

APLICAÇÃO DO MODELO TUCKER-3 PARA A ANÁLISE DA BIODEGRADAÇÃO DE DIESEL

Marlon M. Reis* e Mariza G. Reis

AgResearch Limited, East Street, Private Bag 3123, Hamilton, New Zealand

Cleber C. Luz

Departamento de Química, Faculdade de Ciências Humanas, Exatas e Letras de Rondônia, Instituto João Neóricio, 78900-000 Porto Velho – RO, Brasil

Wanderley R. Bastos e Miyuki Yamashita

Universidade Federal de Rondônia, BR 364, km 9,5, 78900-000 Porto Velho – RO, Brasil

Recebido em 12/9/09; aceito em 5/4/10; publicado na web em 20/7/10

APPLICATION OF THE TUCKER-3 MODEL TO THE STUDY OF DIESEL BIODEGRADATION. Tucker-3 model offers several advantages for analysis of environmental data but its interpretation is still challenging. A Tucker-3 model was applied to a biodegradation experiment involving a large number of overlapped chromatographic peaks and a temporal variation. The Tucker-3 model allowed the data to be decomposed in two processes: evaporation and biodegradation. The results suggest that linear hydrocarbons were those biodegraded first and demonstrate that the data analysis can be simplified by interpreting the elements of the core array. The approach discussed in this work can be applied in similar problems involving multi-way data in other areas of chemistry.

Keywords: multi-way analysis; Tucker-3 models; biodegradation process.

INTRODUÇÃO

O uso de métodos cromatográficos para o monitoramento da biodegradação de substratos complexos, tais como derivados de petróleo, apresenta como maior desafio a sobreposição de picos cromatográficos. Esta sobreposição é resultado da mistura de um grande número de compostos, incluindo aqueles produzidos pela biodegradação e os originais da fonte de carbono sob degradação.

O objetivo do monitoramento, em geral, é identificar quais compostos, ou classes de compostos, estão sendo degradados e como este processo varia com o tempo. Estas informações são empregadas para avaliar a eficiência de agentes biológicos (e.g. consórcio de micro-organismos) no processo de biodegradação. Entretanto, o monitoramento de cada composto degradado e/ou produzido no processo pode se tornar não factível. Mesmo que a identificação dos compostos seja possível, a análise de suas variações no tempo e suas interrelações é de grande complexidade. Nestes casos, o emprego de análise de dados em multimodos permite a identificação dos principais fatores relacionados à biodegradação. Este tipo de análise tem sido empregado em várias áreas da química ambiental,¹⁻⁷ nas quais modelos em multimodos vêm se mostrando úteis na identificação de aspectos não aparentes do conjunto de dados.⁸ Neste trabalho é descrito o monitoramento de um processo de biodegradação por meio de análise de dados em multimodos, onde são identificadas a variação temporal associada ao processo de biodegradação e a provável classe de compostos mais rapidamente biodegradados.

Introdução teórica

O experimento avaliado aqui envolve o monitoramento de dois conjuntos de amostras (controle e amostras inoculadas com consórcio de micro-organismos) ao longo de 21 dias, período em que 4 amostras foram efetuadas para avaliação da composição de voláteis produzidos durante o processo de biodegradação.

O conjunto de dados resultante apresenta uma estrutura em

multimodos caracterizada por: perfis cromatográficos, período de biodegradação e amostras. O primeiro modo representa a variação na composição de compostos voláteis resultantes da perda por evaporação e/ou da biodegradação. O segundo modo reflete como a fração de voláteis variou ao longo do período de monitoramento. Por sua vez, o terceiro modo indica as diferenças entre amostras referentes aos ensaios inoculados (experimento) e não inoculados (controle).

A análise deste tipo de dados pode ser efetuada por meio da decomposição das informações em três classes: composição; variação temporal e discriminação entre controle/experimento. Os modelos Tucker, introduzidos por L. Tucker⁹ durante a década de 60 para a interpretação de estudos psicológicos, podem ser usados para a decomposição do conjunto de dados em multimodos. Estes modelos têm sido aplicados na análise exploratória de dados ambientais ou químicos, na identificação de compostos, calibração de ordem superior e outros.¹⁰⁻¹² O modelo Tucker-3 aplica-se a dados com estrutura em 3-modos e permite a decomposição das informações nas três classes desejadas (i.e. A “composição”, B “variação temporal” e C “discriminação entre controle/experimento”) sendo descrito pela Equação 1,

$$x_{ijk} = \left(\sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R a_{ip} b_{jq} c_{kr} g_{pqr} \right) + e_{ijk} \quad (1)$$

onde a_{ip} , b_{jq} e c_{kr} descrevem os elementos das matrizes componentes A (para o modo A), B (para o modo B) e C (para o modo C) de ordem $I \times P$, $J \times Q$ e $K \times R$, respectivamente; g_{pqr} é o elemento (p, q, r) do núcleo \mathbf{G} de dimensão $P \times Q \times R$; e_{ijk} descreve o erro para o elemento x_{ijk} , quando decomposto pelo modelo Tucker-3. Sendo um elemento do arranjo \mathbf{E} de dimensão $I \times J \times K$, este arranjo representa a parte do conjunto de dados que não pode ser descrita pela estrutura em 3-modos.

Outro modelo de decomposição de dados com estrutura em multimodos amplamente usado em Quimiometria é o PARAFAC (Parallel Factor Analysis) introduzido por Harshman, em 1970, também para estudos em psicometria. Vale notar que Carroll e Chang introduziram independentemente o modelo CANDECOMP (Canonical Decomposition) que apresenta grande similaridade ao PARAFAC.¹¹ O PARAFAC pode ser interpretado como um caso especial do modelo Tucker-3

*e-mail: marlon.dosreis@agresearch.co.nz

(a mesma afirmação é válida para um número de modos maior que 3), onde os elementos g_{pqr} são restritos aos valores ‘1’ para $g_{111}, g_{222}, g_{333}, \dots, g_{p=p=r}$ e ‘0’ para os outros elementos de g_{pqr} . A outra restrição é que as matrizes **A**, **B** e **C** possuam o mesmo número de fatores (i.e. $P=Q=R$). Estas restrições fazem do PARAFAC um modelo menos complexo que o Tucker-3, pois os elementos g_{pqr} não precisam ser interpretados. No entanto, estas restrições também limitam a análise de dados mais complexos como, por exemplo, em situações onde há interações entre fatores, ou que cada modo apresente complexidade variada. Por exemplo, nos dados descritos neste trabalho, o modo referente à composição apresenta uma complexidade maior que o modo variação temporal, pois há um grande número de compostos sendo biodegradados, mas são esperados apenas dois tipos de variação temporal: biodegradação e perda por evaporação. Por sua vez, o modo das amostras apresenta dois grupos: controle e experimento, ou seja, dois tipos de variação. Isto reflete nas matrizes **A**, **B** e **C** que devem possuir números diferentes de fatores (i.e. $P \neq Q = R$) o que justifica o emprego do modelo Tucker-3 neste trabalho.

Modelos Tucker-3 apresentam vantagens em relação a outros modelos de análise multivariada e em multimodos, pois permitem a decomposição de informações com alta complexidade. Isso possibilita sua aplicação em diversas áreas da Quimimétrie, entretanto, a análise dos parâmetros resultantes desta decomposição é complexa, restringindo seu amplo emprego na interpretação de dados químicos.^{7,8} Este trabalho é uma contribuição para o uso de modelos Tucker-3 como ferramenta Quimimétrica para análise de dados em multimodos encontrados em várias áreas da Química, em especial química ambiental, sendo as interpretações do modelo aqui discutidas extrapoláveis para outros problemas envolvendo estruturas em multimodos.

A formulação matricial do modelo Tucker-3 é dada pela Equação 2 ou por meio da decomposição do modelo Tucker-3 em *R* blocos como demonstrado pela Equação 3. O conceito de blocos facilita a interpretação dos resultados aqui apresentados.

$$\underline{\mathbf{X}} = \mathbf{A}\mathbf{G}(\mathbf{C}^T \otimes \mathbf{B}^T) + \mathbf{E} \tag{2}$$

$$\hat{\underline{\mathbf{X}}} = \mathbf{A}\mathbf{G}_1(\mathbf{c}_1^T \otimes \mathbf{B}^T) + \mathbf{A}\mathbf{G}_2(\mathbf{c}_2^T \otimes \mathbf{B}^T) + \dots + \mathbf{A}\mathbf{G}_R(\mathbf{c}_R^T \otimes \mathbf{B}^T) \tag{3}$$

sendo $\underline{\mathbf{X}} = (X_1|X_2| \dots |X_0)$ e $\underline{\mathbf{G}} = (G_1|G_2| \dots |G_R)$ representações matriciais de arranjos em 3 modos ($N \times M \times O$) e ($P \times Q \times R$). Nestes casos, $\underline{\mathbf{X}}$ corresponde a uma matriz resultante da justaposição de *O* matrizes X_i ($N \times M$). $\hat{\underline{\mathbf{X}}}$ corresponde à parte de $\underline{\mathbf{X}}$ descrita pelo modelo Tucker-3. O mesmo é válido para $\underline{\mathbf{G}}$. O arranjo \mathbf{E} , de dimensão $I \times J \times K$, representa a parte do conjunto de dados que não pode ser descrita pela estrutura em 3-modos definida pelo modelo Tucker-3. $(\mathbf{c}_i^T \otimes \mathbf{B}^T) = (\mathbf{c}_{i1}^T \mathbf{B}^T \quad \mathbf{c}_{i2}^T \mathbf{B}^T \quad \dots \quad \mathbf{c}_{iR}^T \mathbf{B}^T)$ representa o produto tensorial de Kronecker entre o vetor linha \mathbf{c}_i^T e a matriz \mathbf{B}^T .

Os modelos Tucker são, em geral, ajustados aos dados experimentais por meio de um algoritmo de Quadrados Mínimos Alternantes - QMA (*Alternating Least Squares*)¹³ onde a função “perda” dada na Equação 4 é minimizada.

$$perda = \|\underline{\mathbf{X}} - \mathbf{A}\mathbf{G}(\mathbf{C}^T \otimes \mathbf{B}^T)\|^2 \tag{4}$$

sendo $\|\cdot\|^2$ empregado para descrever a soma dos quadros dos elementos em questão.

O processo de otimização do QMA, em princípio, faz a busca pelo mínimo da função *perda* fixando todas as direções, exceto uma. Assim, o QMA para a otimização da função dada na Equação 4 tem como primeiro passo a determinação da matriz **A**, onde a “direção” (o termo direção refere-se à direção de otimização), o modo, a ser fixado, é dada pela Equação 5 (se for a primeira interação do QMA, as matrizes **B**, **C** e $\underline{\mathbf{G}}$ devem ser iniciadas seja com valores aleatórios

ou por algum outro critério) e a solução é dada pela Equação 6.

$$\mathbf{Z} = \underline{\mathbf{G}}(\mathbf{C}^T \otimes \mathbf{B}^T) \tag{5}$$

$$\mathbf{A} = \underline{\mathbf{X}}\mathbf{Z}^T(\mathbf{Z}\mathbf{Z}^T)^{-1} \tag{6}$$

No passo seguinte, a matriz **B** é calculada e para tal, a matriz $\underline{\mathbf{X}}$ é remodelada, como mostrado na Equação 7, e a matriz a ser mantida fixa é dada pela Equação 9. A solução deste passo é dada pela Equação 10.

$$\underline{\mathbf{X}}_B = (\mathbf{X}_1^T \quad | \quad \mathbf{X}_2^T \quad \dots) \tag{7}$$

onde

$$\underline{\mathbf{X}}_B = \mathbf{B}\mathbf{G}(\mathbf{C}^T \otimes \mathbf{A}^T) \tag{8}$$

$$\mathbf{Z}_B = \underline{\mathbf{G}}(\mathbf{C}^T \otimes \mathbf{A}^T) \tag{9}$$

$$\mathbf{B} = \underline{\mathbf{X}}\mathbf{Z}_B^T(\mathbf{Z}_B\mathbf{Z}_B^T)^{-1} \tag{10}$$

Para o cálculo da matriz **C**, a matriz $\underline{\mathbf{X}}$ também é remodelada, como mostrado na Equação 11, e a matriz a ser mantida fixa é dada pela Equação 12. A solução deste passo é dada pela Equação 13.

$$vec\underline{\mathbf{X}} = (\mathbf{A} \otimes \mathbf{B})vec\underline{\mathbf{G}}\mathbf{C}^T \tag{11}$$

sendo $vec\underline{\mathbf{X}} = (vecX_1 \quad | \quad vecX_2 \quad | \quad \dots \quad | \quad vecX_0)$ e $vec\underline{\mathbf{G}} = (vecG_1 \quad | \quad vecG_2 \quad | \quad \dots \quad | \quad vecG_R)$ onde *vec* representa o operador de vectorização, a partir do qual a matriz $\underline{\mathbf{X}}$, $I \times J$, é transformada em um vetor coluna IJ com seus vetores coluna sendo justapostos na direção coluna.

$$\mathbf{Z}_C = (\mathbf{A} \otimes \mathbf{B})vec\underline{\mathbf{G}} \tag{12}$$

$$\mathbf{C} = \left[(\mathbf{Z}_C^T \mathbf{Z}_C)^{-1} \mathbf{Z}_C^T vec\underline{\mathbf{X}} \right]^T \tag{13}$$

No último passo, para o cálculo da matriz $\underline{\mathbf{G}}$, a equação usada para **C** é empregada, como mostrado na Equação 11. As matrizes **C** e aquela dada pela Equação 14 são mantidas fixas. A solução deste passo é dada pela Equação 15.

$$\mathbf{Z}_G = (\mathbf{A} \otimes \mathbf{B}) \tag{14}$$

$$vec\underline{\mathbf{G}} = (\mathbf{Z}_G^T \mathbf{Z}_G)^{-1} \mathbf{Z}_G^T vec\underline{\mathbf{X}}\mathbf{C}(\mathbf{C}^T \mathbf{C})^{-1} \tag{15}$$

Após o cálculo de todas as matrizes, **A**, **B**, **C** e $\underline{\mathbf{G}}$, o valor da função dada na Equação 4 é avaliado. Caso a diferença entre o valor atual da função *perda* e o de um passo anterior seja menor que determinado critério, o algoritmo converge, caso contrário, ele é reiniciado com os valores atuais das matrizes **A**, **B**, **C**, e $\underline{\mathbf{G}}$. O algoritmo apresentado pelas Equações 5 a 15 é uma forma genérica; na prática, existem formas alternativas de implementação com maior eficiência computacional.¹⁴ Também é possível aplicar restrições no ajuste das matrizes representando cada modo, bem como o núcleo.^{15,16} Estas restrições são, em geral, baseadas em propriedades do conjunto de dados como, por exemplo, unimodalidade e não negatividade para cromatogramas, não negatividade para espectros, etc.

Os modelos Tucker podem ser avaliados por meio da “Função de Inércia”, Equação 16, que é baseada na função, minimizada pelo algoritmo QMA, usada no cálculo dos modelos Tucker. A Função inércia mostra a porção dos dados descrita pelo modelo. A função inércia também pode ser empregada em cada bloco da Equação 3, o que facilita a identificação dos blocos mais importantes para descrever o conjunto de dados.

$$f = \left(1 - \frac{\| \mathbf{X} - \mathbf{A}\mathbf{G}(\mathbf{C}^T \otimes \mathbf{B}^T) \|^2}{\| \mathbf{X} \|^2} \right) \cdot 100\% \quad (16)$$

O emprego do modelo Tucker-3 tem como etapa crítica a determinação do número de fatores das matrizes componentes. Em geral, esta determinação é efetuada para identificar aquela combinação de números de fatores que descreva a maior variação possível dos dados, ou seja, àquela que obtenha o maior valor de f na Equação 16. Vale notar que para modelos em multimodos nem sempre é possível atingir um ajuste de 100% como, por exemplo, na análise de componentes principais. Esta é uma vantagem, pois só aquela parte do conjunto de dados que apresente uma estrutura em multimodos é descrita pelo modelo. A determinação do número de fatores das matrizes componentes pode ser efetuada de forma sistemática,⁶ ou seja, testando todas as possibilidades de número de fatores e selecionando aquela com maior valor de f na Equação 16. Naqueles casos onde vários ajustes apresentem valores de f que não sejam significativamente diferentes (e.g. $f=97\%$, $f=95\%$, $f=98\%$, $f=96\%$) deve-se optar pela solução com o menor número de fatores possível.

A análise em multimodos tem recebido atenção significativa em Quimiometria, resultando em uma série de recursos desde pacotes computacionais a cursos on-line. O pacote computacional empregado neste trabalho é o PTAK^{17,18} (*Principal Tensor Analysis on k modes*) desenvolvido para o programa R que, dentre várias outros atributos, é uma versão na licença GNU (uso livre). Outros pacotes de interesses têm sido desenvolvidos para o programa computacional MATLAB, com destaque para os desenvolvidos pelos grupos de Bro,¹⁹ Smilde²⁰ e Kiers.²¹ Além disto, textos introdutórios e monografias descrevendo aplicações em Química de análise em multimodos oferecem variados exemplos que facilitam a compreensão destes métodos.²²⁻²⁶

PARTE EXPERIMENTAL

Ensaio de biodegradação

Para a investigação da biodegradação do diesel foram conduzidos ensaios inoculados (na presença de um consórcio de micro-organismos) acompanhados de controles químicos (não inoculados). Os experimentos foram realizados em frascos de 50 mL contendo 20 μ L de diesel e 20 mL de meio mineral inorgânico contendo em g L⁻¹ de água destilada: 1,36 de KH₂PO₄, 1,40 de Na₂HPO₄, 0,30 de (NH₄)₂SO₄, 0,05 de MgSO₄·7H₂O; e em mg L⁻¹ de água destilada: 5,8 de CaCl₂·H₂O, 2,75 de FeSO₄·7H₂O, 1,7 de ZnSO₄·7H₂O, 0,325 de CoCl₂, 0,235 de CuSO₄·5H₂O e 0,17 de Na₂MoO₄·2H₂O.

Os micro-organismos foram obtidos a partir de amostra de água residuária coletada nas proximidades de uma distribuidora de combustíveis situada no bairro Nacional, zona norte da cidade de Porto Velho-RO. Para os ensaios inoculados, foram empregados 1 mL de um consórcio de micro-organismos enriquecido, utilizando diesel como única fonte de carbono após 7 transferências sucessivas.

Os frascos foram esterilizados em autoclave (30 min, 120 °C) e incubados 28 °C sem agitação. O monitoramento foi efetuado ao longo de 21 dias e todos os ensaios foram realizados em triplicata.

Amostragem

Amostragens do *headspace* estático foram realizadas no 3^a, 7^a, 14^a e 21^a dias de experimento, por microextração em fase sólida (SPME) utilizando uma fibra de polidimetilsiloxane/divinil benzeno (65 μ m). Brevemente, os experimentos foram transferidos para um banho-maria a 40 °C e mantidos por 5 min; em seguida, a fibra de SPME foi conectada e mantida por 5 min. Este procedimento foi repetido

para todas as amostras e todos os ensaios realizados em triplicata. Os compostos absorvidos e adsorvidos pela fibra de SPME foram desorvidos no injetor do CG a 230 °C.

Cromatografia em fase gasosa com detecção por ionização em chama

As análises foram realizadas em um cromatógrafo CG17A (Shimadzu) equipado com uma coluna capilar DB5 (5% fenilmetilsiloxano, 30 m x 0,25 mm x 0,25 μ m). O programa de temperatura foi iniciado a 50 °C. A temperatura foi elevada 6 °C/min até 200 °C e, então, elevada em 30 °C/min até 290 °C. O hélio foi usado como fase móvel com fluxo de 1,0 mL/min. O injetor foi usado no modo *splitless* por 1 min com temperatura de 230 °C, e o detector a 250 °C.

Análise de dados

O modelo usado foi ajustado por meio da função '*Principal Component Analysis on n modes*' do pacote PTAK do programa R.²⁷ O roteiro empregado na aplicação desta função é apresentado brevemente no Apêndice e o conjunto de dados pode ser obtido a partir do autor correspondente. O modelo Tucker-3 foi aplicado nos dados originais sem o emprego de pré-tratamentos.

RESULTADOS E DISCUSSÃO

Cada ensaio discutido é descrito por uma matriz com 1101 tempos de retenção e 4 amostragens referentes a um período de 21 dias de monitoramento. O modo representado pela matriz **A** descreve os perfis cromatográficos referentes àqueles compostos presentes originalmente na fonte de carbono (i.e. diesel) e àqueles resultantes do processo de biodegradação deste substrato. O modo representado pela matriz **B** corresponde às 4 amostragens descrevendo a variação temporal. A matriz **C** representa o modo amostras e indica a diferença entre amostras na ausência (controle) e presença do consórcio de micro-organismos (experimento).

O modelo Tucker-3 ajustado considera o posto da matriz **B**, i.e., número de fatores, igual a 2, ou seja, um fator para degradação e outro para perda por evaporação. Alternativamente, o número de fatores da matriz **B** poderia ser maior, implicando em assumir mais de um fator para descrever o processo de evaporação ou degradação (por exemplo, fatores descrevendo o desaparecimento de produtos e outros mostrando a formação de produtos da degradação). Estas possibilidades foram testadas, mas não resultaram em variação significativa no ajuste do modelo (91,17; 92,25 e 92,82% para número de fatores iguais a 2, 3 e 4, respectivamente) e os fatores adicionais, 3 e/ou 4, não apresentam significado químico como mostrado na Figura 2S, material suplementar. Neste caso, a formação de produtos da degradação e o aparecimento dos produtos desta degradação devem ser descritos por um fator na variação temporal. A matriz **C** também foi ajustada com 2 fatores para diferenciação entre amostras controle e experimento. Finalmente, o posto da matriz **A** foi ajustado para que o modelo Tucker-3 descreva a maior variação possível dos dados.

As matrizes **A**, **B** e **C** foram consideradas como semi-ortogonais, isto é, $\mathbf{A}^T\mathbf{A}=\mathbf{I}$, $\mathbf{B}^T\mathbf{B}=\mathbf{I}$ e $\mathbf{C}^T\mathbf{C}=\mathbf{I}$, sendo **I** a matriz identidade de dimensão adequada. A restrição de semi-ortogonalidade oferece uma solução mais geral para o modelo Tucker-3, sem que suposições, além daquela que exista uma estrutura em multimodos, sejam feitas em relação ao conjunto de dados.

A análise por meio do modelo Tucker-3 é efetuada visando a decomposição do conjunto de dados em três fontes de informação: composição; variação temporal e controle/experimento. A Equação 17 representa a decomposição dos dados de biodegradação discuti-

dos aqui, onde **A** descreve os perfis cromatográficos, **B** representa o tempo de degradação e **c**₁ e **c**₂ os vetores colunas descrevendo o modo correspondente às amostras. Neste caso, pode ser dito que o conjunto de dados é decomposto em dois blocos: um relacionado ao processo de perda por evaporação, i.e., $AG_1(c_1^T \otimes B^T)$; e o outro ao processo de biodegradação, i.e., $AG_2(c_2^T \otimes B^T)$. Esta interpretação facilita a compreensão dos resultados do modelo Tucker-3, como discutido a seguir.

$$\hat{X} = AG_1(c_1^T \otimes B^T) + AG_2(c_2^T \otimes B^T) \quad (17)$$

O arranjo original de dados contendo medições em 1101 tempos de retenção, 4 amostragens, em um período de 21 dias, para 6 diferentes amostras foi decomposto em 8 fatores que descrevem o modo dos tempos de retenção e 2 fatores para cada um dos outros dois modos (Figura 1S, material suplementar).

O processo de biodegradação resulta no desaparecimento e na produção de um conjunto de compostos cuja concentração não varia linearmente com o tempo. O desaparecimento de compostos também está confundido com a perda por evaporação, o que dificulta a avaliação direta da biodegradação apenas pela avaliação direta do decaimento da intensidade dos cromatogramas. A análise dos dados com modelo Tucker-3 possibilita a decomposição destes efeitos, i.e., biodegradação vs. perda por evaporação. A Figura 1a apresenta os dois fatores de **B** relacionados à variação temporal. A variação descrita pelos fatores de **B** indica que há dois processos temporais independentes, caso contrário um ou mais fatores desta matriz componente não estaria correlacionado com o tempo. O primeiro destes fatores está diretamente associado a um processo de decaimento e o segundo também contínuo, mas apresentando um crescimento seguido por decaimento. O primeiro fator deve estar ligado diretamente ao processo de perda por evaporação, enquanto o segundo deve indicar o processo de biodegradação. Neste caso, o crescimento inicial observado no segundo fator estaria ligado a um rápido processo de biodegradação associado com a presença de compostos facilmente biodegradáveis (n-alcenos) e o decaimento associado à diminuição do processo de biodegradação devido à maior concentração de compostos de difícil biodegradação. Vale notar que **b**₁>0 ocorreu como uma solução natural para o Tucker-3 sem a aplicação de restrições de não negatividade, enquanto o elemento negativo de **b**₂ é resultado da restrição de semiortogonalidade. A decomposição por meio do modelo Tucker-3 foi efetuada para obtenção de **B** semiortogonal, i.e. $B^T B = I$, (i.e., para que **B** seja ortogonal é necessário que $B^T B = I$ e $BB^T = I$) sendo **I** a matriz identidade de dimensão apropriada. Isto implica que $b_1^T b_1 = 1$, $b_2^T b_2 = 1$ e $b_1^T b_2 = b_2^T b_1 = 0$, sendo **b**₁ e **b**₂ os vetores colunas de **B**. Os elementos de **b**₁ são maiores que zero, impedindo que o segundo componente, **b**₂, possua apenas valores positivos, pois $b_1^T b_2 = \sum_{j=1}^J b_{j1} b_{j2}$ (i.e., para que a igualdade $\sum_{j=1}^J b_{j1} b_{j2} = 0$ seja verdadeira é necessário que

b₂=0 ou pelo menos um termo desta soma seja menor que 0, como **b**₂≠0, um ou mais elemento deste vetor deve ser menor que zero).

Os fatores **c**₁ e **c**₂, Equação 17, informam como os fatores de **A**, **B** e **G**_i estão relacionados ao controle e ao experimento de biodegradação. O primeiro fator **c**₁ pode ser associado ao processo de evaporação, pois todas as amostras apresentam valores muito similares. O segundo fator representa o processo de biodegradação onde as amostras controle e experimento apresentam valores opostos neste fator, como mostrado na Figura 1b.

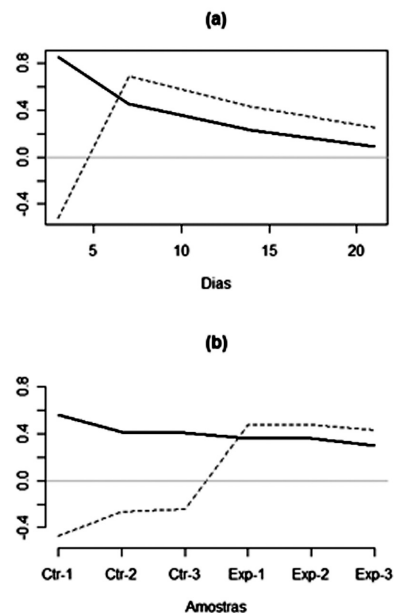


Figura 1. Fatores resultantes da decomposição pelo modelo Tucker-3 para os modos da variação temporal (a) e amostras (b). Linha contínua representa o fator 1 e tracejada, fator 2. ‘Ctr’ representa amostras controle e ‘Exp’ aquelas inoculadas com micro-organismos

O primeiro termo da soma apresentada na Equação 17 está associado à perda por evaporação e o segundo diferencia as amostras inoculadas daquelas usadas como controle. Assim, **G**₁ informa quais os fatores de **A** e **B** estão ligados à perda por evaporação e à biodegradação.

A Tabela 1 apresenta os elementos das matrizes **G**₁ e **G**₂, sendo cada elemento das colunas destas matrizes correspondente a um fator de **A**. Os elementos mais importantes para descrever o processo de perda por evaporação descrita por **G**₁ são: elemento 1 da coluna **g**₁₁ correspondente ao fator 1 de **A**; os elementos 2 e 3 da coluna **g**₂₁ correspondentes aos fatores 2 e 3, respectivamente. O produto AG_1

Tabela 1. Matrizes componentes do arranjo núcleo resultante da decomposição pelo modelo Tucker-3

G ₁				G ₂			
g ₁	g ₂₁	g ₁₁ (%)	g ₂₁ (%)	g ₁₂	g ₂₂	g ₁₂ (%)	g ₂₂ (%)
-1.268.742,82	-16.616,34	-94,67	-2,59	-39.915,07	160.832,26	-9,48	24,10
25.400,98	-408.339,74	1,90	-63,60	3.036,06	174.377,94	0,72	26,13
-17.278,68	-138.568,82	-1,29	-21,58	-76.402,26	-154.239,26	-18,14	-23,11
-6.133,40	28.015,90	-0,46	4,36	-204.614,99	-68.189,04	-48,57	-10,22
-4.679,39	24.783,81	-0,35	3,86	11.941,36	4.389,98	2,83	0,66
-7.013,86	-20.549,05	-0,52	-3,20	56.765,17	-78.010,42	13,47	-11,69
5.416,28	3.356,76	0,40	0,52	-16.194,79	13.177,41	-3,84	1,97
-5.444,50	1.814,20	-0,41	0,28	12.395,29	-14.076,46	2,94	-2,11

‘g_{ij}’ indica o vetor coluna ‘i’ da matriz **G**_j. ‘g_{ij}(%)’ corresponde à fração percentual do elemento em relação à soma de todos os elementos da coluna correspondente.

(Equação 17) resulta em uma matriz com dois fatores associados aos fatores 1 e 2 de **B**, respectivamente. O elemento 1 da coluna \mathbf{g}_{11} é cinquenta vezes maior (em magnitude) que o segundo maior elemento da coluna \mathbf{g}_{11} e três vezes maior que o maior elemento de \mathbf{g}_{21} . Isto mostra que o primeiro termo da Equação 17 é largamente dominado pelo fator 1 de **A** e **B**, ou seja, a perda por evaporação deve ser descrita, principalmente, pelo decaimento temporal dos compostos representados pelo fator 1 de **A**. Enquanto para descrever o processo de biodegradação, \mathbf{G}_2 , um número maior de fatores é necessário. Neste caso, os elementos 1, 3, 4 e 6 da coluna \mathbf{g}_{12} e elementos 1-4 e 6 da coluna \mathbf{g}_{22} , correspondentes aos fatores 1-4 e 6 de **A**.

A análise por cromatografia gasosa com detecção por ionização de chama não permite a identificação direta de compostos como, por exemplo, no caso de detecção por espectrometria de massas, mas em estudos de derivados de petróleo há padrões cromatográficos característicos, como a série homóloga de hidrocarbonetos lineares. Neste caso, o emprego de fatores para descrever o modo dos perfis cromatográficos é útil para a identificação da variação destes padrões. Embora não seja possível identificar os produtos da degradação, a interpretação destes fatores mostra onde estes compostos aparecem na análise cromatográfica. A Figura 2 mostra os fatores de **A**, sendo o fator 1 dominado por uma série homóloga de hidrocarbonetos. Os fatores 2, 3, 4, e 6 mostram padrões diferentes de composição da série homóloga e um número maior de compostos significantes, onde também é observado um deslocamento da linha base, que representa um maior número de compostos sobrepostos conhecidos como mistura complexa não resolvida, que inclui alcanos ramificados e cíclicos e subprodutos das transformações dos hidrocarbonetos. Como \mathbf{G}_2 está associada à biodegradação, estes fatores representam a variação, em concentração e em tipo, dos compostos produzidos durante a biodegradação. A comparação entre os fatores 1, 2, 3, 4, e 6 mostra que a série homóloga de hidrocarbonetos apresenta diferentes padrões em sua composição e deve compreender aqueles compostos mais impactados pelo processo de biodegradação.

CONCLUSÃO

A interpretação do experimento de biodegradação discutida neste trabalho apresenta como maior desafio a presença de pelo menos duas variações temporais associadas a um conjunto de compostos detectados sobre intensa sobreposição cromatográfica.

O Modelo Tucker-3 possibilitou a decomposição das informações associadas ao processo de perda por evaporação *versus* biodegradação e permitiu identificar, ou pelo menos sugerir, a classe de compostos mais rapidamente biodegradada, i.e., a série homóloga de hidrocarbonetos. A metodologia aplicada pode também ser empregada para comparação entre consórcios de micro-organismos e/ou condições de biodegradação permitindo, assim, a identificação das melhores condições de biodegradação para um dado substrato. Embora a análise dos resultados da decomposição pelo modelo Tucker-3 não seja direta como, por exemplo, na análise de componentes principais (PCA), a interpretação aqui discutida empregando o conceito de blocos representa uma forma prática de identificação das informações mais importantes do conjunto de dados. Esta interpretação é importante por ser extrapolável a outros tipos de dados em multimodo, pois não foram feitas suposições particulares para este conjunto de dados.

MATERIAL SUPLEMENTAR

O material suplementar, disponível em <http://quimicanova.sbq.org.br>, na forma de arquivo .PDF com acesso livre, apresenta a representação esquemática de um arranjo em três modos e uma solução para o modelo Tucker-3 com quatro fatores para o modo B.

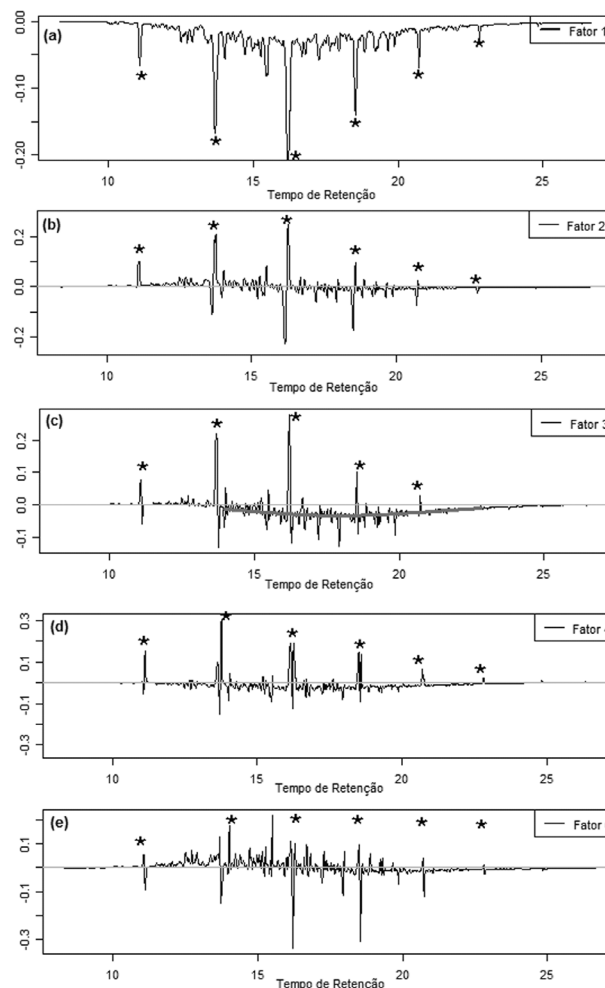


Figura 2. Fatores resultantes da decomposição pelo modelo Tucker-3 para o modo composição correspondentes aos fatores da matriz **A**. A série homóloga de hidrocarbonetos é indicada por '*' e o deslocamento da linha base mostrado em (c) pela linha com maior espessura

AGRADECIMENTOS

Ao CNPq pelo apoio financeiro através dos processos 476232/2004-0 e 553269/05-4, FINEP CT-Proinfra (# 0105056700) e à polícia civil de Rondônia pelo convênio que possibilitou o uso do equipamento de CG-DIC.

APÊNDICE

Rotação livre

O modelo Tucker-3 (e para maiores dimensões) quando ajustado sem restrições aplicadas ao núcleo e às matrizes componentes, ou usando a restrição de seminormalidade para estas matrizes (i.e., $\mathbf{A}^T\mathbf{A}=\mathbf{I}$, $\mathbf{B}^T\mathbf{B}=\mathbf{I}$, $\mathbf{C}^T\mathbf{C}=\mathbf{I}$), permite a aplicação de rotações a estas matrizes. Esta característica também é conhecida como rotação livre (*rotational ambiguity*). Esta rotação é resultado da transformação linear descrita nas Equações 18 e 19.

$$\hat{\mathbf{X}} = \mathbf{A}\mathbf{W}^{-1}\mathbf{W}\mathbf{G}(\mathbf{U}^T \otimes \mathbf{T}^T) \left([\mathbf{U}^{-1}]^T \mathbf{C}^T \otimes [\mathbf{T}^{-1}]^T \mathbf{B}^T \right) \quad (18)$$

$$\hat{\mathbf{X}} = \tilde{\mathbf{A}}\tilde{\mathbf{G}}(\tilde{\mathbf{C}}^T \otimes \tilde{\mathbf{B}}^T) \quad (19)$$

sendo $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{W}^{-1}$, $\tilde{\mathbf{B}} = \mathbf{B}\mathbf{T}^{-1}$, $\tilde{\mathbf{C}} = \mathbf{C}\mathbf{U}^{-1}$, $\tilde{\mathbf{G}} = \mathbf{W}\mathbf{G}(\mathbf{U}^T \otimes \mathbf{T}^T)$, $\mathbf{W}^{-1}\mathbf{W} = \mathbf{I}$, $\mathbf{T}^{-1}\mathbf{T} = \mathbf{I}$, $\mathbf{U}^{-1}\mathbf{U} = \mathbf{I}$, $(\mathbf{U}^T \otimes \mathbf{T}^T)([\mathbf{u}^{-1}]^T \otimes [\mathbf{t}^{-1}]^T) = (\mathbf{u}^T [\mathbf{u}^{-1}]^T \otimes \mathbf{t}^T [\mathbf{t}^{-1}]^T) = ([\mathbf{u}^{-1}]^T \otimes [\mathbf{t}^{-1}]^T)$, e \mathbf{I} representando a matriz identidade de dimensão apropriada. Os termos, '-I' e 'T', em sobrescrito indicam as operações de inversão e transposição, respectivamente.

A transformação das matrizes componentes (i.e., $\mathbf{A} \rightarrow \tilde{\mathbf{A}}$, $\mathbf{B} \rightarrow \tilde{\mathbf{B}}$, $\mathbf{C} \rightarrow \tilde{\mathbf{C}}$) não afeta o ajuste do modelo e representa um rotação dos vetores colunas destas matrizes, ou seja, cada vetor coluna da nova matriz é resultado de combinação linear dos vetores colunas das matrizes originais correspondentes. As Equações 20 e 21 exemplificam esta transformação para a matriz componente A.

$$\tilde{\mathbf{A}} = \mathbf{A}\mathbf{W}^{-1} \quad (20)$$

$$\tilde{\mathbf{a}}_z = \mathbf{A}\mathbf{v}_z = \mathbf{a}_1 v_{1z} + \mathbf{a}_2 v_{2z} + \dots + \mathbf{a}_p v_{pz} + \dots + \mathbf{a}_p v_{pz} \quad (21)$$

$\tilde{\mathbf{a}}_z$ corresponde ao vetor coluna z da matriz transformada $\tilde{\mathbf{A}}$, \mathbf{V}_z é o vetor coluna z da matriz \mathbf{W}^{-1} , v_{pz} é o elemento p do vetor coluna \mathbf{V}_z .

A transformação da matriz componente A pode ser ortogonal ou não ortogonal desde que a matriz inversa \mathbf{W}^{-1} exista.

A matriz transformada descreve o mesmo espaço coluna da matriz original, por ser resultado de uma combinação linear dos vetores da coluna da matriz original, o que significa que o mesmo conjunto de informações descrito pela matriz original é descrito pela matriz transformada, como ilustrado na Figura 2S, material complementar. Desta forma, a rotação das matrizes componentes pode facilitar ou dificultar a interpretação das informações descritas pelas matrizes componentes, mas esta interpretação deve sempre levar ao mesmo conjunto de conclusões, uma vez que estas matrizes descrevem o mesmo conjunto de informações. Neste caso, a interpretação dos elementos do arranjo núcleo é importante, pois indica a importância de cada fator (i.e., vetor coluna nas matrizes componentes) e como eles se interrelacionam. Desta forma, a interpretação em conjunto das matrizes componentes e núcleo evita que a rotação livre do modelo Tucker-3 afete a interpretação do conjunto de dados.

Função Principal Component Analysis on n modes

A seguir, é apresentado um roteiro para uso da função 'PCAn' do pacote 'PTAK' do programa computacional R. O caracter '#' é usado para introdução de comentários, ou seja, o texto após '#' não é interpretado. Aos leitores pouco familiares com o R é sugerida uma consulta a textos descrevendo a instalação²⁸ e introdução à programação R.²⁹⁻³¹

```
# Mudar para o diretório contendo os conjuntos de dados
setwd("C:/.../Dados")
# Ler as tabela de dados usando a função 'read.table'
Controle<-read.table("dieselControle.dat",header=TRUE,dec=',')
Biod<-read.table("dieselTeste.dat",header=TRUE,dec=',')
library(PTAK) # Carrega biblioteca PTAK para análise em multimodos.
```

```
# Criar um bloco de dados IxKxJ
# I: 1101 tempos de retenção
# K: 4 dias de monitoramento
# J: 6 conjuntos de amostras, 3 controles e 3 experimentos de biodegradação.
X.par<-array(NA,c(1101,4,6)) # Cria o bloco
# Preencher o bloco com as matrizes de dados
X.par[,,1]<-as.matrix(Controle[,seq(2,13,3)])
X.par[,,2]<-as.matrix(Controle[,seq(3,13,3)])
X.par[,,3]<-as.matrix(Controle[,seq(4,13,3)])
X.par[,,4]<-as.matrix(Biod[,seq(2,13,3)])
X.par[,,5]<-as.matrix(Biod[,seq(3,13,3)])
```

```
X.par[,,6]<-as.matrix(Biod[,seq(4,13,3)])
# A função 'seq(i,j,k)' cria um vetor de i a j com intervalos de k,
# e.g. seq(2,13,3) gera 2,5,8,11
# As tabelas de dados 'Controle' e 'Biod' possuem em cada coluna
um cromatograma.
# A coluna 1 destas matrizes possui os tempos de retenção. As tri-
plicatas estão dispostas
# em sequência, desta forma, as colunas 2, 5, 8, 11 correspondem
aos 4 dias de
# monitoramento.
# Efetuar a análise Tucker-3 decompondo os dados em 8 fatores para
# o modo cromatográfico e dois fatores para os outros modos
X.cand<-PCAn(X.par,dim=c(8,2,2))
summary(X.cand) # Resumo do cálculo
# O bloco núcleo é dado por:
X.cand[[3]]$scoremat
Nucleo.1<-X.cand[[3]]$scoremat[,1]
Nucleo.2<-X.cand[[3]]$scoremat[,2]

# Estimar a importância dos termos do núcleo
Nucleo.1perc<-100*Nucleo.1%*%diag(1/
apply(abs(Nucleo.1),2,sum))
Nucleo.2perc<-100*Nucleo.2%*%diag(1/
apply(abs(Nucleo.2),2,sum))
# "%*%" : Multiplicação de matrizes
# "diag()" : Cria uma matriz diagonal
# "apply(X,2,sum): Aplica a operação soma nas nos vetores colunas de X.
```

REFERÊNCIAS

1. Stanimirova, I.; Kita, A.; Malkowski, E.; John, E.; Walczak, B.; *Chemom. Intell. Lab. Syst.* **2009**, *96*, 203.
2. Cocchi, M.; Durante, C.; Grandi, M.; Manzini, D.; Marchetti, A.; *Talanta* **2008**, *74*, 547.
3. Pardo, R.; Vega, M.; Debán, L.; Cazorro, C.; Carretero, C.; *Anal. Chim. Acta* **2008**, *606*, 26.
4. Smolinski, A.; Falkowska, L.; Pryputniewicz, D.; *Oceanol. Hydrobiol. Stud.* **2008**, *37*, 49.
5. Astel, A.; Małek, S.; *J. Chemom.* **2008**, *22*, 738.
6. Stanimirova, I.; Simeonov, V.; *Chemom. Intell. Lab. Syst.* **2005**, *77*, 115.
7. Flåten, G. R.; Grung, B.; Kvalheim, O. M.; *Chemom. Intell. Lab. Syst.* **2005**, *77*, 104.
8. Stanimirova, I.; Zehl, K.; Massart, D. L.; Vander Heyden, Y.; Einax, J. W.; *Anal. Bioanal. Chem.* **2006**, *385*, 771.
9. Tucker, L. R.; *Psychometrika* **1966**, *31*, 279.
10. Smilde, A. K.; *Chemom. Intell. Lab. Syst.* **1992**, *15*, 143.
11. Bro, R.; *Crit. Rev. Anal. Chem.* **2006**, *36*, 279.
12. Escandar, G. M.; Olivieri, A. C.; Faber, N. M.; Goicoechea, H. C.; Poppi, R. J.; *Trend Anal. Chem.* **2007**, *26*, 752.
13. Kroonenberg, P. M.; de Leeuw, J.; *Psychometrika* **1980**, *45*, 69.
14. Kiers, H. A. L.; Kroonenberg, P. M.; Ten Berge, J. M. F.; *Psychometrika* **1992**, *57*, 415.
15. Smilde, A. K.; Tauler, R.; Saurina, J.; Bro, R.; *Anal. Chim. Acta* **1999**, *398*, 237.
16. Kiers, H. A. L.; Smilde, A. K.; *J. Chemom.* **1998**, *12*, 125.
17. <http://cran.r-project.org/web/views/ChemPhys.html>, acessada em Janeiro 2010 e Julho 2010.
18. <http://cran.r-project.org/web/packages/PTAK/index.html>, acessada em Janeiro 2010 e Julho 2010.
19. <http://www.models.kvl.dk/source/nwaytoolbox/index.asp>, acessada em Janeiro 2010 e Julho 2010.
20. <http://www.bdagroup.nl/content/Downloads/software/>, acessada em Janeiro 2010 e Julho 2010.

21. <http://www.ppsw.rug.nl/~kiers/Threeway%20m-files.zip>, acessada em Janeiro 2010 e Julho 2010.
22. http://www.bdagroup.nl/content/Publications/presentations/multiway_data_analysis.pdf, acessada em Janeiro 2010 e Julho 2010.
23. <http://www.models.kvl.dk/users/rasmus>, acessada em Janeiro 2010 e Julho 2010.
24. <http://www.ppsw.rug.nl/~kiers/>, acessada em Janeiro 2010.
25. Sena, M. M.; *Tese de Doutorado*, Universidade Estadual de Campinas, Brasil, 2004.
26. Reis, M. M.; *Tese de Doutorado*, Universidade Estadual de Campinas, Brasil, 2002.
27. R Development Core Team (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>, acessada em Janeiro 2010.
28. <http://cran.r-project.org/doc/contrib/Itano-installation.pdf>, acessada em Janeiro 2010.
29. <http://cran.r-project.org/doc/contrib/Torgo-ProgrammingIntro.pdf>, acessada em Janeiro 2010.
30. <http://cran.r-project.org/doc/contrib/Beasley-BioestatisticaUsandoR.pdf>, acessada em Janeiro 2010.
31. <http://cran.r-project.org/doc/contrib/biometria.pdf>, acessada em Janeiro 2010.