

## DIVULGAÇÃO

## AVALIAÇÃO DO USO DE MÉTODOS QUIMIOMÉTRICOS EM ANÁLISE DE SOLOS

**Marcelo M. de Sena e Ronei J. Poppi\***

Departamento de Química Analítica - Instituto de Química - UNICAMP - CP 6154 - 13083-970 - Campinas - SP

**Rosa T. S. Frighetto e Pedro J. Valarini**

Embrapa Meio Ambiente - CP 069 - 13820-000 - Jaguariúna - SP

Recebido em 31/5/99; aceito em 13/10/99

**EVALUATION OF THE USE OF CHEMOMETRIC METHODS IN SOIL ANALYSIS.** One of the major interests in soil analysis is the evaluation of its chemical, physical and biological parameters, which are indicators of soil quality (the most important is the organic matter). Besides there is a great interest in the study of humic substances and on the assessment of pollutants, such as pesticides and heavy metals, in soils. Chemometrics is a powerful tool to deal with these problems and can help soil researchers to extract much more information from their data. In spite of this, the presence of these kinds of strategies in the literature has obtained projection only recently. The utilization of chemometric methods in soil analysis is evaluated in this article. The applications will be divided in four parts (with emphasis in the first two): (i) descriptive and exploratory methods based on Principal Component Analysis (PCA); (ii) multivariate calibration methods (MLR, PCR and PLS); (iii) methods such as Evolving Factor Analysis and SIMPLISMA; and (iv) artificial intelligence methods, such as Artificial Neural Networks.

**Keywords:** chemometrics; soil analysis; principal component analysis.

## INTRODUÇÃO

Com a crescente sofisticação das técnicas instrumentais de análise química, impulsionada pela invasão de microprocessadores e microcomputadores no laboratório, técnicas de tratamentos de dados mais complexas do ponto de vista matemático e estatístico tornaram-se necessárias<sup>1</sup>.

A análise de dados é uma parte essencial em todo experimento, sendo univariada quando somente uma variável é medida sistematicamente para várias amostras. Há muito tempo a estatística univariada vem sendo aplicada a problemas químicos, mas a sua utilização tornou-se limitada. Nas últimas décadas, a análise multivariada foi introduzida no tratamento de dados químicos, ganhando rapidamente popularidade e dando origem a uma nova disciplina, batizada de *Quimiometria*. O modelo estatístico dos métodos multivariados considera a correlação entre muitas variáveis analisadas simultaneamente, permitindo a extração de uma quantidade muito maior de informação.

Neste artigo será avaliada a utilização de métodos quimiométricos em análise de solos. O solo é uma estrutura complexa formada pela influência de fatores geológicos, topográficos, climáticos, temporais e antropogênicos; sendo a parte mais importante da geosfera e de onde provêm a maior parte dos nossos alimentos. Além disso, ele funciona como uma espécie de filtro ambiental para a purificação da água e do ar<sup>2</sup>.

O interesse maior na análise de solos está na avaliação de seus parâmetros químicos (concentração de metais e nutrientes, pH, etc.), físicos (compactação, umidade, etc.) e biológicos (microrganismos, etc.), os quais são um indicador de sua qualidade. Dentre os constituintes do solo, o mais estudado é a matéria orgânica (MO). Esta controla muitas de suas propriedades e é um fator central na manutenção de sistemas agrícolas sustentáveis. Práticas que favoreçam a conservação da MO melhoram as propriedades do solo e ajudam a reduzir o risco de erosão. Nos chamados sistemas de manejo sustentável, microrganismos

benéficos são incorporados à camada superficial do solo juntamente com restos de colheita e outros resíduos orgânicos, em substituição ao uso de pesticidas e fertilizantes, levando a um aumento na estabilidade e na conservação do solo. As substâncias húmicas<sup>3</sup>, polímeros naturais formados pela atividade de microrganismos durante a decomposição de plantas e animais, são uma parte dessa MO que exerce importante papel no transporte, biodisponibilidade e complexação de nutrientes e metais pesados no solo. Além destes, outro grande interesse na análise de solos está na medida de sua contaminação por pesticidas e metais pesados. Para o acompanhamento de todos esses parâmetros, a química analítica tem contado nos últimos anos com um grande avanço no desenvolvimento de técnicas instrumentais, tais como espectroscópicas e cromatográficas.

Os métodos quimiométricos possuem um enorme potencial no tratamento das questões descritas acima. Entretanto, até a segunda metade dos anos 80, a sua presença na literatura era escassa. A partir dos anos 90 nota-se um aumento significativo de suas aplicações, levando os pesquisadores da área de solos a uma grande elevação na sua capacidade de extrair informações dos dados. Pode-se afirmar que está em curso um processo de substituição de muitos métodos tradicionais de análise univariados por métodos multivariados. Mas esse processo ainda está em seu início e há necessidade de uma ampla divulgação para que cada vez mais maior número de pesquisadores tome contato com esta ferramenta de tratamento de dados multivariados que é a quimiometria.

O objetivo deste trabalho é dar um panorama geral destas aplicações, com ênfase na literatura mais recente. As aplicações serão divididas em quatro partes: na primeira serão apresentados os métodos descritivos e exploratórios baseados na Análise de Componentes Principais (PCA); na segunda, os métodos de calibração e previsão (Regressão Linear Múltipla, Regressão em Componentes Principais - PCR e Mínimos Quadrados Parciais - PLS); na terceira, um conjunto de trabalhos de um mesmo grupo de pesquisa, usando os métodos "Evolving Factor Analysis" e SIMPLISMA ("Simple-to-use Interactive Self-Modelling Mixture Analysis"); e na última parte, métodos de inteligência artificial, como Redes Neurais Artificiais.

e-mail: ronei@iqm.unicamp.br

## MÉTODOS DESCRITIVOS E EXPLORATÓRIOS

### Análise de Componentes Principais (PCA)

A Análise dos Componentes Principais, PCA<sup>4-6</sup> (do inglês “Principal Component Analysis”), tem por finalidade básica a redução de dados a partir de combinações lineares das variáveis originais. A PCA decompõe uma matriz de dados  $X$  (onde as  $m$  linhas são as amostras e as  $n$  colunas, as variáveis) de posto (“rank”)  $h$ , em uma soma de  $h$  matrizes de posto igual a 1, como na equação 1:

$$X = M_1 + M_2 + M_3 + \dots + M_h \quad (1)$$

onde o posto expressa o número de vetores linearmente independentes de uma matriz.

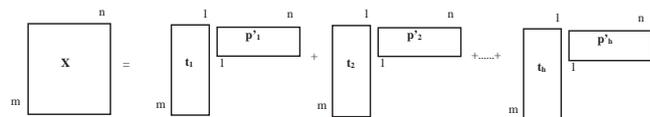
Essas novas matrizes de posto um, são produtos de vetores chamados “scores”,  $t_h$ , e “loadings”,  $p_h$ . Estes “scores” e “loadings” podem ser calculados por um ajuste de mínimos quadrados. A operação é equivalente ao cálculo de autovetores e autovalores de uma matriz pela Decomposição em Valores Singulares (SVD, “Singular Value Decomposition”). A equação pode ser representada na forma vetorial,

$$X = t_1 p_1' + t_2 p_2' + \dots + t_h p_h' \quad (2)$$

na forma matricial,

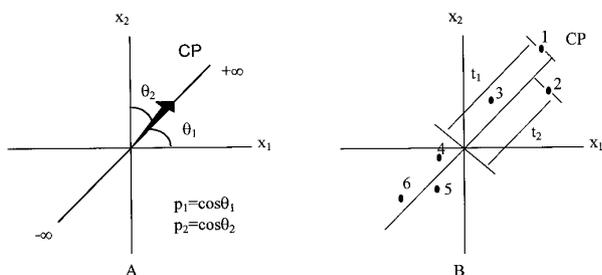
$$X = TP' \quad (3)$$

ou graficamente como na Figura 1.



**Figura 1.** Representação gráfica da somatória dos produtos dos vetores “scores e” loadings”.

Para exemplificar  $t_h$  e  $p_h'$ , a Figura 2 ilustra no plano bidimensional duas variáveis  $x_1$  e  $x_2$ . A Figura 2A mostra uma componente principal (CP), a reta que aponta na direção de maior variabilidade das amostras (pontos 1 a 6) da Figura 2B. Os “scores”  $t_h$  são as projeções das amostras na direção da CP (Fig. 2B) e os “loadings”  $p_h$  são os cossenos dos ângulos formados entre a CP e cada variável (Fig. 2A).



**Figura 2.** Representação de uma Componente Principal (CP) no caso de duas variáveis: (A) os “loadings” são os ângulos do vetor direção; (B) os “scores” são as projeções das amostras 1-6 na direção da CP. Notar que os dados estão centrados na média.

As novas variáveis, as CPs, são ortogonais entre si e, portanto, não correlacionadas. Normalmente, as primeiras CPs explicam a maior parte da variância total contida nos dados e podem ser usadas para representá-los.

A Análise de Fatores é adotada em boa parte da literatura

como sinônimo de PCA. Porém alguns autores<sup>7,8</sup> definem esses termos como métodos diferentes, com base no modo como os fatores (ou CPs) são definidos. Na PCA os fatores devem explicar o máximo da variância contida em todas as variáveis observadas, enquanto que na análise de fatores, eles devem explicar o máximo da intercorrelação entre as variáveis. Neste artigo, de acordo com a maioria da literatura, ambos os termos, fatores e CPs, serão tratados como equivalentes.

### Aplicações de Métodos Exploratórios em Análise de Solos

A PCA é um método exploratório porque ajuda na elaboração de hipóteses gerais a partir dos dados coletados, contrastando com estudos direcionados nos quais hipóteses prévias são testadas. Ela é capaz de separar a informação importante da redundante.

Para a análise exploratória é de fundamental importância entender que os “scores” fornecem a composição das CPs em relação aos objetos (amostras) enquanto os “loadings” fornecem essa mesma composição em relação às variáveis. Como as CPs são ortogonais, é possível examinar as relações entre os objetos através dos gráficos dos “scores” projetados nas primeiras CPs, e entre as variáveis através dos gráficos dos “loadings”. Esses gráficos apresentam informações úteis sobre tendências subjacentes e características da estrutura dos dados, tais como, agrupamentos de objetos ou variáveis e caracterização dos chamados “outliers” (amostras mal comportadas, que não se encaixam no modelo). O estudo conjunto de “scores” e “loadings” permite estimar a influência de cada variável em cada objeto.

A seguir, as aplicações de métodos exploratórios serão divididas em dois grupos, de acordo com sua finalidade.

### Avaliação de propriedades do solo

Nenhuma medida isolada, seja biológica, física ou química, poderá dar uma ampla visão da qualidade do solo. Existe, portanto, a necessidade de uma avaliação integrada dos diversos aspectos dessa qualidade<sup>2</sup>.

Como exemplo de aplicação deste tipo de estratégia serão comparados a seguir, sistemas de manejo do solo que utilizam microrganismos benéficos, aqui chamados sustentáveis, com sistemas que fazem uso intensivo de herbicidas e fungicidas, chamados convencionais<sup>9,10</sup>. Espera-se que a incorporação de MO e de restos de culturas, combinada com a inoculação de microrganismos benéficos em substituição ao uso de agrotóxicos, proporcione uma melhora nos parâmetros químicos, biológicos e bioquímicos, além de aumentar a estabilidade do solo em comparação com práticas de manejo convencionais.

O primeiro estudo foi realizado em três áreas no município de Suzano (região leste da Grande São Paulo)<sup>9</sup>. Na primeira área foram separadas quatro faixas, uma de agricultura sustentável (F1S), uma de pastagem (F2P), uma de agricultura convencional (F3C) e uma de mata nativa (F4M); na segunda área foram separadas duas faixas, uma de agricultura sustentável (F5S) e uma de agricultura convencional (F6C); na terceira área foram separadas três faixas, uma de agricultura sustentável (F7S), uma de mata nativa (F8M) e uma de agricultura convencional (F9C); Foi analisado o conjunto das nove faixas. Para cada faixa foram coletadas 10 amostras compostas de solo. Para cada amostra foram medidos 21 parâmetros químicos e biológicos: fósforo (P), matéria orgânica (MO), pH, potássio (K), cálcio (Ca), magnésio (Mg), acidez total (H+Al), alumínio (Al), soma de bases (S), capacidade de troca catiônica (CTC), volume de saturação por bases (V), boro (B), cobre (Cu), ferro (Fe), manganês (Mn), zinco (Zn), bactérias esporulantes (Bae), actinomicetos (Ac), microrganismos celulolíticos (Cel), teor de carbono em polissacarídeos (Ps) e atividade de desidrogenase (Dh).

A PCA foi aplicada aos valores médios das 10 amostras de

cada faixa. Foi tratada uma matriz 9x21 (9 faixas x 21 variáveis) sendo obtido um modelo que com duas CPs explica 76% da variância dos dados originais. A terceira CP, que explica 11,3% da variância total, foi considerada sem importância para a interpretação de tendências nos dados. É importante notar que, neste tipo de análise os dados devem ser sempre autoescalados (média zero e variância um) para assegurar que todas as variáveis contribuam igualmente para o modelo, independente da escala. Os cálculos foram feitos no “software” “MATLAB 5.2” utilizando o pacote “PLS Toolbox 2.0”<sup>11</sup>.

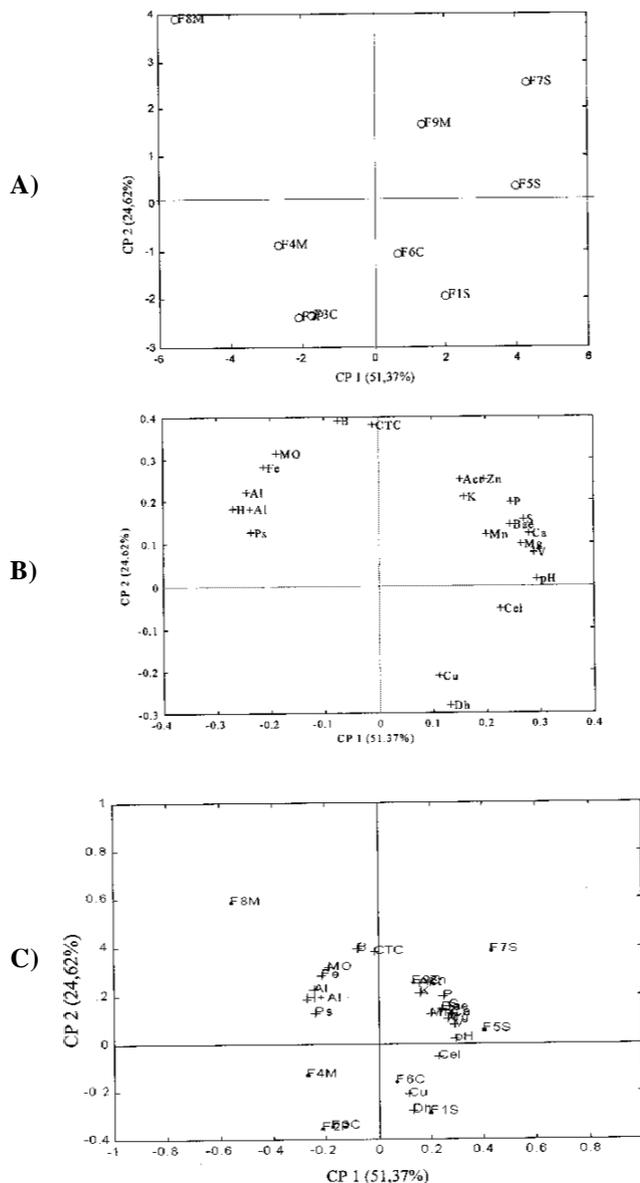
A Figura 3 mostra os gráficos das duas primeiras CPs. O gráfico de “scores” (Fig. 3A) permite a caracterização de tendências entre as faixas. A primeira CP (51,4%) é responsável pela diferenciação entre os tipos de manejo do solo, enquanto a segunda CP (24,6%) separa as faixas de acordo com a área de origem. Na CP1 as faixas de agricultura sustentável, F1, F5 e F7, apresentam valores positivos, contrastando com as faixas de mata, F4 e F8, que apresentam valores negativos; as faixas de agricultura convencional e pastagem, F2, F3, F6 e F9 apresentam valores intermediários. A CP2 possui valores positivos para as faixas da terceira área (F7-F9), intermediários para a segunda área (F5 e F6) e negativos para a primeira área (F1-F4). A faixa de mata F4M mostra comportamento um pouco diferenciado das demais faixas da primeira área (F1-F3), apresentando valores menos negativos em CP2.

O gráfico de “loadings” (Fig. 3B) permite a caracterização de tendências entre as variáveis. Observando-se ao longo do eixo da CP1, as variáveis que mais influenciam diretamente as faixas sustentáveis (valores positivos) são P, pH, Ca, Mg, S, V, Mn, Zn, Ba e Cel; enquanto as variáveis que caracterizam a mata (valores negativos) são MO, H+Al, Al<sup>3+</sup>, Fe e Ps. Pode-se notar, ao longo do eixo da CP2, que as variáveis que mais contribuem para a diferenciação entre as áreas são MO, CTC e B (influenciam diretamente a terceira fazenda – valores positivos na CP2); e Cu e Dh (influenciam a primeira área – valores negativos na CP2). Através dos chamados gráficos “biplots” (Fig. 3C) é possível visualizar conjuntamente as relações entre as CPs, as faixas e as variáveis.

Em um segundo estudo foram escolhidas três fazendas no município de Guaíra (região norte do Estado de São Paulo): Macaúba, Lagoa do Fogão e Cuiabanos<sup>10</sup>. Em cada fazenda foram separadas cinco faixas (correspondente a um quarto da área sob pivô central), duas de sistema sustentável, duas de sistema convencional e uma de mata nativa. Para cada faixa foram coletadas 10 amostras compostas do solo. Para cada amostra foram medidos 8 parâmetros biológicos e bioquímicos: pH, Leveduras (Lev), Actinomicetos (Act), Microrganismos Celulolíticos (Cel), Microrganismos Solubilizadores (Sol), teor de carbono em Polissacarídeos (Ps) em mg/g de solo, atividade de Desidrogenase (Dh) em  $\mu$ L de H/g de solo e Biomassa Microbiana (BiM) em  $\mu$ g de C/g de solo.

Como no estudo anterior, a PCA foi aplicada aos valores médios das 10 amostras de cada faixa. Para cada fazenda partiu-se de matrizes 5x8 (5 faixas x 8 variáveis) obtendo-se um modelo que com duas CPs explicou entre 85 e 91% da variância dos dados originais (Macaúba: 87,37%; Lagoa do Fogão: 85,35%; Cuiabanos: 91,17%). A Figura 4 apresenta os gráficos “biplots” de CP1 x CP2 para as três fazendas: Macaúba (Fig. 4A), Lagoa do Fogão (Fig. 4B) e Cuiabanos (Fig. 4C).

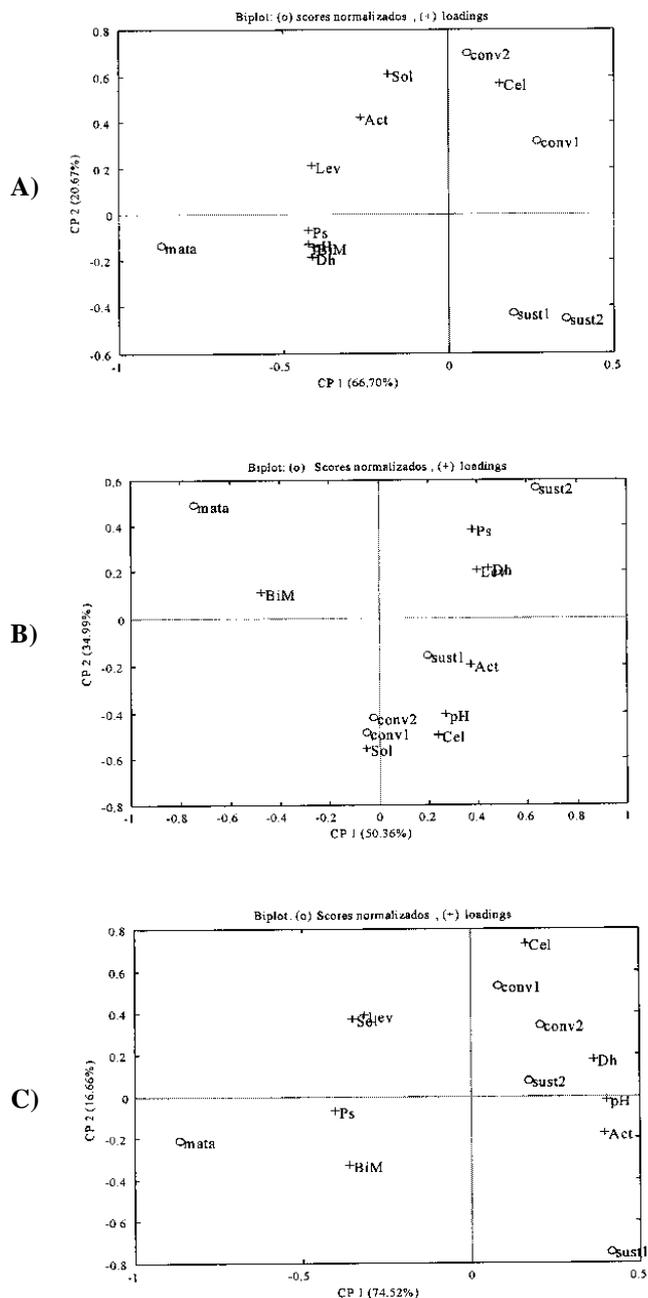
Em todos os casos, o primeiro fator a ser modelado foi a separação da faixa de mata (valores negativos em CP1) das demais faixas (valores positivos na CP1). As faixas de mata caracterizaram-se por uma alta biomassa microbiana (BiM) e uma baixa composição de celulolíticos (Cel). A CP2 foi interpretada como sendo um fator de separação entre as faixas de sistema sustentável e convencional. Os parâmetros bioquímicos, polissacarídeos (Ps), desidrogenase (Dh) e biomassa microbiana (BiM), se correlacionam positivamente com as faixas de tratamento



**Figura 3.** Resultado da Análise de Componentes Principais (PCA) mostrando as duas primeiras Componentes Principais para os dados de Suzano: (A) Gráfico dos “scores”; (B) Gráfico dos “loadings” e (C) Gráfico “Biplot”.

sustentável, enquanto microrganismos celulolíticos e solubilizadores (Sol) correlacionam-se negativamente. A Fazenda Macaúba apresentou a melhor separação entre faixas de sistemas sustentáveis e convencionais. Em todos os casos, a CP3 (que explicava entre 6 e 11% da variância) foi descartada como sendo um fator de diferenciação entre faixas de mesmo tratamento. A PCA permitiu portanto, distinguir quais as variáveis ou fatores que mais influenciaram na separação entre amostras de solos submetidas a diferentes tipos de manejo.

Um exemplo importante deste tipo de aplicação encontrado na literatura é o artigo de Stenberg *et al.*<sup>12</sup>. Nele, as variáveis químicas, físicas e biológicas foram medidas para 26 amostras de diferentes solos cultivados, abrangendo todo o território da Suécia. A aplicação da PCA possibilitou a identificação das principais tendências nos dados (de acordo com as PCs) e relacionou-as com três fatores: matéria orgânica, pH e atividade microbiana. Além disso, foi possível visualizar a separação das amostras em função da granulometria, do tipo de cultura do solo e de sua localização geográfica.



**Figura 4.** Resultado da Análise de Componentes Principais (PCA) mostrando "Biplots" das duas primeiras Componentes Principais para os dados de Guaíra: (A) Fazenda Macaúba; (B) Fazenda Lagoa do Fogo e (C) Fazenda Cuiabanos.

Muitos trabalhos recentes procuraram aplicar esse tipo de estratégia para avaliar a fertilidade do solo e comparar as condições de manejo. Estes estudos permitiram por exemplo, classificar solos de acordo com a sua fertilidade<sup>13</sup>, comparar nutrientes de plantas e solos<sup>14</sup> e determinar os fatores que controlam a mineralização de nitrogênio em solos úmidos<sup>15</sup>, a denitrificação em solos cultivados e não cultivados<sup>16</sup> e a deterioração de solos semi-áridos pelo uso<sup>17</sup>. E mais, relacionar propriedades físicas com a formação de turfa, visando melhorar a qualidade do solo<sup>18</sup>; e constatar que a composição da solução do solo é a principal responsável pela acidificação de solos na região meio-oeste da Suécia<sup>19</sup>. Outra aplicação interessante foi o estudo dos parâmetros que controlam a atividade de microrganismos benéficos, visando estabelecer um controle biológico<sup>20</sup>.

A utilização de modernas técnicas analíticas instrumentais

gera dados com grande potencial de informação abrindo novos horizontes para as estratégias multivariadas. A Espectroscopia no Infravermelho (IV) é a técnica mais usada, principalmente na determinação de MO. Em análises exploratórias, a PCA pode ser usada para a seleção de bandas espectrais visando a classificação do solo de acordo com o teor de MO<sup>21</sup>, ou na seleção de amostras de solo que tenham grande variabilidade de propriedades, partindo de uma grande população<sup>22</sup>. A Espectrometria de Massas com Pirólise (EM-P) é outra opção de técnica instrumental, mais adequada ao estudo de solos com baixo teor de carbono orgânico. Schulten e Hempfling<sup>23</sup> aplicaram PCA a dados de EM-P visando à classificação de solos sob diferentes condições de manejo, em função da composição húmica.

### Estudo de poluentes no solo

O outro grande objetivo das estratégias multivariadas exploratórias é o estudo da contaminação de solos por metais pesados e outros poluentes.

As CPs (ou fatores) obtidas na PCA podem ser usadas para caracterizar fontes poluentes de metais pesados<sup>24-27</sup>. Uma boa ilustração desta aplicação é o artigo de Krieg e Einax<sup>24</sup>, que estudaram este tipo de contaminação em diferentes locais da Alemanha, próximos a indústrias. Neste trabalho, um modelo com duas CPs explicou 92,4% da variância total. Nos gráficos de "scores" as duas primeiras CPs foram traçadas em função da profundidade do solo. A primeira CP foi identificada como um fator de contaminação antropogênica (Cu, Pb, Zn e MO), diminuindo com a profundidade, enquanto a segunda CP foi associada a fontes geogênicas (Na, K e Sr), aumentando com a profundidade.

A Análise de Fatores do Modo Q<sup>7,28</sup>, uma variante da PCA mais indicada para descrever relações entre objetos, também pode ser usada na caracterização da distribuição espacial de metais pesados em solos<sup>29</sup>. Outro método de grande potencial é o Tucker<sup>30</sup>, uma generalização da PCA para dados de ordem superior. Esse tipo de método permite tratar matrizes de três dimensões ou mais. As vantagens são evidentes, pois existe a possibilidade do estudo simultâneo de muitas variáveis para muitas amostras sob a variação de alguma outra condição (profundidade, tempo, etc.). Contudo, esta metodologia parece ainda ser desconhecida por muitos dos pesquisadores de solos. O único artigo<sup>31</sup> encontrado na literatura, descreve o uso do Tucker3, seguido de rotação dos fatores, no estudo de 6 metais contaminantes em solos de 4 locais, sob 6 tipos diferentes de cultura.

Finalizando esta seção menciona-se ainda que, a PCA, além de ser usada no estudo direto dos contaminantes no solo, pode ser útil na escolha das melhores condições para métodos de determinação destes mesmos contaminantes. Kokot *et al.*<sup>32</sup> compararam qual o melhor dentre 18 extratores, para a digestão em forno de microondas, de 5 metais determinados por Análise de Absorção Atômica (AAS). Pyle *et al.*<sup>33</sup> estudaram qual a melhor dentre 4 técnicas analíticas instrumentais para a determinação de chumbo e cádmio em 50 amostras de solo.

Além de metais pesados, foram encontradas na literatura aplicações de métodos quimiométricos no estudo de compostos aromáticos<sup>34</sup>, pesticidas<sup>35,36</sup> e radionuclídeos<sup>37</sup> em solos.

### MÉTODOS DE CALIBRAÇÃO MULTIVARIADA

#### Regressão Linear Múltipla (RLM), PCR e PLS<sup>4,11,38</sup>

A Calibração Multivariada tem como princípio básico a utilização simultânea de muitas variáveis  $x_1, x_2, \dots, x_n$  (como valores de absorbância a vários comprimentos de onda), para quantificar alguma outra variável de interesse  $y$  (como concentração).

O método mais simples de calibração multivariada é a Regressão Linear Múltipla (RLM). Na RLM a variável a ser prevista,  $y$ , é estimada por uma combinação linear das variáveis independentes contidas na matriz  $X$ ,

$$\mathbf{Y} = \mathbf{X}\mathbf{b} \text{ ou } \mathbf{y} = \mathbf{b}_1\mathbf{x}_1 + \mathbf{b}_2\mathbf{x}_2 + \dots + \mathbf{b}_n\mathbf{x}_n \quad (4)$$

onde  $\mathbf{b}$  é o vetor com os coeficientes de regressão. A solução de mínimos quadrados para encontrar  $\mathbf{b}$  é:  $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ .

A RLM é o melhor método para sistemas bem comportados (respostas lineares, sem interferentes, sem interações entre analitos, com baixo ruído e nenhuma colinearidade). A limitação da RLM está no fato de ela usar toda a informação contida na matriz  $\mathbf{X}$ , não importando se é relevante ou não, para construir o modelo. Com isso, quantidade significativa de variância (informação) irrelevante é incorporada. Outro problema está na etapa de inversão da matriz  $(\mathbf{X}'\mathbf{X})$ . Se a matriz  $\mathbf{X}$  contiver menos linhas que colunas (menos amostras que variáveis), ela não poderá ser invertida e a solução não existirá.

O PCR ("Principal Components Regression"), ou Regressão em Componentes Principais, é um método que fornece uma maneira de superar os problemas descritos. A regressão, ao invés de ser feita nas variáveis originais, é feita nos "scores" obtidos em uma PCA.

O PLS ("Partial Least Squares"), ou Mínimos Quadrados Parciais, é o método mais usado em calibração multivariada e difere do PCR por usar a informação de  $\mathbf{y}$  no cálculo das chamadas variáveis latentes (equivalentes às CPs). As matrizes  $\mathbf{X}$  e  $\mathbf{Y}$  são decompostas simultaneamente em uma soma de "h" variáveis latentes, como nas equações a seguir:

$$\mathbf{X} = \mathbf{T}\mathbf{P}' + \mathbf{E} = \sum \mathbf{t}_h\mathbf{p}'_h + \mathbf{E} \quad (5)$$

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}' + \mathbf{F} = \sum \mathbf{u}_h\mathbf{q}'_h + \mathbf{F} \quad (6)$$

onde  $\mathbf{T}$  e  $\mathbf{U}$  são as matrizes de "scores" das matrizes  $\mathbf{X}$  e  $\mathbf{Y}$ , respectivamente;  $\mathbf{P}$  e  $\mathbf{Q}$  são as matrizes de "loadings" das matrizes  $\mathbf{X}$  e  $\mathbf{Y}$ , respectivamente; e  $\mathbf{E}$  e  $\mathbf{F}$  são os resíduos. A correlação entre os dois blocos  $\mathbf{X}$  e  $\mathbf{Y}$  é simplesmente uma relação linear obtida pelo coeficiente de regressão linear, tal como descrito abaixo:

$$\mathbf{u}_h = \mathbf{b}_h\mathbf{t}_h \quad (7)$$

para "h" variáveis latentes, sendo que os valores de  $\mathbf{b}_h$  são agrupados na matriz diagonal  $\mathbf{B}$ , que contém os coeficientes de regressão entre a matriz de "scores"  $\mathbf{U}$  de  $\mathbf{Y}$  e a matriz de "scores"  $\mathbf{T}$  de  $\mathbf{X}$ . A melhor relação linear possível entre os "scores" desses dois blocos é obtida através de pequenas rotações das variáveis latentes dos blocos de  $\mathbf{X}$  e  $\mathbf{Y}$ .

A matriz  $\mathbf{Y}$  pode ser calculada de  $\mathbf{u}_h$ ,

$$\mathbf{Y} = \mathbf{T}\mathbf{B}\mathbf{Q}' + \mathbf{F} \quad (8)$$

e a concentração de novas amostras prevista a partir dos novos "scores",  $\mathbf{T}^*$ , substituídos na equação (8)

$$\mathbf{Y} = \mathbf{T}^*\mathbf{B}\mathbf{Q}' \quad (9)$$

Nesse processo é necessário achar o melhor número de variáveis latentes, o que normalmente é feito por um procedimento chamado validação cruzada ("cross validation"), no qual o erro mínimo de previsão é determinado. Existe ainda a diferenciação entre PLS1, em que a regressão é feita para uma variável dependente de cada vez (a matriz  $\mathbf{Y}$  é um vetor coluna), e PLS2, onde todas são calculadas simultaneamente.

## APLICAÇÕES DE RLM E PLS EM ANÁLISE DE SOLOS

### Determinação de matéria orgânica e outras propriedades do solo por Infravermelho

O uso da Espectroscopia de Refletância Difusa no Infravermelho com Transformada de Fourier, DRIFT ("Diffuse Reflectance Fourier Transformed Infrared")<sup>39</sup>, combinada com

métodos de calibração multivariada para prever MO e outras variáveis do solo, merece grande destaque. O objetivo é substituir as técnicas tradicionais de análise via-úmida<sup>40</sup>, mais caras e trabalhosas. A determinação de MO, por exemplo, é feita por oxidação com Cr(VI), baseada no Método de Walkey & Black<sup>40</sup>. Este método consome muito tempo, utiliza reagentes tóxicos e produz resíduos que podem agredir o meio ambiente. A DRIFT é um método não destrutivo (a amostra não é consumida) mais simples, "limpo" e barato. Os espectros são calibrados contra as determinações feitas por métodos tradicionais.

Sudduth e Hummel<sup>41</sup> forneceram um histórico das tentativas feitas para correlacionar propriedades espectrais de solos com MO. Os primeiros estudos datam do final dos anos 60/começo dos 70. O artigo de Condit<sup>42</sup> (1970) é citado como a primeira aplicação de quimiometria (PCA) a espectros de refletância.

Entretanto, a falta de um conhecimento mais amplo dos métodos quimiométricos tem limitado as aplicações da DRIFT. Ainda recentemente, vários autores<sup>43-46</sup> têm procurado caracterizar MO, substâncias húmicas e outras propriedades do solo, de forma apenas qualitativa, a partir de simples correlações ou proporções entre picos espectrais. Niemeier *et al.*<sup>46</sup> afirmaram não ser possível quantificar MO a partir de simples intensidade de picos.

Em 1983, Wetzel<sup>39</sup> descrevia o potencial da DRIFT para vários tipos de análise e a classificava como uma "adormecida entre as técnicas espectrais". Ele sugeria o uso de calibração multivariada através de RLM, com seleção de variáveis baseada nas informações espectrais disponíveis para o analito. Apesar de representarem um avanço, as tentativas de aplicação da RLM também se mostraram limitadas. Dalal e Henry<sup>47</sup> determinaram umidade, MO e nitrogênio total no IV próximo a partir de uma seleção de apenas três variáveis. O procedimento é trabalhoso porque várias combinações de variáveis são testadas para se escolher o melhor modelo. Além da limitação do número de variáveis, a RLM, quando utilizada para análise espectral, tende a incorporar quantidade significativa de informação desnecessária (devido à alta correlação entre as variáveis) levando a modelos pouco robustos.

O artigo de Ben-Dor e Banin<sup>48</sup> é um bom exemplo dessa limitação na análise. Eles procuraram prever 6 propriedades (MO, argila, etc.) de solos semi-áridos de Israel, no IV próximo. Para isto foi usada uma compressão espectral que reduzia as mais de 3000 bandas medidas a conjuntos que variavam entre 25 e 300 novas bandas (o espectro era dividido e as novas bandas calculadas como simples médias). Foram testados vários tamanhos de resolução do espectro e para cada propriedade o melhor modelo foi selecionado. O coeficiente de correlação ( $r$ ) para o conjunto de amostras previstas versus o seu valor calculado pelo método de referência variou entre 0,71 e 0,83. Além da previsão, foi realizada uma análise qualitativa com base nas atribuições espectrais. Os solos foram classificados em 11 grupos e as equações para cada modelo de RLM foram interpretadas. Os autores deste artigo claramente desconheciam o potencial de técnicas como PCA (para interpretação de tendências nos dados) e PLS (para previsão).

A utilização do PLS juntamente com a DRIFT<sup>41,49-56</sup>, no IV próximo ou médio, representa o último avanço na previsão de propriedades do solo. Sudduth e Hummel<sup>41</sup> construíram modelos com RLM, PCR e PLS, para prever MO no IV próximo, e concluíram que o PLS fornecia os melhores resultados. Baseados nisto, eles propuseram a construção<sup>50</sup> e a utilização<sup>51</sup> de um espectrofotômetro portátil para medidas de campo.

Masserschmidt *et al.*<sup>52</sup> obtiveram ótimos resultados para a previsão de MO no IV médio. Este artigo chama a atenção para a importância da escolha da região espectral correta, a

fim de eliminar interferências espectrais drásticas, e para o pré-tratamento dos dados. Os melhores resultados foram obtidos com alisamento seguido de derivação (primeira derivada) e normalização dos dados (erro de previsão de 0,63 e  $r = 0,98$ ). Fidêncio *et al.*<sup>53</sup> usaram estratégia semelhante para a previsão de MO (na faixa entre 0,49 e 3,77%) no IV próximo, em amostras de solos fornecidas pelo Instituto Agronômico de Campinas (IAC). Neste trabalho foi feito um pré-tratamento dos dados utilizando filtro digital e derivada de segunda ordem, e então, construído um modelo PLS (erro de previsão de 0,21 e  $r = 0,94$ ). Este modelo é mostrado na Figura 5 (MO determinada pelo método via úmida versus MO prevista utilizando espectros no IV próximo e calibração multivariada).

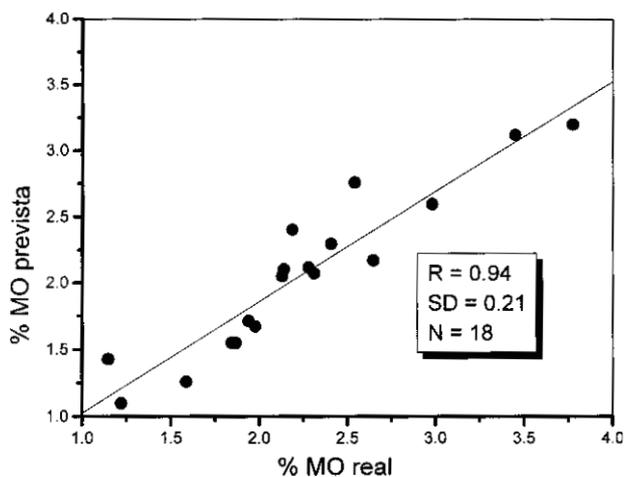


Figura 5. Calibração Multivariada. Porcentagem de Matéria Orgânica (MO) medida pelo método de referência (MO real) versus porcentagem de MO prevista por PLS. Amostras de solos fornecidas pelo IAC.

Janik *et al.*<sup>54</sup> também obtiveram bons resultados para a previsão de elementos no IV médio, em amostras de solo da Austrália. Os elementos foram determinados por Fluorescência de Raios-X. Os autores demonstraram as vantagens da DRIFT sobre a Fluorescência de Raios-X, a qual requer uma etapa complexa de preparação das amostras, apesar de ser uma técnica rápida e permitir determinar vários elementos simultaneamente. Os resultados foram considerados bons para a previsão de silício ( $r=0,99$ ), alumínio ( $r=0,96$ ) e ferro ( $r=0,97$ ), e insatisfatórios para titânio ( $r=0,88$ ), magnésio ( $r=0,89$ ) e cálcio ( $r=0,85$ ).

Em um segundo artigo, os mesmos autores<sup>55</sup> tentaram prever outras propriedades de solos medidas em laboratório. Os resultados foram considerados, levando-se em conta o erro de previsão e o coeficiente de correlação, ruins para soma de cátions e conteúdo de argila, razoáveis para MO, nitrogênio e pH, e bons para carbonato. O uso de PLS localmente linear (os dados são divididos em faixas de composição e para cada uma é construído um modelo) melhorou as previsões. Através da observação dos "loadings" das variáveis latentes foi possível extrair mais informações sobre o modelo, estabelecendo correlações entre variáveis e atribuindo bandas. Regiões espectrais com maior informação apareceram com valores positivos nos "loadings" das primeiras variáveis latentes (CPs), enquanto interferentes apareceram com valores negativos.

Variáveis biológicas, geralmente medidas com uma imprecisão bem maior do que variáveis químicas, também podem ser previstas a partir do IV próximo. Para isso, Hämäläinen e Albano<sup>56</sup> propuseram uma complexa estratégia multivariada envolvendo PCA e PLS.

## Determinação de pesticidas em solos

Vários trabalhos (todos de autores espanhóis) propuseram a determinação por calibração multivariada de misturas de pesticidas e seus produtos de transformação no solo. Neles, o PLS foi usado em dados de Fluorescência de Excitação<sup>57</sup>, Espectrofotometria no UV/Visível<sup>58</sup>, Cromatografia Líquida de Alta Eficiência (CLAE)<sup>59</sup> e CLAE com detecção no UV por arranjo de diodos (CLAE-UV)<sup>60-62</sup>. O objetivo foi, sempre, determinar simultaneamente espécies que apresentavam picos sobrepostos. A grande vantagem desta metodologia, no caso da espectroscopia, é não necessitar da aplicação prévia de métodos de separação. Comparativamente, a seleção das melhores condições de análise (fase estacionária, fase móvel, vazão) em CLAE é um procedimento trabalhoso e que consome muito tempo.

Todas as aplicações foram teóricas. O modelo, depois de construído, foi testado em misturas sintéticas e, em amostras de água e de solo, "contaminadas" com quantidades conhecidas das espécies a serem determinadas. No final, a porcentagem de recuperação foi calculada. Não foram determinadas amostras reais.

Os estudos comparativos entre PLS1, PLS2 e PCR estão descritos nos vários artigos e, de um modo geral, o PLS1 foi superior. Na maioria dos casos, os resultados foram melhorados com o uso de alisamento e derivadas no pré-tratamento dos dados. O uso de derivadas pode reduzir problemas de linha base e diminuir (mas não eliminar) o efeito de interferentes<sup>62</sup>. A porcentagem de recuperação para cada espécie variou entre 85 e 115%, na grande maioria das amostras de solo.

Em alguns artigos<sup>60-62</sup> o potencial de uma técnica como a CLAE-UV, que gera simultaneamente dados em duas dimensões (superfícies), não foi completamente explorado. Foram propostos diferentes modelos que não levaram em conta toda a sinergia dos dados (ex: ao invés de analisar conjuntamente os espectros e os cromatogramas gerados, foram comparados modelos construídos somente com dados cromatográficos em diferentes comprimentos de onda<sup>60</sup>). Técnicas hífenadas, como CLAE-UV ou Fluorescência de Emissão/Excitação, geram dados mais adequados ao tratamento por métodos de ordem superior. Estes métodos consideram todas as correlações entre os dados, nas diferentes dimensões. Entre eles estão o GRAM ("Generalized Rank Annihilation Method")<sup>63</sup>, o PARAFAC<sup>64</sup> e o N-Way PLS<sup>65</sup>. O PARAFAC é um método de decomposição para dados de ordem superior (multidimensionais) que pode ser considerado uma versão mais restrita do Tucker3 (ver seção II.2.2), fornecendo solução única (independente de rotação). Esta última característica o torna muito útil na modelagem de dados espectrais e/ou cromatográficos. O N-Way PLS é um método de calibração para esse tipo de dados.

## Outras aplicações de RLM e PLS

Esta seção apresenta aplicações de RLM e PLS, que possuem um caráter mais qualitativo do que quantitativo. Em nenhuma delas buscou-se construir um modelo de previsão de rotina. O objetivo maior foi sempre estudar a influência de algumas variáveis independentes (x) em alguma variável dependente (y). Em vários casos, a PCA poderia ser sugerida como um método mais adequado para esse tipo de estudo.

A RLM foi usada para prever a correlação de parâmetros do solo com a concentração de selênio em alfafa<sup>66</sup> e com a adsorção de cádmio no solo<sup>67</sup>; e também a correlação dos fatores que controlam a extração de triazinas do solo por fluido super-crítico<sup>68</sup>, em um experimento planejado. Robertson<sup>69,70</sup> usou RLM e PLS para estudar os fatores do solo responsáveis pela emissão de N<sub>2</sub>O para a atmosfera, mas não obteve bons resultados. Em uma das mais interessantes aplicações de RLM<sup>(71)</sup>, foi estabelecida a influência das diversas frações de enxofre extraídas do solo sobre a concentração de enxofre em plantas. Isso permitiu determinar em quais frações o enxofre estaria mais disponível para ser absorvido pela planta.

Palmborg *et al.*<sup>72</sup> usaram o PLS em uma análise exploratória para estudar a correlação entre variáveis, tais como metais pesados, nutrientes e parâmetros de qualidade da MO, com a taxa de respiração do solo (um parâmetro da atividade microbiana). As variáveis latentes obtidas (ao contrário das CPs da PCA) foram influenciadas diretamente pela variável de maior interesse (taxa de respiração do solo).

## ANÁLISE EVOLUCIONÁRIA DE FATORES E SIMPLISMA

### Análise Evolucionária de Fatores (EFA) e Simplisma

A Análise Evolucionária de Fatores (EFA, "Evolving Factor Analysis")<sup>28,73</sup> é um método que acompanha a evolução do posto de uma matriz de dados  $\mathbf{X}$ , em função de alguma variação (ex: tempo). O posto define o número de vetores linearmente independentes da matriz e, portanto o número de componentes do sistema (CPs). O acompanhamento é feito examinando as submatrizes  $\mathbf{X}_i$ , de  $\mathbf{X}$ , formadas pelos primeiros  $i$  resultados (ex: espectros ou cromatogramas). À medida que uma nova espécie surge, o posto de  $\mathbf{X}$  aumenta de uma unidade, podendo ser acompanhado graficamente. Uma varredura no sentido inverso permite o conhecimento do ponto em que cada espécie desaparece. A Figura 6 ilustra o acompanhamento do posto em uma eluição cromatográfica com picos sobrepostos.

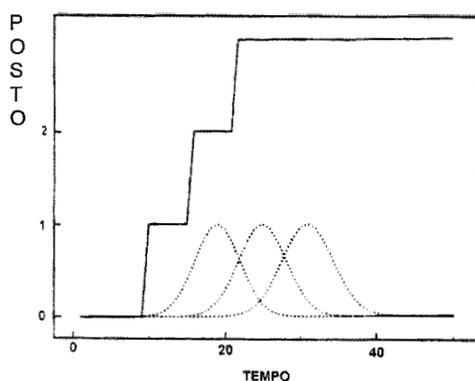


Figura 6. Evolução do posto ("rank") de uma matriz de dados coletados em um processo de eluição cromatográfica.

A EFA secundária<sup>74</sup> é um método iterativo que, partindo de uma matriz discriminante, permite decompor uma matriz de dados de misturas em, por exemplo, espectros puros e perfis de concentração.

O SIMPLISMA ("SIMPLe-to-use Interactive Self-modelling Mixture Analysis")<sup>75</sup> é um procedimento para a decomposição da matriz baseado na determinação de variáveis puras (ele não utiliza CPs). Uma variável pura é, por exemplo, um comprimento de onda em que o valor de intensidade depende apenas de um dos componentes da mistura. O método usa diagramas de variância e critérios de dissimilaridade para encontrar as variáveis puras. Em seguida elas são extraídas uma a uma da matriz original, até que esta contenha só ruído.

### Aplicação de EFA e SIMPLISMA em Análise de Solos

Uma série de artigos<sup>74,76-88</sup> de um mesmo grupo de pesquisa (português) vem utilizando EFA e SIMPLISMA no estudo das interações que ocorrem no solo entre ácidos fúlvicos (AFs) e íons metálicos ou  $H^+$ . Os AFs são a parte das substâncias húmicas de menor peso molecular. Eles são misturas complexas de macromoléculas com diferentes constituições e fornecem espectros difíceis de serem interpretados. Existe muito

interesse no seu estudo porque eles atuam, por complexação, no transporte de nutrientes e de elementos tóxicos<sup>3</sup>.

Em todos os artigos, AFs extraídos do solo foram titulados potenciométricamente com metais ou base. O objetivo foi caracterizar sítios de ligação com metais ou sítios ácidos. A titulação foi acompanhada por espectros de Fluorescência Sincronizada ("Synchronous Fluorescence" - SyF). A SyF permite um aumento da resolução e da quantidade de informação extraídas em relação à Emissão e à Excitação. A EFA foi usada para saber quantos são os tipos de sítios ligantes e em que ponto cada um está sendo complexado. A EFA secundária<sup>74,76-84</sup> ou o SIMPLISMA<sup>85-88</sup> foram usados para obter os espectros e os perfis de concentração dos componentes. Os experimentos foram feitos em pH constante ou variável e os grupos de sítios ligantes foram caracterizados espectralmente e, às vezes classificados. Em alguns casos foram calculadas as constantes de estabilidade condicional dos complexos.

Em um dos artigos<sup>76</sup> foi estudada a complexação entre o íon Al(III) e AFs (em pH=4) provenientes do solo e de plantas em decomposição. O Al(III) é importante por ser fitotóxico e formar complexos estáveis com AFs. Através da EFA foram detectados quatro componentes (tipos de sítios ligantes) nas amostras de solo. Uma Análise de Clusters (HCA)<sup>89</sup> permitiu agrupar os tipos componentes dentro do conjunto das amostras. Um componente comum a todas as amostras foi interpretado como sendo do tipo ácido salicílico.

Em outros artigos, além do alumínio<sup>74,84,88</sup> foram estudados berílio<sup>77,82</sup>, cobre<sup>80,81,87</sup>, níquel<sup>80</sup>, cobalto<sup>80</sup> e o íon urânio ( $UO_2^{2+}$ )<sup>86</sup>. Um artigo<sup>83</sup> interessante estudou as propriedades de AFs provenientes de lodo de esgoto, visando aplicá-lo como fertilizante. O uso dos métodos quimiométricos permitiu avaliar uma série de aspectos físico-químicos destes sistemas.

## MÉTODOS DE INTELIGÊNCIA ARTIFICIAL

Os métodos de inteligência artificial são baseados no desenvolvimento de sistemas computacionais inteligentes, isto é, que exibem características associadas à inteligência no comportamento humano<sup>6</sup>. Entre eles podem ser citados Redes Neurais Artificiais, Algoritmos Genéticos e Lógica Difusa (Lógica "Fuzzy"). Todas as aplicações deste tipo, encontradas em análise de solos, referem-se a Redes Neurais.

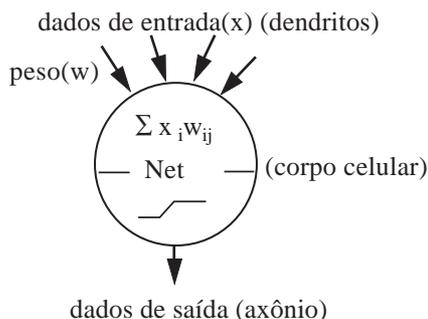
### Redes Neurais Artificiais<sup>6,90</sup>

As Redes Neurais Artificiais vêm despertando um interesse crescente em muitas áreas da química. Elas podem ser aplicadas como um método de calibração multivariada, eficiente na modelagem de sistemas analíticos complexos, ou como um método de classificação.

O objetivo da rede neural é estabelecer uma relação, linear ou não, entre um conjunto de dados de entrada e uma correspondente saída ou resposta. Para isto são utilizadas transformações matemáticas na entrada, de modo a produzir a saída desejada. A rede realiza este trabalho "aprendendo" como o sistema em estudo se comporta, e então, aplica este conhecimento a novos dados de entrada para prever a saída apropriada.

O seu comportamento é baseado no funcionamento do cérebro humano, no qual o sistema nervoso detecta estímulos (informações de entrada) e os envia para o cérebro, através dos neurônios. Nos neurônios, a informação (recebida através dos dendritos) é processada e uma resposta é fornecida através dos axônios. Em um neurônio artificial, as interconexões de entrada (interconexões peso) correspondem aos dendritos; os cálculos efetuados correspondem ao que ocorre no corpo celular de um neurônio biológico; e as interconexões de saída correspondem aos axônios. Os dados de entrada,  $\mathbf{x}$ , são multiplicados pelos respectivos pesos,  $\mathbf{w}$ , fornecendo a função somatória  $Net_j = \sum \mathbf{x}_i \mathbf{w}_{ji}$ . A saída do

neurônio é obtida a partir de uma função de transferência (linear, sigmoide, etc., dependendo do comportamento dos dados) aplicada sobre  $\text{Net}_j$ . A Figura 7 representa os componentes de um neurônio artificial.



**Figura 7.** Representação dos componentes de um neurônio artificial: entradas e interconexões peso (dendritos), núcleo com função de transferência (corpo celular) e saída (axônio).

A seqüência de entradas e saídas repete-se tantas vezes quantas forem as camadas de neurônios de uma rede. O conjunto de neurônios forma a Rede Neural com uma camada de entrada, uma ou mais camadas internas ou escondidas, e uma camada de saída. O número de neurônios assim como o de camadas, é determinado por tentativa e erro, de forma a minimizar os erros na saída da rede.

O treinamento da Rede Neural pode ser dividido em duas fases: uma de aprendizagem e outra de validação. Na primeira fase, um conjunto conhecido de dados de entrada  $\mathbf{X}$  e saída  $\mathbf{Y}$ , é apresentado à rede. Em seguida os pesos (estabelecidos aleatoriamente no início) entre os neurônios são ajustados até a entrada produzir a saída desejada. Este ajuste é realizado através de métodos de otimização dos pesos em função dos erros, dos quais o mais utilizado é a Retropropagação dos Erros.

### Aplicações de Redes Neurais Artificiais em Análise de Solos

As Redes Neurais têm sido utilizadas em análise de solos, tanto como método de previsão, quanto como método de classificação. O método de correção dos pesos mais empregado vem sendo a Retropropagação dos Erros, em que os erros são corrigidos a partir da última até a primeira camada.

Chao *et al.*<sup>91</sup> construíram um modelo não linear para prever o coeficiente de absorção de poluentes orgânicos no solo, usando Redes Neurais. Esses coeficientes normalmente são calculados por uma relação linear logarítmica com algum parâmetro do composto orgânico. Um conjunto de 119 compostos foi usado para “treinar” a rede, enquanto outros 14 compostos foram usados na previsão. Valores de coeficientes tabelados foram usados no treinamento. A rede apresentou uma arquitetura 3-2-1, com três neurônios na camada de entrada, dois na camada intermediária e um na camada de saída. Os logaritmos dos parâmetros solubilidade em água e coeficiente de partição octanol/água, além do “bias” (o “bias” é uma constante adicionada à função de transferência de cada camada, deslocando-a para um melhor ajuste dos pesos) foram usados como entradas. O neurônio de saída forneceu o log do coeficiente de absorção. A performance do modelo não linear fornecido pela Rede Neural foi comparada com modelos lineares (regressões contra cada parâmetro e contra os dois conjuntamente). A Rede Neural forneceu os melhores resultados: menor soma de resíduos e menor desvio na previsão das amostras.

As Redes Neurais foram usadas também em aplicações mais próximas as áreas de agronomia e engenharia. Foram construídos modelos de simulação da concentração de pesticidas em

solos<sup>92</sup>, e modelos de aclimação microbiana em solos contaminados com hidrocarbonetos de petróleo<sup>93</sup>. Nessas aplicações, as redes empregam um número bem maior de neurônios e camadas internas do que nas aplicações usuais em química.

Para utilizar as redes como método de classificação, os neurônios de saída são treinados para fornecer respostas binárias (1 quando pertence à classe, e 0 quando não pertence). Baseado nisso, Levine *et al.*<sup>94</sup> classificaram uma grande quantidade de solos em três categorias, de acordo com sua granulometria. Conteúdos de silte, argila e carbono orgânico foram usados como entradas. Em outra aplicação<sup>95</sup>, microrganismos do solo foram corretamente classificados a partir de seus espectros de massas.

Song *et al.*<sup>96</sup> propuseram um novo tipo de arquitetura para Redes Neurais baseada em outros métodos de inteligência artificial, a “Teoria da Ressonância Adaptada” e a Lógica Difusa. O objetivo é criar um método que seja ao mesmo tempo robusto e flexível. Este método foi usado com sucesso na classificação de amostras de ácidos graxos extraídas do solo. As amostras foram classificadas de acordo com a cultura de seu solo de origem. A nova rede foi comparada com a tradicional Rede Neural com Retropropagação de Erros, se mostrando superior por fornecer previsões melhores e apresentar um processo de treinamento mais rápido.

Em alguns destes trabalhos, as Redes Neurais foram superiores a métodos de previsão e classificação mais comuns, devendo ser destacada a sua importância na modelagem de sistemas não lineares. Redes que utilizam um número excessivo de neurônios podem tornar-se muito rígidas, memorizando apenas a informação para a qual foram treinadas. Mas como em qualquer aplicação, não existe um método ideal, sendo sempre necessário examinar as condições para se escolher o melhor. De qualquer maneira, é possível afirmar que a utilização de Redes Neurais, tanto na análise de solos quanto na química, possui um amplo horizonte pela frente.

Encerrando esta seção, destaca-se a possibilidade de aplicação de Algoritmos Genéticos em análise de solos. Este método poderia ser usado, por exemplo, na seleção de variáveis de dados espectroscópicos em calibração multivariada<sup>97-99</sup>.

### CONCLUSÃO

Este trabalho apresentou uma visão ampla e crítica da aplicação de métodos quimiométricos em análise de solos. Os tipos de aplicação foram divididos em quatro partes, de acordo com os métodos empregados. No entanto as aplicações contidas nas duas primeiras partes podem ser consideradas as de maior importância: a avaliação integrada de parâmetros químicos, físicos, biológicos e bioquímicos buscando a determinação de tendências e o reconhecimento de padrões; e a previsão de matéria orgânica e outras propriedades do solo a partir de métodos de calibração multivariada e técnicas instrumentais analíticas.

Foi sugerida, a partir do cenário atual, a utilização de novas técnicas como os Métodos de Ordem Superior e os Algoritmos Genéticos, que poderão ampliar ainda mais este campo de pesquisa. Por fim, é preciso destacar a necessidade de uma maior interdisciplinaridade na abordagem dos problemas do solo. Pesquisadores e profissionais das áreas, tais como agronomia, biologia, geologia, química, matemática, estatística e informática, devem manter um intercâmbio permanente para que seja sempre possível uma visão global e atualizada dos problemas e de suas soluções. A Quimiometria terá um papel de destaque na extração e na interpretação das informações.

### REFERÊNCIAS

1. Bruns, R. E.; Faigle, J. F. G.; *Quim. Nova* **1985**, *4*, 84.
2. Torstensson, L.; Pell, M.; Stenberg, B.; *Ambio* **1998**, *27*, 4.
3. Piccolo, A. (Ed.); *Humic Substances in Terrestrial Ecosystems*; Elsevier Science B. V.; Amsterdam, 1996.

4. Martens, H.; Naes, T.; *Multivariate Calibration*; John Wiley & Sons; New York, 1989; p 97-108.
5. Geladi, P.; Kowalski, B. R.; *Anal. Chim. Acta* **1986**, *185*, 1.
6. Zampronio, C. G.; *Dissertação de Mestrado*, UNICAMP, Campinas, 1998; p 30-40.
7. Jöreskog, K. G.; Klovan, J. E.; Reyment, R. A.; *Geological Factor Analysis*; Elsevier; Amsterdam, 1976; p 59, 86-115.
8. Forina, M.; Armanino, C.; Lanteri, S.; Leardi, R.; *J. Chemom.* **1988**, *3*, 115.
9. Sena, M. M.; Poppi, R. J.; Frighetto, R. T. S.; Valarini, P. J.; Livro de Resumos, AB-073, 22<sup>a</sup> R. A. da SBQ, Poços de Caldas, 1999.
10. Sena, M. M.; Poppi, R. J.; Frighetto, R. T. S.; Valarini, P. J.; 38<sup>o</sup> Congresso Brasileiro de Química, São Luís, 1998; p 107-108.
11. Wise, B. M.; Gallagher, N. B.; *PLS Toolbox 2.0*; Eigenvector Research Inc., Manson, 1998; p16-24.
12. Stenberg, B.; Pell, M.; Torstensson, L.; *Ambio* **1998**, *27*, 9.
13. Nolin, M. C.; Wang, C.; Caillier, M. J.; *Can. J. Soil Sci.* **1989**, *69*, 525.
14. Németh, T.; Borszédi, J.; Gegus, E.; *Can. J. Appl. Spectrosc.* **1993**, *38*, 121.
15. González-Prieto, S. J.; Villar, M. C.; Carballas, M.; Carballas, T.; *Plant Soil* **1992**, *144*, 31.
16. Horwath, W. R.; Elliot, L. F.; Steiner, J. J.; Davis, J. H.; Griffith, S. M.; *J. Environ. Qual.* **1998**, *27*, 225.
17. Quiroga, A. R.; Buschiazzo, D. E.; Peinemann, N.; *Soil Sci.* **1998**, *163*, 591.
18. McCoy, E. L.; *Agron. J.* **1998**, *90*, 411.
19. Lundstrom, U. S.; Nyberg, L.; Danielsson, R.; van Hees, P. A. W.; Andersson, M.; *Ambio* **1998**, *27*, 551.
20. Duffy, B. K.; Ownley, B. H.; Weller, D. M.; *Phytopathology* **1997**, *87*, 1118.
21. Henderson, T. L.; Szilagy, A.; Baumgardner, M. F.; Chen, C. T.; Landgrebe, D. A.; *Soil Sci. Soc. Am. J.* **1989**, *53*, 1778.
22. Stenberg, B.; Nordkvist, E.; Salomonsson, L.; *Soil Sci.* **1995**, *159*, 109.
23. Schulten, H. F.; Hempfling, R.; *Plant Soil* **1992**, *142*, 259.
24. Krieg, M.; Einax, J.; *Fresenius J. Anal. Chem.* **1994**, *348*, 490.
25. Davies, B. E.; *Wat. Air Soil Poll.* **1997**, *94*, 85.
26. Wang, P. X.; Qu, E. F.; Li, Z. B.; Shuman, L. M.; *J. Environ. Qual.* **1997**, *26*, 795.
27. Carlosena, A.; Andrade, J. M.; Prada, D.; *Talanta* **1998**, *47*, 753.
28. Sena, M. M.; *Dissertação de Mestrado*, UNICAMP, Campinas, 1996, p 17-32.
29. Juang, K. W.; Lee, D. Y.; *J. Environ. Qual.* **1998**, *27*, 355.
30. Kroonenberg, P. M.; *Three Mode Principal Component Analysis: Theory and Applications*; DSWO, Leiden, 1983.
31. Henrion, R.; *J. Chemom.* **1993**, *7*, 477.
32. Kokot, S.; King, G.; Keller, H. L.; Massart, D. L.; *Anal. Chim. Acta* **1992**, *259*, 267.
33. Pyle, S. M.; Nocerino, J. M.; Deming, S. N.; Palasota, J. A.; Palasot, J. M.; Miller, E. L.; Hillman, D. C.; Kuharic, C. A.; Cole, W. H.; Fitzpatrick, P. M.; Watson, M. A.; Nichols, K. D.; *Environ. Sci. Technol.* **1996**, *30*, 204.
34. Vogt, N. B.; Brakstad, F.; Thrane, K.; Nordenson, S.; Krane, J.; Aamot, E.; Kolset, K.; Esbensen, K.; Steinnes, E.; *Environ. Sci. Technol.* **1987**, *21*, 35.
35. Steinheimer, T. R.; Pfeiffer, R. L.; Scoggin, K. D.; *Anal. Chem.* **1994**, *66*, 645.
36. Almendros, G.; *Eur. J. Soil Sci.*; **1995**, *46*, 287.
37. Elejalde, C.; Herranz, M.; Romero, F.; Legarda, F.; *Wat. Air Soil Poll.* **1996**, *89*, 23.
38. Beebe, K. R.; Kowalski, B. R.; *Anal. Chem.*; **1987**, *59*, 1007A.
39. Wetzel, D. L.; *Anal. Chem.* **1983**, *55*, 1165A.
40. Tomé Jr., J. B.; *Manual para a Interpretação de Análise de Solo*; Liv. e Ed. Agropecuária; Guaíba, 1997; p 59-87.
41. Sudduth, K. A.; Hummel, J. W.; *Trans. ASAE* **1991**, *34*, 1900.
42. Condit, H. R.; *Photogrammetric Eng.* **1970**, *36*, 955. Citado em: (39).
43. Arocena, J. M.; Pawluk, S.; Dudas, M. J.; Gajdostik, A.; *Can. J. Soil Sci.* **1995**, *75*, 327.
44. Wander, M. M.; Traina, S. J.; *Soil Sci. Soc. Am. J.* **1996**, *60*, 1087.
45. Kaiser, K.; Guggenberger, G.; Haumaier, L.; Zech, W.; *Eur. J. Soil Sci.* **1997**, *48*, 301.
46. Niemeyer, J.; Chen, Y.; Bollag, J. M.; *Soil Sci. Soc. Am. J.* **1992**, *56*, 135.
47. Dalal, R. C.; Henry, R. J.; *Soil Sci. Soc. Am. J.* **1986**, *50*, 120.
48. Ben-Dor, E.; Banin, A.; *Soil Sci. Soc. Am. J.* **1995**, *59*, 364.
49. Couillard, A.; Turgeon, A. J.; Shenk, J. S.; Westerhaus, M. O.; *Crop Sci.* **1997**, *37*, 1554.
50. Sudduth, K. A.; Hummel, J. W.; *Trans. ASAE* **1993**, *36*, 185.
51. Sudduth, K. A.; Hummel, J. W.; *Trans. ASAE* **1993**, *36*, 1571.
52. Masserschmidt, I.; Cuelbas, C. J.; Poppi, R. J.; Andrade, J. C.; Abreu, C. A.; Davanzo, C. U.; *J. Chemom.* **1999**, *13*, 265.
53. Fidêncio, P. H.; Poppi, R. J.; Cantarella, H.; Andrade, J. C.; *Uso de Espectroscopia NIR e Calibração Multivariada para a Determinação de Matéria Orgânica em Solos*; 1<sup>a</sup> Escola de Verão em Quimiometria na PUC-Rio, Rio de Janeiro, 1999; p 8.
54. Janik, L. J.; Skjemstand, J. O.; Raven, M. D.; *Aust. J. Soil Res.* **1995**, *33*, 621.
55. Janik, L. J.; Skjemstand, J. O.; *Aust. J. Soil Res.* **1995**, *33*, 637.
56. Hämäläinen, M.; Albano, C.; *J. Chemom.* **1992**, *6*, 41.
57. del Olmo, M.; Díez, C.; Molina, A.; de Orbe, I.; Vilchez, J. L.; *Anal. Chim. Acta* **1996**, *335*, 23.
58. Galera, M. M.; Vidal, J. L. M.; Frenich, A. G.; Parrilla, P.; *Analyst* **1994**, *119*, 1189.
59. Galera, M. M.; Vidal, J. L. M.; Frenich, A. G.; Garcia, M. D. G.; *J. Chromatogr. A* **1996**, *727*, 39.
60. Galera, M. M.; Vidal, J. L. M.; Frenich, A. G.; Garcia, M. D. G.; *J. Chromatogr. A* **1997**, *778*, 139.
61. Garcia, M. D. G.; Frenich, A. G.; Vidal, J. L. M.; Galera, M. M.; de la Peña A. M.; Salinas, F.; *Anal. Chim. Acta* **1997**, *348*, 177.
62. Vidal, J. L. M.; Garcia, M. D. G.; Galera, M. M.; Frenich, A. G.; *Anal. Lett.* **1997**, *30*, 2409.
63. Sánchez, E.; Kowalski, B. R.; *Anal. Chem.* **1986**, *58*, 496.
64. Bro, R.; *Chemom. Intell. Lab. Syst.* **1997**, *38*, 149.
65. Bro, R.; *J. Chemom.*; **1996**, *10*, 47.
66. Severson, R. C.; Gough, L. P.; *J. Environ. Qual.* **1992**, *21*, 353.
67. Bolton, K. A.; Evans, L. J.; *Can. J. Soil Sci.* **1996**, *76*, 183.
68. van der Velde, E. G.; Ramlal, M. R.; van Beuzekom, A. C.; Hoogerbrugge, R.; *J. Chromatogr. A* **1994**, *693*, 125.
69. Robertson, K.; *J. Environ. Qual.* **1994**, *23*, 805.
70. Robertson, K.; Klemmedtsson, L.; *Plant Soil* **1996**, *178*, 49.
71. Shen, X. Q.; Chen, B.; Zhang, T. H.; Li, F. L.; Wen, B.; Qian, J.; *Sci. Total Environ.* **1997**, *199*, 237.
72. Palmborg, C.; Bringmark, L.; Bringmark, E.; Nordgren, A.; *Ambio* **1998**, *27*, 53.
73. Maeder, M.; Zilian, A.; *Chemom. Intell. Lab. Syst.* **1988**, *3*, 205.
74. Silva, C. S. P. C. O.; da Silva, J. C. G. E.; Machado, A. A. S. C.; *Appl. Spectrosc.* **1994**, *48*, 363.

75. Windig, W.; Heckler, C. E.; Agblevor, F. A.; Evans, R. J.; *Chemom. Intell. Lab. Syst.* **1992**, *14*, 195.
76. da Silva, J. C. G. E.; Ferreira, M. A.; Machado, A. A. S. C.; Rey, F.; *Anal. Chim. Acta* **1996**, *333*, 71.
77. da Silva, J. C. G. E.; Machado, A. A. S. C.; Ramos, M. A.; Arce, F.; Rey, F.; *Environ. Sci. Technol.* **1996**, *30*, 3155.
78. da Silva, J. C. G. E.; Machado, A. A. S. C.; Silva, C. S. P. C. O.; *Anal. Chim. Acta* **1996**, *318*, 365.
79. da Silva, J. C. G. E.; Machado, A. A. S. C.; *Analyst* **1997**, *122*, 1299.
80. Machado, A. A. S. C.; da Silva, J. C. G. E.; *Chemom. Intell. Lab. Syst.* **1993**, *19*, 155.
81. da Silva, J. C. G. E.; Machado, A. A. S. C.; *Chemom. Intell. Lab. Syst.* **1995**, *27*, 115.
82. Machado, A. A. S. C.; da Silva, J. C. G. E.; Garcia, T. M. O.; *Appl. Spectrosc.* **1995**, *49*, 1500.
83. da Silva, J. C. G. E.; Machado, A. A. S. C.; Silva, M. A. B. A.; *Wat. Res.* **1998**, *32*, 441.
84. da Silva, J. C. G. E.; Machado, A. A. S. C.; Ramos, M. A.; Arce, F.; Rey, F.; *Environ. Toxicol. Chem.* **1997**, *16*, 1845.
85. da Silva, J. C. G. E.; Machado, A. A. S. C.; *Talanta* **1994**, *41*, 2095.
86. da Silva, J. C. G. E.; Machado, A. A. S. C.; Oliveira, C. J. S.; *Analyst* **1996**, *121*, 1373.
87. Machado, A. A. S. C.; da Silva, J. C. G. E.; Maia, J. A. C.; *Anal. Chim. Acta* **1994**, *292*, 121.
88. da Silva, J. C. G. E.; Machado, A. A. S. C.; Oliveira, C. J. S.; *Anal. Chim. Acta* **1997**, *349*, 23.
89. Bratchell, N.; *Chemom. Intell. Lab. Syst.* **1989**, *6*, 105.
90. Zupan, J.; Gasteiger, J.; *Neural Networks for Chemists: An Introduction*; VCH; New York, 1993; p 119-148.
91. Gao, C.; Govind, R.; Tabak, H. H.; *Environ. Toxicol. Chem.* **1996**, *15*, 1089.
92. Yang, C. C.; Prasher, S. O.; Sreekanth, S.; Patni, N. K.; Masse, L.; *Trans. ASAE* **1997**, *40*, 1285.
93. Suchorsky-Tremblay, A. M.; Kok, R.; *Can. Agr. Eng.* **1997**, *39*, 123.
94. Levine, E. R.; Kimes, D. S.; Sigillito, V. G.; *Ecol. Model.* **1996**, *92*, 101.
95. Chun, J.; Atalan, E.; Ward, A. C.; Goodfellow, M.; *FEMS Microbiol. Lett.* **1993**, *107*, 321.
96. Song, X. H.; Hopke, P. K.; Bruns, M. A.; Bossio, D. A.; Scow, K. M.; *Chemom. Intell. Lab. Syst.* **1998**, *41*, 161.
97. Lucasius, C. B.; Kateman, G.; *Chemom. Intell. Lab. Syst.* **1993**, *19*, 1.
98. Costa F<sup>o</sup>, P. A.; *Dissertação de Mestrado*, UNICAMP, Campinas, 1998.
99. Costa F<sup>o</sup>, P. A.; Poppi, R. J.; *Quim. Nova* **1999**, *22*, 405.