

## 25 ANOS DE QUIMIOMETRIA NO BRASIL

### Benício de Barros Neto

Departamento de Química Fundamental, Universidade Federal de Pernambuco, 50740-540 Recife - PE, Brasil

### Ieda S. Scarminio

Departamento de Química, Universidade Estadual de Londrina, CP 6001, 86051-990 Londrina - PR, Brasil

### Roy E. Bruns\*

Instituto de Química, Universidade Estadual de Campinas, CP 6154, 13083-970 Campinas - SP, Brasil

Recebido em 1/9/05; aceito em 20/1/06; publicado na web em 11/8/06

25 YEARS OF CHEMOMETRICS IN BRAZIL. Chemometric activities in Brazil are described according to three phases: before the existence of microcomputers in the 1970s, through the initial stages of microcomputer use in the 1980s and during the years of extensive microcomputer applications of the '90s and into this century. Pioneering activities in both the university and industry are emphasized. Active research areas in chemometrics are cited including experimental design, pattern recognition and classification, curve resolution for complex systems and multivariate calibration. New trends in chemometrics, especially higher order methods for treating data, are emphasized.

Keywords: chemometrics; microcomputers; multivariate analysis.

## INTRODUÇÃO

Na ciência as idéias freqüentemente surgem bem antes da tecnologia necessária para explorá-las. Foi o que aconteceu em todos os ramos da Química onde a teoria indicava que a resolução de determinado problema exigiria muitos cálculos. Na química quântica, por ex., metodologias como Hartree-Fock, interação de configurações e teoria de perturbação Möller-Plesset, apesar de concebidas nas primeiras décadas do século 20, somente puderam ser postas em prática muitas décadas depois, quando os recursos computacionais necessários se tornaram disponíveis<sup>1</sup>. Da mesma forma, os métodos de matemática e estatística multivariada demoraram a ser usados, não só nas ciências naturais como em outras áreas, pela falta de tecnologia computacional capaz de suportar a pesada carga de cálculos necessários na maioria das aplicações de interesse prático. A importante análise de componentes principais, proposta por Pearson<sup>2</sup>, já em 1901, e desenvolvida por Hotelling 30 anos depois<sup>3</sup>, somente nos anos 70 começou a ser usada como método de classificação em química<sup>4</sup>, sendo posteriormente introduzida na calibração multivariada, onde hoje é quase indispensável<sup>5</sup>. Um exemplo ainda mais dramático da distância entre a concepção e a aplicação prática é o da transformada de Fourier, que foi formulada na primeira metade do século XIX mas levou cerca de um século e meio até se incorporar à instrumentação analítica.

A quimiometria<sup>6,7</sup>, no sentido estrito, começou formalmente na primeira metade da década de 70, mas só se firmou definitivamente quando o computador – por si só ou como componente de instrumentos de medição – invadiu o laboratório químico. No sentido amplo, há quem argumente que a combinação de química com estatística começou muito antes, com os trabalhos do químico cervejeiro que se denominava Student, o do famoso teste t. Apesar da conservadora política de informática adotada pelo governo brasileiro na época, a invasão dos computadores não demorou a alcançar os laboratórios nacionais.

Neste trabalho pretendemos registrar um pouco da história dos

primeiros anos da atividade quimiométrica no Brasil e refletir sobre o desenvolvimento da disciplina até os dias atuais. Esperamos, com isto, não apenas traçar um esboço histórico desse período, mas também ilustrar a dinâmica evolutiva da ciência, especialmente da química. Nossa intenção não é fazer uma revisão completa dos trabalhos quimiométricos no Brasil, e sim mostrar como avanços reais na pesquisa podem ser favorecidos quando as novas idéias são examinadas sem preconceito. Dada a importância dos recursos computacionais para os problemas quimiométricos, é conveniente dividir o desenvolvimento da quimiometria no Brasil em três estágios: como era antes do microcomputador, como ficou após a chegada dos micros aos laboratórios, e como está atualmente.

## ANTES DO MICROCOMPUTADOR

No final da década de 70, os microcomputadores ainda não tinham chegado ao Brasil e os cálculos muito extensos eram feitos em computadores de grande porte, que funcionavam em centros de computação. Na época, a Unicamp tinha um dos computadores mais avançados das universidades brasileiras, um PDP-10, que era multiusuário e tinha nada menos que 1 Mb (isto mesmo: um megabyte) de memória RAM, a ser compartilhada por todos os usuários da universidade. Pelo menos uma vez por mês, o centro de computação fechava um dia ou mais para processar a folha de pagamento da universidade.

Naquela época, era preciso programar o software usando estratégias cuidadosas para poupar espaço na memória RAM e nas mídias magnéticas, além de diminuir o tempo de computação. Os algoritmos usados para fazer os cálculos eram bastante sofisticados, mas a produtividade dos programas era limitada pela tecnologia. O acesso aos resultados de qualquer cálculo era feito exclusivamente através de listagens impressas, produzidas por impressoras de linhas de 132 caracteres. Mas havia um ponto positivo: naquele tempo o software era de graça.

O primeiro programa quimiométrico produzido pelo grupo de Prof. Bruce Kowalski chamava-se ARTHUR. Segundo consta, em homenagem ao funcionário da agência financiadora responsável

\*e-mail: bruns@iqm.unicamp.br

pelo apoio ao desenvolvimento do programa. Esse programa chegou ao Instituto de Química da Unicamp em 1978, via correio, em uma fita magnética de 2400 pés (730 m) de extensão. Depois que o programa finalmente foi compilado no PDP-10, só era possível executá-lo entre meia-noite e 7:00 h da manhã, porque ele ocupava a maior parte da memória RAM da máquina. Os cálculos eram realizados em uma seqüência de sub-rotinas, cada uma das quais executava um método de análise exploratória ou classificação: análise de componentes principais, análise de agrupamentos hierárquica, regra dos vizinhos mais próximos e SIMCA, todos ainda amplamente utilizados. Depois de enviar um conjunto de dados para análise, o usuário precisava esperar pelo menos um dia para receber os resultados.

Em 1980 o Prof. Kowalski, que estava em licença sabática, aceitou um convite para ministrar um curso intensivo de quimiometria no IQ da Unicamp. O curso ocorreu no mês de novembro e foi o primeiro ministrado no país sobre esse assunto.

A primeira dissertação de mestrado envolvendo a aplicação de métodos quimiométricos – um estudo sobre a classificação de amostras de água mineral – foi defendida por Ieda S. Scarminio<sup>8</sup>, em fevereiro de 1981. As concentrações de 18 metais em amostras de águas minerais coletadas nos municípios de Serra Negra, Lindóia e Valinhos foram analisadas por espectrometria de emissão atômica, com fonte de plasma indutivamente acoplada. As análises foram feitas no CENA, em Piracicaba, com a colaboração de Antonio Jacintho, pesquisador do grupo de química analítica. O grupo era liderado pelo Prof. Henrique Bergamin, que encorajava entusiasticamente os trabalhos quimiométricos. Os resultados da tese foram publicados<sup>9</sup>, e várias tabelas e figuras foram mais tarde reproduzidas no livro *Data Analysis for Chemists*, David Livingstone<sup>10</sup>. Ainda em 1981, Elias Zagatto, que também integrava o grupo do CENA, defendeu a primeira tese de doutorado usando técnicas quimiométricas no Brasil<sup>11</sup>. Além de aplicar métodos quimiométricos para classificar solos do estado de São Paulo, o Prof. Zagatto pesquisou o uso do método generalizado de adições padrão em sistemas de análise em fluxo com espectrometria de emissão atômica, com fonte de argônio indutivamente acoplada<sup>12,13</sup>. Nessa época – início da década de 80 – os primeiros microcomputadores começavam a chegar aos laboratórios brasileiros.

## A CHEGADA DOS MICROCOMPUTADORES

### Os microcomputadores de 8 bits

Embora continuasse sendo possível fazer cálculos quimiométricos em computadores de grande porte, o processo estava longe de ser satisfatório. Os pesquisadores ainda tinham de esperar muito para receberem os resultados, principalmente quando o computador era utilizado para processar a folha de pagamento da Unicamp ou para executar outras tarefas consideradas prioritárias pela administração da universidade.

Com a reserva do mercado de informática em pleno vigor, nem se podia pensar em importar um microcomputador. Os primeiros micros vendidos no país eram feitos por pequenas firmas ou por algumas empresas associadas aos grandes bancos, como, por ex., a Itatec. Uma das pequenas firmas era a DICON, que tinha desenvolvido parte de sua tecnologia no Instituto de Física da Unicamp e vendia um dos micros mais avançados no mercado brasileiro. O micro vinha com dois “drives” para disquetes flexíveis de oito polegadas (20 cm), cada um com capacidade de 32 kbytes. Um disquete continha o sistema operacional e o outro armazenava o programa que se desejava executar, mais os dados de entrada. Como não havia ainda programas comerciais de quimiometria que pudessem ser

executados nesse tipo de microcomputador, o grupo da Unicamp teve de segmentar o programa ARTHUR, para transformar suas sub-rotinas em programas independentes, pequenos o bastante para caberem no disquete disponível. Parte desse trabalho foi realizada entre 1984 e 1986, com apoio da Docegeo – Rio Doce Geologia e Mineração S.A. – que tinha interesse em aplicar métodos de análises exploratória aos seus resultados de pesquisas de campo. Graças a esse esforço, conseguiu-se resolver o problema computacional dos cálculos quimiométricos do pessoal da Unicamp. Infelizmente, os programas compilados em um determinado micro dificilmente funcionavam em outros (mesmo se os dois fossem da mesma marca), o que tornava difícil implantar o uso de métodos quimiométricos em outros laboratórios. Quase sempre era necessário fazer modificações nas linhas do código de programação, e a simples compilação de um programa com apenas mil instruções FORTRAN demorava tanto que os interessados tinham tempo de sair e tomar um café tranqüilamente antes que ela terminasse. Muito café foi consumido nessa época.

Pouco tempo depois da compra da máquina DICON, Mário César U. de Araújo e Celio Pasquini conseguiram montar um sistema de aquisição de dados para um fotômetro de chama Micronal B262 de dois canais, usando uma placa protótipo 8085-A com interface serial por “loop de corrente”. Essa interface foi escolhida “porque os dados, no nosso caso, seriam transmitidos para um micro (DICON) localizado a mais de 30 m da placa”<sup>14</sup>. O micro era utilizado para fazer os cálculos necessários para os métodos de adições padrão simples e generalizado.

### O micro de 16 bits – o PC

O grande avanço computacional deu-se por volta de 1985, com o aparecimento dos microcomputadores com palavras de 16 bits, isto é, os primeiros PCs. Essas máquinas tinham uma característica fundamental: a padronização que faltava nos micros de 8 bits. Graças a ela, tornou-se possível executar em um determinado micro um programa que tinha sido compilado em outro, facilitando a transferência dos programas computacionais de quimiometria de um grupo de pesquisa para outro. Essa facilidade contribuiu para a implantação da quimiometria como disciplina acadêmica, já que, além de se oferecer uma base teórica ao aluno, também passava a ser possível colocar os conhecimentos em prática. Os programas computacionais disponíveis incluíam SCAL (manipulação de dados e auto-escalonamento), CORREL (cálculo de matrizes de correlação), KARLOV (componentes principais), VARVAR (gráficos), DISTAN (cálculo de distâncias entre pontos), KNN (regra dos vizinhos mais próximos), HIER (agrupamento hierárquico), SIMCA (classificação usando o método Simca) e WEIGHT (cálculo de índices de poderes de separação). Esses programas, que originalmente eram sub-rotinas do ARTHUR, foram adaptados para o microcomputador de 16 bits por Ieda S. Scarminio, no início de seu trabalho de doutorado, no IQ da Unicamp<sup>15</sup>.

O segundo curso intensivo de extensão em quimiometria, quatro anos depois do curso pioneiro de Kowalski, foi ministrado por Roy E. Bruns em 1984, no Departamento de Química Fundamental da Universidade Federal de Pernambuco. Nessa ocasião, o departamento estava modificando a grade curricular de seu curso de graduação e incluiu a disciplina Quimiometria como eletiva. No ano seguinte, o Prof. Ivano G. R. Gutz defendeu, no Instituto de Química da Universidade de São Paulo, uma tese de livre-docência tratando de vários aspectos da quimiometria<sup>16</sup>.

Em 1985, o primeiro curso regular de 60 h semestrais em quimiometria foi oferecido na pós-graduação do IQ-Unicamp, atraindo um total de dois alunos. Nesse mesmo ano, um trabalho de

revisão sobre quimiometria foi publicado em *Química Nova*<sup>4</sup>. Ao mesmo tempo, algumas indústrias nacionais começavam a se interessar pelo uso dos métodos quimiométricos. Em março de 1985, um curso intensivo de quimiometria foi ministrado na Souza Cruz S.A. e, logo em seguida, outros cursos foram ministrados nas empresas Docege e 3M do Brasil. O interesse que essas firmas mostraram pela aplicação de métodos eficientes de otimização despertou nossa atenção para o assunto.

O caso da 3M é um bom exemplo de como a interação com a indústria pode abrir novas linhas de pesquisa na academia. Em 1987, a empresa solicitou um curso sobre otimização de misturas. O primeiro livro sobre estatística de misturas tinha sido publicado em 1980<sup>17</sup>, mas nenhuma universidade brasileira ainda tinha explorado o assunto, pelo menos na área de química. A 3M forneceu o material bibliográfico para que o curso pudesse ser ministrado na própria empresa, em 1988. A partir daí, a otimização de misturas passou a fazer parte dos cursos de quimiometria da Unicamp e da UFPE, vindo a se tornar uma de nossas linhas de pesquisa em otimização em Química.

Nos últimos anos da década de 80, a demanda das universidades e empresas por cursos de quimiometria aumentou. Novos cursos intensivos foram ministrados na Copersucar, na Pirelli (cabos), na Oxiten (então Ultraquímica), na Universidade Federal do Pará (dois cursos) e na Universidade Estadual de Londrina.

O trabalho com a indústria estimulou o grupo da Unicamp a desenvolver outros programas computacionais, especialmente para tratar dados gerados por experimentos estatisticamente planejados, como planejamentos fatoriais completos ou fracionários, planejamentos compostos centrais e planejamentos de misturas. O primeiro programa de otimização de experimentos usando o método SIMPLEX foi desenvolvido no trabalho de mestrado de Romeu Virgili, cuja dissertação<sup>18</sup> foi defendida em junho de 1988. Alguns desses programas ainda se encontram disponíveis ao público, <http://www.chemomatrix.iqm.unicamp.br>, mas hoje o mercado dispõe de muitos programas comerciais para cálculos quimiométricos, tanto de planejamento e otimização como de classificação e calibração multivariada. Além de incluir novas metodologias quimiométricas que não constavam dos programas pioneiros, os programas de hoje apresentam avanços espetaculares na parte gráfica. O volume de dados que os programas atuais conseguem manipular praticamente só é limitado pela memória disponível. Do ponto de vista computacional, fazer cálculos quimiométricos com centenas de amostras e milhares de variáveis hoje é coisa rotineira.

Os primeiros programas comerciais de quimiometria para microcomputadores apareceram em meados da década de 80. O programa comercial GSAM (para cálculos com o método generalizado de adições padrão), do grupo de Kowalski, que desde 1981 era usado na Unicamp em um computador de grande porte, mais tarde foi adaptado para microcomputador por Mário C. U. de Araújo.

Em 1986, o Prof. Benício de Barros Neto juntou-se ao grupo de Campinas para um estágio de pós-doutorado e começou a fazer aplicações quimiométricas a problemas de natureza teórica. Os resultados desse trabalho contribuíram para mudar o enfoque das investigações sobre tensores polares<sup>19</sup>. Pouco tempo depois, refletindo sobre a crescente demanda por cursos de quimiometria, os autores do presente artigo concluíram que a ocasião era propícia para escreverem um texto didático de quimiometria. O livro começou a tomar forma no início da década de 90, sendo finalmente publicado em 1995, com o título *Planejamento e Otimização de Experimentos*<sup>20</sup>. Uma versão completamente revisada e ampliada saiu em 2001, com o título *Como Fazer Experimentos*<sup>21</sup>.

Exceto pelas aplicações do método generalizado de adições padrão (essencialmente o método da matriz **K**), os trabalhos brasilei-

ros de calibração multivariada só começaram a ser publicados no final dos anos 80. Ronei J. Poppi, em seu mestrado, usou modelos de regressão em componentes principais e regressão por mínimos quadrados parciais para determinar tolueno, isoctano e etanol a partir de picos cromatográficos<sup>22, 23</sup>. Os cálculos foram feitos com o programa comercial para microcomputadores SIMCA-3X, do grupo do Prof. Svante Wold.

## ÁREAS QUIMIOMÉTRICAS CONSOLIDADAS NO BRASIL

Os trabalhos em quimiometria no Brasil podem ser agrupados em três áreas principais: planejamento e otimização de experimentos, reconhecimento de padrões (métodos de análise exploratória e classificação) e calibração multivariada. A seguir, destacamos algumas das contribuições nessas áreas.

### Planejamento de experimentos

Com o advento da quimiometria, o uso de experimentos estatisticamente planejados cresceu rapidamente em vários campos de pesquisa no Brasil, especialmente Química, Engenharia Química, Engenharia de Alimentos e Biotecnologia. Além disso, a utilização dos métodos quimiométricos de planejamento e análise não ficou restrita às atividades acadêmicas. Indústrias como Oxiten, Pirelli, Braskem, Clariant, Nitroquímica, 3M do Brasil, Unilever, Petrobrás, Petroflex e Masterfoods do Brasil têm empregado – e continuam empregando – técnicas de planejamento e otimização de experimentos.

Uma busca no ISI, o Instituto de Informação Científica, mostrou que, dos 4099 trabalhos científicos localizados com a palavra-chave “factorial design”, 225 (isto é, 5,5%) tiveram participação de autor(es) brasileiro(s). Dos 3617 trabalhos com a palavra-chave “response surface”, 137 (3,8%) foram feitos, pelo menos parcialmente, no Brasil. Estas proporções são superiores à participação geral do Brasil na produção científica mundial, que fica entre 1 e 2%<sup>24</sup>.

Na Figura 1 apresentamos a evolução, nos últimos anos, do número de trabalhos brasileiros usando planejamentos fatoriais ou análise de superfícies de resposta. A partir de 1997-1998 os números aumentaram significativamente. Nos últimos três anos, a média anual foi de 33 trabalhos com planejamentos fatoriais e 20 trabalhos com análise de superfícies de resposta.

Os planejamentos experimentais mais usados no Brasil são os fatoriais, completos ou fracionários, e os planejamentos compostos

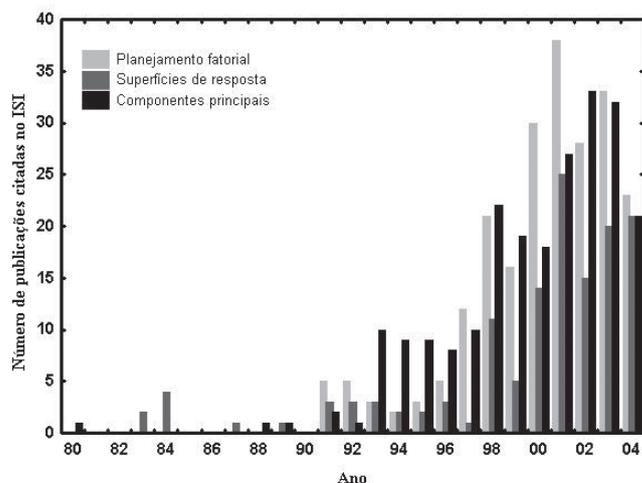


Figura 1. Publicações em três áreas da quimiometria, entre 1980 e 2004, com a participação de pesquisadores brasileiros

centrais. Exemplos de utilização destes métodos são amplamente encontrados em revistas científicas. Neste trabalho vamos enfatizar algumas áreas e metodologias que ainda são pouco usadas no Brasil, e mesmo em nível internacional.

O número de trabalhos encontrados com planejamentos estatísticos de misturas, por ex., é relativamente pequeno: ISI – busca com “mixture design” apenas 220 no total, 7 dos quais com a participação de pesquisadores brasileiros. Os trabalhos tratam de assuntos bem distintos, como melhoria das propriedades de produtos cerâmicos<sup>25</sup>, desempenho de produtos de concreto<sup>26</sup>, determinação de molibdênio em aço<sup>27</sup>, otimização das propriedades nutricionais de misturas de proteínas<sup>28</sup> e atividades de catalisadores à base de Cu-ZnO-Al<sub>2</sub>O<sub>3</sub><sup>29</sup>.

A triagem de um grande número de fatores foi encontrada em 3 trabalhos. Dois deles utilizam planejamentos Plackett-Burman<sup>30, 31</sup>. O terceiro faz uma avaliação de 15 fatores usando um planejamento fracionário 2<sup>15-11</sup> (dos 15 fatores, 11 foram analisados e 15-11 = 4 foram usados para estimar erro)<sup>32</sup>.

Recentemente, Ferreira e colaboradores<sup>33,34</sup> vêm se dedicando à otimização de métodos analíticos usando planejamentos Doehlert. Esses planejamentos se distinguem dos mais tradicionais para o ajuste de superfícies de resposta por varrerem o espaço experimental de forma muito econômica e, principalmente por admitirem o uso de diferentes números de níveis para diferentes fatores.

A maioria dos planejamentos baseados em princípios de estatística multivariada envolve de uma meia dúzia até várias dezenas de experimentos. É recomendável que os experimentos sejam executados em ordem aleatória, para se obter uma estimativa confiável da extensão do erro experimental. No entanto, químicos ou engenheiros muitas vezes trabalham com fatores que são difíceis de variar frequentemente. Mudar os níveis desses fatores cada vez que um experimento fosse executado poderia inviabilizar um projeto experimental ou causar demoras inaceitáveis na obtenção de resultados. Problemas deste tipo podem ser abordados com planejamentos do tipo lote repartido (“split-plot”), que permitem que variáveis difíceis de ajustar sejam mudadas em ordem aleatória somente em blocos de experimentos, enquanto os níveis dos fatores menos problemáticos são variados em ordem aleatória para todos os experimentos. Este procedimento facilita os aspectos operacionais no laboratório, mas tem seu preço. A estatística envolvida na análise dos dados é bem mais complexa que a dos planejamentos com ensaios feitos em ordem totalmente aleatória. Planejamentos “split-plot” têm sido muito empregados em estudos agrícolas. Para os quais, aliás, foram inventados, o que explica a terminologia. O “lote repartido”, originalmente, era um lote de terreno mesmo. Quase todos esses estudos, porém, se limitam a fazer uma análise da variância (ANOVA) dos dados experimentais. Para um químico, normalmente é mais interessante ir além da ANOVA e desenvolver um modelo mostrando como os resultados da pesquisa ou as propriedades do sistema variam em função das mudanças nos níveis dos fatores experimentais. Alguns trabalhos recentemente publicados ilustram a utilidade dos planejamentos “split-plot” para estudar sistemas químicos<sup>35-37</sup>. Nesses trabalhos, buscou-se otimizar ao mesmo tempo variáveis de processo e de mistura.

### Reconhecimento de padrões

Os métodos de Reconhecimento de Padrões (RP) podem ser aplicados com diferentes finalidades, entre elas a análise exploratória de dados, a classificação de amostras e a resolução de curvas.

Um conjunto de dados químicos consiste em um certo número de objetos, descritos por um determinado número de variáveis. Tipicamente, os objetos químicos são compostos, amostras analíticas, es-

pectros ou cromatogramas. As variáveis são características ou propriedades dos objetos, muitas vezes associadas à sua composição, como, por ex., concentrações dos elementos mais importantes, alturas de picos em perfis cromatográficos, ou intensidades e absorvâncias em diferentes comprimentos de onda. A análise exploratória é usada para se tentar detectar padrões de associação no conjunto de dados, a partir dos quais se pode estabelecer relações entre objetos e variáveis, descobrir objetos anômalos ou agrupar objetos. Os dois métodos de análise exploratória mais usados são a Análise de Componentes Principais (ACP) e a Análise de Agrupamentos Hierárquica (AAH). O principal objetivo de uma ACP é reduzir a dimensionalidade do conjunto de dados, preservando ao mesmo tempo o máximo da informação. Isto é feito calculando-se combinações lineares das variáveis originais. A primeira componente principal, PC1, é a combinação linear de máxima variância (isto é, de máxima informação). A segunda componente, PC2, também é de máxima variância, porém ortogonal a PC1. A terceira é de máxima variância e ortogonal às duas primeiras PCs, e assim por diante. Como esses eixos são calculados em ordem decrescente de importância, muitas vezes a informação relevante fica concentrada nas duas ou três primeiras PCs, que podem ser então examinadas à procura de padrões. A análise de agrupamentos hierárquica, por outro lado, é usada para analisar a estrutura do conjunto de dados em termos de grupos definidos de maneira hierárquica, de acordo com a similaridade entre objetos ou variáveis. Embora existam muitas variantes matemáticas, o que define a similaridade entre dois pontos é a distância entre eles no espaço multidimensional. Quanto menor essa distância, maior a similaridade entre os objetos (ou variáveis) representados pelos dois pontos.

Outro estudo típico de RP é a classificação de amostras de acordo com uma propriedade específica, baseada em medidas indiretamente relacionadas com a propriedade. A regra de classificação é desenvolvida para um conjunto de amostras cujas propriedades e medidas são conhecidas (o conjunto de treinamento). Os métodos de classificação que usam esse processo são conhecidos como métodos supervisionados de reconhecimento de padrões. Os mais usados em química são a regra dos vizinhos mais próximos (KNN) e o SIMCA, acrônimo correspondente a diversas expressões em inglês, entre elas “Soft Independent Modelling of Class Analogy”. O método SIMCA usa a ACP para construir separadamente um modelo para cada classe do conjunto de treinamento. Uma vez estabelecidos, os modelos são usados para classificar novas amostras (as amostras do conjunto de teste). O método KNN faz a classificação de acordo com a similaridade que a amostra desconhecida apresenta em relação às amostras do conjunto de treinamento. Quando não se dispõe de informação prévia sobre a classificação das amostras, procura-se identificar no conjunto de treinamento possíveis padrões de agrupamento, usando-se um método de reconhecimento não supervisionado. A técnica mais popular para isto é a análise de agrupamentos hierárquica.

No Brasil, os trabalhos usando métodos classificatórios em química ainda são poucos, sendo a maior parte voltada para a busca de relações entre estrutura molecular e atividade biológica, usando valores calculados por química quântica ou propriedades empíricas das moléculas<sup>38-40</sup>. Por outro lado, o interesse cada vez maior pela análise de amostras para fins de caracterização de ecossistemas e suas modificações antrópicas tem provocado um aumento significativo no uso dos métodos de análise exploratória, principalmente ACP e AAH, para estudar amostras naturais, como rochas, solos, ar, água e sedimentos.

Estudos com plantas medicinais em várias áreas, por ex. etnobotânica, etnofarmacologia, farmacologia, farmacognosia, fitoquímica, agronomia, biotecnologia e análise da qualidade, requerem que as espécies envolvidas estejam corretamente identificadas e classificadas, o que significa que essas áreas estão intimamente

associadas à taxonomia das plantas. Mesmo realizada por especialistas, porém, a diferenciação de espécies nem sempre é conclusiva, porque baseia-se em dados da literatura, análises histológicas e comparação com exsiccatas depositadas em herbários. Nos últimos anos, o interesse na utilização da química em sistemática vegetal (quimiotaxonomia ou quimiosistemática) tem aumentado bastante, principalmente pela disponibilidade de métodos analíticos que permitem análises mais fáceis e rápidas de um grande número de plantas. Mesmo assim, sistemas classificatórios que levem em conta as características químicas dos vegetais ainda são raros.

O uso de dados químicos em sistemática vegetal tem sido limitado por vários fatores, dos quais o mais importante é a pequena proporção de plantas investigadas quimicamente. Além disso, parece existir uma tendência de buscar apenas metabólitos com maior probabilidade de serem encontrados e com presença relevante nos táxons. Outros problemas são o custo dos processos de isolamento e identificação e a dificuldade de isolar produtos presentes em pequenas concentrações. Com tudo isto, a contribuição da análise química para a quimiotaxonomia tem ficado restrita ao desenvolvimento de métodos para isolar compostos específicos.

Alguns métodos quimiométricos de análise exploratória de dados, classificação e planejamento experimental foram aplicados com êxito na identificação, caracterização e busca da impressão digital cromatográfica de espécies trialadas do gênero *Baccharis*<sup>41-44</sup>. Quando feita da forma tradicional, a diferenciação entre as espécies deste gênero nem sempre é conclusiva, devido à sua grande semelhança e sinonímia.

### Métodos de resolução de curvas

São métodos quimiométricos utilizados para determinar corretamente os perfis de resposta (perfis espectrais ou de concentrações) de compostos presentes em misturas complexas. Esses métodos pressupõem que a matriz original pode ser representada pelo produto de duas matrizes menores, uma com o perfil das concentrações e a outra com os espectros puros de cada composto. No Brasil, os estudos de resolução de curvas têm sido muito pouco explorados. Uma busca no ISI mostrou que, dos 572 trabalhos científicos selecionados com a palavra-chave “curve resolution”, somente 4 tiveram a participação de autores brasileiros, sendo um deles um trabalho de revisão<sup>45-48</sup>.

Entre as técnicas propostas para se obter bons resultados em problemas de resolução de curvas estão a ACP com rotação, o uso de propriedades geométricas de espaço multivariado com restrições e a redução de posto (“rank annihilation”). Várias das técnicas propostas usam os resultados de uma ACP como ponto de partida. Os escores de uma análise de componentes principais podem ser usados para detectar similaridades entre amostras, mas a ACP, sendo um método do tipo R, tem seu foco nas relações entre as variáveis. A análise fatorial do tipo Q, por outro lado, é voltada expressamente para o estudo de relações entre objetos, o que torna os pesos dos fatores em uma análise do tipo Q mais apropriados para descrever as composições das amostras (isto é, dos objetos) que os escores obtidos em uma análise do tipo R. A descrição das composições depois pode ser melhorada, por meio de rotações aplicadas aos fatores calculados inicialmente.

As análises do tipo Q têm sido muito pouco exploradas em Química, mas alguns trabalhos que utilizam rotações varimax e oblíqua de Imbrie demonstram sua utilidade na resolução de curvas e em estudos cinéticos de equilíbrios complexos. Recentemente, uma análise do tipo Q foi usada para estudar a degradação de antocianinas em função do pH do meio. A partir das concentrações relativas das espécies presentes e dos perfis espectrais, foi possível inferir o mecanismo da transformação estrutural das antocianinas<sup>49</sup> e calcular as constantes de equilíbrio associadas a essas transformações<sup>50</sup>.

### Calibração multivariada

O problema principal da química analítica quantitativa é estimar a concentração de uma ou mais espécies, a partir dos valores de determinadas propriedades físico-químicas do sistema de interesse. Para isto, é necessário construir um modelo de calibração, ou seja, determinar a relação entre propriedades medidas e concentrações. A calibração multivariada é uma das mais bem sucedidas combinações de métodos estatísticos com dados químicos, tanto na química analítica quanto na química teórica. Os métodos mais usados incluem as regressões em componentes principais (PCR) e por quadrados mínimos parciais (PLS), o algoritmo genético (GA), as redes neurais artificiais (ANN) e os mínimos quadrados parciais não lineares (N-PLS). Recentemente foram publicados os primeiros trabalhos sobre figuras analíticas de mérito em calibração multivariada<sup>51,52</sup>. Devido ao número e à diversidade das publicações em calibração multivariada, fica difícil descrever cada contribuição nestes 25 anos, pelo que destacaremos apenas o trabalho do grupo do Prof. Ronei J. Poppi, o pioneiro na aplicação destes métodos no Brasil.

### Novas tendências: métodos para dados de ordem superior

Os dados de ordem superior são produzidos por instrumentação acoplada (ou hifenada), como cromatografia a gás/espectrometria de massa (GC/MS) e espectrometria de excitação/emissão. Nesta categoria, os métodos utilizados no Brasil incluem os modelos de Tucker, a calibração multivariada de ordem superior e o PARAFAC (“parallel factor analysis”). O modelo PARAFAC é usado tanto para resolução de curvas para dados de ordem superior quanto para análise exploratória e classificação. Os grupos mais atuantes nesta área estão em Campinas (por ex., os grupos dos Profs. Ronei J. Poppi e Márcia M. C. Ferreira) e Londrina (Profa. Ieda S. Scarminio).

Passados 25 anos, parece-nos que a quimiometria no Brasil finalmente chegou à maioria. Tanto é que a gente quase já não precisa esclarecer que não se trata de quimioterapia. Para isto, certamente contribuíram os incansáveis e entusiásticos esforços de divulgação feitos por vários pesquisadores, bem como as três escolas de verão organizadas pelo Depto. de Química da Pontifícia Universidade Católica do Rio de Janeiro, sob a coordenação dos Profs. Pércio Farias (1999), José Godoy (2000) e Ricardo Aucélio (2004). Hoje podemos encontrar grupos de pesquisa e desenvolvimento em quimiometria atuando em diversos estados da federação, não só nas universidades como em algumas empresas de grande porte. Além disso, diversos cursos de pós-graduação – e mesmo de graduação – já incluem a quimiometria em seus currículos. Mas talvez o sinal mais claro da consolidação das atividades quimiométricas no Brasil seja a realização do encontro internacional *Chemometrics in Analytical Chemistry Conference*, mais conhecido pela sigla CAC, que pela primeira vez viaja ao sul do Equador. A décima edição, que terá como tema “A Quimiometria nos Trópicos”, está programada para o período 10-14 de setembro de 2006, na Unicamp, sob a coordenação da Profa. Márcia Ferreira.

### REFERÊNCIAS

1. Szabo, A.; Ostlund, N. S.; *Modern Quantum Chemistry*, Dover Publications, Inc.: Mineola, New York, 1982.
2. Pearson, K.; *Philos. Mag.* **1901**, *2*, 559.
3. Hotelling, H.; *J. Educ. Psych.* **1933**, *24*, 417.
4. Bruns, R. E. Faigle, F.; *Quim. Nova* **1985**, *8*, 84
5. Martens, H.; Naes, T.; *Multivariate Calibration*, Wiley and Sons: Chichester, GB, 1989.
6. Kowalski, B. R.; *J. Chem. Info. Comp. Sci.* **1975**, *15*, 201.
7. Wold, S.; *Patt. Recog.* **1976**, *8*, 127.

8. Scarminio, I. S.; *Dissertação de Mestrado*, Universidade Estadual de Campinas, Brasil, 1981.
9. Scarminio, I. S.; Bruns, R. E.; Zagatto, E. A. G.; *Energ. Nucl. Agric.* **1982**, *4*, 99.
10. Livingstone, D. L.; *Data Analysis for Chemists*, Oxford University Press: Oxford, GB, 1995.
11. Zagatto, E. A. G.; *Tese de Doutorado*, Universidade Estadual de Campinas, Brasil, 1981.
12. Jacintho, A. O.; Zagatto, E. A.; Bergamin F<sup>o</sup>, H.; Krug, F. J.; Reis, B. F.; Bruns, R. E.; Kowalski, B. R.; *Anal. Chim. Acta* **1981**, *130*, 243.
13. Zagatto, E. A. G.; Jacintho, A. O.; Krug, F. J.; Reis, B. F.; Bruns, R. E.; Araújo, M. C. U.; *Anal. Chim. Acta* **1983**, *145*, 169.
14. Pasquini, C.; Araújo, M. C. U.; Bruns, R. E.; *Laboratory Microcomputer* **1990**, *9*, 44.
15. Scarminio, I. S.; Bruns, R. E.; *Trends Anal. Chem.* **1989**, *8*, 326.
16. Gutz, I. G. R.; *Tese de Livre-Docência*, Universidade de São Paulo, Brasil, 1985.
17. Cornell, J. A.; *Experiments with mixtures, designs, models and the analysis of mixture data*, 3<sup>rd</sup> ed. Wiley: New York, 2002.
18. Virgili Jr., R.; *Dissertação de Mestrado*, Universidade Estadual de Campinas, Brasil, 1988.
19. Neto, B. B.; Ferreira, M. M. C.; Scarminio, I. S.; Bruns, R. E.; *J. Phys. Chem.* **1989**, *93*, 1728.
20. Neto, B. B.; Scarminio, I. S.; Bruns, R. E.; *Planejamento e otimização de experimentos*, Ed. da Unicamp: Campinas, 1995.
21. Neto, B. B.; Scarminio, I. S.; Bruns, R. E.; *Como Fazer Experimentos – Pesquisa e Desenvolvimento na Ciência e na Indústria*, Ed. da Unicamp: Campinas, 2001.
22. Poppi, R. J.; *Dissertação de Mestrado*, Universidade Estadual de Campinas, Brasil, 1989.
23. Faigle, J. F.; Poppi, R. J.; Scarminio, I. S.; Bruns, R. E.; *J. Chromatogr.* **1991**, *539*, 123.
24. Marques, F.; *Pesquisa FAPESP* **2004**, *102*, 24.
25. Nardi, J. V.; Acchar, W.; Hotza, D.; *J. Eur. Ceram. Soc.* **2004**, *24*, 375.
26. Ribeiro, A. C. B.; de Almeida, I. R.; *Materials and Structure* **2000**, *33*, 398.
27. Andrade, J. C. de; de Almeida, A. M.; Coscione, A. R.; Aleixo, L. M.; *Analyst* **2001**, *126*, 892.
28. Castro, I. A.; Tirapegui, J.; Siva, R. S. S. F.; *Nutr. Res.* **2000**, *20*, 1341.
29. Lima, A. A. G.; Nele, M.; Moreno E. L.; Andrade, H. M. C.; *Appl. Catal., A* **1998**, *171*, 31.
30. Kalil, S. J.; Maugeri, F.; Rodrigues, M. I.; *Process Biochem.* **2000**, *35*, 539.
31. Mendees, F. D.; Martins, A. H.; *Miner. Metall. Process.* **2002**, *19*, 165.
32. Mendes, A. S.; de Carvalho, J. E.; Duarte, M. C. T.; Duran, N.; Bruns R. E.; *Biotechnol. Lett.*, **2001**, *23*, 1963.
33. Ferreira, S. L. C.; dos Santos, W. N. L.; Quintella, C. M.; Neto, B. B.; Bosque-Sendra, J. A.; *Talanta* **2004**, *63*, 1061.
34. Ferreira, A. C.; Korn, M. D. A.; Ferreira, S. L. C.; *Microchim. Acta* **2004**, *37*, 1437.
35. Bortoloti, J. A.; Bruns, R. E.; de Andrade, J. A.; Vieira, R. K.; *Chemom. Intell. Lab. Syst.* **2004**, *70*, 113.
36. Bortoloti, J. A.; de Andrade, J. C.; Bruns, R. E.; *J. Braz. Chem. Soc.* **2004**, *15*, 241.
37. Reis, C.; de Andrade, J. A.; Bruns, R. E.; Moran, R. C. C. P.; *Anal. Chim. Acta* **1998**, *369*, 455.
38. Subramanian, S.; Ferreira, M. M. C.; Trsic, M. A.; *Struct. Chem.* **1998**, *9*, 47.
39. Menezes, F. A. S.; Montanari, C. A.; Bruns, R. E.; *J. Braz. Chem. Soc.* **2000**, *11*, 393.
40. Kirali, R.; Takahata, Y.; Ferreira, M. M. C. *Quantitative Structure-Activity Relationships* **2003**, *22*, 430.
41. Lonni, A. A. S. G.; *Dissertação de Mestrado*, Universidade Estadual de Londrina, Brasil, 2002.
42. Silva L. M. da C.; *Dissertação de Mestrado*, Universidade Estadual de Londrina, Brasil, 2005.
43. Lonni, A. A. S. G.; Scarminio, I. S.; Silva, L. M. da C.; Ferreira, D. T.; *Anal. Sci.* **2003**, *19*, 1013.
44. Lonni, A. A. S. G.; Scarminio, I. S.; Silva, L. M. da C.; Ferreira, D. T.; *Anal. Sci.* **2005**, *21*, 235.
45. Scarminio, I. S.; Ishikawa, D. N.; Barreto, W. J.; Paczkowski, E. L.; Arruda, I. C.; *Quim. Nova* **1998**, *21*, 590.
46. Reis, M. M. dos; Ferreira, M. M. C. *Quim. Nova* **1999**, *22*, 11.
47. Sena, M. M.; Collins, C. H.; Collins, K. E.; Scarminio, I. S.; *Quim. Nova* **2001**, *24*, 331.
48. Saldanha, T. C. B.; Araújo, M. C. U. de; Barros Neto, B.; *Quim. Nova* **1999**, *22*, 847.
49. Levi, M. A. B.; *Dissertação de Mestrado*, Universidade Estadual de Londrina, Brasil, 2003.
50. Março, P. H.; *Dissertação de Mestrado*, Universidade Estadual de Londrina, Brasil, 2004.
51. Braga, J. W. B.; Poppi R. J.; *Quim. Nova* **2004**, *27*, 1004.
52. Braga, J. W. B.; Poppi R. J.; *J. Pharm. Sci.* **2004**, *93*, 2124.