidade estão associados às concentrações das espécies absorventes e, como citado anteriormente, modelos PLS ou PCR conseguem contornar estes efeitos. Por exemplo, se o caminho óptico utilizado for muito grande, mesmo para soluções diluídas, a absorvância poderá ser maior que a unidade e efeitos de não linearidade não serão observados, no entanto, os valores de s_c/c serão elevados. Portanto, os resultados da Figura 2 mostram que as incertezas relativas nas medidas de I e I_0 são fatores importantes a serem considerados na escolha da região espectral e mostram, claramente, que a escolha de regiões espectrais onde a absorvância é maior que a unidade pode levar a modelos de calibração pouco precisos e inexatos.

Como exemplo, a Tabela 2 mostra os parâmetros de modelos de calibração PLS para a determinação de MON em gasolina e teor alcoólico em AEHC. Para cada tipo de combustível analisado foram escolhidas duas regiões, uma, onde a absorvância é maior do que a unidade, e outra, onde a absorvância é menor que a unidade. Embora se tratem de regiões espectrais distintas, sua relação com a composição química da amostra é semelhante, uma vez que modos vibracionais envolvendo as ligações C-H entram na composição espectral destas regiões. No caso do AEHC incluímos também os resultados para um modelo PLS onde toda a faixa espectral, de 4200 a 9000 cm^{-1} , foi selecionada.

Os resultados apresentados na Tabela 2 mostram que os erros de previsão ($RMSEP$) são menores quando se usam regiões espectrais onde a absorvância é menor que a unidade, uma vez que a incerteza relativa de concentração cresce exponencialmente para absorvâncias maiores que a unidade (ver Figura 2).

Para se comparar a qualidade de previsão de dois modelos de calibração pode-se usar o teste F , desde que se conheça a variância de cada modelo⁴³. Considere-se a expressão^{32,33}:

$$RMSEP = \sqrt{\frac{1}{N} \sum_i (dif_i)^2} \quad (5)$$

onde N é o número de espectros das amostras do grupo de teste e dif_i é a diferença entre o valor obtido pelo modelo de calibração para a i -ésima amostra e o valor usado como referência (verdadeiro). Esta expressão, portanto, fornece o valor de desvio padrão. A variância é dada pelo quadrado do desvio padrão e o valor de F é dado pela razão entre as variâncias⁴³. Assim, considerando o quadrado dos valores de $RMSEP$ como uma medida da variância, a Tabela 2 também apresenta os valores de F obtidos comparando os modelos onde a absorvância é menor do que a unidade com os outros modelos. Os valores entre parênteses são os valores de F esperados para 95% de nível de confiança. Como os valores de F obtidos são maiores do que o esperado, conclui-se que a diferença entre as variâncias dos modelos é significativa, dentro de 95% de probabilidade, ou seja, o modelo em que se usa uma região espectral com absorvâncias menores que a unidade é mais exato do que os modelos que incluem regiões onde a absorvância é maior que a unidade, considerando o universo de amostras selecionadas para o grupo de teste.

Uma vez que a incerteza relativa de concentração é importante na escolha da região espectral, este parâmetro poderia ser usado como método de auxílio na escolha das regiões espectrais no uso de modelos de calibração baseados em análise multivariada. Para exemplificar a utilização deste método, a Figura 3 mostra os espectros de absorvância e da distribuição espectral da incerteza relativa de concentração (s_c/c) para gasolina e AEHC. Cabe destacar que, para a gasolina, a resolução espectral utilizada foi de 4 cm^{-1} , enquanto para AEHC foi de 8 cm^{-1} . Estes espectros mostram que, para as regiões espectrais onde a absorvância é maior do que a unidade, os valores de s_c/c aumentam bruscamente. Além disso, existem regiões, como aquela onde aparece a banda de absorção referente ao modo de combinação de estiramento e deformação OH (4900-5500 cm^{-1} , Figura 3b), onde se observam valores significativamente elevados de s_c/c . Note que os valores de s_c/c para a gasolina são, em geral, o dobro dos valores observados para AEHC. Isto se deve ao fato de as resoluções espectrais nos dois casos serem diferentes, ou seja, a resolução espectral para AEHC (8 cm^{-1}) é o dobro da usada para a gasolina (4 cm^{-1}). Para um mesmo número de varreduras, quanto menor a resolução espectral (em cm^{-1}) pior a relação sinal/ruído. Observe-se que nas curvas da Figura 2, onde a resolução espectral foi a mesma, 8 cm^{-1} , tanto para

Tabela 2. Parâmetros de validação/calibração (PLS) para determinação de MON em gasolina e teor alcoólico em álcool combustível (AEHC), considerando regiões com absorvância maior e menor do que a unidade e toda a faixa espectral

Combustível	Região espectral* (cm^{-1})	Validação grupo de teste		Calibração		Teste F
		R^2	$RMSEP$	R^2	$RMSEP$	
Gasolina ^a	4200-4550 com $A > 1$	70,3 (2)	0,678	81,9	0,658	1,49 (1,35) ^d
	5378-5992 com $A < 1$	80,2 (2)	0,555	80,9	0,578	
AEHC ^b	4200-4550 com $A > 1$	69,58 (8)	0,127	82,09	0,122	6,61 (1,22) ^e
	5378-5992 com $A < 1$	95,37 (5)	0,0494	96,90	0,0788	
AEHC ^c	4200-9000	92,7 (4)	0,060	93,20	0,07	1,48 (1,22) ^e

*O pré-processamento espectral para gasolina foi normalização vetorial seguido de 1ª derivada (alisamento com janela de 13 pontos); no caso do AEHC foi 1ª derivada (alisamento com janela de 13 pontos); ^a número de amostras para validação e calibração foram 39 e 37, respectivamente; ^{b,c} número de amostras para validação e calibração foram 53 e 59, respectivamente; ^d valor em parênteses refere-se a F para 111 graus de liberdade (37 amostras x 3 espectros) e 95% de nível de confiança; ^e valores em parênteses referem-se a F para 159 graus de liberdade (53 amostras x 3 espectros) e 95% de nível de confiança⁴³.

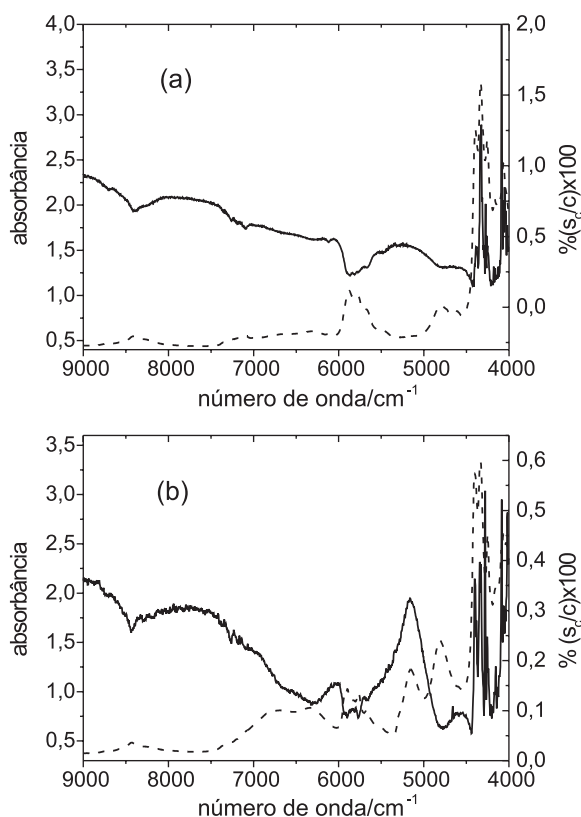


Figura 3. Espectros de absorção (---) e distribuição espectral de s_e/c (%) (—) para gasolina (a) e AEHC (b)

as amostras de gasolina como para as misturas etanol/água, os valores mínimos de s_e/c são muito próximos, 0,15 e 0,12%, respectivamente.

Os resultados discutidos acima, no que diz respeito à distribuição espectral de s_e/c , sugerem que as melhores regiões espectrais são aquelas onde os valores de s_e/c passam por mínimos não descontínuos. Para fundamentação desta conclusão é necessário nos reportarmos aos espectros da Figura 4.

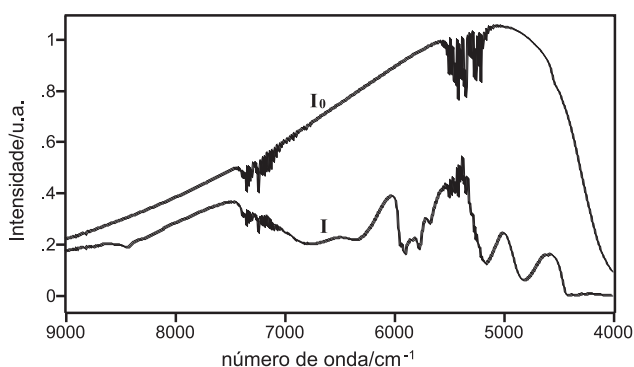


Figura 4. Espectros de I_0 e I para uma amostra de AEHC

Primeiramente, cabe destacar que o aumento nos valores de s_e/c para números de onda crescentes se deve ao fato de a resposta do detector (Ge no presente caso) diminuir para números de onda maiores que 5000 cm⁻¹ (ver curva para I_0 , Figura 4). Note também que para a região abaixo de 4500 cm⁻¹, a distribuição espectral de s_e/c é muito ruidosa, apresentando aumentos bruscos (descontínuos) em

seus valores. Isto se deve ao fato de que a quantidade de luz que chega ao detector é muito baixa, ou seja, o sinal que está sendo medido é muito baixo (ver espectro de I na Figura 4), comprometendo a relação sinal-ruído nesta região. Coincidentemente, é nesta região que as absorbâncias são maiores do que a unidade. A região entre 4900 e 5500 cm⁻¹ também apresenta um máximo no espectro de s_e/c . Note, na Figura 4, que nesta região os espectros de I_0 e I apresentam contribuições das absorções (rotações vibrações) de vapor de água. Portanto, diferenças de compensação para I_0 e I , nesta região, levam a uma maior incerteza nos valores de s_e/c .

Poder-se-ia argumentar que, ao invés de se usar o espectro de s_e/c para escolha da região espectral, se usasse diretamente o espectro do desvio padrão relativo de absorbância (s_a), como feito por Chung *et al.*⁴². Para efeito de comparação, a Figura 5 mostra a distribuição espectral de s_e/c e do desvio padrão relativo da absorbância (s_a). Os valores de s_a foram obtidos a partir da média de 20 medidas independentes de espectros de absorção de uma amostra de AEHC.

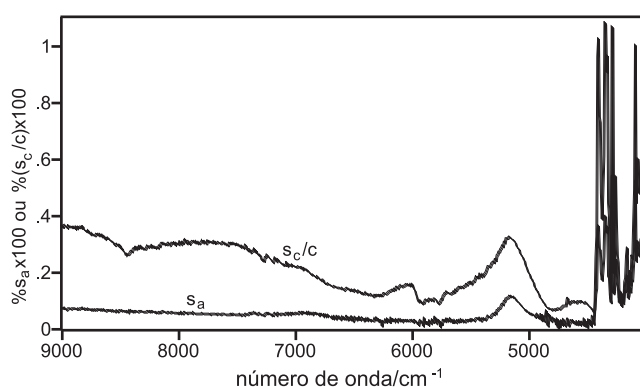


Figura 5. Distribuição espectral de s_e/c e do desvio padrão relativo de absorbância (s_a)

Observe-se que as duas curvas mostram valores elevados de s_e/c e s_a nas regiões onde as absorbâncias são maiores que a unidade. Ambas as curvas também apresentam um máximo para a região entre 5000 e 5500 cm⁻¹. No entanto, para as outras regiões (acima de 5500 cm⁻¹), a curva de s_a não fornece nenhuma indicação sobre qual região poderia ser selecionada, uma vez que a curva de s_a não apresenta praticamente qualquer variação significativa com o número de onda. Ou seja, o uso deste método levaria à escolha de toda a região espectral, de 5500 a 9000 cm⁻¹. Por outro lado, a distribuição espectral de s_e/c apresenta mínimos, entre 5500 e 6000 cm⁻¹ e 8200 e 8700 cm⁻¹, coincidentes com bandas de absorção do etanol. Portanto, a escolha destas regiões espectrais, para a construção de modelos de calibração multivariada, deve apresentar erros menores de previsão, uma vez que a incerteza relativa de concentração nestas regiões passa por um mínimo.

Usando-se o método proposto acima, para a escolha da região espectral, foram construídos modelos de calibração PLS-FTNIR para a determinação de densidade e MON em gasolina C, e teor alcoólico e densidade em AEHC. As Tabelas 3 e 4 apresentam os parâmetros PLS para os casos investigados. Comparando os resultados obtidos para MON (gasolina) e teor alcoólico (AEHC), apresentados nestas tabelas, com aqueles da Tabela 2, observa-se uma melhora significativa na qualidade dos modelos. Por exemplo, os valores de *RMSEP* caíram praticamente à metade com uma maior aproximação de *R*² do valor de 100%.

Alguns manuais de aplicativos de modelos PLS³³ e artigos⁴⁴ recomendam o uso do espectro de correlação (*R*² vs cm⁻¹) para a escolha da região espectral. A Figura 6 mostra o espectro de correlação

Tabela 3. Parâmetros de validação/calibração (PLS) para determinação de densidade e MON em gasolina, considerando a seleção da região espectral por diferentes métodos

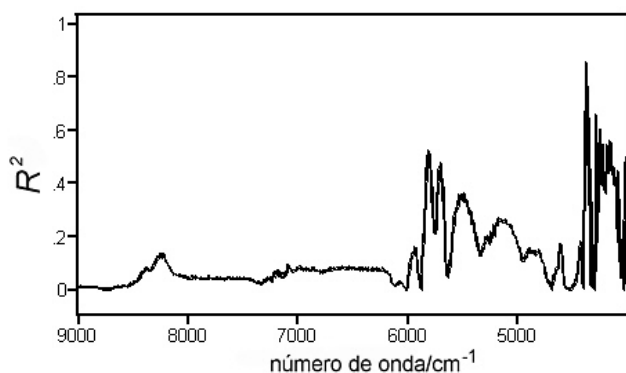
Propriedade do combustível	Método*	Região espectral (cm ⁻¹)	Validação		Teste <i>F</i>
			<i>R</i> ²	<i>RMSEP</i>	
Densidade	<i>R</i> ²	4210-4478 e 5550-5911	67,2 (4) ^a	0,005	3,70 (1,35) ^b
	<i>s/c</i>	4740-7483 e 8282-8502	91,0 (9)	0,0026	
MON	<i>R</i> ²	4170-5100 e 5300-5900	91,8 (4)	0,36	1,19 (1,35)
	<i>s/c</i>	4140-4200, 4550-7400 e 8100-8600	92,9 (7)	0,33	

* Para a comparação dos métodos, foi utilizado o mesmo tipo de pré-processamento dos espectros nos dois métodos. No caso de *R*² foram consideradas as regiões para as quais *R*² > 50%; ^a valores em parênteses referem-se ao número de componentes principais utilizados; ^b valores em parênteses referem-se a *F* para 111 graus de liberdade e 95% de nível de confiança⁴³.

Tabela 4. Parâmetros de validação/calibração (PLS) para determinação de teores de etanol e densidade em álcool combustível (AEHC), considerando a seleção da região espectral por diferentes métodos

Propriedade do combustível	Método	Região espectral (cm ⁻¹)	Validação grupo de teste*		Teste <i>F</i>
			<i>R</i> ²	<i>RMSEP</i>	
Teor alcoólico	<i>R</i> ²	4200-5360	91,39 (4) ^a	0,0673	2,87 (1,22) ^b
	<i>s/c</i>	4630-5005 5358-6017 6523-7171 7969-8826	96,97 (4)	0,0399	
Densidade	<i>R</i> ²	4200-5360	85,56 (4)	2,5x10 ⁻⁴	1,73 (1,22) ^b
	<i>s/c</i>	4630-5005 5358-6017 6523-7171 7969-8826	91,45 (3)	1,9x10 ⁻⁴	

* Para a comparação dos métodos, foi utilizado o mesmo tipo de pré-processamento dos espectros nos dois métodos; ^a valores em parênteses referem-se ao número de componentes principais utilizados; ^b valores em parênteses referem-se ao valor de *F* para 159 graus de liberdade e 95% de nível de confiança⁴³.

**Figura 6.** Espectro de correlação (*R*²) para o conjunto de 50 amostras de gasolina C, referente à densidade

obtido para o conjunto de 50 amostras de gasolina C, referentes à densidade. De acordo com este método, as regiões a serem escolhidas são aquelas onde os valores de *R*² se aproximam de 100%. Note que os maiores valores de *R*² são obtidos para regiões onde a absorvância é maior do que a unidade e que a região correspondente ao modo de combinação estiramento e deformação angular OH do etanol, importante para a determinação da densidade da gasolina, não aparece com valores significativos para *R*².

Para comparação com o método proposto neste trabalho, foram construídos modelos PLS-FTNIR, onde as regiões espectrais foram

escolhidas com base nas indicações dos espectros de correlação. Neste caso, foram escolhidas as regiões espectrais onde os valores de *R*² são maiores que 50%. Os parâmetros PLS para estes modelos também são apresentados nas Tabelas 3 e 4.

Uma comparação entre os dois métodos para escolha da região espectral mostra que o método baseado na distribuição espectral dos valores de *s/c* fornece valores de erros de previsão (*RMSEP*) sempre menores que aqueles obtidos pelo método de correlação. A ferramenta estatística escolhida para a comparação dos dois conjuntos de resultados é o teste *F*. Usando os mesmos argumentos utilizados para a comparação dos modelos da Tabela 2, os valores de *F* obtidos (ver Tabelas 3 e 4) mostram que a diferença entre as variâncias dos dois métodos é significativa, para 95% de nível de confiança. Já para os valores de MON, o valor de *F* obtido mostra que não existe diferença significativa entre os dois modelos para o nível de confiança estipulado (95%).

Para o caso do AEHC discutiremos o significado das diferenças dos dois métodos apenas para a densidade, uma vez que o teor alcoólico é obtido a partir da medida de densidade. Os dados da Tabela 4 mostram que o valor de *F* obtido é superior ao esperado, quando se compara o método baseado em *R*² ao método baseado em *s/c* para a escolha da região espectral. Portanto, a diferença dos dois métodos, quanto à capacidade de prever valores de densidade é significativa, para um nível de confiança de 95%. Portanto, o modelo de calibração onde a região espectral foi selecionada com base nos valores de *s/c* fornece valores recuperados de densidade mais exatos, considerando o universo de amostras separadas para a validação.

O teste F pode também ser usado para comparar a precisão dos dois métodos, com a precisão do método de referência. A Tabela 5 apresenta os valores de desvio padrão para determinações em triplicata, considerando os métodos baseados em s_c/c e R^2 . A Tabela 5 também fornece os valores de repetibilidade e reprodutibilidade das normas de referência. A repetibilidade e a reprodutibilidade servem como parâmetros para avaliação da precisão de uma medida³⁵. Para efeitos de comparação, usamos os valores de repetibilidade, uma vez que não temos como avaliar a reprodutibilidade dos métodos propostos. Assim, com base nos valores de repetibilidade para as normas ASTM usadas, foram obtidos os valores de desvio padrão para cada um dos métodos baseados em normas ASTM. A partir desses valores, foram obtidas as respectivas variâncias. O valor de F , em cada caso, foi calculado pela razão entre a variância do método proposto (s_c/c ou R^2) e a variância da norma ASTM. Para um nível de confiança de 95%, os métodos para os quais os valores de F (obtidos) são menores que o valor esperado (ver Tabela 5) indicam não existir diferença significativa quanto à precisão, em relação à norma de referência. A exceção é a determinação da densidade da gasolina pelo método R^2 , onde o valor de F obtido foi bem superior ao esperado, sugerindo a existência de uma diferença significativa entre a precisão deste método e aquela da norma de referência. A explicação

para esta diferença reside no fato de os valores de s_c/c para as regiões selecionadas serem, em média, bem superiores àqueles das regiões selecionadas para determinação da densidade em AEHC. Como discutido acima, esta diferença nos valores de s_c/c deve-se ao fato de que os espectros NIR das gasolinas foram obtidos com uma resolução espectral que é a metade da usada na obtenção dos espectros NIR das amostras de AEHC, ou seja, o ruído associado aos espectros de AEHC é a metade do ruído associado aos espectros das gasolinas.

Os resultados obtidos para determinação de MON em gasolina, usando os métodos baseados em s_c/c e R^2 , fornecem erros relativos de previsão da ordem de 0,39 e 0,42%, respectivamente. Estes valores estão próximos do encontrado na literatura, 0,7%, para a determinação de MON por NIR, usando um instrumento dispersivo com detector multicanal⁴⁵. Os erros relativos de previsão para densidade da gasolina são de 0,35 e 0,7% para s_c/c e R^2 , respectivamente. Neste caso, não encontramos valores na literatura para comparação. Nas determinações de densidade e teor alcoólico de AEHC estes valores são de 0,02 e 0,03% e 0,04 e 0,07%, respectivamente. Os valores para determinação de teor alcoólico (0,04 e 0,07% para s_c/c e R^2 , respectivamente) são comparáveis aos obtidos por Guchardi e Pasquini⁴⁶, 0,07%, para determinação de teor alcoólico em AEHC.

Tabela 5. Valores de desvio padrão para medidas em triplicata usando os métodos das Tabelas 3 e 4. Para comparação são apresentados também os valores de repetibilidade e reprodutibilidade das normas de referência e os valores para o teste F

Combustível	Gasolina				AEHC			
	MON		Densidade		Teor alcoólico		Densidade	
	R^2	s_c/c	R^2	s_c/c	R^2	s_c/c	R^2	s_c/c
Desvio padrão ^a	0,1	0,09	3×10^{-3}	2×10^{-4}	0,01	2×10^{-3}	1×10^{-4}	2×10^{-5}
Repetibilidade [*]	$\pm 0,2$		$\pm 1 \times 10^{-4}$		$\pm 0,01$		$\pm 1 \times 10^{-4}$	
Reprodutibilidade [*]	$\pm 0,9$		$\pm 6 \times 10^{-4}$		$\pm 0,06$		$\pm 6 \times 10^{-4}$	
Desvio padrão ^b	0,1		2×10^{-4}		0,02		2×10^{-4}	
Teste F (2,60) ^c	1	0,81	225	1	0,25	0,01	0,25	0,01

* os valores de repetibilidade e reprodutibilidade referem-se às normas ASTM D-2700 e ASTM D-4052, para determinação de MON e densidade; ^a desvio padrão obtido para média de três determinações; ^b desvio padrão obtido considerando $z = 1,96$ e um número infinito de medidas, ou seja, entende-se que a repetibilidade é o limite fiducial (ou intervalo de confiança) para um nível de confiança de 95%; ^c o valor em parênteses corresponde ao valor de F para variâncias obtidas com base em três e infinitas medidas, ou seja, para 3 e infinitos graus de liberdade, considerando nível de confiança de 95%⁴³.

CONCLUSÕES

Os resultados apresentados neste trabalho mostram a importância da escolha da região espectral na construção de modelos de calibração por PLS e levam a uma metodologia simples para esta escolha, a qual se baseia na análise da distribuição espectral da incerteza relativa de concentração. A metodologia proposta permite excluir da análise por PLS as regiões espectrais onde as absorvâncias não sofrem alteração com a variação da concentração, de regiões onde as absorvâncias são maiores do que a unidade e regiões onde os aumentos da incerteza relativa de concentração se devem a diferenças nas compensações das absorções do vapor de água e se aplica a qualquer técnica espectroscópica baseada em medidas de absorção.

A comparação do método proposto com o método do espectro de correlação para a escolha da região espectral mostrou que o método proposto, em três dos casos investigados, levou à obtenção de modelos de calibração PLS com maior confiabilidade (menores erros de previsão) e maior precisão.

Finalmente, cabe destacar que os resultados aqui apresentados

são de fundamental importância para a questão da análise de combustíveis realizada por laboratórios onde são usados equipamentos portáteis, os quais não permitem a variação do caminho óptico, nem tampouco a modificação das regiões espectrais usadas na calibração. Como mostram os resultados da Tabela 3, a região espectral de 4740 a 7483 cm^{-1} (onde aparecem bandas de absorção características do etanol, ver Tabela 1) deve ser selecionada para a construção de modelos PLS para a determinação de densidade da gasolina. Com o aumento do teor de etanol na gasolina, alguns equipamentos portáteis, que não permitem a variação do caminho óptico, podem se tornar imprecisos e inexatos para a determinação de teores de etanol em gasolina tipo C, por exemplo, uma vez que este parâmetro é obtido a partir da medida da densidade.

AGRADECIMENTOS

Os autores agradecem à FINEP/CTPetro pelo apoio financeiro, ao CNPq pelas bolsas e à Copersucar e ao CEPAT (Centro de Pesquisas e Análises Tecnológicas) pelos padrões certificados de AEHC e gasolinas A e C.

REFERÊNCIAS

1. de Groot, P. J.; Postma, G.J.; Melssen, W. J.; Buydens, L. M. C.; *Anal. Chim. Acta* **2002**, *453*, 117.
2. Furukawa, T.; Kita, Y.; Sasao, S.; Matsukawa, K.; Watari, M.; Sasic, S.; Siesler, H. W.; Ozaki, Y.; *J. Near Infrared Spectrosc.* **2002**, *10*, 195.
3. Scafi, S. H. F.; Pasquini, C.; *Analyst* **2001**, *126*, 2218.
4. Vaidyanathan, S.; Arnold, S. A.; Matheson, L.; Mohan, P.; McNeil, B.; Harvey, L. M.; *Biotechnol. Bioeng.* **2001**, *74*, 376.
5. Barges, N.; Conan, H.; Poisson, N.; *Analisis* **1998**, *26*, M43.
6. Sena, M. M. de; Poppi, R. J.; Frighetto, R. T. S.; Valarini, P. J.; *Quim. Nova* **2000**, *23*, 547.
7. Sena, M. M.; Frighetto, R. T. S.; Valarini, P. J.; Tokeshi, H.; Poppi, R. J.; *Soil Till. Res.* **2002**, *67*, 171.
8. Togersen, G.; Isaksson, T.; Nilsen, B.N.; Bakker, E. A.; Hildrum, K. I.; *Meat Sci.* **1999**, *51*, 97.
9. Peirs, A.; Scheerlinck, N.; Touchant, K.; Nicolai, B. M.; *Biosyst. Eng.* **2002**, *81*, 305.
10. Cleve, E.; Bach, E.; Schollmeyer, E.; *Anal. Chim. Acta* **2000**, *420*, 163.
11. Buttner, G.; *Process Control Qual.* **1997**, *9*, 197.
12. Kim, M.; Lee, Y. H.; *Comput. Chem. Eng.* **2000**, *24*, 513.
13. Geladi, P.; Kolwalski, B. R.; *Anal. Chim. Acta* **1986**, *185*, 1.
14. Sjoström, M.; Wold, S.; Lindberg, W.; Persson, J. A.; Martens, H.; *Anal. Chim. Acta* **1983**, *150*, 61.
15. Stone, M.; Brooks, R. J.; *J. Roy. Stat. Soc. B Met.* **1990**, *52*, 237.
16. de Jong, S.; Kiers, H. A. L.; *Chemom. Intell. Lab. Syst.* **1992**, *14*, 155.
17. de Jong, S.; *Chemom. Intell. Lab. Syst.* **1993**, *18*, 251.
18. Wentzell, P. D.; Andrews, D. T.; Kowalski, B. R.; *Anal. Chem.* **1997**, *69*, 2299.
19. Kalivas, J. H.; *J. Chemom.* **1999**, *13*, 111.
20. Lober, A.; Kowalski, B. R.; *J. Chemom.* **1988**, *2*, 67.
21. Jiang, J.; Berry, R. J.; Siesler, H. W.; Ozaki, Y.; *Anal. Chem.* **2002**, *74*, 3555.
22. Xu, L.; Schechter, I.; *Anal. Chem.* **1996**, *68*, 2392.
23. Spiegelman, C. H.; McShane, M. J.; Goetz, M. J.; Motamedi, M.; Yue, Q. L.; Cote, G. L.; *Anal. Chem.* **1998**, *70*, 35.
24. Kalivas, J. H.; Roberts, N.; Sutter, J. M.; *Anal. Chem.* **1989**, *61*, 2024.
25. Rimbaud, D. J.; Walczak, B.; Massart, D. L.; Last, I. R.; Prebble, K. A.; *Anal. Chim. Acta* **1995**, *304*, 185.
26. Martens, H.; Næs, T.; *Multivariate Calibration*, Wiley: New York, 1989.
27. Brown, P. J.; *J. Chemom.* **1992**, *6*, 151.
28. Liang, Y. -Z.; Xie, Y. -L.; Yu, R. -Q.; *Anal. Chim. Acta* **1989**, *222*, 347.
29. Horchner, U.; Kalivas, J. H.; *Anal. Chim. Acta* **1995**, *311*, 1.
30. Lucasius, C. B.; Kateman, G.; *Trends Anal. Chem.* **1991**, *10*, 254.
31. Costa, P. A. D.; Poppi, R. J.; *Quim. Nova* **2002**, *25*, 46.
32. *Spectroscopic Software – OPUS NT Quant 2*, Bruker Optik GmbH, 2000, vol. 3.
33. *PLSplus IQ™ User's Guide*, Thermo Galactic, 2002.
34. Rothman, L. D.; Crouch, S. R.; *Anal. Chem.* **1975**, *47*, 1226.
35. Skoog, D. A.; Leary, J. J.; *Principles of Instrumental Analysis*, Saunders College Publishing, 4.^a ed., 1992.
36. McGlone, V. A.; Jordan, R. B.; Seelye, R.; Martinsen, P. J.; *Postharvest Biol. Tec.* **2002**, *26*, 191.
37. Rambla, F. J.; Garrigues, S.; de la Guardia, M.; *Anal. Chim. Acta* **1997**, *344*, 41.
38. Buerck, J.; Roth, S.; Kraemer, K.; Scholz, M.; Klaas, N.; *Hazard J. Mater.* **2001**, *83*, 11.
39. Kim, J. Y.; Woo, Y. A.; Kim, H. J.; Kim, J. D.; *J. Pharm. Biomed. Anal.* **2001**, *26*, 73.
40. Brennan, D.; Alderman, J.; Sattler, L.; O'Connor, B.; O'Mathuna, C.; *Measurement* **2003**, *33*, 67.
41. Chung, H.; Lee, J.-S.; Ku, M.-S.; *Appl. Spectrosc.* **1998**, *52*, 885.
42. Chung, H.; Ku, M.-S.; Lee, J.-S.; *Vib. Spectrosc.* **1999**, *20*, 155.
43. Freund, J. E.; Simon, G. A.; *Modern Elementary Statistics*, Prentice-Hall, Inc., 1997.
44. Qin, D.; Kean, R. T.; *Appl. Spectrosc.* **1998**, *52*, 488.
45. Wang, Z. M.; Hua, W. Y.; Wei, Z. K.; Zhang, H. H.; Wu, H. Z.; *Spectrosc. Spect. Anal.* **1999**, *19*, 684.
46. Guchardi, R.; Pasquini, C.; *Appl. Spectrosc.* **2001**, *55*, 454.