

CALIBRAÇÃO MULTIVARIADA PARA SISTEMAS COM BANDAS SOBREPOSTAS ATRAVÉS DA ANÁLISE DE FATORES DO TIPO Q

Ieda Spacino Scarminio, Dilson Norio Ishikawa, Wagner José Barreto, Édio Luiz Paczkowski e Iara Cintra Arruda
Departamento de Química - Universidade Estadual de Londrina (UEL) - CP 6001 - 86051-970 - Londrina - PR

Recebido em 15/8/97; aceito em 22/4/98

Q MODE FACTOR MULTIVARIATE CALIBRATION FOR SYSTEMS WITH OVERLAPPED BANDS. A multivariate calibration method to determine chemical compositions of systems with severely overlapped bands is proposed. Q mode factors are determined from the spectral data and subsequently rotated using the varimax and oblique transformation of Imbrie. The method is applied to two sets of simulated data to test the sensitivity of analytical results to random experimental error. The chemical concentrations of alanine and threonine mixture are determined from spectral data of the 302,5 - 548,5 nm region.

Keywords: factor analysis; aminoacids; overlapped bands.

INTRODUÇÃO

Um dos mais importantes problemas na química analítica é a estimativa do número e concentração das espécies em misturas através de espectros ou cromatogramas. Uma série de técnicas estatísticas tem sido utilizadas para se desenvolver metodologias multivariadas e para extrair informação dos espectros e cromatogramas, com o objetivo de identificar as espécies presentes e determinar quantitativamente as concentrações de algumas ou todas elas. A aplicabilidade de cada metodologia depende do conjunto de dados (informação experimental) submetidos a análise. Do ponto de vista da química analítica¹ estes métodos multivariados de uma forma geral podem ser classificados em três categorias: I) *quando se tem informação espectral de todas as espécies existentes na amostra*; II) *quando a informação dos espectros das espécies químicas são incompletos ou parciais*; III) *quando não existe nenhuma informação sobre a composição química das amostras*. A primeira categoria pode ser dividida em quatro subcategorias, dependendo da escolha do planejamento experimental: a) métodos de calibração direta: regressão linear múltipla^{2,3} (MLR), "Kalman Filtering"^{4,5} (KF); b) métodos de calibração indireta: método da matriz k ⁶⁻⁸, método da matriz- p ^{7,8}, regressão por mínimos quadrados parciais^{2,9-12} (PLS), regressão nas componentes principais¹²⁻¹⁴ (PCR); c) métodos de adição de padrão generalizada¹⁵⁻¹⁷ (GSAM); "Iterative target testing factor analysis technique"^{18,19} (ITTFA); d) métodos de referência interna generalizada²⁰ (GIRM); Na segunda categoria o método mais utilizado é o "generalized rank annihilation factor analysis"²¹ (GRAFA), enquanto que na terceira categoria incluem-se: o método para análise de espectros correlacionados²² (DATAN), "self-modeling curve resolution"²³ (SMCR), "envolving factor analysis"²⁴ (EFA), "iterative self-modeling multivariate analysis"²⁵⁻²⁸ (SIMPLISMA) e "self-modeling multivariate analysis"²⁵⁻²⁸ (ISMA).

Destes métodos a regressão nas componentes principais e métodos similares têm sido usados com maior frequência. A análise de componentes principais é uma metodologia baseada na decomposição de um conjunto de dados em autovalores e autovetores. Na química tem sido quase que exclusivamente usada a análise de componentes principais do tipo R, embora exista uma metodologia matematicamente equivalente, que é a do tipo Q.

Os métodos GSAM, GIRM, MLR, matriz k , ISMA e SIMPLISMA, não utilizam a análise de componentes principais no cálculo da calibração, enquanto os demais são métodos baseados na análise de componentes principais do tipo R. Na literatura

química são praticamente inexistentes aplicações de métodos baseados na análise de componentes principais do tipo Q, mais conhecida como análise de fatores do tipo Q, embora para os químicos esta metodologia possa oferecer várias vantagens em relação à análise do tipo R. Além de apresentar facilidades computacionais como cálculo mais rápido e matrizes menores para o tratamento de diferentes conjuntos de dados, a análise de fatores do tipo Q fornece resultados das análises químicas de uma forma especialmente conveniente para interpretação.

Neste artigo propomos um método de calibração multivariada para determinar a composição química de misturas complexas, baseado na análise de fatores do tipo Q de Imbrie³⁰, seguido pelas rotações varimax^{30,31} e oblíqua de Imbrie³⁰. Este método pode ser enquadrado tanto na categoria I como na categoria III, dependendo da informação disponível sobre o sistema químico a ser analisado. Para aplicação na categoria I, este método não requer um conjunto de calibração extenso como o PLS, PCR, matriz k , e MLR; bastam apenas os espectros das espécies absorventes puras ou espectros de referência. Na categoria III, o método é capaz de identificar o número de espécies e suas concentrações relativas a partir de um conjunto de espectros. Os perfis dos espectros das espécies puras também podem ser obtidos combinando-se as concentrações relativas calculadas com o método da matriz k .

Um requisito importante para aplicação de todos os métodos citados acima é que os espectros das misturas devem ser uma combinação linear dos espectros das espécies, ponderados por suas concentrações. Em outras palavras, que a Lei de Beer ou alguma expressão semelhante seja obedecida.

Para demonstrar a aplicabilidade do método nas categorias I e III foram simulados dois conjuntos de dados experimentais de um sistema de misturas ternárias, com a restrição que as concentrações das espécies somam para um valor constante. Estes exemplos mostram como a metodologia pode ser usada com ou sem dados espectrais de um conjunto de amostras de calibração. A sua aplicação a um caso real é ilustrada com uma análise espectrofotométrica simultânea de misturas binárias dos aminoácidos alanina e treonina. Os aminoácidos, em geral, são dificilmente analisados utilizando-se métodos univariados, porque seus espectros na região UV-VIS são severamente sobrepostos.

METODOLOGIA

Mistura é um termo usado para definir uma formulação ou composição³². Geralmente admite-se que a resposta medida

para uma mistura é uma função das proporções das espécies e não uma função da quantidade total da mistura. A soma das composições percentuais de todas as espécies presentes numa mistura é unitária³¹. Matematicamente isto pode ser expresso através da equação

$$\sum_{i=1}^q c_i = 1, \quad 0 \leq c_i \leq 1, \quad (1)$$

onde c_i é a fração da i -ésima espécie na mistura de q espécies. Estas frações podem ser expressas como a fração em mol, fração da massa ou de volume, ou mesmo percentagem.

Muitos problemas de interesse químico podem produzir um conjunto de dados consistindo essencialmente em proporções. Este tipo de dados ocorre quando os dados originais são normalizados de alguma maneira que torne a soma sempre uma constante. Se os valores das respostas das q espécies puras são conhecidas é possível projetar o problema em um simplex q -dimensional³³. Como a soma de todas as frações é constante, somente as concentrações de $q-1$ espécies podem ser independentes, a outra será uma variável dependente.

Análise de fatores é um termo genérico usado para descrever uma classe de métodos apropriados para analisar a inter-relação dentro de um conjunto de variáveis (altura de picos cromatográficos, valores de absorvâncias, etc...) ou objetos³⁴ (amostras analíticas, compostos químicos, espectros, etc...). O objetivo da análise é separar a matriz \mathbf{A} ($n \times p$), formada por n objetos (amostras ou misturas) e p variáveis (absorvâncias em diferentes comprimentos de onda, tempo de retenção, etc...) no produto de duas matrizes:

$$\mathbf{A}_{(n \times p)} = \mathbf{P}_{(n \times q)} \mathbf{T}_{(q \times p)}^t + \mathbf{E}_{(n \times p)}, \quad (2)$$

onde \mathbf{T} é a matriz dos escores dos fatores, \mathbf{P} é a matriz dos pesos dos fatores, \mathbf{E} é a matriz dos resíduos, q é um escalar e indica o número de fatores que descreve a maior parte da variância dos dados e o sobrescrito t representa a transposta da matriz. Cada fator corresponde a uma linha ou coluna das matrizes \mathbf{T} e \mathbf{P} .

Existem duas definições diferentes para a análise de fatores, dependendo da ordem da multiplicação das matrizes \mathbf{A} e \mathbf{A}^t : a análise do tipo Q ($\mathbf{A}\mathbf{A}^t$) e a análise do tipo R ($\mathbf{A}^t\mathbf{A}$). Na realidade a direção da multiplicação não altera a solução obtida. As diferentes soluções dependem da forma como a normalização dos dados é realizada, ou seja, normalização por linhas, por colunas ou ambos. Normalmente as matrizes usadas no tipo R são normalizadas para obter as matrizes de correlação ou covariância, enquanto que na análise de tipo Q existem três métodos, dependendo de como a matriz de similaridade ou associação é definida³⁰: o método de tipo Q de Imbrie, o método de Gowers, e o método de Benzécri^{30,35}. Dependendo do conjunto de dados estudado, a multiplicação $\mathbf{A}\mathbf{A}^t$ do tipo Q pode diminuir as exigências computacionais relativas àquelas necessárias quando é usado o produto $\mathbf{A}^t\mathbf{A}$ (tipo R).

A análise de fatores do tipo Q de Imbrie³⁰ define a similaridade dos objetos considerando as proporções das espécies. Este método procura na matriz de dados \mathbf{A} , os objetos mais divergentes, representados pelas espécies puras, ou constituídos por uma proporção significativa destas espécies, os quais representam os vértices do simplex. Os demais membros devem ser combinações lineares destes objetos. O número mínimo de vértices é determinado pela aproximação da matriz de dados por uma matriz de ordem menor. Isto definirá um número de vetores linha linearmente independentes, bem como a dimensionalidade do sistema. A contribuição de cada objeto é obtida pelo uso da auto-análise. Esta análise é feita sobre uma matriz real e simétrica obtida da matriz de dados.

Imbrie e Purdy definiram um índice de similaridade proporcional referido como co-seno de θ . Para dois objetos, n e m , o $\cos\theta$ é determinado por

$$\cos\theta_{nm} = \frac{\sum_{j=1}^q a_{nj} a_{mj}}{\sqrt{\sum_{j=1}^q a_{nj}^2 \sum_{j=1}^q a_{mj}^2}} \quad (3)$$

Para dados positivos este índice varia de zero, isto é nenhuma similaridade, para um, identidade.

O procedimento matemático do método começa com o cálculo dos $\cos\theta$ para todos os objetos, a partir da matriz \mathbf{A} . O primeiro passo é normalizar a matriz \mathbf{A} por linha, pré-multiplicando-a por uma matriz diagonal \mathbf{D}^{-1} ($n \times n$), em que \mathbf{D}^{-1} representa a inversa da matriz \mathbf{D} . A diagonal principal da matriz \mathbf{D} é formada pelas raízes quadradas dos comprimentos dos vetores linhas de \mathbf{A} ,

$$\mathbf{W}_{(n \times p)} = \mathbf{D}^{-1}\mathbf{A}. \quad (4)$$

A normalização por linha não afeta a relação de proporcionalidade entre as variáveis e remove o efeito da diferença das proporções entre os objetos. A matriz de similaridade ou associação é definida por

$$\mathbf{H} = \mathbf{W}\mathbf{W}^t = \mathbf{D}^{-1}\mathbf{A}\mathbf{A}^t\mathbf{D}^{-1}, \quad (5)$$

e pode ser expressa aproximadamente como o produto de uma matriz dos pesos $\mathbf{P}_{(n \times q)}$, e uma matriz dos escores, $\mathbf{T}_{(p \times q)}^t$, sendo q é o posto aproximado da matriz \mathbf{W} . A matriz \mathbf{H} é quadrada e simétrica e pode ser fatorada como $\mathbf{H} = \mathbf{U}\mathbf{U}^t$, onde \mathbf{U} é a matriz dos autovetores e \mathbf{A} a matriz diagonal dos autovalores associados. Assim, $\mathbf{H} = \mathbf{U}\mathbf{U}^t$, $\mathbf{P} = \mathbf{U}\mathbf{A}^{1/2}$ e $\mathbf{T} = \mathbf{W}^t\mathbf{P}\mathbf{A}^t$. Devemos enfatizar que as matrizes \mathbf{P} e \mathbf{T} obtidas assim não são as mesmas determinadas na análise do tipo R.

A matriz \mathbf{P} determinada pelo método de Imbrie não fornece um conjunto de objetos composicionalmente distintos. Uma maneira de resolver o problema é através das rotações varimax e oblíqua.

Rotação Varimax^{30,31} - A rotação varimax gira rigidamente os vetores da matriz \mathbf{P} até que eles coincidam com os vetores mais divergentes no espaço. Isto é feito maximizando-se a variância dos pesos dos fatores em cada vetor, sujeitos à restrição que os fatores retenham sua ortogonalidade, ou seja, que os fatores permaneçam não correlacionados entre si. A ordem da matriz \mathbf{P} para a rotação é igual ao número de vértices, o qual é determinado pelo número de autovalores significativos. A matriz \mathbf{P} é então rodada para produzir uma nova matriz, \mathbf{F} . A relação entre as duas matrizes é dada por

$$\mathbf{F} = \mathbf{P}\mathbf{R} \quad (6)$$

onde $\mathbf{R}_{(q \times q)}$ é a matriz de transformação e $\mathbf{F}_{(n \times q)}$ é a matriz dos pesos varimax. Cada linha da matriz \mathbf{F} corresponde a um objeto ou uma combinação linear deles e cada coluna representa um fator.

Projeção Oblíqua³⁰ - Imbrie descreveu um procedimento para projeção oblíqua que gira os fatores varimax ortogonais até que eles também coincidam com os vetores mais divergentes. Desta forma todos os demais membros do conjunto são definidos como proporções destes objetos. A diferença dessa rotação é que o ângulo entre os fatores é oblíquo e portanto eles são correlacionados entre si. Isto é feito construindo-se uma matriz $\mathbf{V}_{(q \times q)}$, que contém os mais altos valores absolutos dos pesos varimax em cada coluna dos objetos. A matriz da projeção oblíqua é dada por

$$\mathbf{C} = \mathbf{F}\mathbf{V}^{-1} \quad (7)$$

onde \mathbf{V}^{-1} é a inversa de \mathbf{V} . Os vetores linha da matriz \mathbf{C} fornecem as contribuições proporcionais de todos os objetos em termos dos objetos de referência. Para recalculer os valores em

termos de dados originais é necessário dividir cada vetor coluna da matriz **C** pelo comprimento do vetor do objeto correspondente, o que desnormaliza os vetores coluna da matriz **C**.

Método da matriz k- calibração⁶⁻⁸: Este método baseia-se na lei de Beer-Lambert,

$$A_j = \sum_{i=1}^q a_{ij} b_{ci} \quad (8)$$

onde A_j é a absorvância de uma amostra multicomponente (ou referência) em um comprimento de onda j , a_{ij} é a absorvância do componente químico i no comprimento de onda j , b é o comprimento do caminho óptico e c_i é a concentração do i -ésima espécie, com q espécies possíveis em cada mistura. Definindo K_{ij} , como o produto $a_{ij}b$ e adicionando o erro aleatório e_j no espectro em cada comprimento de onda j , temos

$$A_j = \sum_{i=1}^q k_{ij} c_i + e_j \quad (9)$$

Se um conjunto de misturas com proporções conhecidas for usado como referência na fase de calibração das análises, uma série de equações simultâneas podem ser usadas para descrever os espectros das misturas. Em notação matricial estas equações são dadas por

$$\mathbf{A} = \mathbf{K}\mathbf{C} + \mathbf{E} \quad (10)$$

A matriz **A** neste caso é $p \times n$, $\mathbf{C}_{(q \times n)}$ é a matriz das concentrações das q espécies nas amostras ($n \times q$), **K** é uma matriz $p \times q$ cujos elementos K_{ij} são estimados pelo método dos mínimos quadrados e $\mathbf{E}_{(p \times n)}$ contém os erros experimentais para cada espectro.

Se a lei de Beer for válida as colunas da matriz **K** correspondem aos espectros das espécies puras, na concentração unitária e comprimento do caminho óptico unitário, nos comprimentos de onda usados na análise. Se a matriz $\mathbf{C}\mathbf{C}^t$ não for singular, a solução dos mínimos quadrados ponderados para a equação acima torna-se:

$$\mathbf{K}^* = \mathbf{A}\mathbf{C}^t(\mathbf{C}\mathbf{C}^t)^{-1} \quad (11)$$

onde \mathbf{K}^* corresponde aos espectros estimados. Esta matriz pode ser usada para determinar as concentrações das amostras de composições não conhecidas, ou seja, amostras de previsão:

$$\mathbf{A}^s = \mathbf{K}\mathbf{C} + \mathbf{E} \quad (12)$$

onde a matriz \mathbf{A}^s contém as absorvâncias das amostras de composição não conhecida e, a matriz \mathbf{K}^* contém os espectros das espécies puras. As concentrações estimadas para as amostras de previsão pelos mínimos quadrados não ponderados torna-se

$$\mathbf{C}^* = (\mathbf{K}^*\mathbf{K}^*)^{-1}\mathbf{K}^*\mathbf{A}^s \quad (13)$$

admitindo-se que a matriz inversa exista.

EXPERIMENTAL

Materiais e reagentes: Uma solução de p-benzoquinona (PBQ) (Riedel) $0,1 \text{ mol dm}^{-3}$ foi preparada em dimetilsulfóxido (DMSO) (Merck). A PBQ utilizada foi purificada através de sublimação. Foram preparadas soluções padrões $1,5 \text{ g dm}^{-3}$ de alanina e treonina (ambas da Sigma) em tampão fosfato $0,1 \text{ mol dm}^{-3}$, pH 6,0).

Preparação dos padrões puros: Foram preparadas duas soluções padrões com os aminoácidos alanina e treonina. A um tubo teste contendo um volume de solução tampão fosfato $0,1 \text{ mol dm}^{-3}$ pH 6,0 adicionou-se uma alíquota de $200,0 \times 10^{-6} \text{ dm}^3$ do padrão aminoácido. Acrescentou-se $50,00 \times 10^{-6} \text{ dm}^3$ de

PBQ $0,1 \text{ mol dm}^{-3}$ e completou-se o volume final com solução tampão a $5,0 \times 10^{-3} \text{ dm}^3$.

Preparação das misturas: A três tubos contendo um volume de solução tampão fosfato $0,1 \text{ mol dm}^{-3}$, pH 6,0, foram transferidas alíquotas de soluções padrões de alanina e treonina de 100,0 e 100,0; 150,0 e 50,0; 50,0 e 150,0 10^{-6} dm^3 , respectivamente. A cada um desses tubos acrescentaram-se $50,00 \times 10^{-6} \text{ dm}^3$ de PBQ $0,1 \text{ mol dm}^{-3}$ e completou-se o volume final a $5,0 \times 10^{-3} \text{ dm}^3$ com solução tampão. O branco foi preparado de forma semelhante aos padrões diluindo-se $50,00 \times 10^{-6} \text{ dm}^3$ de PBQ $0,1 \text{ mol dm}^{-3}$ em solução tampão para um volume final de $5,0 \times 10^{-3} \text{ dm}^3$. As proporções em massa resultantes foram 50% e 50%, 75% e 25% e 25% e 75% de alanina e treonina, respectivamente.

Procedimento experimental: Após a adição da solução de PBQ os tubos foram agitados durante 1 minuto com agitador ultra-sônico, vedados e submetidos a aquecimento à 100°C em banho-maria por 18 minutos. Os tubos foram resfriados a temperatura ambiente e os espectros dos padrões puros e das misturas foram registrados contra o branco, no intervalo de 548,5 nm a 302,5 nm. Foi utilizado um espectrofotômetro Milton Roy, modelo Genesys 2, interfaceado com um microcomputador.

Programas computacionais: Todos os programas foram desenvolvidos no nosso laboratório em linguagem FORTRAN, para microcomputador tipo PC.

RESULTADOS E DISCUSSÃO

Os dados espectrais simulados foram preparados para demonstrar a aplicabilidade do método. Estes espectros simulam um sistema de misturas ternárias. Os espectros das espécies puras foram compostos somando-se duas bandas Gaussianas de diferentes intensidades e larguras, e representados por cem pontos ($p=100$).

Primeira aplicação: No primeiro conjunto de dados experimentais simulados, os espectros das espécies puras são conhecidas. Dez espectros de misturas foram obtidos como combinações lineares de diferentes proporções dos espectros puros:

$$\mathbf{A} = \mathbf{C}\mathbf{V} + \mathbf{E}$$

onde **C** é a matriz das proporções ou concentrações, (tabela 1), **V** é a matriz dos espectros puros e **E** representa a adição de um ruído artificial de 10% da intensidade do sinal médio. A figura 1 mostra a sobreposição espectral dos espectros puros bem como das misturas. O objetivo do estudo é determinar a composição química das misturas a partir dos espectros das espécies puras.

Tabela 1. Composição original das misturas^a.

Amostra n ^o	Espécie 1	Espécie 2	Espécie 3
1	1,0000	0	0
2	0	1,0000	0
3	0	0	1,0000
4	0	0,2500	0,7500
5	0,0769	0,3077	0,6154
6	0,1428	0,3572	0,5000
7	0,2000	0,4000	0,4000
8	0,2500	0,4375	0,3125
9	0,3125	0,4375	0,2500
10	0,4000	0,4000	0,2000
11	0,5000	0,3572	0,1428
12	0,6154	0,3077	0,0769
13	0,7500	0,2500	0

a) As amostras 1-3 correspondem as espécies puras; 4-12 as misturas ternárias e a amostra 13 uma mistura binária com as espécies 1 e 2.

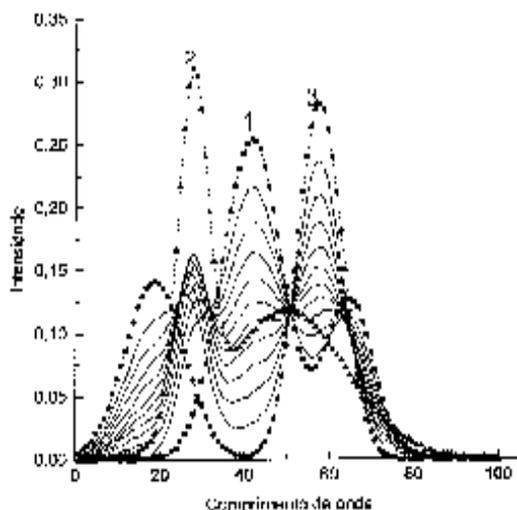


Figura 1. Treze espectros simulados com 10% de ruído. Os símbolos correspondem aos espectros puros com adição de ruído, 1■, 2▲ e 3●. As intensidades foram simuladas usando bandas gaussianas e a escala de comprimento de onda é relativa.

A dificuldade associada a este tipo de análise está relacionada com o número de fatores que devem ser girados. Uma aproximação bastante utilizada supõe que as espécies químicas têm uma grande contribuição para a variância dos dados, e o número de espécies no sistema químico é igual ao número de autovalores significativos mais um.

A análise de fatores do tipo Q de Imbrie mostra que dois fatores explicam em torno de 97% da variância dos dados, cujos autovalores são $\lambda_1 = 11,315$ e $\lambda_2 = 1,341$. Este resultado é consistente com o fato de que este sistema tem dois graus de liberdade, isto é, tem três espécies químicas cuja soma das proporções é sempre constante. O terceiro fator explica menos que 3% da variância total nos espectros ($\lambda_3 = 0,342$) e pode ser atribuído ao ruído introduzido nos espectros. Para obter os resultados para as três espécies químicas utilizou-se rotações destes três fatores.

A tabela 2 mostra a matriz dos pesos dos fatores. Estes resultados não são facilmente interpretáveis porque: 1) o primeiro fator constitui essencialmente da média de todas as misturas e as espécies puras e 2) os pesos nos fatores para cada mistura não somam um e portanto não podem ser interpretados facilmente em termos de composição. Na inspeção direta desta matriz não é óbvia a identidade dos objetos mais divergentes, que entretanto se tornam visíveis com a rotação dos fatores pelo método varimax. A tabela 3, que contém a matriz dos pesos dos fatores varimax, mostra que as três primeiras linhas possuem os maiores pesos nas colunas correspondentes; portanto estes serão os eixos de referência. Os valores encontrados nesta tabela não refletem as proporções relativas das espécies em cada mistura. A estrutura exata das proporções é revelada depois de se fazer uma projeção oblíqua destes fatores. A tabela 4 mostra o resultado da composição calculada para cada mistura com a adição de 10% de ruído, bem como as espécies puras (peso igual a 1). Este resultado mostra um excelente ajuste com os valores originais na tabela 1 ($\pm 0,001$ ou menor).

Para recuperar os espectros puros, o método da matriz k pode ser aplicado aos resultados da tabela 4 e os dados espectrais do conjunto de calibração da figura 1. A figura 2 mostra os espectros originais sem adição de ruídos e os espectros recuperados. Verifica-se que os espectros se superpõem completamente.

Segunda aplicação: Esta simulação pode ser considerada como a representação de uma reação cujas espécies variam de concentração com o tempo ou o pH, e podem ser formados produtos intermediários ou novos produtos. Nestes casos, muitas

Tabela 2. Os pesos dos três primeiros fatores do tipo Q de Imbrie.

Amostra nº	Fator 1	Fator 2	Fator 3
1	0,2448	-0,4321	-0,4569
2	0,2659	-0,0624	0,7539
3	0,2321	0,5050	-0,3755
4	0,2637	0,3951	-0,1034
5	0,2794	0,2943	-0,0390
6	0,2894	0,1973	0,0215
7	0,2947	0,1075	0,0719
8	0,2964	0,0260	0,1164
9	0,2963	-0,0436	0,1096
10	0,2944	-0,1156	0,0536
11	0,2898	-0,1924	-0,0141
12	0,2813	-0,2760	-0,0878
13	0,2681	-0,3634	-0,1667

Tabela 3. Composição calculada através da rotação varimax.

Amostra nº	Fator 1	Fator 2	Fator 3
1	0,9703	0,1065	0,2174
2	0,4996	0,7636	0,4094
3	0,1937	0,0891	0,9777
4	0,2987	0,2760	0,9136
5	0,4003	0,3370	0,8522
6	0,4872	0,3882	0,7825
7	0,5597	0,4273	0,7103
8	0,6189	0,4579	0,6384
9	0,6743	0,4779	0,5796
10	0,7371	0,4293	0,5222
11	0,7996	0,3909	0,4561
12	0,8599	0,3447	0,3770
13	0,9137	0,2895	0,2854

Tabela 4. Composição calculada a partir dos três primeiros fatores da rotação oblíqua.

Amostra nº	Espécie 1	Espécie 2	Espécie 3
1	1,0000	0	0
2	0	1,0000	0
3	0	0	1,0000
4	0,0034	0,2495	0,7471
5	0,0786	0,3084	0,6130
6	0,1427	0,3593	0,4980
7	0,1999	0,4009	0,3992
8	0,2507	0,4378	0,3115
9	0,3137	0,4375	0,2488
10	0,4008	0,4019	0,1973
11	0,5000	0,3578	0,1422
12	0,6148	0,3077	0,0775
13	0,7499	0,2488	0,0013

vezes os espectros das espécies puras ou de referências não podem ser obtidos separadamente ou não podem ser isolados por meios químicos.

Para simular esta situação partiu-se de uma espécie pura, a amostra nº 1 na tabela 5, e duas misturas, amostra nº 6, contendo as espécies 1 e 2 e a amostra nº 14, contendo as espécies 2 e 3, onde 2 representa um intermediário e 3 o produto final da reação. Estas duas espécies têm proporções que variam entre zero e 0,90, como mostrado na tabela 5. Como na primeira aplicação, foi tratado um conjunto de 14 espectros, ilustrados na figura 3. Neste caso considera-se que não existe nenhuma informação sobre o número de espécies, bem como sobre suas concentrações.

A análise de fatores mostra que dois fatores explicam em torno de 96% da variância total dos dados espectrais. Este

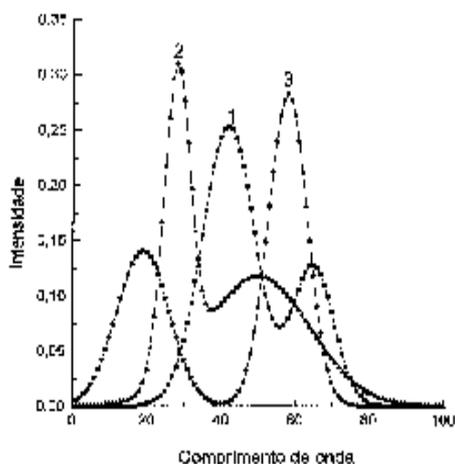


Figura 2. Perfil dos espectros para as três espécies químicas, calculados com 10% de ruído. Os símbolos representam os espectros originais destas espécies sem ruído. Ver legenda da figura 1.

Tabela 5. Composição original das misturas^a.

Amostra n ^o	Espécie 1	Espécie 2	Espécie 3
1	1,0000	0	0
2	0,8000	0,2000	0
3	0,6000	0,4000	0
4	0,5000	0,5000	0
5	0,3000	0,7000	0
6	0,0980	0,9000	0,0020
7	0,0014	0,8800	0,1186
8	0,0012	0,8000	0,1988
9	0,0110	0,7000	0,2989
10	0,0010	0,6000	0,3990
11	0,0006	0,5000	0,4940
12	0,0005	0,4000	0,5950
13	0	0,2000	0,8000
14	0	0,1000	0,9000

a) As amostras 1 é uma espécie pura; a amostra 6 tem 9,8% da espécie 1 e 90% da espécie 2 e a amostra 14 tem 10% e 90% das espécies 2 e 3 respectivamente.

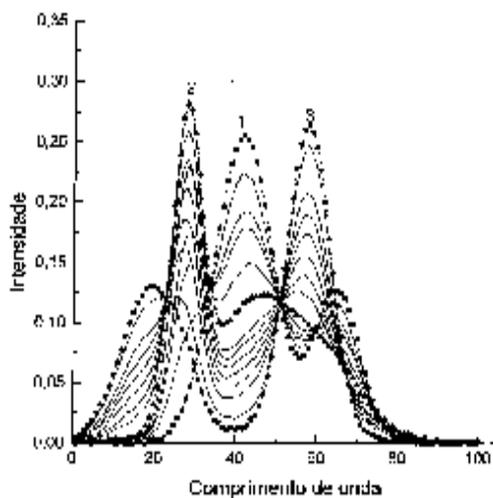


Figura 3. Espectros simulados com 10% de ruído. O espectro 1 corresponde a 100% da espécie 1, o espectro 2 com 90% da espécie 2 e 9,8% da espécie 3 e o espectro 3 com 10% e 90% das espécies 2 e 3, respectivamente. Ver legenda da figura 1.

resultado é consistente com a existência de três espécies químicas na mistura.

A tabela 6 mostra o resultado final da composição das misturas após as rotações, bem como os eixos de referência. Os eixos de referência determinados pelo método correspondem aos objetos 1, 6 e 14, exatamente como na matriz original. As composições calculadas com a adição de ruídos também estão muito próximas das originais, tabela 5.

Tabela 6. Composição calculada através da rotação oblíqua^a.

Amostra n ^o	Espécie 1	Espécie 2	Espécie 3
1	1,0000	0	0
2	0,7923	0,2065	0,0012
3	0,5782	0,4212	0,0006
4	0,4700	0,5300	0
5	0,2409	0,7591	0
6	0	1,0000	0
7	0	0,8853	0,1147
8	0	0,8051	0,1949
9	0	0,7013	0,2987
10	0	0,5961	0,4039
11	0	0,4868	0,5132
12	0	0,3708	0,6292
13	0	0,1278	0,8722
14	0	0	1,0000

a) As amostras 1, 6 e 14, que são as mais divergentes, são arbitrariamente identificadas pelo algoritmo como representando uma espécie diferente.

A figura 4 mostra os espectros puros originais e os calculados através da matriz k, com adição de 10% de ruído. Neste caso houve um pequeno deslocamento do espectro calculado, mas que manteve o mesmo perfil do original.

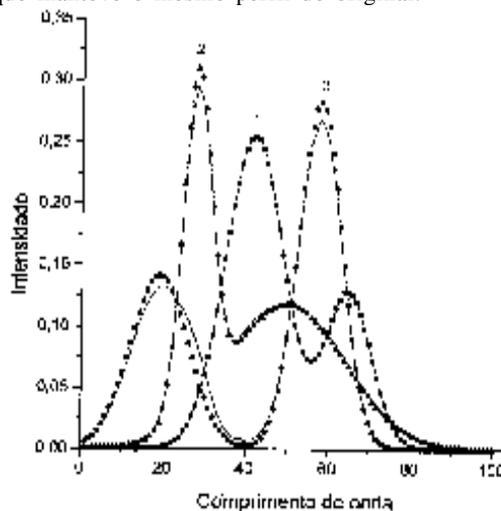


Figura 4. Perfil dos espectros para as três espécies químicas calculadas com 10% de ruído e os originais, sem ruído, 1■, 2▲ e 3●. Ver legenda da figura 1.

É importante salientar que os objetivos das duas análises exemplificadas com os dados simulados são diferentes. Os resultados na tabela 4 foram obtidos sem informação sobre as concentrações das espécies nas misturas. Como os espectros das espécies puras foram incluídos na análise, o método do tipo Q conseguiu identificá-los e atribuir as concentrações corretas para todas as espécies químicas em todas as amostras. Se a informação sobre as concentrações é incluída no tratamento do tipo Q e as curvas de calibração multivariadas são determinadas como

normalmente é feito em análises de rotina, não é necessário incluir amostras das espécies puras no conjunto de calibração. As curvas de calibração são simplesmente aplicadas para os dados espectrais de amostras com concentrações desconhecidas para fazer a análise quantitativa das espécies químicas.

As curvas de calibração não podem ser determinadas em situações onde existam somente dados espectrais, como na figura 3. A metodologia tipo Q identifica os espectros correspondentes às amostras com quantidades máximas das três espécies, ou seja, espectros das amostras 1, 6, e 14. Na ausência de informações sobre as verdadeiras concentrações destas amostras, a metodologia aproxima estas concentrações como sendo de espécies puras. Mesmo assim informações relevantes para a cinética química de todas as espécies podem ser extraídas dos resultados das outras amostras.

Terceira aplicação: São relatados na literatura vários trabalhos dedicados à determinação simultânea de aminoácidos e proteínas, destacando-se a metodologia da p-benzoquinona (PBQ)³⁶, que permite a determinação quantitativa simultânea de aminoácidos totais e proteínas totais. O objetivo desta aplicação foi utilizar a metodologia proposta na determinação quantitativa simultânea e individual de misturas binárias de aminoácidos com sobreposição espectral quase total.

O sistema estudado consistiu de misturas binárias de alanina e treonina nas proporções em massa, 75% e 25%, 50% e 50%, 25% e 75% dos aminoácidos puros, respectivamente. Todos os ensaios foram realizados em duplicatas autênticas. Os espectros das misturas e dos aminoácidos puros foram registrados de 548,5 a 302,5 nm, em 83 comprimentos de onda igualmente espaçados.

A figura 5 mostra a sobreposição espectral dos espectros puros, bem como das misturas. Pode-se observar que esta sobreposição é quase total, pela não especificidade do reagente cromogênico utilizado.

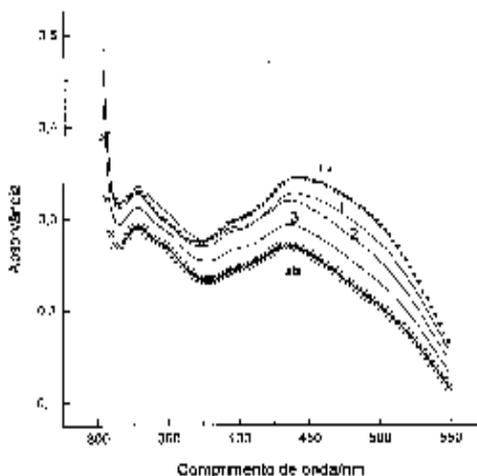


Figura 5. Espectros experimentais dos aminoácidos: 1) 25% de alanina e 75% de treonina, 2) 50% de alanina e 50% de treonina e 3) 75% de alanina e 25% de treonina. Os símbolos correspondem aos espectros das espécies puras.

A análise de fatores do tipo Q mostra que um único fator explica em torno de 99,9% da variância total dos dados espectrais. Este resultado é consistente com a existência de duas espécies químicas na mistura.

A tabela 7 mostra a composição original e o resultado final da composição calculada das misturas após as rotações. Estes resultados também podem ser observados na figura 6, onde os valores previstos para um dos aminoácidos (alanina), é apresentado em função dos valores originais. Os valores calculados estão próximos da bissetriz, o que representa um bom ajuste com a composição original, mostrando um desvio previsto em torno de 10%. Resultados similares foram obtidos para a treonina.

Devido à sobreposição das bandas espectrais, seria impossível determinar a composição química destas misturas, sem utilizar métodos baseados em estatística multivariada.

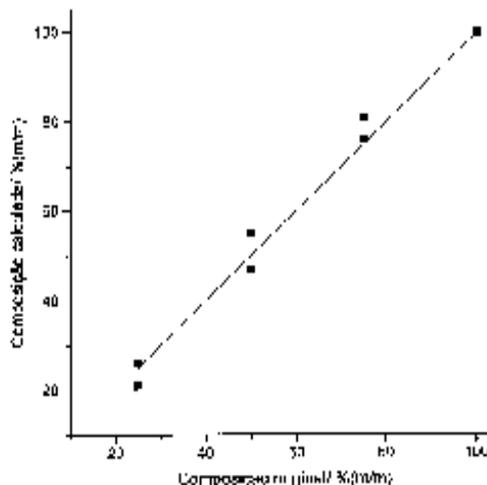


Figura 6. Gráfico da composição calculada em função da composição original (25%, 50% e 75%, m/m) para o aminoácido alanina.

A figura 7, mostra os espectros puros originais e os calculados através da matriz k. Verifica-se que os espectros se superpõem completamente. Estes espectros foram normalizados de tal forma que a soma das absorvâncias dos 83 comprimentos de onda fosse unitária.

CONCLUSÕES

A metodologia de fatores do tipo Q de Imbrie apresenta várias vantagens em relação à análise do tipo R. Do ponto de vista computacional, a análise dos aminoácidos por esta metodologia exige a diagonalização de uma matriz 10x10, enquanto que na do tipo R envolveria a diagonalização de uma matriz 83x83. A vantagem no uso do tipo Q é que um aumento na faixa de comprimento de onda ou na resolução do espectro não implicará mais tempo de computação.

A metodologia proposta permite a determinação das concentrações das espécies em misturas sem fazer análise de regressão,

Tabela 7. Composição original e calculada para as misturas de aminoácidos, (% em massa).

Mistura n ^o	Composição original		Composição calculada	
	Alanina, %	Treonina, %	Alanina, %	Treonina, %
1	25	75	26	74
2	25	75	21	79
3	50	50	47	53
4	50	50	55	45
5	75	25	81	19
6	75	25	76	24

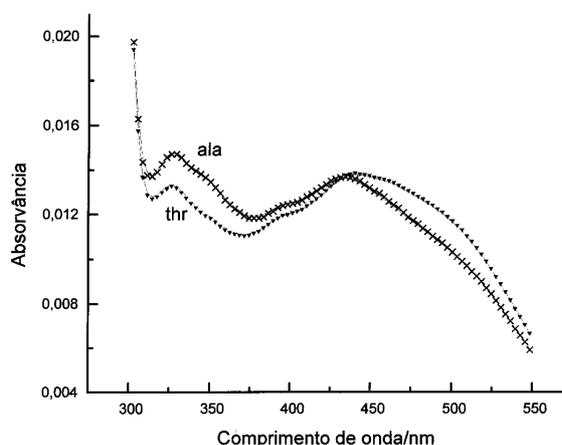


Figura 7. Perfil dos espectros recuperados para as duas espécies químicas (representado por linhas). Os símbolos representam os espectros originais, alanina (x) e treonina (▼).

enquanto que métodos usando fatores do tipo R normalmente envolvem o uso de critério por mínimos quadrados para a determinação das concentrações. Além disto, a presença de "outliers" no conjunto de dados não prejudica os resultados obtidos para as outras amostras.

Um dos aspectos mais importante na metodologia proposta neste trabalho é a não necessidade de um conjunto de calibração para a análise química das amostras. Por isto pode ser aplicada a inúmeros problemas que não poderiam ser resolvidos por métodos convencionais como por exemplo, a determinação das concentrações em equilíbrios complexos e das constantes de velocidade de reações, onde existem sobreposição espectral.

Finalmente, este trabalho mostra também que o erro devido a sobreposição espectral propagado nas concentrações determinadas são pequenos. Além disto, como esta metodologia não utiliza processos de separação química, os erros experimentais daí provenientes não estão presentes.

AGRADECIMENTOS

Os autores agradecem o apoio financeiro recebido do CNPq, bem como bolsas do programa PIBIC, CNPq-UEL (E.L.P e I.C.A).

REFERÊNCIAS

- Liang, Y. Z.; Kvalheim, O. M.; Manne, R.; *Chemom. Intell. Lab. Syst.* **1993**, *18*, 235.
- Geladi, P.; Kowalski, B. R.; *Anal. Chim. Acta* **1986**, *185*, 1.
- Draper N.; Smith, H.; *Applied Regression Analysis*; John Wiley; New York, 1981.
- Poullisse, H. N. J.; *Anal. Chim. Acta* **1979**, *112*, 361.
- Shi, L.; Li, Z.; Xu, Z.; Pan, Z.; Wang, L.; *J. Chemometr.* **1991**, *5*, 193.

- Fuller, M. P.; Ritter, G. L.; Draper, C. S.; *Appl. Spectrosc.* **1988**, *42*, 217.
- Haaland, D. M.; Easterling, G.; Vopicka, D. A.; *Appl. Spectrosc.* **1985**, *39*, 73.
- Brow, C. W.; Lynch, P. F.; Obremski, R. J.; Lavery, D. S.; *Anal. Chem.* **1982**, *54*, 1472.
- Höskuldsson, A.; *J. Chemometr.* **1988**, *2*, 211.
- Galera, M. M.; Vidal, J. L. M.; Frenich, A. G.; Parrilla, P.; *Analyst* **1994**, *119*, 1189.
- Peña, M. S.; Peña, A. M. de la; Salinas, F.; Mahedero, M. C.; Aaron, J. J.; *Analyst* **1994**, *119*, 1177.
- Lorber, A.; Wangen, L.; Kowalski, B. R.; *J. Chemometr.* **1987**, *1*, 19.
- Naes, T.; Martens, H.; *J. Chemometr.* **1988**, *2*, 155.
- Arrabal, M. J.; González, P. V.; Gámez, C. C.; Misugo, A. S.; Peña, A. M. de la; *Analyst* **1994**, *119*, 1537.
- Hoyer, B.; Kryger, L.; *Anal. Chim. Acta* **1985**, *167*, 11.
- Kalivas, J. H.; Kowalski, B. R.; *Anal. Chem.* **1981**, *53*, 2207.
- Vandeginste, B.; Klaessens, J.; Kateman, G.; *Anal. Chim. Acta* **1983**, *150*, 71.
- Gemperline, P. J.; *Anal. Chem.* **1986**, *58*, 2656.
- Vandeginste, B. G. M.; Derks, W.; Kateman, G.; *Anal. Chim. Acta* **1985**, *173*, 253.
- Lorber, A.; Goldbart, Z.; *Anal. Chem.* **1984**, *56*, 37.
- Öhman, J.; Geladi, P.; Wold, S.; *J. Chemometr.* **1990**, *4*, 135.
- Scarmínio, I. S.; Kubista, M.; *Anal. Chem.* **1993**, *65*, 409.
- Lawton, W. E.; Sylvestre, E. A.; *Technometrics*, **1971**, *13*, 617.
- Maeder, M.; Zilian, A.; *Chemom. Intell. Lab. Syst.* **1988**, *3*, 205.
- Widing, W.; Stephenson, D. A.; *Anal. Chem.* **1992**, *64*, 2735.
- Widing, W.; Heckler, C. E.; Agblevor, F. A.; Evans, R. J.; *Chemom. Intell. Lab. Syst.* **1992**, *14*, 195.
- Widing, W.; Guilment, J.; *Anal. Chem.* **1991**, *63*, 1425.
- Widing, W.; Lippert, J. L.; Robbins, M. J.; Kresinske, K. R.; Twist, K. P.; Snyder, A. P.; *Chemom. Intell. Lab. Syst.* **1990**, *9*, 7.
- Paatero, P.; Tapper, U.; *Chemom. Intell. Lab. Syst.* **1993**, *18*, 183.
- Jöreskog, K. G.; Klován, J. E.; Reymont, R. A.; *Geological Factor Analysis*; Elsevier, Amsterdam, 1976.
- Davis, J. C.; *Statistics and Data Analysis in Geology*; John Wiley, New York, 1983.
- Palasota, J. M.; Deming, S. N.; *Anal. Chim. Acta* **1992**, *270*, 13.
- Deane, J. M.; MacFIE, H. J. H.; King, A. G.; *J. Chemometr.* **1989**, *3*, 359.
- Hopke, K. P.; *Chemom. Intell. Lab. Syst.* **1989**, *6*, 7.
- Greenacre, M.; Hastie, T.; *Journal of the American Statistical Association* **1987**, *82*, 437.
- Zaia, D. A. M.; Barreto, W. J.; Santos, N. J.; Endo, A. S.; *Anal. Chim. Acta* **1993**, *277*, 89.