

Aplicação do algoritmo genético na análise de medicamentos anti-inflamatórios industrializados

Valdir Bruxel Junior (UNISC) hagnat@bol.com.br
Graciele Parisotto (UNISC) gracieleparisotto@yahoo.com.br
Tainá Flôres da Rosa (UNISC) m42365@mx.unisc.br
Jorcem Simon de Souza (UNISC) jorcem@bol.com.br
Marco Flôres Ferrão (UNISC) ferrao@unisc.br
João Carlos Furtado (UNISC) jcarlosf@unisc.br

Resumo

Amostras de medicamentos industrializados e manipulados contendo anti-inflamatórios (diclofenaco de potássio, diclofenaco de sódio, piroxicam e prednisona) e amostras de amido, foram estudadas através de espectroscopia por reflexão difusa no infravermelho médio com transformada de Fourier (DRIFTS) combinada à análise multivariada empregando o método de agrupamento hierárquico (HCA). Inicialmente 1038 sinais de cada espectro foram modelados utilizando-se o software Pirouette® v.2.7 da Infometrix para a construção dos dendogramas (HCA). Visando a redução da matriz de dados foi empregada a seleção das variáveis (frequências espectrais) em ambiente MATLAB, através da implementação de uma rotina de algoritmo genético, sendo o fitness baseado no cálculo da similaridade das amostras de mesma constituição. Os resultados obtidos para a matriz de dados original (1038 variáveis) foram semelhantes àqueles que empregaram apenas 100 destas. Com isso constatou-se que a redução no número de variáveis (frequências do espectro) através do emprego do algoritmo genético resulta numa alternativa adequada para avaliar a similaridade entre as replicatas dos espectros, resgatando a porção da informação mais representativa. Modelos simplificados para o reconhecimento de padrões, utilizando DRIFTS, podem então facilmente serem desenvolvidos para a análise de medicamentos, sem necessariamente empregar todo o sinal espectral, construindo-se assim uma ferramenta analítica mais robusta e mais adequada ao controle industrial de diversos medicamentos. Palavras chave: Medicamentos anti-inflamatórios, Algoritmos genéticos, DRIFTS.

1. Introdução

Frente a crescente evolução nos processos produtivos da indústria farmacêutica, necessita-se cada vez mais de ferramentas analíticas capazes de satisfazer a estas exigências produtivas do mercado mundial, bem como agregar vantagens operacionais e lucrativas frente às técnicas já utilizadas para o controle de qualidade de fármacos.

A dosagem de princípios ativos em fármacos encontra-se inteiramente relacionada com a atuação terapêutica destas substâncias no organismo humano. Em consequência, a incorreta dosagem, se administrada, poderá tornar o medicamento inócuo, e, no caso de uma super dosagem, poderá levar a intoxicações graves.

Por tais razões, é de enorme importância o desenvolvimento de métodos eficazes e rápidos para um maior controle de qualidade dos medicamentos comercializados, tornando-se imprescindível o desenvolvimento de rotinas de análises que, além de servirem para este propósito, favoreçam sua implantação em comparação aos métodos usuais.

A identificação destas substâncias pode ser alcançada tanto quantitativamente como qualitativamente através da espectroscopia na região do infravermelho, associada ao crescente desenvolvimento das técnicas de reflexão. Esta metodologia garante a caracterização de fármacos, com extrema rapidez (em poucos minutos), além disso, ressalta-se a possibilidade

da quantificação de mais de um parâmetro por espectro obtido. Os espectros gerados possuem um grande número de informações sobre a funcionalidade das moléculas orgânicas tidas como amostras, mesmo sem uma prévia separação, ou seja, sem identificação individual dos componentes que compõem determinado medicamento. Assim pode-se gerar conjuntos de dados passíveis de um monitoramento quando são empregados métodos multivariados de classificação e regressão.

Conforme Ribeiro et al.(2002), no Brasil, uma característica evidente dos anti-inflamatórios não esteroidais é o fato de serem medicamentos de venda livre e estarem em grande parte das associações medicamentosas irracionais disponíveis no mercado, além do fácil acesso devido ao grande número de apresentações disponíveis. O que acaba por caracterizar esta classe de medicamentos, como a mais prescrita por médicos e dentistas, e conseqüentemente uma das mais consumidas. Fato este confirmado por diversos estudos sobre o uso de medicamentos no país, nos quais os anti-inflamatórios não esteroidais figuram entre os primeiros colocados no *ranking* dos fármacos mais consumidos pela população.

Portanto, ainda que estes medicamentos causem complicações significativas somente em pequena porcentagem dos pacientes, uma vez que se constituem fármacos seguros, o grande número de usuários e o tempo prolongado de sua utilização fazem que estas complicações adquiram proporção preocupante, o que reforça a necessidade de se intensificarem os estudos sobre a questão do seu intenso controle de qualidade.

2. Espectroscopia no infravermelho

Segundo Workman Jr. (2002), os métodos espectroscópicos estão entre os mais apropriados para o monitoramento de processos e verificação da qualidade de produtos industriais. Estes destacam-se principalmente pelo baixo tempo necessário para realização das análises e pela grande quantidade de informação que estas técnicas podem obter. Cada vez mais vem sendo aplicadas metodologias baseadas nas técnicas por reflexão no infra-vermelho associadas as ferramentas quimiométricas disponíveis.

A técnica de reflexão difusa está bem fundamentada na literatura (ABDULLAH & SHERMAN, 1997; OSBORNE, FEARN & HINDLE, 1993; PASQUINI, 2003), sendo amplamente associada aos equipamentos que operam no infravermelho próximo, com os quais geralmente é designada de espectroscopia de reflexão no infravermelho próximo (NIRS ou NIR); ou associada àqueles que operam na região do infravermelho médio, sendo conhecida por espectroscopia por reflexão difusa no infravermelho médio com transformada de Fourier (DRIFTS).

A reflexão difusa ocorre em superfícies não totalmente planas, podendo o substrato ser contínuo ou fragmentado (na forma de pó). Neste processo de reflexão o feixe incidente penetra a superfície da amostra interagindo com a matriz, retornando à superfície da mesma, após absorção parcial e múltiplos espalhamentos, conforme ilustrado pela Figura 1.

Na reflexão difusa, a radiação incidente entra em contato diversas vezes com as partículas da amostra sendo conseqüentemente atenuada. Com isto a radiação que sofre a reflexão difusa fornece informações qualitativas e quantitativas, sobre a natureza química da amostra.

Um maior detalhamento desta técnica, bem como de outras técnicas de reflexão empregadas em análises que potencialmente aplicam informações químicas obtidas via espectroscopia no infravermelho com transformada de Fourier, pode ser encontrado em recente revisão publicada por Ferrão (2001).

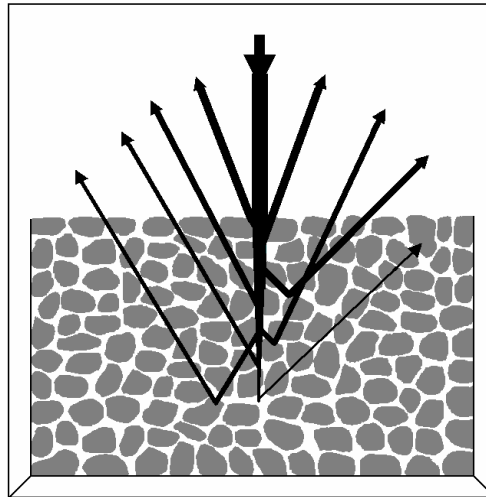


Figura 1: Representação da reflexão especular e difusa de uma onda eletromagnética em uma amostra particulada (WETZEL, 1983).

3. Análise multivariada

Métodos de análise multivariada tem sido cada vez mais utilizados em química analítica, principalmente quando os componentes presentes em uma mistura necessitam ser determinados, mas a informação analítica disponível não apresenta seletividade, ou seja, em uma mistura não se é possível identificar os componentes individuais de cada espécie a partir da resposta instrumental (ZAMORA, 1997).

A análise multivariada é a rigor qualquer abordagem analítica que considere o comportamento de duas ou mais variáveis simultaneamente, num vasto campo do conhecimento que envolve uma grande multiplicidade de conceitos estatísticos e matemáticos (PEREIRA, 1999).

Assim, no caso em que empregamos análise multivariada associada a técnicas espectroscópicas no infravermelho, é possível manipular dados de absorvância espectral associados a mais de uma frequência ao mesmo tempo. Williams (1987) apresenta aplicações destes métodos na modelagem de propriedades químicas e físicas de sistemas simples e complexos a partir de seus dados espectroscópicos.

Uma das dificuldades da modelagem de dados espectroscópicos é o elevado número de variáveis que os espectros costumam apresentar. Conforme Mallet et. al. (1996) diversos métodos de análise discriminante vem sendo desenvolvidos, necessitando a redução da dimensão da matriz de dados. Uma alternativa encontrada a utilização da análise de componentes principais (PCA) estudada por Vogt e Tacke (2001) ou da transformada wavelet discreta (DWT), apresentada por Mallet et. al. (1996).

Neste trabalho sugere-se a utilização da análise hierárquica de agrupamentos precedida pela seleção das variáveis empregando-se algoritmos genéticos. A proposta do HCA é agrupar dados em conjuntos que têm atributos similares, apresentando-os de uma maneira que enfatize os agrupamentos naturais, analogamente à classificação taxonômica realizada em ciências biológicas (VALENTIN, 2000). Para tanto, as distâncias entre pares de amostras são calculadas e comparadas. Quando as distâncias entre as amostras são relativamente pequenas, isto implica que as mesmas são similares; já amostras diferentes serão separadas por distâncias relativamente grandes. A apresentação dos resultados dessa classificação hierárquica se dá na forma de um gráfico denominado dendograma, facilitando o reconhecimento visual de tais categorias.

Para efetuar este agrupamento é necessário definir matematicamente o conceito de similaridade, ou seja, a distância entre dois objetos. Entre as medidas mais usuais para estabelecer o conceito de distância entre duas amostras, destacam-se o coeficiente de correlação de Pearson, a distância Euclidiana e a distância de Manhattan. Neste trabalho, no entanto, foi utilizada somente a distância Euclidiana, que é matematicamente a maneira mais simples de realizar o cálculo, conforme a expressão 1, onde $m=2$:

$$d_{ab} = \sqrt{\sum_{i=1}^m (x_{ai} - x_{bi})^2} \quad (1)$$

Após as distâncias entre todos os pares das amostras terem sido calculadas, as duas amostras mais similares são ligadas para formar um novo ponto. Depois que as distâncias entre esse novo ponto e todos os outros existentes são determinadas, a menor distância entre pontos é novamente calculada e outra ligação é formada. Esse processo é repetido até que todos os pares de pontos tenham sido agrupados.

Vários métodos são utilizados para calcular as distâncias entre as amostras e os agrupamentos hierárquicos. Porém, neste trabalho foi utilizado somente o método de conexão flexível, cuja distância pode ser calculada conforme expressão 2.

$$d_{ab \rightarrow c} = \left(0,625d_{ac}^2 + 0,625d_{bc}^2 + 0,25d_{ab}^2 \right)^{1/2} \quad (2)$$

4. Algoritmo genético na seleção de variáveis

Ultimamente têm sido crescente a aplicação de algoritmos genéticos (AG) na seleção das variáveis mais representativas de espectros no infra-vermelho. Lavine et. al. (2002) aplicam o AG em análises de reconhecimento de padrões de dados multivariados contendo informação química. Assim como recentemente Konzen et. al. (2003) utilizam o AG na seleção de variáveis espectrais visando otimizar métodos de regressão multivariada.

Algoritmo Genético é uma técnica de busca aleatória direcionada, desenvolvida por Holland (1975), capaz de obter a solução ótima global num espaço de busca complexo multi-dimensional. O AG é baseado na evolução natural das espécies, usando operadores inspirados no processo de evolução natural. Estes operadores, conhecidos como operadores genéticos, manipulam indivíduos de uma população, através de gerações, para melhorar (aperfeiçoar) a adaptação (*fitness*) gradativamente. Os indivíduos numa população, também denominados de cromossomos, são representados por cadeias (*strings*) de números binários. A função de avaliação (*fitness*) estabelece a relação entre o AG e o problema de otimização.

O AG revela-se útil na resolução de problemas complexos (LORENA & FURTADO, 2001), que podem ser divididos em três classes, segundo Lucasius e Kateman,(1993) na primeira classe de problemas em que os AG's são passíveis de aplicação, encontra-se a otimização de modelos que descrevam relações entre informações coletadas de um sistema (espectros de substâncias, por exemplo) a valores de propriedades (respostas do sistema) que se queira estudar. Um segundo tipo de aplicação para os AG's seria na resolução de problemas envolvendo relações, capacidade de planejamento e tomada de decisões. A seleção de subconjuntos de dados de um sistema ilustra a terceira classe de aplicações a que os AG's se destinam.

Nos AG's, as condições em que o experimento é conduzido são consideradas como sendo o material genético dos seres vivos, e a resposta experimental é considerada como medida de

adaptação ao ambiente. O material genético é assumido como sendo um simples cromossomo, onde cada gene corresponde a uma variável do processo.

Basicamente, pode-se descrever o funcionamento dos AG's nos passos:

- Criação da população inicial: nesta etapa, é selecionado um determinado número de condições experimentais a partir dos valores das variáveis (cromossomos);
- Cada indivíduo da população é avaliado usando-se o conceito de similaridade, conforme a expressão 3, onde, d_{ab} = distância entre as amostras a e b e d_{\max} = maior distância entre dois pontos no conjunto de dados.

$$\text{Similaridade}_{ab} = 1 - \frac{d_{ab}}{d_{\max}} \quad (3)$$

- No passo seguinte, inicia-se um processo que será repetido até que uma condição seja satisfeita, que poderá ser a resposta determinada (o mínimo global, por exemplo), um certo número de gerações ou um determinado tempo de execução.
- Da população, são escolhidos alguns indivíduos para realizar as operações de recombinação e mutação:
 - Recombinação: a partir de um determinado número de cruzamentos entre os cromossomos, selecionados aleatoriamente a partir das melhores respostas experimentais, estes combinam-se formando novos indivíduos;
 - Mutação: os componentes de uma população podem sofrer perturbações em seus genes.
- Os indivíduos são novamente avaliados;
- Os indivíduos com piores avaliações são eliminados;

A avaliação da população é feita através de uma função que avalia a resposta de determinado cromossomo em relação ao modelo. Esta função é também chamada de *fitness*. No presente trabalho, o algoritmo de AG implementado busca manter a similaridade das amostras de mesma natureza baseando-se na equação 3.

5. Metodologia

Neste trabalho foram adquiridas amostras de medicamentos industrializados ou manipulados contendo diclofenaco de sódio, diclofenaco de potássio, piroxicam, predinizona e excipientes comumente empregados nas farmácias de manipulação da cidade de Santa Cruz do Sul - RS, podendo variar a concentração do fármaco em questão, bem como o excipiente utilizado.

Os espectros de todas as amostras foram adquiridos empregando-se um espectrofotômetro Magna-550 da Nicolet acoplado a um acessório de reflexão difusa Pike, realizando-se 32 varreduras e utilizando-se resolução de 4 cm^{-1} .

A seleção das variáveis (frequências espectrais) foi realizada em ambiente MATLAB, empregando-se uma rotina de algoritmo genético e posteriormente utilizou-se o *software* Pirouette para a aplicação do método de análise hierárquica de agrupamento (HCA).

Primeiramente selecionou-se o conjunto de regiões do espectro com o maior número de informações de interesse, isto é, os números de onda que apresentam melhor correlação com a quantificação do princípio ativo em estudo. As regiões espectrais selecionadas foram de 801 a 1799 cm^{-1} e de 2801 a 3799 cm^{-1} .

À matriz de dados construída foi submetida ao AG visando a otimização da similaridade entre os espectros de mesma natureza, utilizando-se como parâmetros de entrada: (a) Tamanho da população: 100 cromossomos; (b) Total de números de onda selecionados: 50; (c) Percentual de seleção na população: 25; (d) Percentual de mutação na população: 5; (e) Número de iterações: 500.

Na seqüência construiu-se os dendogramas, considerando-se as amostras otimizadas e também a matriz completa, com todos os números de onda. O pré-processamento utilizado foi o auto-escalamento, que consiste em centrar os dados na média e dividi-los, um a um, pelo desvio-padrão, de forma que todos as variáveis passam a ter a mesma importância, ou seja, o mesmo peso. Além disso, apesar de terem sido testados os métodos de construção simples, incremental e flexível, selecionou-se somente o tipo de conexão flexível para aplicar a rotina de otimização via AG, empregando-se ainda, à matriz de dados, a correção do espalhamento de luz (MSC).

6. Resultados

Na Tabela 1 são apresentados os resultados da otimização empregando algoritmo genético.

Teste	Fitness	Total nº onda sel.	Iterações
A	1,1848	100	500
B	1,1885	100	500
C	1,1914	100	500
D	1,2038	100	500
E	1,1893	100	500
F	1,2227	100	500

Tabela 1 – Valores de *fitness* para os testes realizados na otimização dos dados DRIFTS.

Nas Figuras 1 e 2 são apresentados os dendogramas para o modelo não otimizado e para o teste C respectivamente. Observa-se que o modelo não otimizado, empregando os valores de absorvância referentes a 1038 números de onda apresenta um dendograma distinguindo a maior parte das diferentes amostras ensaiadas. De forma semelhante o dendograma do modelo otimizado, Figura 2, apresenta um perfil semelhante de clusterização, apesar de utilizar os valores de absorvância referentes a somente 100 números de ondas originais.

7. Conclusões

Como apresentado anteriormente, os resultados de HCA com 1038 variáveis e com 100 variáveis (otimizado com AG) mostram semelhantes estruturas de clusterização distinguindo os principais grupos dos diferentes medicamentos analisados, bem como das amostras que contém apenas amido. Esta evidência indica que grande parte da informação original do espectro não se refere aos compostos presentes nas formulações de medicamentos aqui estudadas. Assim, muitas das variáveis adicionais presentes no modelo não otimizado devem conter ruído experimental ou informações de natureza não discriminatória.

Adicionalmente podemos afirmar que a metodologia aqui empregada serve como excelente alternativa para refinar (otimizar) o estudo quimiométrico de medicamentos, sejam eles manipulados ou industrializados, visando um controle de qualidade através de métodos não destrutivos e não geradores de resíduos nocivos tanto ao ambiente industrial quanto ao meio ambiente.

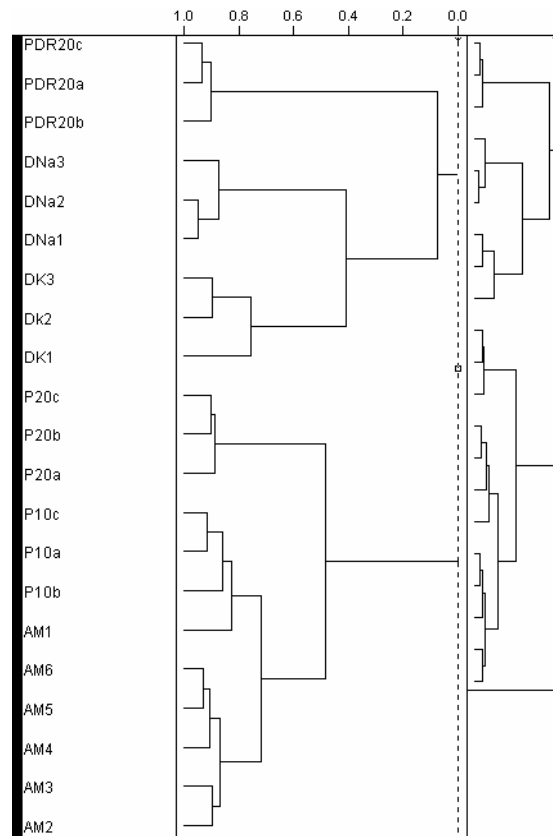


Figura 1: Dendrograma obtido a partir dos espectros DRIFT de amostras para o modelo não otimizado.

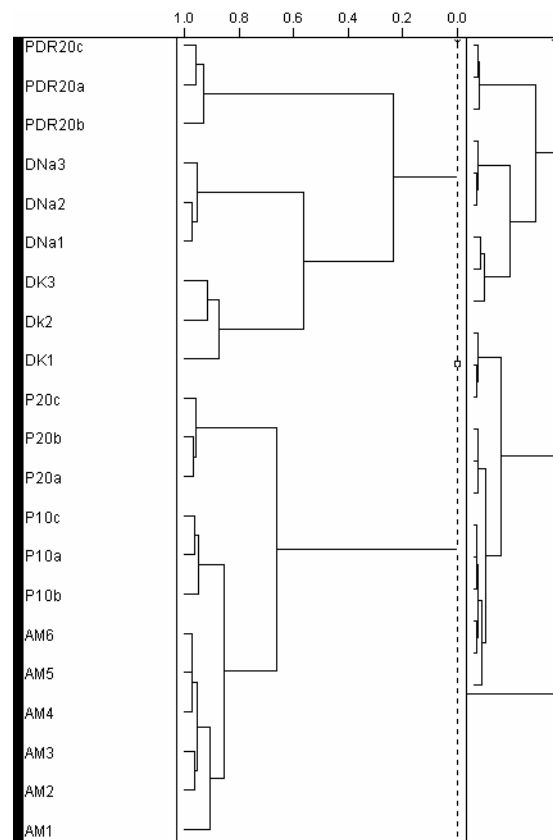


Figura 2: Dendrograma obtido a partir dos espectros DRIFT de amostras para o modelo otimizado com algoritmo genético selecionando 100 variáveis.

8. Referências

- ABDULLAH, A.H. & SHERMAN, W.F. (1997) - Kramers-Kronig Type Analysis of Short Spectral Range Reflection Spectra. *Vibrational Spectroscopy*, Vol.13, p. 133-142.
- FERRÃO, M.F. (2001) - Técnicas de Reflexão no Infravermelho Aplicadas na Análise de Alimentos. *Tecnológica*, Vol. 5, n.1, p. 63-85.
- HOLLAND, J.H. (1975) - *Adaptation in Natural and Artificial Systems*. MIT Press, p. 11-147.
- KONZEN, P.H.A.; FURTADO, J.C.; CARVALHO, C.W.; FERRÃO, M.F.; MOLZ, R.F.; BASSANI, I.A. & HÜNING, S.L. (2003) - Otimização de Métodos de Controle de Qualidade de Fármacos Usando Algoritmo Genético e Busca Tabu. *Pesquisa Operacional*, Vol. 23, n. 1, p. 189-207.
- LAVINE, B.K.; DAVIDSON, C.E. & MOORES, A.J. (2002) - Innovative Genetic Algorithms for Chemoinformatics. *Chemometrics and Intelligent Laboratory Systems*, Vol. 60, p. 161-171.
- LORENA, L.A.N. & FURTADO, J.C. (2001) - Constructive Genetic Algorithm for Clustering Problems. *Evolutionary Computation*, Vol. 9, p. 309-327.
- LUCASIUS, C.B. & KATEMAN, G. (1994) - Understanding and Using Genetic Algorithms Part 2. Representation, Configuration and Hybridization. *Chemometrics and Intelligent Laboratory Systems*, Vol. 25, p. 99-145.
- MALLET, Y.; COOMANS, D. & DE VEL, O. (1996) - Recente Developments in Discriminant Analysis on High Dimensional Spectral Data. *Chemometrics and Intelligent Laboratory Systems*, Vol. 35, p. 157-173.
- OSBORNE, B.G.; FEARN, T. & HINDLE, P.H. (1993) - *Practical NIR Spectroscopy With Applications in Food and Beverage Analysis*. Longman Scientific & Technical, Singapore.
- PASQUINI, C. (2003) - Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications. *Journal of Brazilian Chemical Society*, Vol.14, n.2, p.198-219.
- PEREIRA, J.C. (1999) - *Análise de Dados Qualitativos: Estratégias Metodológicas para as Ciências da Saúde, Humanas e Sociais*. 2. ed. São Paulo: Editora da Universidade de São Paulo.
- RIBEIRO, A. Q. et al. (2002) - Reações Adversas Gastrointestinais aos Anti-inflamatórios Não-esteróides – Uma Abordagem em Farmacoepidemiologia. *Rev. Ciênc. Farm.*, Vol.23, n.2, p.175-198.
- VALENTIN, J.L. (2000) - *Ecologia Numérica: Uma Introdução à Análise Multivariada de Dados Ecológicos*. Rio de Janeiro: Interciência.
- VOGT, F. & TACKE, M. (2001) - Fast Principal Component Analysis of Large Data Sets. *Chemometrics and Intelligent Laboratory Systems*, Vol. 59, p. 1-18.
- WETZEL, D.L. (1983) - Near-Infrared Reflectance Analysis Superior Among Spectroscopic Techniques. *Analytical Chemistry*, Vol. 55, p. 1165A-1176A.
- WILLIAMS, P.C. (1987) - *Commercial Near-infrared Reflectance Analyzers*. In: *Near-infrared Technology in the Agricultural and Food Industries*, by Phil Williams & Karl Norris. American Association of Cereals Chemists, Saint Paul.
- WORKMAN JR., J. (2002) - The State of Multivariate Thinking for Scientists in Industry: 1980-2000. *Chemometrics and Intelligent Laboratory Systems*, Vol. 60, p.13-23.
- ZAMORA, P. P. et al. (1997) - Alternativas Quimiométricas para a Resolução de Problemas Analíticos Clássicos. Determinação Espectrofotométrica de Misturas de Zircônio e Háfnio. *Química Nova*. Vol.20, n.5. p.469-474.

9. Agradecimentos

Os autores agradecem a Central Analítica da UNISC pela colaboração e à Fundação de Amparo a Pesquisa do Rio Grande do Sul e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico pelas bolsas concedidas.