Universidade Estadual de Campinas

Instituto de Química

Desenvolvimento e Aplicações de Modelos de Calibração Multivariada em Espectroanalítica e Eletroanalítica

Tese de Doutorado

Eduardo Osório de Cerqueira

Orientador: Prof. Dr. Ronei Jesus Poppi

Co-orientador: Prof. Dr. Lauro Tatsuo Kubota

Campinas – Maio / 2002

Para aquelas pessoas que fazem sorrir meu coração.....

Para as pessoas que fizeram a diferença em minha vida!!

Para as pessoas que quando olho para trás sinto muitas saudades .....

Para as pessoas que me deram uma força quando eu não estava muito animado para o trabalho!!!

Para as pessoas que amei...

Para as pessoas que abracei...

Para aqueles que encontro apenas em meus sonhos...

Para aqueles que encontro todos os dias e não tenho a chance de dizer tudo que sinto olhando nos olhos...

Para aqueles que já se esqueceram de quanto foram importantes para mim....

A Vida é um caminho cheio de surpresas, e o final dele ninguém conhece.

O que importa não é o que você tem na vida, mas QUEM você tem na vida.

Por isso que guardo todas as pessoas importantes na minha vida em uma caixinha dentro do meu coração !!!!

Dedico à minha mãe Zélia e aos meus irmãos Ricardo e Osorinho, que sempre me apoiaram em todos os momentos de minha vida. Que mesmo sem compreender o meu trabalho, tinham consciência de sua importância, por acreditarem em mim e em minha capacidade.

Agradecimentos

Ao professor Ronei, pela oportunidade, pela orientação, por todo apoio e pela amizade ao longo dessa etapa importante em minha vida;

Ao professor Lauro pela orientação;

À FAPESP pelo financiamento do projeto do (processo nº 97/06988-3) e pela bolsa concedida;

Ao Lídio, da Empresa FEMTO, pelo intercâmbio e pela possibilidade de obtenção dos espectros NIR.

À Usina Alta Mogiana, pelo fornecimento dos dados utilizados nos modelos não lineares de calibração multivariada propostos.

Aos amigos do Laboratório de Quimiometria em Química Analítica - LAQQA: Cleidiane Zamprônio, Marcelo Sena, Silvio Lima, Jacqueline Duenas, Rosângela Barthus, Alessandra Borin, Fernando Barbosa, Waldomiro Borges, Marcello Trevisan, Paulo Fidêncio, Jez Willian e Chiquinho (Leoberto Balbinot), pela troca de idéias, incentivo, companheirismo e amizade;

Aos amigos dos laboratórios adjacentes: Fernanda Almeida, Pillar, Rosângela, Miyuki, Jequié (Antônio), Arnaldo, Beth, Antenor, Carin, Simone, Lucilene, Lauter, Walter, Tânia e Perci, entre outros que provavelmente esqueci de citar, mas que foram tão importantes quanto, pela agradável convivência e amizade;

Ao Instituto de Química da UNICAMP e aos seus funcionários, pelo apoio técnico e prontidão;

Aos amigos da república onde morei: Marcelo Moreira, Josinira Antunes, Fábio, Márcio Paredes, Álvaro, Wedson, Binho (Fábio Ribeiro), Frede, Marcos Antônio, Onzinho (Marcelo) , Pedro Guedes e Wally, entre outros que por lá passaram, pela convivência diária e amizade.

Para a Daniela Millares, pelo carinho e companheirismo durante o final dessa empreitada.

Resumo

Título: **"Desenvolvimento e Aplicações de Modelos de Calibração Multivariada em Espectroanalítica e Eletroanalítica"**

Autor: Eduardo Osório de Cerqueira

Orientador: Ronei Jesus Poppi

Co-orientador: Lauro Tatsuo Kubota

Palavras-chave: Calibração multivariada, Algoritmos genéticos, redes neurais e adição de padrão.

Neste trabalho propôs-se modelos de calibração multivariada baseados em transformações não-lineares para a mudança de base, seguidas de uma truncagem da nova base otimizada por algoritmo genético. Isso é necessário porque devido à quantidade de interferências e ruídos encontrados em análises reais, nem sempre é possível encontrar uma relação linear de modo a obter previsões com erros aceitáveis.

No primeiro modelo utilizou-se um algoritmo genético para selecionar quais freqüências de Fourier são relevantes numa calibração multivariada utilizando-se o método dos mínimos quadrados parciais - PLS para a determinação dos parâmetros de qualidade brix e pol em amostras de caldo de cana concentrado a partir de espectros no infravermelho próximo. Para a determinação do brix obteve-se erros médios de 1,3 % enquanto que para a determinação do pol obteve-se erros médios de 1,4 %.

No segundo modelo utilizou-se um algoritmo genético para otimizar uma rede neural com funções radiais de base (RBFs). Foram otimizados os raios, a posição dos centros, o número de RBF's, o tipo de RBF utilizada e as conexões de cada RBF de uma rede de RBFs. Esse modelo foi aplicado na determinação de açúcares redutores totais em caldo de cana concentrado a partir do espectro no infravermelho próximo. Obteve-se erros médios de previsão de 1,5% na determinação do ART.

Adicionalmente, propôs-se um modelo para calibração multivariada por adição de padrão simultânea para mais de uma espécie química. Esse modelo foi aplicado na determinação de ácido ascórbico e ácido úrico por voltametria de pulso diferencial, utilizando um potenciostato construído no laboratório, controlado por microcomputador. Obteve-se erro médio de previsão para o ácido úrico de 4,6% e para o ácido ascórbico de 3,0 %.

Abstract

Title: "Development and applications of Multivariate Calibration Models in Spectroanalitical and eletroanalytical "

Author: Eduardo Osório de Cerqueira

Advisor: Prof. Dr. Ronei Jesus Poppi

Co-Advsor: Lauro Tatsuo Kubota

Keywords : Multivariate Calibration; Genetic algorithms; Neural Networks; Standard Addition

In this work it was proposed multivariate calibration models based on non-linear transformations to basis change, followed by the truncation of the new base optimized by genetic algorithm. This procedure is necessary due the quantity of interferents and noise presents in real samples, making not possible to create a linear relationship in way to obtain experimental predictions with acceptable errors.

In the first model, it was used a genetic algorithm to select the relevant Fourier frequencies for the determination of the quality parameters birx and pol in concentrated sugar cane juice samples from near infrared spectra using the PLS model. For brix determination it was obtained average errors of 1.3%, while for pol determination it was obtained average errors of 1.4%.

In the second model, it was used a genetic algorithm to optimize a Radial Basis Function Network. It was optimize the radius, center position, number of base functions, type and the connections in each network neuron. This model was applied in the total reducing sugar determination in concentrated sugar cane juices from its near infrared spectra. It was obtained average errors of 1.5% for total reducing sugar determination.

Also, it was proposed a model to simultaneous multivariate calibration standard addition for more than one chemical specie. This model was applied in the determination of ascorbic acid and uric acid by differential pulse voltametry. It was obtained average errors of 4.6% for uric acid and 3.0% for ascorbic acid.

Índice Geral

Índice de figuras

Índice de figuras (continuação)

Índice de figuras (continuação)

Índice de figuras (continuação)

Índice de figuras (continuação)

Índice de tabelas

CURRICULUM VITAE

1. DADOS PESSOAIS

Nome: Eduardo Osorio de Cerqueira
Nascimento: 26/02/1973, Recife/PE - Brasil

2. FORMAÇÃO ACADÊMICA/TITULAÇÃO

2.1. Doutorado em Química
Universidade Estadual de Campinas, UNICAMP, São Paulo, Brasil, 2002.
Título: Desenvolvimento e Aplicações de modelos de Calibração Multivariada em espectroanalítica e eletroanalítica.

2.2. Mestrado em Química
Universidade Estadual de Campinas, UNICAMP, São Paulo, Brasil, 1997.
Título: Desenvolvimento de um Espectrofotôetro Multicanal para Análise Multivariada em Sistema de Fluxo.

2.3. Graduação em Química

Bacharelado em Química
Universidade Federal de Pernambuco, UFPE, Pernambuco, Brasil, 1994.

3. Trabalhos Publicados em Periódicos

3.1. E. O de Cerqueira e R. J. Poppi
"Using DDE (Dynamic Data Exchange) to exchange information between Visual Basic and Matlab. A photodiode Array application."
TRENDS IN ANALYTICAL CHEMISTRY 15, 500-503 (1996)

3.2. E. O. de Cerqueira, R. J. Poppi, L. T. Kubota e C. Mello
"Utilização de filtro de transfomada de Fourier para minimizaçãoe ruídos em sinais analíticos"
QUÍMICA NOVA 23, 690-698 (2000)

3.3. E. O. de Cerqueira, J. C. de Andrade, R. J. Poppi e C. Mello
"Redes Neurais e suas aplicasções em calibração multivariada"
QUÍMICA NOVA 24, 864-873 (2001)

4. Trabalhos apresentados em Congressos e Encontros Científicos

4.1. E. O de Cerqueira e R. J. Poppi
"Desenvolvimento de hardware e software para gerenciamento de um espectrofotômetro com um arranjo linear de fotodiodos"
8o. Encontro Nacional de Química Analítica - Belo Horizonte – MG, 1995

4.2. E. O. de Cerqueira e R. J. Poppi

"Determinação espetrofotométrica de pH em fluxo utilizando calibração multivariada"
19a. Reunião Anual da SBQ - Poços de Caldas – MG, 1996

4.3. P. A . da Costa Filho, E . O . Cerqueira e R. J. Poppi
"Genetic algorithm for simultaneous selection of variables in generalized rank annihilation method applied to flow analysis"
7th International Conference on Flow Analysis – Piracicaba – Brasil, 1997

4.4. C. Mello, E. O. de Cerqueira, R. J. Poppi, L. T. Kubota e P. A. da Costa Filho
"Aplicações do filtro de transformada de Fourier (FFT) na minimização de ruídos em sinais analíticos"
10o. Encontro Nacional de Química Analítica - Santa Maria – RS, 1999

4.5. E. O. de Cerqueira, R. J. Poppi, M. J. Saeki e L. T. Kubota
"Utilização do PARAFAC para a modelagem da caracterização voltamétrica de azul de metileno e azul de meldola adsorvidos em peroviskita"
23a. Reunião Anual da SBQ - Poços de Caldas – MG, 2000

4.6. C. Mello, E. O. de Cerqueira e R. J. Poppi
"Uma nova abordagem baseada em FFT, algoritmo genático e PLS para o pré processamento e modelagem simultânea de dados analíticos"
23a. Reunião Anual da SBQ - Poços de Caldas – MG, 2000

4.7. E. O. de Cerqueira e R. J. Poppi
"Construção de um modelo de calibração multivariada baseado em uma rede de funções de base radial (RBF) otimizada por algoritmo genético"
24a. Reunião Anual da SBQ – Poços de Caldas – MG, 2001

4.8. K. Y. Chumbimuni-Torres, E. O. de Cerqueira, R. J. Poppi e L. T. Kubota
"Determinação simultânea de I^- e Cl^- usando eletrodo íon seletivo com calibração multivariada em sistema FIA"
11o. Encontro Nacional de Química Analítica – Campinas – SP, 2001

4.9. F. V. Almeida, E. O. de Cerqueira, R. J. Poppi e W. F. Jardim
"Otimização quimiométrica da determinação de arocloros em sedimentos do sistema Tietê"
11o. Encontro Nacional de Química Analítica – Campinas – SP, 2001

5. Atividades Didáticas

5.1. Auxiliar didático no Instituto de química - UNICAMP
Disciplina QA- 214 - Química Analítica II, 2° Semestre de 1997

5.2. Programa de Estágio de Capacitação a Docência (PECD) no Instituto de Química – UNICAMP
Disciplina QA – 415 – Química Analítica Instrumental Experimental, 2° Semestre de 1998

5.3. Programa de Capacitação a Docência (PED) no Instituto de Química – UNICAMP
Disciplina QA - 581 – Métodos em Química Analítica Instrumental, 1° Semestre de 2001

Introdução e objetivos

Com a crescente necessidade de aumento de produtividade do setor industrial, o controle do processo fabril bem como o controle de qualidade do produto final necessitam de sistemas de análise que minimizem custos, acelerem a velocidade de análise e que sejam confiáveis. Essa necessidade impulsionou a migração das técnicas de análises convencionais, realizadas em laboratório, para técnicas que pudessem ser automatizadas diretamente na linha de produção industrial, ou mesmo realizadas em maior quantidade sem a necessidade de parar a produção. Dentre essas técnicas, as baseadas em análise do espectro de radiação eletromagnética (espectroscopia) ganharam bastante popularidade.

Das técnicas espectroscópicas, a análise de luz na região do infravermelho próximo tem tomado grande destaque devido à sua natureza não destrutiva. Aliadas ao desenvolvimento computacional e a métodos quimiométricos de calibração multivariada, essas técnicas vêm sendo implantadas em um número crescente de indústrias.

A maioria das aplicações da espectroscopia na região do infravermelho próximo utiliza modelos lineares de calibração multivariada, baseada em modelagem de fatores (variáveis latentes ou componentes principais). Porém, não se pode garantir que um modelo linear ajuste satisfatoriamente dados de espectroscopia na região do infravermelho próximo. Essa região apresenta sobretudo informações a cerca de harmônicos superiores de transições vibracionais fundamentais e de combinações de transições vibracionais fundamentais. Dependendo do grau de sobreposição espectral e/ou do "modo" (absorbância, transmitância ou reflectância) de análise, comumente há desvios do comportamento linear.

Nos modelos lineares de calibração multivariada considerados "padrões", (PLS e PCR), basicamente utiliza-se uma transformação linear de mudança da base para diminuir

sua dimensão (número de variáveis), de modo a eliminar ou diminuir a redundância de informação (variáveis correlacionadas). Esse procedimento gera uma melhor separação dos padrões de informação. Truncando-se o tamanho dessa base (segundo o número de fatores utilizados) pode-se eliminar grande parte do ruído nos dados. Desse modo facilita-se a determinação de uma relação linear entre os dados (na nova base) e a(s) propriedade(s) que se deseja determinar.

Devido à quantidade de interferência e ruídos encontrados em análises reais, nem sempre é possível encontrar uma relação linear (segundo essas técnicas mencionadas) de modo a obter previsões com erros aceitáveis. Entende-se por ruído as variações da medida analítica não correlacionadas com a propriedade que se deseja determinar, e interferência é a influência de um ou mais espécies sobre a determinação de uma terceira espécie (sinergismo ou antagonismo).

Baseado no teorema de Cover [01] sobre a separabilidade dos padrões "Um problema complexo de classificação de padrões disposto não linearmente em um espaço de alta dimensão tem maior probabilidade de ser linearmente separável do que em um espaço de baixa dimensionalidade", propôs-se modelos baseados em transformações não-lineares para a mudança e expansão da base, seguidas de uma truncagem (seleção de dados) da nova base otimizada por algoritmo genético para posterior calibração multivariada linear.

No primeiro modelo proposto, utilizou-se a transformada de Fourier para realizar a mudança de base, eliminando redundância de informação (covariância entre as variáveis originais) separando assim os padrões de informação de forma não linear. Após isso, um algoritmo genético foi utilizado para selecionar quais freqüências de Fourier são relevantes para a determinação da propriedade de interesse por um modelo linear de regressão por

mínimos quadrados parciais (PLS) entre os respectivos coeficientes de Fourier (selecionados pelo AG) e a propriedade de interesse. Ou seja, o algoritmo genético otimizou a truncagem da nova base de modo a minimizar o erro de previsão.

A teoria dos algoritmos genéticos utilizados é explicada no capítulo 1. O capítulo 2 explica a implementação do algoritmo genético para a seleção de freqüências de Fourier.

Para o segundo modelo proposto, utilizou-se um algoritmo genético para otimizar as funções radiais de base (RBF) e a arquitetura de uma rede de funções radiais de base (RBFN). Uma vez determinados os parâmetros de cada RBF, o conjunto de RBF's utilizadas forma uma base (não linear) que separa os padrões de informação dos dados originais e elimina a redundância de informação entre as variáveis originais. Em seguida a rede realiza uma regressão linear múltipla, por mínimos quadrados, para encontrar uma relação linear entre os dados (na base formada pelas RBF's) e a propriedade de interesse.

O Capítulo 3 explica o funcionamento de uma rede de RBF's e a implementação de um algoritmo genético para otimizar a arquitetura da rede de RBF's, bem como as próprias funções radiais.

As aplicações dos modelos propostos são mostradas no Capítulo 4. O Modelo de seleção de freqüências de Fourier (FFT-PLS-AG) foi aplicado na determinação do brix e do pol em amostras de caldo de cana concentrado (xarope) a partir de seus respectivos espectros de absorção na região do infravermelho próximo. O modelo de otimização de uma RBFN por algoritmo genético foi aplicado na determinação de açúcar redutores totais (ART) em amostras de caldo de cana concentrado (xarope) partindo-se de seus respectivos espectros de absorção na região do infravermelho próximo.

Geralmente os modelos quimiométricos são amplamente utilizados para a análise e modelagem de dados de espectroscopia vibracional e eletrônica. Porém sua utilização para calibração multivariada em dados provenientes de técnicas eletroquímicas ainda é alvo de poucos trabalhos. Com o intuito de explorar essa grande potencialidade, construiu-se um potenciostato e através de um modelo de adição de padrões multivariada simultaneamente para mais de uma espécie, determinou-se a concentração de ácido úrico e ácido ascórbico em uma matriz de urina humana a partir de voltamogramas de pulso diferencial. Tradicionalmente, essas substâncias apresentam voltamogramas muito sobrepostos, o que dificulta a determinação simultânea.

No capítulo 5 são mostradas a construção do potenciostato utilizado e sua avaliação. O capítulo 6 explica o modelo de adição de padrão multivariada para determinação simultânea de duas espécies. No capítulo 7 é mostrada a utilização do modelo de adição de padrão simultânea para a determinação multivariada de ácido úrico e ácido ascórbico a partir da análise por voltametria de pulso diferencial das adições de padrão sobre uma matriz de urina humana.

Por fim, são apresentados uma conclusão do trabalho de tese realizado e algumas perspectivas de trabalhos futuros.

Objetivos

Esse trabalho teve como objetivos principais:

O desenvolvimento de um modelo não-linear para calibração multivariada, utilizando seleção de frequências de Fourier por algoritmo genético para calibração com o PLS (modelo FFT-PLS-AG).

A aplicação do modelo proposto (FFT-PLS-AG) na determinação do brix e do pol a partir do espectro NIR de amostras de caldo de cana concentrado.

O desenvolvimento de um modelo não linear de calibração multivariada, baseado na otimização de uma rede de funções radiais de base por algoritmo genético (modelo RBFN-AG).

A aplicação do modelo proposto (RBFN-AG) na determinação dos açúcares redutores totais a partir do espectro NIR de amostras de caldo de cana concentrado.

A construção de um potenciostato mono-canal para aplicações eletroquímicas utilizando calibração multivariada.

A aplicação de um método de adição de padrão para análises simultâneas multivariadas na determinação de ácido úrico e ácido ascórbico em amostras de urina humana.

Capítulo 1

Algoritmos Genéticos

Os Algoritmos Genéticos (AGs) são métodos evolucionários adaptativos que podem ser usados para resolver problemas de busca e otimização. Os AGs estão baseados no processo de seleção e adaptação, bem como, na transmissão de caracteres hereditários dos organismos vivos. Ao longo de gerações, as populações evoluem na natureza de acordo com os princípios da seleção natural e a sobrevivência dos mais aptos, postulados por Darwin [2]. Por imitação deste processo, os Algoritmos Genéticos são capazes de criar soluções para problemas do mundo real. A evolução dessas soluções até valores ótimos do problema depende em grande parte de uma adequada codificação das mesmas.

Os princípios básicos dos Algoritmos Genéticos foram estabelecidos por Holand [3], e se encontram bem descritos em vários textos: Goldberg [4], Davis [5], Michaoewicz [6] e Reeves [7], por exemplo.

Na natureza os indivíduos de uma população competem entre si na busca de recursos tais como comida, água e refúgio. Inclusive os membros de uma mesma espécie competem na busca de um companheiro. Aqueles indivíduos que tem mais êxito em sobreviver e em atrair companheiros têm maior probabilidade de gerar um grande número de descendentes. Por outro lado indivíduos pouco dotados produzirão um menor número de descendentes. Isto significa que os genes dos indivíduos mais adaptados se propagarão em sucessivas gerações em um número crescente de indivíduos. A combinação de características boas, provenientes de diferentes ancestrais, pode produzir descendentes "superindivíduos", cuja adaptação é muito maior que a de qualquer de seus ancestrais. Desta maneira, as espécies evoluem com características cada vez melhor adaptadas ao meio em que vivem.

Os Algoritmos Genéticos usam uma analogia direta com o comportamento natural. Trabalham com uma população de indivíduos, em que cada um representa uma possível solução a um problema dado. A cada individuo associa-se um valor ou pontuação, relacionado com a qualidade da respectiva solução codificada no indivíduo. Na natureza isto equivaleria ao grau de adaptação de um indivíduo para competir por alguns determinados recursos. Quanto maior for a adaptação de um individuo ao problema, maior será a probabilidade de que o mesmo seja selecionado para reproduzir-se, cruzando seu material genético com outro indivíduo selecionado de forma análoga. Este cruzamento produzirá novos indivíduos, descendentes dos anteriores, os quais compartilham algumas das características de seus pais. Quanto menor for a adaptação de um indivíduo, menor será a probabilidade de que este indivíduo seja selecionado para a reprodução e, portanto, de que seu material genético se propague em gerações sucessivas.

Desta maneira produz-se uma nova população de possíveis soluções, a qual substitui a anterior. Essa nova população apresenta uma maior proporção de características boas em comparação com a população anterior. Assim, ao longo das gerações, as boas características se propagam através da população. Favorecendo o cruzamento dos indivíduos mais adaptados, vão sendo exploradas as áreas mais promissoras do espaço de busca, ou seja, do espaço com as possíveis repostas para o problema em questão. Se o Algoritmo Genético tiver sido bem montado, a população convergirá para a solução ótima do problema.

Os Algoritmos Genéticos são técnicas bastante robustas, e podem tratar com êxito uma grande variedade de problemas provenientes de diferentes áreas, incluindo aquelas em

que outros métodos encontram dificuldades. Embora não se possa garantir que o Algoritmo Genético encontre a solução ótima do problema, existe evidência empírica de que se encontram soluções em um nível aceitável, em um tempo competitivo quando comparado com outros algoritmos de otimização combinatória. Nos casos em que existam técnicas especializadas para resolver um determinado problema, o mais provável é que superem o Algoritmo Genético, tanto em rapidez como em eficácia. O grande campo de aplicação dos Algoritmos Genéticos relaciona-se com aqueles problemas para os quais não existem técnicas especializadas. Porém, mesmo nos casos em que essas técnicas existam, e funcionem bem, pode-se utilizar o AG para efetuar melhorias nessas técnicas especializadas melhorando o modelo como um todo.

Os modelos de otimização baseados em AG's contemplam: evolução, seleção natural, reprodução e ausência de memória.

A vantagem principal dos AG's ao trabalharem com o conceito de população, ao contrário de muitos outros métodos de otimização que trabalham com um só ponto, é que eles encontram segurança na quantidade. Os AGs trabalham com vários pontos do espaço de busca (indivíduos da população) simultaneamente a cada geração, e a transmissão de informações entre as gerações não depende apenas de um indivíduo. Tendo uma população de indivíduos bem adaptados, é reduzida a possibilidade de alcançar um falso ótimo (ótimo local). Os AG's conseguem grande parte de sua aplicabilidade simplesmente ignorando informação que não constitua parte do objetivo, enquanto outros métodos se apóiam fortemente nesse tipo de informação e, em problemas nos quais a informação necessária não está disponível ou se apresenta de difícil acesso, estes outros métodos falham. Em suma, os AG's podem ser aplicados praticamente em qualquer problema. Apesar de

aleatórios, eles não são caminhadas aleatórias não-direcionadas, ou seja, buscas totalmente sem rumo, pois exploram informações históricas para encontrar novos pontos de busca onde são esperados melhores desempenhos. Isto é feito através de processos iterativos, em que cada iteração é chamada de geração.

Na prática, pode-se implementar facilmente um AG com o simples uso de vetores de "bits" ou caracteres para representar os cromossomos e, com simples operações de manipulação de "bits" pode-se implementar cruzamento, mutação e outros operadores genéticos.

1.1 - O Algoritmo Genético básico

Antes de qualquer coisa, faz-se necessário uma codificação adequada do problema a ser resolvido. Os parâmetros, ou variáveis, do problema em questão (a serem otimizados) são codificados formando cromossomos. A seqüência de cromossomos forma um indivíduo. O conjunto destes, por sua vez, forma uma população. A Figura 1.1 mostra a representação de uma população de indivíduos que codificam variáveis contínuas. Nesse caso, cada segmento de vetor (cromossomo) representa a codificação de uma variável do problema.

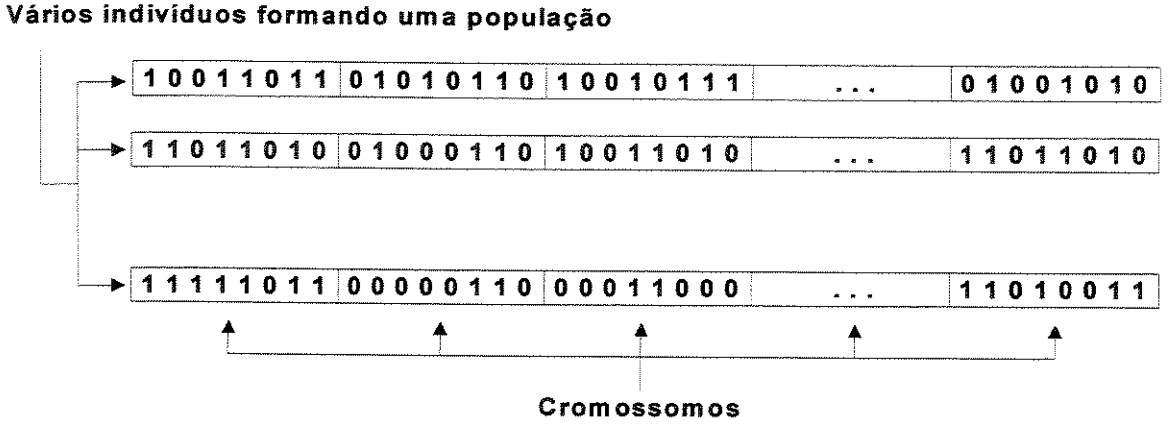


Figura 1.1 – Esquema de codificação do problema para o AG.

O Algoritmo Genético básico, também denominado Canônico, é mostrado no diagrama da Figura 1.2.

A partir do problema em questão define-se uma função de ajuste ou adaptação ao problema, a qual associa um número real a cada possível solução codificada (indivíduo). Durante a execução do AG, os indivíduos "pais" (membros da população corrente) devem ser selecionados para a reprodução, levando em consideração o resultado da avaliação de cada indivíduo. O cruzamento deles originará os indivíduos filhos, e sobre cada um dos indivíduos da nova população atuará um operador de mutação.

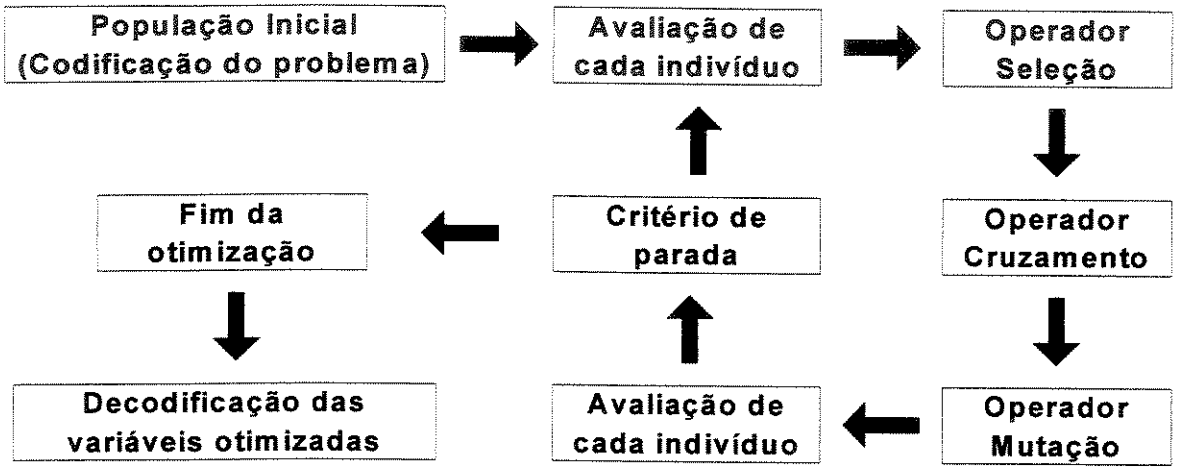


Figura 1.2 - Esquema simplificado de um Algoritmo genético simples.

O resultado dessas operações sobre a população é uma nova população de indivíduos que substituirá a população corrente, completando assim uma geração ou passo, da evolução no Algoritmo genético. Esse processo segue continuamente até atingir o nível ótimo desejado ou um número específico de gerações (ambos previamente estabelecidos). Em seguida os indivíduos da população final (resultante da última geração) são

decodificados, de acordo com a codificação utilizada para o problema em questão. Desse modo as variáveis otimizadas são obtidas.

1.1.1 – Codificação

Os indivíduos (possíveis soluções do problema) podem ser representados como um conjunto de parâmetros (que denominaremos genes), os quais quando agrupados formam uma lista de valores (denominados cromossomos) que codifica cada parâmetro ou variável do problema em questão.

Embora a notação utilizada para representar os indivíduos não seja necessariamente constituída por "0" (zeros) e "1" (uns), boa parte da teoria na qual se fundamentam os Algoritmos Genéticos utiliza esta notação. Pode-se utilizar a codificação em termos que caracteres alfa-numéricos, porém, nesse caso deve-se definir toda uma "nova álgebra", com todos os operadores utilizados para a codificação em questão

Os Algoritmos genéticos podem ser usados para otimização de variáveis contínuas ou discretas, e em cada caso a codificação deve ser feita diferentemente.

No caso de utilização do AG para a otimização de valores contínuos, cada cromossomo é a codificação de um parâmetro real, e é composto por um vetor binário cujos limites (mínimo e máximo) são padronizados entre os extremos de uma faixa dentro da qual está a respectiva variável que se deseja otimizar. Desse modo o número de bits do cromossomo determina a precisão com que se deseja otimizar a variável codificada pelo

cromossomo. Por exemplo: vejamos um cromossomo com 8 bits codificando uma variável que pode variar entre 0 e 2

[0 0 0 0 0 0 0 0] = 0 → 0

[1 1 1 1 1 1 1 1] = 255 → 2

[1 0 0 1 0 1 0 0] = 148 → ? = 1,161

Ou seja, o vetor [1 0 0 1 0 1 0 0] representa um cromossomo que codifica o valor 1,161 para a variável em questão. A precisão utilizada foi de 8 bits, ou seja, a menor variação possível na variável codificada é de $2/2^8$ ou 0,008

No caso de utilização do AG para a otimização de variáveis discretas (seleção de variáveis), cada cromossomo é formado por apenas um gene (bit) e a codificação é feita diretamente, ou seja, se o valor do gene for igual a 1 implica na utilização da variável na respectiva posição. Caso contrário essa variável não é utilizada. O indivíduo possui o número de cromossomos igual ao número máximo de variáveis que podem ser utilizadas. A Figura 1.3 ilustra a codificação de um indivíduo para seleção de variáveis para utilização em calibração multivariada.

No exemplo da Figura 1.3 o indivíduo em questão codificou a seleção das variáveis 4, 5, 6, 7, 10, 11, 12 e 13 dentre as 19 variáveis possíveis.

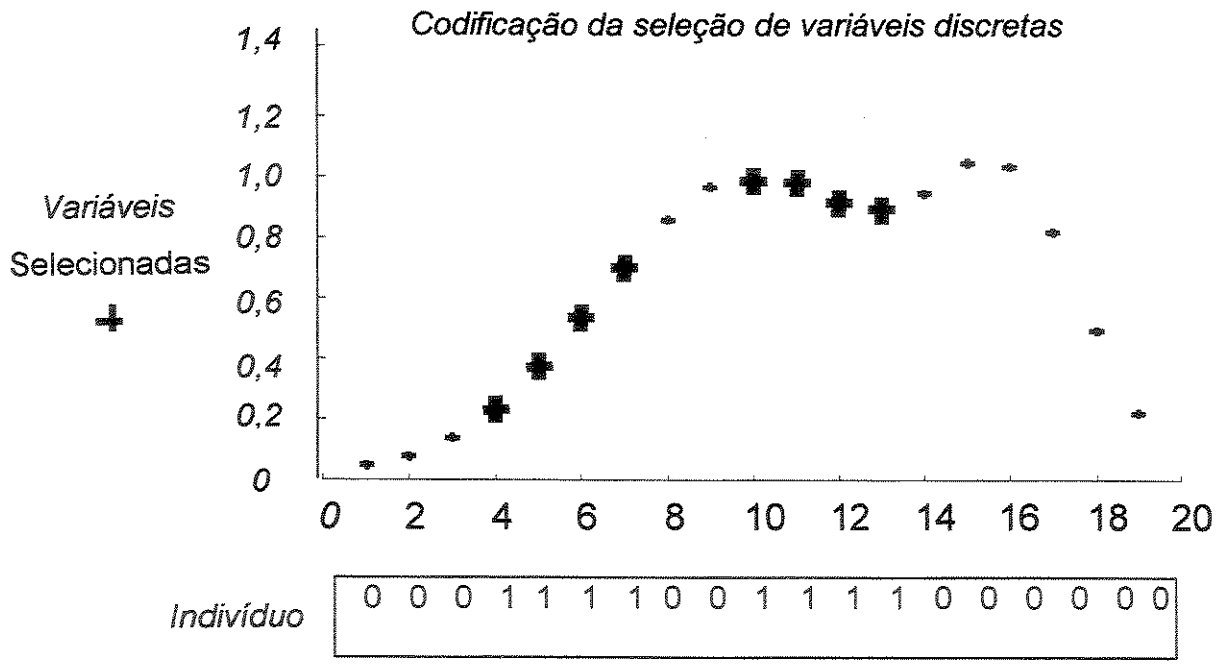


Figura 1.3 – Representação de um indivíduo para a otimização de variáveis discretas (seleção de variáveis).

1.1.2 - População inicial

A população inicial do AG normalmente é construída aleatoriamente ou utilizando uma heurística simples e rápida. Uma vez determinada a codificação do problema, gera-se um conjunto de indivíduos de modo que se possa explorar uniformemente o espaço de busca. Pode-se também iniciar o AG partindo de uma posição previamente otimizada. Nesse caso pode-se utilizar a população final de outra otimização por AG para o mesmo problema. Também pode-se adicionar, à população inicial, indivíduos que codificam pontos específicos do espaço de busca, baseando-se em conhecimento prévio sobre o sistema.

1.1.3 - Avaliação dos indivíduos

A adaptação ao problema, em um indivíduo, depende da avaliação do sistema com os parâmetros codificados pelo mesmo. A função de avaliação deve ser designada para cada problema de maneira específica. Dado um indivíduo particular, a função de avaliação retorna um número real, que reflete o nível de adaptação ao problema para o individuo em questão. Por exemplo: No caso de seleção de variáveis para calibração multivariada a função de avaliação pode ser o erro quadrático médio de previsão para o modelo construído com as variáveis selecionadas. Desse modo, os melhores indivíduos são os que apresentam os menores erros quadráticos médios de previsão.

1.1.4 - Seleção dos indivíduos

Após definir a codificação do AG (forma de representação do problema com números binários ou outra base de codificação), e gerar a população inicial, o passo seguinte é decidir como o AG procederá para seleção dos indivíduos, ou seja, como escolher os indivíduos na população que criarão os descendentes para a próxima geração. Os mecanismos de amostragem são bastante variados, sendo que, entre eles, se destacam três grupos principais segundo o grau de influência da aleatoriedade no processo:

- Amostragem direta: seleção de um subconjunto de indivíduos da população mediante um critério fixo, no estilo de "os n melhores", "os n piores", "a dedo", etc...

- Amostragem aleatória simples ou equiprovável: é determinada a mesma probabilidade para todos os elementos da população, de serem selecionados para a execução dos demais operadores (cruzamento, mutação, etc).

- Amostragem estocástica: são atribuídas probabilidades de seleção ou pontuações aos elementos da população com base na sua função de aptidão (avaliação). Nesse caso os indivíduos com melhor aptidão têm maior probabilidade de serem selecionados.

1.1.5 - Cruzamento

Pode-se dizer que a principal característica de distinção do AG em relação a outras técnicas de otimização é o uso do cruzamento.

O operador cruzamento é utilizado após o operador de seleção. Esta fase é marcada pela troca de segmentos entre "casais" de cromossomos selecionados para dar origem a novos indivíduos que formarão a população da próxima geração.

A idéia central do cruzamento é a propagação das características dos indivíduos mais aptos da população por meio de troca de segmentos de informações entre os mesmos o que dará origem a novos indivíduos.

Para a execução do cruzamento determina-se os "casais" (pares de indivíduos) que irão cruzar, de acordo com uma probabilidade de cruzamento previamente estabelecida. A mecânica do operador cruzamento é ilustrada na Figura 1.4.

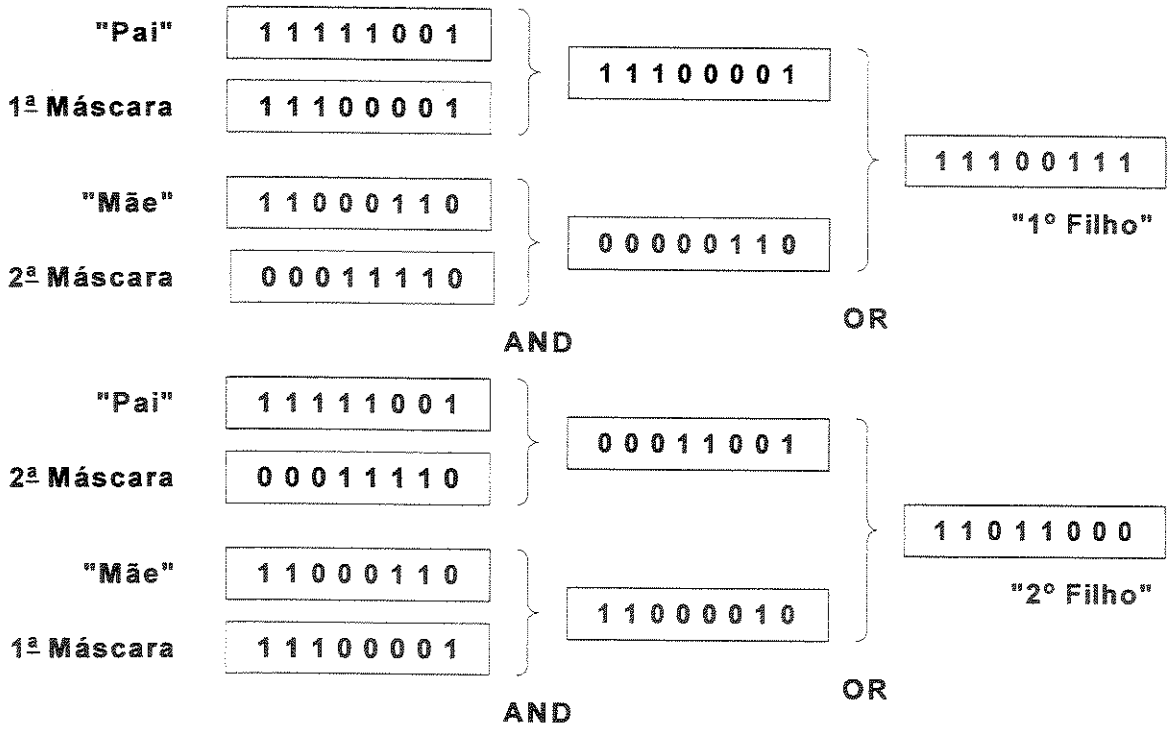


Figura 1.4 - Diagrama esquemático do operador cruzamento.

A cada geração o operador cruzamento define, aleatoriamente, vetores complementares que servem de máscaras para o cruzamento. Por uma operação booleana do tipo "AND" entre os pais e as respectivas máscaras, seguida de uma operação booleana do tipo "OR", gera-se o primeiro indivíduo resultante do cruzamento dos indivíduos em questão. O número de indivíduos geradores (pais) deve ser o mesmo do número de máscaras, e pela permutação entre esses indivíduos e as máscaras, geram-se os diversos indivíduos "filhos", como ilustrado na Figura 1.4.

O número de máscaras complemantares define o número de indivíduos que trocarão informações. Assim com duas máscaras permuta-se informações entre dois indivíduos e consequentemente geram-se dois filhos. Utilizando-se três máscaras permutam-se informações entre três indivíduos e geram-se três filhos, e assim sucessivamente. As

mascaras têm que ser complementares para que não sejam selecionados informações, nas mesmas posições, de mais de um indivíduo.

As duas formas mais comuns de reprodução sexual em AG's são com um ponto de quebra e com dois pontos de quebra. No cruzamento de um ponto, o ponto de quebra de cada cromossomo é escolhido de forma aleatória sobre a longitude da palavra que define a máscara que será usada no cruzamento, e é a partir desse ponto que se realiza a troca de material cromossômico entre os indivíduos. No cruzamento de dois pontos, como mostrado na Figura 1.4, em que os pontos de quebra foram nos elementos 3 e 7 (posições de mudança nas máscaras), procede-se de maneira similar ao cruzamento de um ponto. Na verdade as máscaras são quem definem o tipo de cruzamento, e pode-se também utilizar o cruzamento com um número maior "quebras", ou seja, com a máscara gerada aleatoriamente. Desse modo, antes de se proceder o cruzamento deve-se definir o tipo e o número de máscaras. Na prática, a cada cruzamento em cada geração, criam-se novas máscaras de acordo com parâmetros pré-estabelecidos (número de pontos de quebra e número de indivíduos que permutarão informações em cada cruzamento). Desse modo garante-se uma troca de informações não-tendenciosa ao longo dos cruzamentos.

1.1.6 - Mutação

Para a execução do operador mutação, seleciona-se aleatoriamente uma ou mais posições (genes) de um ou mais cromossomos em um indivíduo e permuta-se o seu(s) valor(es). Esse processo é mostrado na Figura 1.5.

A mutação é geralmente vista como um operador de "background", responsável pela introdução e manutenção da diversidade genética na população Ela trabalha alterando

arbitrariamente um ou mais componentes de um ou mais indivíduos escolhidas entre a descendência, logo após o cruzamento, fornecendo dessa forma meios para a introdução de novos elementos na população. Assim, a mutação assegura que a probabilidade de se chegar a qualquer ponto do espaço de busca nunca será zero. O operador de mutação é aplicado aos indivíduos com uma probabilidade dada pela taxa de mutação. Geralmente se utiliza uma taxa de mutação pequena, justamente por ser um operador genético secundário.

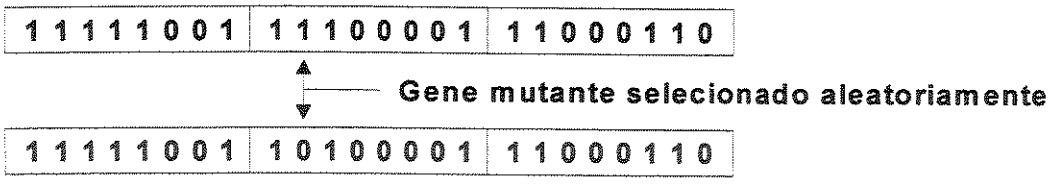


Figura 1.5 – Diagrama esquemático do operador mutação

1.1.7 - Exemplo

Para ilustração de um AG simples, suponhamos o problema de se encontrar o máximo da função $f(X) = -0{,}4\ X^2 + 10\ X + 5$, sobre o domínio dos números inteiros entre 1 e 32 (1, 2, ... 32). Evidentemente para encontrar esse ótimo, numericamente, bastaria aplicar a função em todo o domínio (entre 1 e 32) para ver o valor máximo, porém isso só é possível devido ao reduzido espaço de busca. Ademais, trata-se apenas de um exemplo ilustrativo.

Inicialmente precisa-se determinar a codificação, o tamanho da população inicial e a função de avaliação para o problema em questão. Supondo que o alfabeto utilizado para codificar os indivíduos seja constituído por {0, 1}, necessita-se então de apenas 5 dígitos para representar os 32 pontos do espaço de busca. Vale ressaltar que para esse AG cada indivíduo apresenta apenas um cromossomo, visto que há uma variável a ser otimizada, e esse cromossomo possui cinco genes. Utilizemos uma população com quatro indivíduos e

como função de avaliação a própria função f(X) que se deseja maximizar, como mostrado na Tabela 1.1.

Na Tabela 1.1, Têm-se 4 indivíduos que constituem a população inicial, bem como suas respectivas avaliações e probabilidades de que cada indivíduo seja selecionado (amostragem estocástica). Tomando uma distribuição uniforme no intervalo [0 1] para a determinação dos indivíduos selecionados, determinam-se aleatoriamente quatro números. Suponha que seja 0,31; 0,52; 0,25 e 0,80. De acordo com a probabilidade acumulada de cada indivíduo, por aproximação com a distribuição gerada, tem-se o 1º par formado pelos indivíduos 1 e 3 e o 2º par formado pelos indivíduos 1 e 4.

Pelo diagrama do Algoritmo genético mostrado na Figura 1.1 o próximo passo é executar o cruzamento dos dois pares de indivíduos. A geração subseqüente, formada pelo cruzamento e mutação dos indivíduos da Tabela 1.1, é mostrada na Tabela 1.2. O processo continua até atingir um número pré-determinado de gerações ou até o algoritmo não mais conseguir otimizar os sistema em questão. Nesse exemplo o AG já atingiu o ótimo, com o 1º indivíduo da 2ª geração (Tabela 1.2). A partir desse ponto o AG não mais consegue melhorar o sistema.

Tabela 1.1 – População inicial do exemplo do Algoritmo Genético simples

	População inicial	Valor no espaço de busca	Avaliação f(X)	$\frac{f(X)}{\sum f(X)}$ Probabilidade de seleção	Probabilidade de seleção acumulada
1	0 1 0 1 1	11	66,6	0,3513	0,3513
2	1 1 0 0 1	25	5	0,0264	0,3776
3	0 1 0 0 1	9	62,6	0,3302	0,7078
4	1 0 0 1 0	18	55,4	0,2922	1,0000
Soma			189,6		
Média			47,4		
Melhor			66,6		

No caso em questão a seleção foi feita baseada em uma distribuição uniforme com a probabilidade acumulada. Os melhores indivíduos têm maior probabilidade de serem selecionados mais de uma vez, em detrimento dos piores indivíduos que são excluídos. Esse procedimento também é conhecido por "Roleta". Observa-se claramente, através da média das avaliações, que a população como um todo melhorou significativamente. Isso foi obtido por conseqüência direta da seleção da melhor amostra por duas vezes para o cruzamento.

Esse tipo de seleção é apenas um dentre os tipos possíveis. Outra alternativa muito utilizada é a seleção pelo critério fixo "os n melhores indivíduos", Nesse caso o cruzamento da população e feito por duas vezes (com emparelhamento distinto) e se toma a metade da população resultante (os n melhores indivíduos) para seguir a geração seguinte. Em ambos os casos a convergência do AG é semelhante.

Tabela 1.2 – Continuação do AG com os indivíduos da Tabela 1.1: 2ª geração.

Emparelhamento dos indivíduos selecionados	Ponto de quebra	Descendentes	Decendentes mutados	Valor no espaço de busca	Avaliação f(X)
0 1 0 1 1	2	0 1 0 0 1	0 1 1 0 1	13	67,4
0 1 0 0 1	2	0 1 0 1 1	0 1 0 1 1	11	66,6
0 1 0 1 1	3	0 1 0 1 0	0 1 0 1 0	10	65,0
1 0 0 1 0	3	1 0 0 1 1	1 0 0 0 1	17	59,4
Soma					196,0
Média					64,6
Melhor					67,4

1.2 – Parâmetros dos Algoritmos Genéticos

É importante também, analisar de que maneira alguns parâmetros influem no comportamento dos Algoritmos Genéticos, para que se possa estabelecê-los conforme as necessidades do problema e dos recursos disponíveis. Dentre os operadores do AG básico pode-se destacar os seguintes parâmetros:

a) Tamanho da População: O tamanho da população determina o número de cromossomos na população, afetando diretamente o desempenho global e a eficiência dos AG's. Com uma população pequena o desempenho pode cair, pois deste modo a população fornece uma pequena cobertura do espaço de busca do problema. Uma grande população geralmente fornece uma cobertura representativa do domínio do problema, além de prevenir convergências prematuras para soluções locais ao invés de globais. No entanto, para se trabalhar com grandes populações, são necessários maiores recursos computacionais, ou que o algoritmo trabalhe por um período de tempo maior.

b) Taxa de Cruzamento: Esse parâmetro é também chamado de probabilidade de cruzamento, e define a fração dos indivíduos que irão cruzar em cada geração. Quanto maior for esta taxa, mais rapidamente novas estruturas serão introduzidas na população. Mas se esta for muito alta, a maior parte da população será substituída, e pode ocorrer perda de estruturas de alta aptidão. Com um valor baixo, o algoritmo pode tornar-se muito lento.

c) Tipo de Cruzamento: O tipo de cruzamento a ser utilizado determina a forma como se procederá a troca de segmentos de informação entre os pares de cromossomos selecionados para cruzamento. O cruzamento pode ocorrer separando os cromossomos (ou os indivíduos) em 2, 3 ou um número maior de pedaços com informações genéticas. Além disso, pode-se cruzar 3 ou mais indivíduos simultaneamente, de modo a obter uma maior diversidade. Basicamente, definir o tipo de cruzamento é equivalente a definir o tipo e o número de máscaras utilizadas.

d) Taxa de Mutação: Esse parâmetro é também chamado de probabilidade de mutação, e define a fração dos indivíduos que sofrerão mutação em cada geração. A mutação é utilizada para dar nova informação para a população e também para prevenir que a população se sature com cromossomos semelhantes (convergência prematura). Uma baixa taxa de mutação previne que uma dada posição fique estagnada em um valor, além de possibilitar que se chegue em qualquer ponto do espaço de busca. Com uma taxa muito alta, a busca se torna essencialmente aleatória além de aumentar muito a possibilidade de que uma boa solução seja destruída.

1.3 – Modificações do algoritmo Genético básico

Além dos operadores do AG básico, podem ser utilizados vários outros operadores genéticos para incorporar rotinas específicas ao processo de otimização, aumentar a diversidade da população ou mesmo para refinar a otimização realizada. Tal implementação é denominada algoritmo genético não convencional (AGNC). Não existe um número definido de possíveis operadores genéticos, cabendo adequar os tipos de operadores (além do AG básico) ao problema em questão. Porém, como complementação aos operadores do AG básico, tem-se utilizado com maior freqüência os operadores elitismo (ou dominância), migração, modificações do critério de parada, cruzamento lateral, utilização de algoritmos genéticos paralelos, dentre outros.

1.3.1 – Operador elitismo

Após um determinado número de gerações, é possível observar a convergência da população, com a manutenção de indivíduos com melhor ajuste (maior aptidão), contudo, sem acarretar numa convergência da otimização a um valor ótimo do espaço de busca. Nesse ponto, utiliza-se o operador elitismo,que consiste em submeter os indivíduos com melhor ajuste a um procedimento de busca intensiva, baseado nos próprios operadores genéticos, através de cruzamentos e/ou mutações, ou em outros modelos de otimização. Posteriormente, o resultado dessa otimização é novamente introduzido ao AG e o processo segue normalmente.

1.3.2 – Operador migração

Outra forma de garantir que o AG não fique preso a um ótimo local, dentro do espaço de busca, é com a utilização do operador migração. A migração consiste em alterar o valor do número de indivíduos da população durante a otimização. Desse modo pode-se diminuir a população (emigração) e eliminar indivíduos ruins que prejudicam a convergência a um ótimo global. Pode-se também aumentar o tamanho da população (imigração). Nesse caso, adicionam-se novos indivíduos, gerados aleatoriamente, de modo a introduzir novos pontos do espaço de busca à população corrente, e desse modo aumentar a diversidade. Adicionalmente pode-se combinar os processos de emigração e imigração de modo a aumentar e diminuir o tamanho da população periodicamente. Esse procedimento aumenta

a probabilidade de interação (troca de informação) entre indivíduos bons com outros indivíduos ainda não explorados no espaço de busca, diminuindo a probabilidade de que o processo fique preso a um ótimo local.

1.3.3 - Modificação dos critérios de parada

Um dos critérios de parada, utilizados com sucesso nos Algoritmos Genéticos não convencionais, funciona da seguinte forma: quando uma população passa a indivíduos muito semelhantes (isso pode ser detectado comparando os seus níveis de aptidão), possivelmente teremos uma taxa de renovação baixa. Neste momento aciona-se um critério de parada interna do AG, partindo para uma etapa de diversificação da população, mas aproveitando informações dos melhores indivíduos. Feita a atualização da população, reinicia-se o operador genético para a construção de novas populações. Se, após algumas diversificações consecutivas, o melhor indivíduo não for atualizado, o critério de parada do algoritmo é acionado.

1.3.4 - Operador Cruzamento Lateral

Nos casos de utilização do AG para a seleção de variáveis, em que as diversas variáveis do espaço de busca são muito correlacionadas, não há muita vantagem de se selecionar variáveis escolhidas pontualmente (mesmo pelas regras do AG). Se uma determinada variável é realmente importante, provavelmente as variáveis que possuem alta correlação

com a variável selecionada também são importantes. As variáveis que representam as absorbâncias de um mesmo pico formam um exemplo claro desse caso. Nesses casos pode-se utilizar o operador cruzamento lateral. O cruzamento lateral consiste em selecionar periodicamente variáveis que possuem grande covariância com outras que já foram escolhidas nos melhores indivíduos.

1.3.5 – Algoritmos Genéticos paralelos

O tempo de execução requerido, quando comparado com modelos de otimização com outras heurísticas, é uma das principais deficiências dos Algoritmos Genéticos, visto que os mesmos são seqüenciais, ou seja, as diversas gerações são realizadas uma após a outra. Tal limitação tem em parte se resolvido com a introdução dos Algoritmos Genéticos não Convencionais, onde são incluídas outras técnicas para melhorar o desempenho também no tempo computacional exigido. Contribuições nesta área foram obtidas também através de um melhor controle dos seus parâmetros e de critérios de parada mais eficientes. Mas com certeza o caminho mais promissor para resolver este problema é aproveitar o paralelismo intrínseco existente nos algoritmos genéticos. De fato, os algoritmos genéticos são inspirados em um processo evolutivo paralelo de uma população de indivíduos. Contudo, deve-se ter em mente que o processo de paralelização dos AGs não é imediato pois necessita de um controle global no passo de seleção, o que requer uma interação entre processadores podendo produzir com isso altos custos de comunicação.

Existem vários modelos propostos que procuram reduzir estes problemas viabilizando o processo de paralelização. Estes modelos em geral se agrupam em três grandes categorias:

Modelo ilha ou modelo de granularidade espessa: Neste modelo, várias sub-populações isoladas evoluem em paralelo e periodicamente trocam informações através da migração dos seus melhores indivíduos para sub-populações vizinhas.

Modelo de granularidade fina: Também conhecido como "neighbourhood model", onde uma única população evolui e cada indivíduo é colocado em uma célula de uma grade planar. O processo de seleção e de cruzamento é aplicado somente entre indivíduos vizinhos na grade (de acordo com uma topologia definida).

Modelo Panmítico: São essencialmente versões paralelas de algoritmos genéticos seqüenciais ordinários. Eles operam sobre uma população global, são normalmente síncronos e apresentam granularidade de média para grossa. Este modelo é adequado a arquiteturas paralelas com memória compartilhada.

1.4 – Análise teórica dos Algoritmos genéticos

Até o momento, foi descrita a forma de funcionamento e implementação de um algoritmo genético. Sem uma interpretação matemático/estatística mais formal, pode-se pensar que os AGs apenas processam possíveis soluções em um problema de otimização.

Se apenas fosse esse o caso, não haveria muito sentido na utilização da codificação binária. O AG processaria da mesma forma indivíduos com variáveis representadas com valores decimais (de ponto flutuante), sem ainda a necessidade de ficar convertendo entre binário de decimal toda vez que os indivíduos fossem avaliados. A razão da utilização de elementos binários é que os AGs não processam apenas indivíduos, mas sim padrões de informação entre esses indivíduos. Tendo em vista que, dependendo do problema em questão, cada indivíduo pode conter vários padrões de informação, a eficiência do processo de otimização se multiplica. É interessante observar que se o sistema em questão for relativamente simples, com poucos padrões de informação, a utilização de um AG para resolvê-lo pode não ser a melhor opção. Seria como utilizar um canhão para matar uma formiga.

1.4.1 – Conceitos básicos e definições.

A fundamentação teórica dos AGs geralmente é representada em termos de esquemas [3]. Um esquema representa um conjunto de indivíduos, ou partes destes (cromossomos), que apresentam semelhanças em sua composição. O esquema é um padrão de semelhança no vetor binário (ou string) que compõe os cromossomos de alguns indivíduos em um AG, no qual os elementos que se distinguem entre os diversos indivíduos é representado por um asterisco. Assim o alfabeto que representa os diversos genes passa a ser { 0, 1, *). Por exemplo: O esquema (*10*0101) representa o conjunto de quatro indivíduos: [01000101] , [01010101], [11000101] e [11010101].

Os esquemas permitem escrever, de forma sintética, configurações de indivíduos que apresentam características interessantes para a resolução de um determinado problema. Desse modo o esquema (11001001) representa apenas um indivíduo (não possui o caracter "*"), e o esquema (********) representa todas os indivíduos formados pela permutação de 8 genes. Vale salientar que o esquema é apenas uma representação para que se possa entender o funcionamento do AG. Na prática o AG não trabalha com esquemas, mas sim com os indivíduos. Tendo isso posto, podemos então definir alguns parâmetros que servirão de base para o formalismo matemático.

- Alfabeto utilizado (V): $V = \{0,1\}$

- Indivíduo (A): $A = a_1, a_2, \ldots a_l$ com $a_i \in V$

- População (P): $P(t) = \{A_1,A_2,A_3,\ldots A_N\}$ Conjunto de N indivíduos na geração t

Podemos então expandir o alfabeto para permitir a representação de esquemas:

- Alfabeto expandido (V+): $V+ = \{0, 1, *\}$

- Esquema (H): $H = h_1,h_2,\ldots,h_l$ com $h_i \in V+$

Definida a representação utilizada, pode-se observar as seguintes propriedades relativas aos esquemas:

1- Se um esquema contém k símbolos de indiferença (*) então representa 2^k indivíduos distintos;

2- Um indivíduo binário de comprimento l se encaixa em 2^l esquemas diferentes;

3- Considerando os indivíduos binários de comprimento l, existem 3^l possíveis esquemas;

4- Uma população de n indivíduos de comprimento l contém entre 2^l e $n2^l$ esquemas diferentes;

Definidas as propriedades, é oportuno definir dois conceitos de grande importância para um dado esquema H:

- Ordem de um esquema O(H): Número de valores fixos na representação de um esquema. Por exemplo:

10*110* é de ordem 5: O(10*110*) = 5

0****** é de ordem 1: O(0*******) = 1

- Longitude característica de um esquema d(H): distância em dígitos entre o primeiro e o último valor fixo na representação de um esquema. Por exemplo:

10*110* tem d(H) 5: d(10*110*) = 6-1 = 5

0****** tem d(H) 0: d(0******) = 1-1 = 0

Dado que um esquema H representa $2^{l-O(H)}$ indivíduos, quanto maior for a ordem do esquema, menos indivíduos ele representará. Logo, pode-se calcular a probabilidade de sobrevivência de um esquema em relação aos cruzamentos.

Por exemplo, o esquema S = (****00**1*) possui o(S) = 3 e d(S) = 9-5 = 4. A aptidão média da população na geração t (AptMed(P(t))), é a média das aptidões de todos os

indivíduos da população na geração t, ou em equivalência, a aptidão do esquema(* ...*) em p(t):

$$\text{AptMed(P(t) = aptidão((*...*),P(T)) = } \frac{1}{n}\sum_{V_i \in P(t)} \text{aptidão}\left(V_i\right)$$ onde n é o número de indivíduos de P(t).

A aptidão relativa de S em P(t):

$$\text{AptMed(S,P(t)) = } \frac{aptidão(S,P(t))}{AptMed(P(t))}$$

Por fim, sendo Pcruz e Pmut as probabilidades de cruzamento e mutação de um AG respectivamente, a presença de um certo esquema S na população P(t) evolui em média de acordo com a Equação 1.1:

$$\overline{\xi}\left(S,P(t+1)\right) \geq \overline{\xi}\left(S,P(t)\right) - AptMed\left(S,P(t)\right) - \left(1 - \frac{Ppop\ d(S)}{l-1} - Pmut\ O(S)\right) \qquad 1.1$$

Traduzindo a Equação 1.1, temos que a esperança média da aptidão relativa de um determinado esquema em uma população numa geração específica é sempre maior que a da geração anterior quando S representa indivíduos com alta aptidão, ou no mínimo igual. Desse modo a presença do esquema S evolui estatisticamente de modo exponencial ao longo das sucessivas gerações. Vale ressaltar que as aptidões são normalizadas entre 0 e 1 de modo que a cada geração, os melhores indivíduos possuem aptidão perto de 1 enquanto os piores perto de zero.

Capítulo 2

Seleção de freqüências de Fourier para calibração multivariada com PLS

Uma das primeiras preocupações da química analítica é a modelagem da resposta de um sensor, ou conjunto de sensores, como função das variações das concentrações de espécies químicas, ou ainda como função de uma outra propriedade qualquer. Esse tipo de modelagem é chamada calibração, e após estabelecida, a mesma pode ser usada para a previsão da concentração (ou outra propriedade qualquer) de espécies químicas em outras amostras a partir de suas respostas ao(s) sensor(es) em questão.

A calibração multivariada estende as idéias básicas de modelagem e previsão usadas em calibração univariada, onde há apenas uma variável dependente como função de uma variável independente, para várias variáveis dependentes e independentes.

A análise quimiométrica de dados, em particular as técnicas de calibração multivariada [8 -12] têm sido amplamente utilizadas em análises químicas. Geralmente explora-se o comportamento linear entre alguma propriedade multivariada facilmente monitorável e alguma(s) propriedade(s) que se deseja determinar, como por exemplo: a linearidade entre o espectro de absorção e a concentração de espécies químicas em solução (lei de Beer). Porém, frequentemente existem interferentes que mascaram essa linearidade, ou até mesmo efeito cinérgico (ou antagônico) entre as espécies responsáveis pela(s) propriedade(s) que se deseja determinar, e/ou interferentes. Nesses casos utiliza-se um maior número de parâmetros para construir o modelo de calibração multivariada de modo a modelar essas "não linearidades". Alternativamente, tem-se mostrado promissor a utilização de técnicas não lineares para a calibração multivariada, dependendo do grau de interferência dos dados em questão.

2.1 – Mínimos quadrados Parciais – PLS (Partial Least Squares)

O modelo mais simples em calibração multivariada é tão somente a resolução de um sistema de equações lineares em uma regressão linear múltipla [9 -12] (MLR – Mutiple Linear Regression). Em uma MLR cada variável dependente (resposta y) é expressa como uma combinação linear das variáveis independentes (o conjunto de n absorbâncias x_i das diversas espécies).

$$y = \sum b_i x_i + e \qquad 2.1$$

onde "b_i" são os coeficientes de regressão e "e" é o termo que expressa o erro. Quando se toma os dados relativos às múltiplas análises, obtém-se um conjunto de equações lineares representadas em forma matricial:

$$\mathbf{y = Xb + e} \qquad 2.2$$

ou no caso de haver mais de uma variável dependente:

$$\mathbf{Y = XB + E} \qquad 2.3$$

Essa matriz de coeficientes **B** é então obtida pelo método dos mínimos quadrados:

$$\mathbf{B = (X^tX)^{-1}X^tY} \qquad 2.4$$

O modelo de MLR é muito eficiente quando não se tem correlação entre as variáveis independentes, caso contrário pode-se cair em um sistema de equações mal condicionado. Esse problema pode ser evitado através de um planejamento experimental adequado, com seleção de variáveis independentes ortogonais. Porém, nem sempre é possível trabalhar com variáveis independentes totalmente ortogonais.

Muitas vezes não é possível utilizar o modelo direto de calibração multivariada (resposta instrumental como função da concentração). A grande maioria dos instrumentos

analíticos apresenta o inverso, ou seja, a concentração das espécies (ou as propriedades de interesse) como função da resposta instrumental. Nesses casos a MLR dificilmente poderá ser usada, visto que o número de variáveis (canais das respostas instrumentais, por exemplo, a absorbância em diversos comprimentos de onda) geralmente é superior ao número de ensaios (amostras analisadas) e o sistema de equações a ser resolvido torna-se indeterminado. Nesse caso pode-se utilizar técnicas de ortogonalização baseadas em mudança de base vetorial.

Provavelmente a técnica de ortogonalização mais difundida é a análise em componentes principais – PCA [9 -13] (Principal component analysis). O primeiro passo nesse modelo é a formação de uma matriz de variância/covariância dos dados. Essa matriz isola a fonte de variação dos dados.

$$\mathbf{Z} = \mathbf{X}^t\mathbf{X} \qquad 2.5$$

onde **Z** é a matriz de variância/covariância e **X** a matriz de dados previamente escalonada (subtraída da média em cada coluna). A matriz de covariância é então diagonalizada por uma transformação unitária

$$\Lambda = \mathbf{V}^{-1}\,\mathbf{Z}\,\mathbf{V} \qquad 2.6$$

onde Λ é uma matriz diagonal cujos elementos são autovalores de **Z** e **V** é a matriz de autovetores, denominados "loadings". Basicamente os "loadings" formam uma nova base ortonormal que explica os dados (**X**), e a projeção dos dados nessa base é denominada "scores". Desse modo os dados são decompostos por um conjunto de vetores "loadings" e "scores"

$$\mathbf{X} = \mathbf{T}\,\mathbf{V}^t \qquad 2.7$$

O conjunto "loading" e "scores" é denominado "componente principal". Um método conveniente que tem se tornado padrão para o cálculo dos autovetores na Equação 2.6 é a decomposição em valores singulares, disponível na maioria dos "softwares" para análise numérica. Nela a matriz de dados é decomposta no produto de três matrizes:

$$\mathbf{X} = \mathbf{U}\,\mathbf{S}\,\mathbf{V}^t \qquad 2.8$$

onde **U** é a matriz dos autovetores de $\mathbf{X}\mathbf{X}^t$, **S** é uma matriz diagonal de valores singulares (raiz quadrada dos autovalores) e **V** é a matriz dos autovetores de $\mathbf{X}^t\mathbf{X}$. A matriz de autovetores **V**, obtida por decomposição em valores singulares é equivalente a matriz **V** da Equação 2.6, e o produto da matriz de autovetores **U** pela matriz de valores singulares **S** é equivalente à matriz de "scores" **T** da Equação 2.7.

O número máximo de componentes principais obtidos é igual ao número de vetores de dados utilizados (posto da matriz de dados independentes), sendo que nem todas as componentes principais possuem informações úteis. As últimas componentes principais modelam o ruído inerente aos dados. Desse modo a eliminação das últimas componentes principais freqüentemente aumenta a relação sinal/ruído.

A regressão por componentes principais [9 -13] (PCR – Principal Component Regression) surge então como uma alternativa imediata à MLR. Ela nada mais é do que uma análise por componentes principais (PCA) seguida de uma regressão linear múltipla (MLR). A utilização da PCA restringe o número de variáveis ao número de componentes principais (variáveis ortogonais) utilizadas, e desse modo, ao mesmo tempo que facilita a resolução do sistema de equações lineares, também elimina ruído explicado pelas últimas componentes principais que são truncadas.

Dentre os modelos de calibração multivariada mais utilizados, certamente está o PLS [9-13] ou a regressão por PLS (PLSR). A PLSR estende o conceito do modelo inverso (propriedade como função da medida experimental) trocando as variáveis originais por um sub-conjunto truncado das variáveis latentes dos dados originais.

$$\mathbf{X} = \mathbf{T}\mathbf{P}^t + \mathbf{E}_x = \Sigma\, t_h\, p_h^t + \mathbf{E}_x \qquad 2.9$$

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^t + \mathbf{E}_y = \Sigma\, u_h\, q_h^t + \mathbf{E}_y \qquad 2.10$$

$$\hat{u}_h = b_h\, t_h \qquad 2.11$$

Onde **X** é a matriz de dados, **Y** é a matriz de resposta (concentrações, por exemplo), **T** e **U** são os "scores" para as duas matrizes de dados, **P** e **Q** são os respectivos "loadings", $\mathbf{E}_x$ e $\mathbf{E}_y$ são os respectivos resíduos compostos pelas variáveis latentes descartadas e b_h é o coeficiente do modelo linear para cada variável latente.

O modelo PLSR é obtido através de um processo iterativo, no qual se otimiza ao mesmo tempo a projeção das amostras sobre o(s) "loading(s)", para a determinação dos "scores", e o ajuste por uma função linear dos "scores" da matriz **X** aos "scores" da matriz **Y** de modo a minimizar os desvios. Essa otimização simultânea ocasiona pequenas distorções nas direções dos "loadings", de modo que, rigorosamente eles perdem a ortogonalidade, levando à pequenas redundâncias de informação. Porém são essas pequenas redundâncias que otimizam a relação linear entre os "scores". Essas distorções da ortogonalidade entre os fatores no PLS é que fazem com que os mesmos não sejam componentes principais (ortogonais) mas sim variáveis latentes.

2.2 – Utilização das freqüências de Fourier na calibração multivariada

Nem sempre é possível obter uma relação linear entre a medida instrumental e a concentração (propriedade de interesse). Torna-se então necessário utilizar modelos não-lineares para a calibração multivariada. Em princípio, existe uma infinidade de transformações não-lineares possíveis de serem utilizadas, porém a transformada de Fourier [14,15] tem tomado grande importância devido a sua interpretação física e fácil implementação. A transformada de Fourier converte os dados instrumentais, no domínio do tempo/espaço, em dados no domínio das freqüências/energia. Desse modo, pode-se associar bandas largas de absorção a sinais de baixa freqüência, e ruídos a sinais de alta freqüência. Surge então, quase que espontaneamente, a idéia de utilizar a diferença de freqüência do sinal analítico como elemento de separação da informação relevante dos dados instrumentais.

A forma mais rápida e fácil de se implementar a transformada de Fourier em dados instrumentais discretos (digitalizados) é através da transformada rápida de Fourier (FFT – fast Fourier transform) [14,15], dada pela Equações 2.12, 2.13, 2.14 e 2.15 .

$$F(\omega_p) = \frac{1}{2N} \sum_{k=0}^{2N-1} f(t_k) e^{i\omega_p t_k} \qquad 2.12$$

$$e^{i\theta} = \cos(\theta) + i\,\mathrm{sen}(\theta) \qquad 2.13$$

$$\omega_p = 2\pi\, p/T \qquad 2.14$$

$$t_k = 0, \frac{T}{2N}, \frac{2T}{2N}, \cdots, \frac{(2N-1)T}{2N} \quad k = 0,1,2,\cdots 2N-1 \qquad 2.15$$

Onde ω_p é a freqüência e T é o período, N é o número de pontos do vetor de dados, t_k é o indexador dos dados (sinais discretos) no k-ésimo tempo (medida) e $f(t_k)$ são os dados propriamente ditos.

Pela identidade de Euler (Equação 2.13) explicita-se a interpretação, em termos de freqüência, para os coeficientes da FFT ($F(\omega_p)$).

O algoritmo da FFT admite que cada vetor discreto de dados, obtido com a digitalização das diversas variáveis para cada amostra, representa um período de uma função periódica infinita (para cada amostra) e normaliza os coeficientes de Fourier obtidos entre 0 e 2π. Adicionalmente, se houver uma descontinuidade no vetor de dados (um pico saturado, por exemplo), a FFT adicionará freqüências não existentes aos vetores de coeficientes de Fourier obtidos, na tentativa de modelar a descontinuidade. Na prática isso limita a utilização da FFT, pois nesses casos a seleção de freqüências pode distorcer a informação contida nas amostras

Como as funções seno e cosseno da transformada de Fourier formam uma base vetorial, além de converter os sinais analíticos para o domínio de freqüências, a FFT simultaneamente, também descorrelaciona as variáveis dos dados analíticos, ou seja, os coeficientes de Fourier podem ser truncados de modo a eliminar informação não relevante dos dados originais. Isso pode ser feito removendo-se as freqüências que compõem o ruído experimental do sinal analítico. Esse procedimento compacta a informação relevante em poucos coeficientes de Fourier. Com poucas freqüências, o número de variáveis utilizadas é menor, e conseqüentemente o trabalho computacional para a modelagem de calibração é mais rápido. Adicionalmente, como as freqüências são descorrelacionadas (são ortogonais

entre si), a FFT elimina eventuais problemas de ordem numérica nos procedimentos de modelagem e calibração.

2.3 – Seleção de freqüências por algoritmo genético

Nos modelos lineares de calibração multivariada considerados "padrões", (PLS e PCR), basicamente, utiliza-se uma transformação linear de mudança da base para diminuir a dimensão da base de dados (número de variáveis), de modo a eliminar ou diminuir a redundância de informação (variáveis correlacionadas) gerando uma melhor separação dos padrões de informação, e trunca-se o tamanho dessa base (segundo o número de fatores utilizados) para eliminar ruído nos dados. Desse modo facilita-se a determinação de uma relação linear entre os dados (na nova base) e a(s) propriedade(s) que se deseja determinar.

Devido à quantidade de interferência e ruídos encontrados em análises reais, nem sempre é possível encontrar uma relação linear (segundo essas técnicas mencionadas) de modo a obter previsões experimentais com erros aceitáveis. Assim, propôs-se um modelo baseado em transformações não-lineares para a mudança de base, seguidas de uma truncagem da nova base otimizada por algoritmo genético. Esse modelo otimiza o conceito de um filtro de transformada de Fourier [16], ou filtro de freqüências.

O filtro de transformada de Fourier [16] é uma ferramenta matemática bem estabelecida e amplamente utilizada para a minimização de ruído em sinais analíticos. A principal vantagem do filtro de transformada de Fourier sobre os filtros de suavização de ruído como, por exemplo, o filtro de média móvel e o filtro de Savitzky – Golay [17], é a sua

atuação direta sobre as freqüências que compõem o sinal. Assim, o filtro de transformada de Fourier permite, de um modo simples, atuar seletivamente sobre o ruído. A razão para este fato é que geralmente o ruído está presente nas freqüências altas e o sinal analítico útil, nas freqüências baixas.

O filtro de transformada de Fourier atua em três etapas: na primeira etapa, através da transformada de Fourier direta, identificam-se as freqüências que compõem o sinal, na segunda etapa eliminam-se as freqüências altas e na terceira etapa, através da transformada de Fourier inversa obtém-se o sinal original, com ruído minimizado.

As operações utilizadas no filtro de transformada de Fourier são simples de entender e fáceis de se implementar computacionalmente. Todavia, existem dois fatores que limitam sua eficácia. O primeiro deles é que alguns tipos de ruído, o ruído gaussiano, por exemplo, apresenta freqüências distribuídas sobre todo o espectro de potências do sinal analítico, inclusive nas freqüências baixas, que possuem informação útil sobre o sinal analítico. Devido a esta característica, o filtro de transformada de Fourier só permite minimizar este tipo de ruído, porém sem eliminá-lo completamente. Além disso, mesmo nos casos em que se deseja remover ruído cíclico, os quais possuem freqüência característica e bem definida, a seleção da freqüência de corte para sinais complexos, em que está envolvido um grande número de freqüências, torna-se uma tarefa extremamente difícil. Assim sendo, o grande desafio na aplicação do filtro de transformada de Fourier é a seleção da freqüência de corte.

Até o momento os métodos utilizados para a seleção da freqüência de corte são fundamentados exclusivamente na análise do espectro de potência do sinal e no conhecimento prévio do espectro de potencias do ruído, caso contrário, a aplicação do filtro de transformada de Fourier não é possível.

Tendo estas dificuldades em mente, neste trabalho elaborou-se um novo método para a seleção das freqüências relevantes (que possuem informação com respeito à(s) propriedade(s) de interesse). O fundamento do método é aliar o filtro de transformada de Fourier ao método dos mínimos quadrados parciais PLS, de tal modo que somente as freqüências que possuam informação útil a respeito do sinal analítico, isto é, aquelas freqüências que possuam alta correlação com as variáveis independentes (concentração, por exemplo), sejam conservadas. Idealmente, este método deve remover do sinal analítico não somente ruído, mas toda informação não correlacionada à(s) variável(is) dependente(s), possibilitando reunir em um só método várias etapas do pré-processamento de sinais.

Para realizar a seleção das freqüências relevantes utilizou-se um algoritmo genético. Desse modo, o AG seleciona quais as freqüências são relevantes para uma calibração multivariada com o PLS. Esse procedimento não só identifica a freqüência de corte como também seleciona freqüências em faixas descontínuas, ou seja, determina o conjunto de freqüências, não necessariamente consecutivas, com informação relevante.

Nesse ponto poder-se-ia utilizar qualquer modelo de calibração multivariada, visto que quem faz a otimização das freqüências é o AG (e não o modelo de calibração multivariada). Optou-se, porém, pelo PLS porque o mesmo leva em consideração a informação tanto das variáveis independentes (freqüências selecionadas pelo AG) quanto das variáveis dependentes (propriedades de interesse). Ao selecionar as freqüências o AG leva em consideração o ajuste fino, produzido pelo PLS, entre as freqüências selecionadas e a propriedade de interesse. Adicionalmente, ao aplicar a FFT nos dados originais, quebra-se a linearidade entre as variáveis independentes originais (análises) e as variáveis dependentes (respostas), pois a FFT é uma transformação não linear. O PLS pode então ser

utilizado com um número maior de fatores para modelar as não-linearidades introduzidas pela FFT.

A Figura 2.1 ilustra um diagrama esquemático (em blocos) do procedimento e implementação da calibração multivariada com PLS utilizando freqüências de Fourier selecionadas por algoritmo genético.

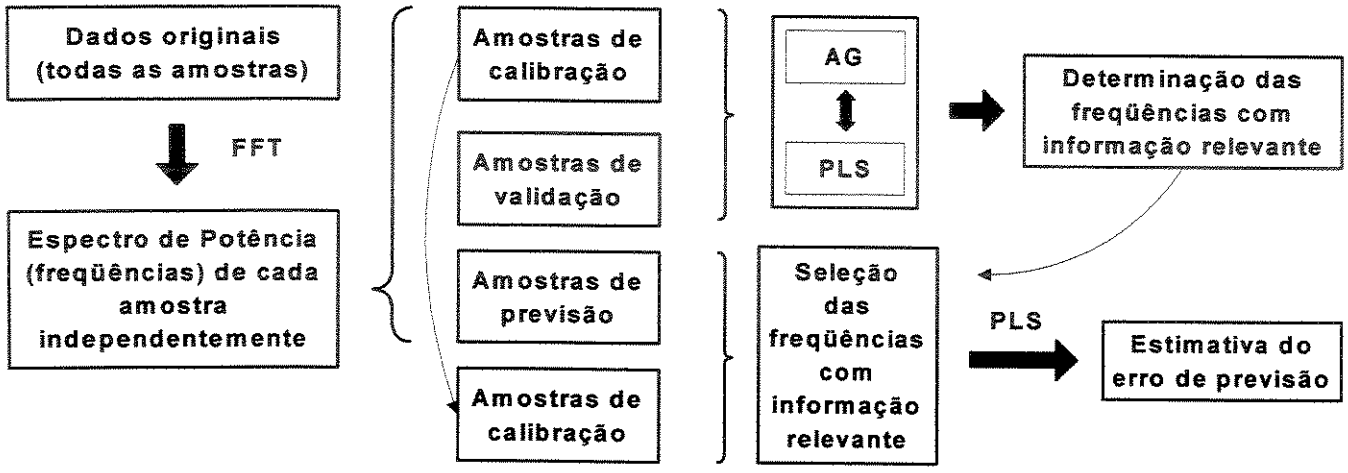


Figura 2.1- Diagrama esquemático do modelo de seleção de freqüências de Fourier por AG para calibração multivariada por PLS.

A partir da digitalização dos dados analíticos, cada amostra forma um vetor (linha) em que cada coluna representa uma variável independente (absorbância em um comprimento de onda específico, por exemplo), doravante chamado simplesmente de amostra para maior simplicidade.

Em cada amostra aplicou-se a FFT de modo a obter os espectros de potência (intensidade em função das freqüências) das mesmas. Nesse ponto existe uma diminuição do número de variáveis pela metade, visto que por simetria da base da FFT (funções senos e cossenos) a informação contida na primeira metade dos coeficientes de Fourier (0 - π), em

termos da quantificação, é a mesma contida na metade subseqüente dos coeficientes. A segunda metade do vetor de coeficientes de Fourier (π - 2π) apenas se torna relevante se a fase, ou seja, a posição relativa dos picos for relevante. No processo de calibração apenas a amplitude traz informações quantitativas. Seria, portanto, redundante trabalhar com todos os coeficientes de Fourier obtidos.

Na seqüência, as amostras são separadas em três conjuntos: calibração, validação e previsão. Os conjuntos de calibração e validação são utilizados pelo AG com o PLS para selecionar as freqüências que contém informação relevante.

Para cada indivíduo de cada geração do AG, constrói-se um modelo de calibração com o PLS para as amostras de calibração utilizando as freqüências determinadas pelo respectivo indivíduo. Com o modelo de calibração construído executa-se a previsão das amostras de calibração e de validação. O erro médio obtido nas amostras de validação é então estimado pelo erro quadrático médio de validação (RMSEV- root mean square error for validation) [8]. O erro médio obtido nas amostras de calibração é estimado pelo erro quadrático médio de calibração (RMSEC - root mean square error for calibration) [8]. Combinando-se o RMSEC e o RMSEV pode-se avaliar o respectivo indivíduo. Idealmente ambos devem ser baixos, indicando as variáveis selecionadas realmente apresentam informação relevante e com mesma ordem de grandeza, indicando que o modelo não é sobre-ajustado para as amostras em questão. Após a avaliação dos indivíduos, o AG segue até convergir e determinar o indivíduo ótimo. Ou seja, no fim do processo, o AG determina quais são as freqüências que possuem informação relevante.

Para avaliar o conjunto de freqüências selecionadas contrói-se um modelo de PLS com as amostras de calibração e estima-se a(s) propriedades de interesse para as amostras

de previsão. O erro médio é então estimado pelo erro quadrático médio para as amostras de previsão (RMSEP – root mean square error for prediction) [8]. Novamente no caso ideal o RMSEP deve ser baixo e ter a mesma ordem de grandeza do RMSEV e do RMSEC. É importante ressaltar que as amostras de calibração não podem participar do AG na seleção de variáveis, caso contrário, provavelmente a seleção de variáveis (freqüências) torna-se "viciada", ou seja, sobre-ajustada para os dados em questão, e não se pode extrapolar o modelo para outras amostras. Elas apenas são utilizadas para avaliar o modelo construído com as freqüências selecionadas pelo AG.

Pode-se também utilizar os erros quadráticos médios na forma relativa percentual. Esse procedimento facilita a utilização de critérios de convergência do AG, bem como da avaliação do modelo construído.

O modelo proposto foi implementado em ambiente computacional matricial Matlab versão 5.2 [18]. Adicionalmente, foram construídas interfaces gráficas (também em ambiente Matlab) para auxiliar a confecção do modelo construído, permitindo ao usuário, a execução fácil e intuitiva do modelo proposto.

Capítulo 3

Calibração multivariada com RBFN

Estendendo o conceito de transformação não-linear como alternativa para os casos em que não é possível encontrar uma relação linear entre os dados da análise instrumental e a(s) propriedade(s) de interesse, pode-se também utilizar uma Rede Neural [19, 20] para executar a calibração multivariada.

Uma rede neural (RN), também conhecido por MLP (multilayer perceptron), é um aproximador universal, ou seja, é um arranjo de funções matemáticas dispostas em rede de modo interconectado, que recebe como argumento um conjunto de dados de entrada e retorna um conjunto de dados de saída. Uma MLP é formada por uma camada de entrada que recebe os dados instrumentais, uma ou mais camadas intermediárias, que recebe os dados da camada de entrada ponderados por um fator de peso, e o resultado de cada camada serve de argumento para a camada seguinte. Por fim, o processo termina em uma camada de saída que idealmente deve retornar a(s) propriedade(s) de interesse. A Figura 3.1 ilustra o modelo de uma rede Neural (MLP) com apenas uma camada intermediária. Nesse exemplo, um espectro vibracional é digitalizado e os valores de absorbância nos diversos comprimentos de onda servem de argumento para a camada de entrada. Cada nó da rede é um "neurônio" artificial e é representado por uma função de transferência. Cada sinapse, ou seja, cada conexão entre "neurônios", toma a saída do neurônio da camada anterior e a pondera por um peso, que define a "força da sinapse" (em analogia aos sistemas biológicos). O somatório do valor de cada sinapse serve de argumento de entrada para os neurônios da camada seguinte. Normalmente os neurônios da camada de entrada não possuem função de transferência, ou melhor, possuem função de transferência identidade (a saída é igual à entrada).

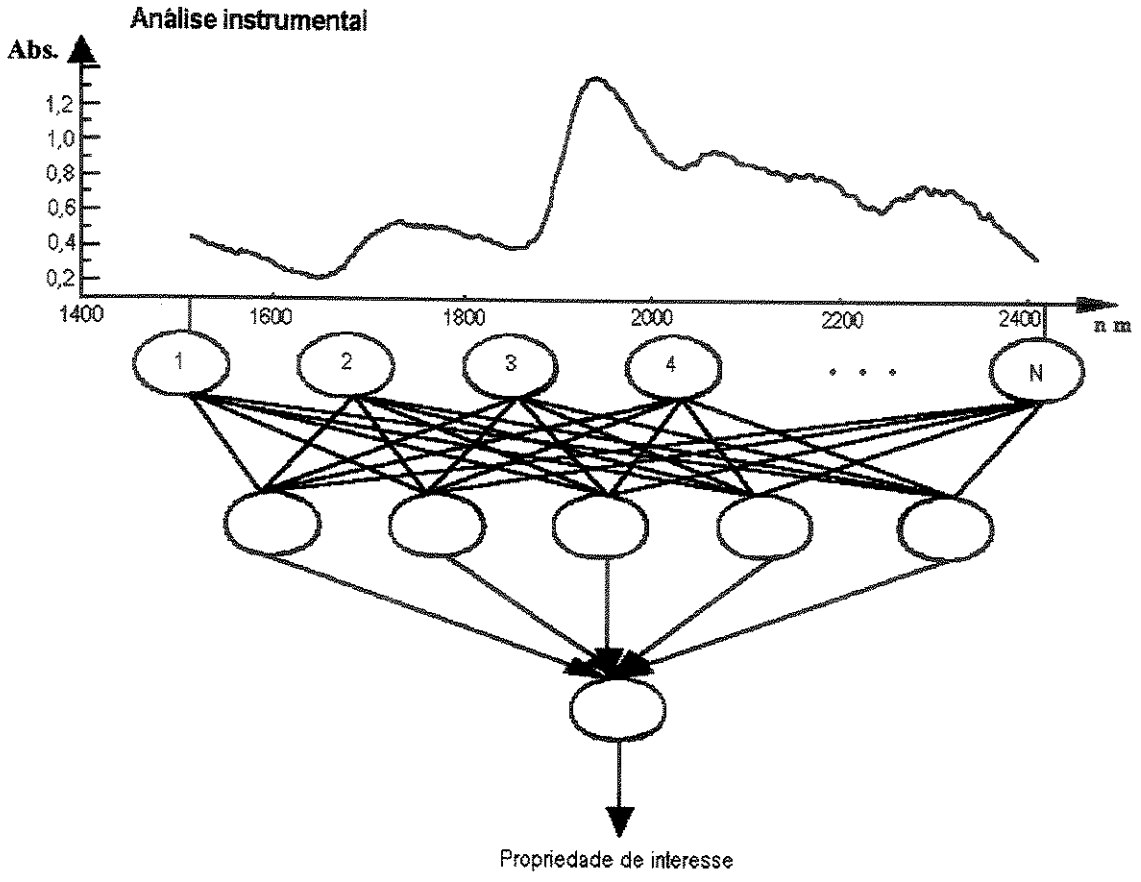


Figura 3.1 – Modelo de uma rede neural artificial MLP

A MLP inicialmente passa por um processo de aprendizado supervisionado em que é executado o treinamento da rede. Nesse processo são administrados conjuntos conhecidos de dados de entradas e de saídas, e são calculados os pesos (fatores de ponderação) de cada conexão em cada camada da rede. Essa etapa é equivalente à fase de calibração em uma calibração multivariada por modelos lineares (PLS, PCR, etc.). Uma vez definidos todos os pesos da rede, ou seja, definida a forma da função do "aproximador universal" pode-se então utilizar a MLP para executar previsões a cerca do valor das propriedades de interesse em novas amostras, como se executa normalmente em calibração multivariada linear. Nesse

caso, tomam-se os valores obtidos na análise instrumental como argumento de entrada da MLP. O resultado de saída é então a previsão da rede para os novos dados em questão.

Existem vários algoritmos para executar o treinamento de uma MLP. Dentre os mais usados estão a retropropagação de erros [21] e o método de Marquardt-Levenberg [22].

Um problema que geralmente ocorre é quanto à definição da arquitetura de rede. Deve-se definir quantas camadas ocultas serão utilizadas, quantos neurônios em cada camada e qual a função de transferência de cada neurônio. Tendo em vista esses problemas, surgiram alguns algoritmos para otimizar a arquitetura da rede. A idéia básica deste método é iniciar a rede neural com um número razoável de neurônios na camada intermediária e, durante a etapa de treinamento cortar as conexões (ou pesos) dos neurônios que possuem pouca influência no erro final. Neurônios que tiverem todas as conexões cortadas serão eliminados e, portanto, ao final dos "cortes", sobrarão somente os neurônios realmente necessários à modelagem. A técnica de apodização [19,23] (pruning) reduz a complexidade da rede neural, melhorando sua capacidade de previsão, pois evita modelos sobre-parametrizados (muitos neurônios) em que a possibilidade de sobre-ajuste é grande.

Existem basicamente dois métodos para a apodização de redes neurais: Optimal Brain Damage [24] (OBD) e Optimal Brain Surgeon [25,26] (OBS). Em ambos os métodos as conexões (ou pesos) são cortadas e a correspondente variação no erro E, chamada de saliência, é avaliada. No método OBD as conexões são cortadas durante a etapa de treinamento e a rede neural não é re-treinada após os cortes. No método OBS, as conexões são cortadas e, após o corte de uma conexão, a rede é re-treinada, permitindo que um número maior de cortes seja efetuado. Além disso, no método OBS a rede neural é re-

treinada, aproximando-se os erros de treinamento por uma função quadrática, de modo a garantir a existência de um mínimo.

Em princípio uma MLP pode modelar qualquer conjunto de dados (análise instrumental) a um conjunto resposta (propriedades de interesse), e quanto mais complexa for a relação entre os dados instrumentais e os dados de resposta, mais parâmetros a rede deve ter para poder modelar satisfatoriamente. Há um compromisso entre o grau de ajuste da rede e o número de parâmetros da mesma. Quanto maior for o número de parâmetros, maior a probabilidade de ocorrer um sobre-ajuste na etapa de treinamento. Isso inviabilizaria a utilização da MLP para calibração multivariada, visto que existem erros experimentais e instrumentais inerentes às análises. Por outro lado, um número pequeno de parâmetros pode fazer com que a MLP não consiga modelar satisfatoriamente os dados em questão. Em ambos os casos ocorrem erros grandes na calibração multivariada. Entende-se por parâmetros os componentes que restringem os graus de liberdade da modelagem, assim, no caso da MLP os parâmetros são os próprios neurônios.

Para eliminar os problemas inerentes às MLP's, optou-se por utilizar outro tipo de rede neural artificial, a rede neural com funções de base radial (RBFN – radial basis function network) [19,27] . As RBFN´s têm tido recente aplicação [28 – 31] em diversas áreas. Nesse tipo de rede há apenas uma camada intermediária e os pesos não são calculados por um processo iterativo, e sim por meios algébricos, baseado no método dos mínimos quadrados. Isso diminui a probabilidade de sobre-ajuste, visto que o método dos mínimos quadrados é relativamente robusto e, portanto, extensível a outras amostras que não as usadas na etapa de treinamento. Adicionalmente, utilizou-se um algoritmo genético para otimizar o tipo, posição, raio e número das funções radiais utilizadas.

3.1 – A Função radial de base – RBF e o arranjo em redes de RBF's

RBF's são funções radiais que decrescem (ou crescem) monotonicamente à medida que se distanciam do ponto central. As funções radiais utilizadas possuem fórmula geral do tipo:

$$h(x)=\phi\left(\frac{(x-c)^t(x-c)}{r^2}\right)=\phi\left(\sum\frac{(x_i-c_i)^2}{r^2}\right) \qquad 3.1$$

em que ϕ é o tipo de função usada, "c" é o vetor de posições em que a função está centrada (uma posição para cada variável) e "r" é o vetor de dispersões da RBF em relação a cada variável de entrada e "x" é o vetor com as variáveis de entrada. Assim, para as RBF's utilizadas, tem-se:

Gaussianoide: $\phi(z)=e^{-z}$; 3.2

Multiquádrica: $\phi(z)=(1+z)^{\frac{1}{2}}$; 3.3

Multiquádrica inversa: $\phi(z)=(1+z)^{-\frac{1}{2}}$; 3.4

Cauchy: $\phi(z)=(1+z)^{-1}$. 3.5

em que: $z=\left((x-c)^t R^{-1}(x-c)\right)$ 3.6

A Figura 3.2 mostra o perfil, em relação a cada variável de entrada, das funções radiais mais utilizadas. Vale salientar que cada RBF, diferentemente dos neurônios de uma MLP, é uma função multidimensional, ou seja, toma como argumentos as diversas entradas na forma de variáveis independentes. O neurônio da MPL toma como argumento de entrada apenas uma variável, dada pelo somatório das sinapses.

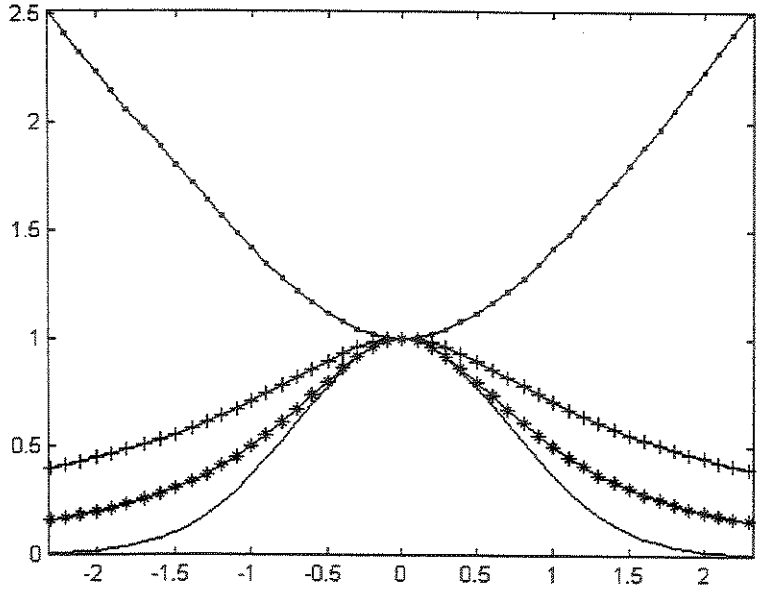


Figura 3.2 - Perfil dos tipos de RBF mais utilizados: gaussiana(–); multiquádrica (-∘); multiquádrica inversa (-+) e cauchy (-*).

RBF's por si só, são apenas uma classe de funções, e podem ser utilizadas em rede sob qualquer arranjo. Porém, o termo RBFN (Radial Basis Function Network) ou rede de funções radiais de base, geralmente se refere a uma rede com uma camada intermediária, conforme mostrado na Figura 3.3.

Cada variável da matriz de dados (X_i) é usada como argumento de todas as RBF's. A saída de cada RBF é ponderada e somada para se determinar uma saída. Adicionalmente pode-se acrescentar um termo independente dos dados de entrada para otimizar o ajuste ("bias").

Uma RBFN não é linear (em relação aos parâmetros) se as RBF's puderem mudar de posição (variação de c) ou tamanho (variação de r) ou caso a rede possua mais de uma camada oculta. Caso contrário, embora a RBFN seja formada por funções não lineares, ela é linear em relação aos parâmetros (w), ou seja, a saída da rede é uma combinação linear das

saídas das funções. Em outras palavras, desde que não haja RBF's com raios e centros iguais simultaneamente, elas formam uma base para um espaço vetorial no qual serão modeladas as amostras em questão.

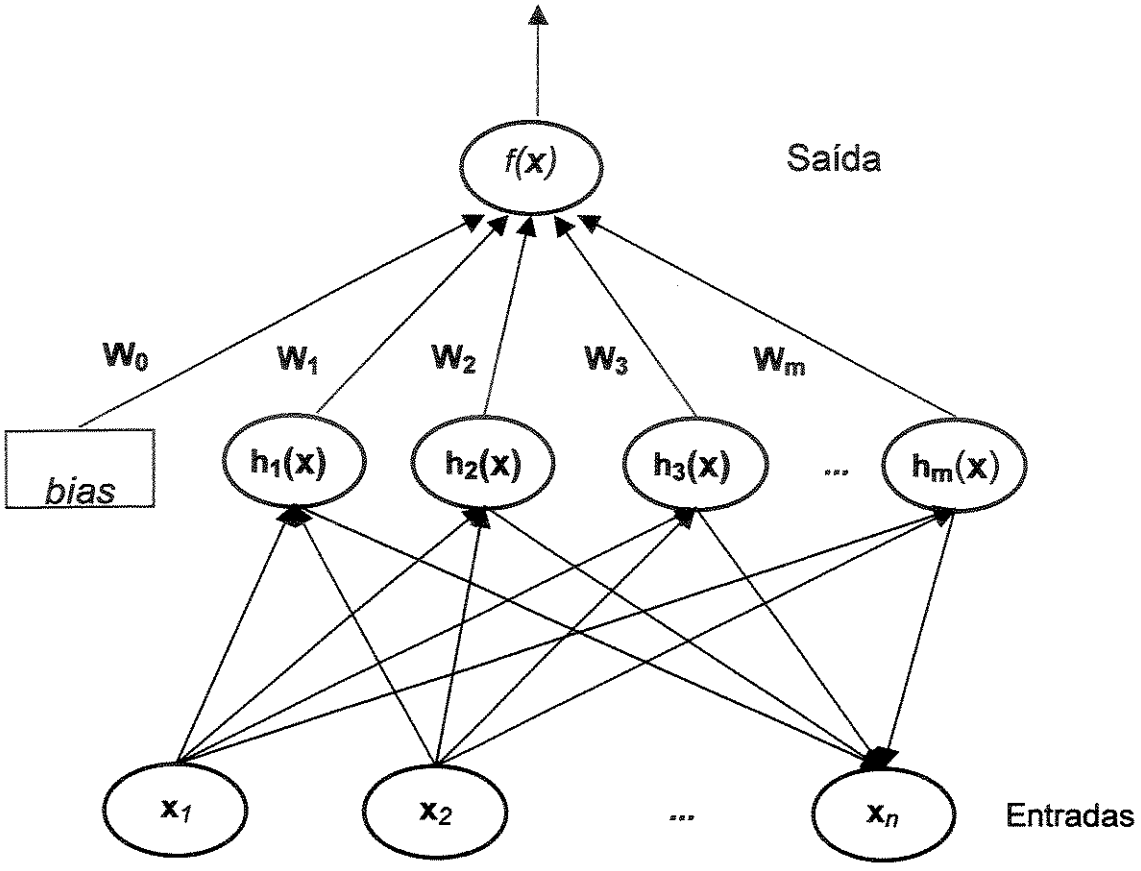


Figura 3.3 - Representação de uma rede de RBF's.

Devido à linearidade das RBFNs com apenas uma camada oculta, uma vez fixados o número, os raios e os centros das RBF's, os pesos (parâmetros de ponderação w) podem ser determinados pelo método dos mínimos quadrados. Isso possibilita o cálculo desses parâmetros de modo algébrico (matricial), o qual é muito mais rápido que os modelos iterativos de otimização de redes neurais.

Para a execução do método de mínimos quadrados, basicamente tem-se que diferenciar a expressão do somatório quadrático em relação a cada parâmetro a ser otimizado, igualar cada expressão a zero e resolver o sistema de equações resultante.

O modelo de RBFN gera uma resposta do tipo:

$$f(\mathbf{x}) = \sum_{j=1}^{m} w_j h_j(\mathbf{x}) \qquad 3.7$$

Para a determinação dos pesos é necessário minimizar o somatório quadrático

$$S = \sum_{i=1}^{p} (y_i - f(\mathbf{x}_i))^2 \qquad 3.8$$

Diferenciando o somatório quadrático em relação a cada parâmetro têm-se

$$\frac{\partial S}{\partial w_j} = 2\sum_{i=1}^{p} (f(\mathbf{x}_i) - y_i) \frac{\partial f}{\partial w_j}(\mathbf{x}_i) \qquad 3.9$$

Devido à linearidade do sistema, as derivadas em relação a parâmetros distintos são independentes.

$$\frac{\partial f}{\partial w_j}(\mathbf{x}_i) = h_j(\mathbf{x}_i) \qquad 3.10$$

Substituindo a Equação 3.10 na derivada do somatório do erro quadrático (Equação 3.9), e igualando a zero têm-se:

$$\sum_{i=1}^{p} f(\mathbf{x}_i) h_j(\mathbf{x}_i) = \sum_{i=1}^{p} y_i h_j(\mathbf{x}_i) \qquad 3.11$$

Colocando em forma matricial, têm-se:

$$\mathbf{h}_j^t\mathbf{f} = \mathbf{h}_j^t\mathbf{y} \tag{3.12}$$

em que

$$\mathbf{h}_j = \begin{bmatrix} h_j(\mathbf{x}_1) \\ h_j(\mathbf{x}_2) \\ \vdots \\ h_j(\mathbf{x}_p) \end{bmatrix}, \qquad \mathbf{f}_j = \begin{bmatrix} f(\mathbf{x}_1) \\ f(\mathbf{x}_2) \\ \vdots \\ f(\mathbf{x}_p) \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}$$

ou seja:

$$\begin{bmatrix} \mathbf{h}_1^t\mathbf{f} \\ \mathbf{h}_2^t\mathbf{f} \\ \vdots \\ \mathbf{h}_m^t\mathbf{f} \end{bmatrix} = \begin{bmatrix} \mathbf{h}_1^t\mathbf{y} \\ \mathbf{h}_2^t\mathbf{y} \\ \vdots \\ \mathbf{h}_m^t\mathbf{y} \end{bmatrix} \tag{3.13}$$

Utilizando as propriedades de multiplicação matricial,

$$\mathbf{H}^t\mathbf{f} = \mathbf{H}^t\mathbf{y} \tag{3.14}$$

onde **H** é a matriz de "design" com os vetores $\{\mathbf{h}_j\}_{j=1}^m$ em suas colunas

$$\mathbf{H} = [\mathbf{h}_1\ \mathbf{h}_2\ \ldots\ \mathbf{h}_m] \tag{3.15}$$

e tem p linhas relativas às amostras do conjunto de treinamento

$$\mathbf{H} = \begin{bmatrix} h_1(\mathbf{x}_1) & h_2(\mathbf{x}_1) & \cdots & h_m(\mathbf{x}_1) \\ h_1(\mathbf{x}_2) & h_2(\mathbf{x}_2) & \cdots & h_m(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ h_1(\mathbf{x}_p) & h_2(\mathbf{x}_p) & \cdots & h_m(\mathbf{x}_p) \end{bmatrix} \tag{3.16}$$

O vetor f pode ser decomposto no produto de dois termos: A matriz de design e o vetor de pesos

$$\mathbf{f}_i = \mathbf{f}(\mathbf{x}_i) = \sum_{j=1}^{m} \hat{w}_j h_j(\mathbf{x}_i) = \overline{h}_i^t \hat{\mathbf{w}} \qquad 3.17$$

onde

$$\overline{\mathbf{h}}_i = \begin{bmatrix} h_1(\mathbf{x}_i) \\ h_2(\mathbf{x}_i) \\ \vdots \\ h_m(\mathbf{x}_i) \end{bmatrix} \qquad 3.18$$

É importante observar que enquanto $\mathbf{h}_j$ é uma das colunas de **H**, $\overline{\mathbf{h}}_i^t$ é uma de suas linhas.

$$\mathbf{f} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_p \end{bmatrix} = \begin{bmatrix} \overline{h}_1^t \mathbf{w} \\ \overline{h}_2^t \mathbf{w} \\ \vdots \\ \overline{h}_p^t \mathbf{w} \end{bmatrix} = \mathbf{H}\hat{\mathbf{w}} \qquad 3.19$$

Finalmente substituindo na Equação 27 tem-se:

$$\mathbf{H}^t\mathbf{y} = \mathbf{H}^t\mathbf{H}\hat{\mathbf{w}} \qquad 3.20$$

Multiplicando a Equação 3.20 à esquerda em ambos lados pela inversa de $\mathbf{H}^t\mathbf{H}$ têm-se:

$$\hat{\mathbf{w}} = \left(\mathbf{H}^t\mathbf{H}\right)^{-1}\mathbf{H}^t\mathbf{y} \qquad 3.21$$

Para adicionar o termo independente (bias) à rede em questão, basta adicionar um vetor coluna unitário na matriz de design (H), de modo a gerar uma RBF cuja resposta independe dos dados na matriz de design.

A fim de eliminar problemas numéricos e acelerar o cálculo dos modelos de uma RBFN, usualmente utiliza-se um modelo de ortogonalização nos dados de entrada. Assim, elimina-se a correlação entre as variáveis de entrada e o número das mesmas. Esse processo condensa a informação relevante dos dados de entrada em uma matriz de menor tamanho, eliminando informações redundantes, facilitando assim, o cálculo com matrizes de menor tamanho.

Em uma rede com RBF's, expande-se o espaço de modelagem com uma base maior que o número de entradas (após a ortogonalização e conseqüente eliminação da correlação das variáveis de entrada). Nesse espaço com dimensão superior a separação dos padrões de informação é feita mais facilmente e posteriormente é feita uma regressão linear múltipla (método dos mínimos quadrados) entre os dados de entrada nessa nova base e a saída esperada (concentração da espécie química de interesse, por exemplo).

O principal problema em se trabalhar com RBFN é definir o número, as posições, raios e tipos das RBF's utilizadas na rede, de modo otimizado, ou seja, definir a base de dimensão superior na qual será feita a separação dos padrões de informação. Não existe uma fórmula geral para definir esses parâmetros, porém, recentemente tem surgido alguns artigos [32,33] com essa finalidade.

Uma estratégia alternativa de otimização dos parâmetros da RBFN é comparar modelos construídos com diferentes subconjuntos dentre um grupo definido de possíveis RBF's. Encontrar o melhor conjunto, por "tentativa e erro" é praticamente inviável, visto que existem $2^n - 1$ sub-conjuntos distintos em um conjunto com n RBF's, de modo que geralmente utiliza-se de métodos heurísticos para realizar essa tarefa. Um dos tipos mais simples dentre os métodos heurísticos de otimização é a "Forward Selection" [33,34].

O modelo de otimização por "Forward Selection" baseia-se em iniciar um sub-conjunto vazio de RBF's e ir adicionando RBF's de modo a diminuir (minimizar) o somatório do erro quadrático ou outro critério de minimização como a validação cruzada generalizada (por exemplo).

Outro modelo heurístico muito usado é fazer o oposto, ou seja, partir de um conjunto maior de RBF's e ir eliminando RBF's de modo a diminuir o erro (critério de minimização), conhecido como "Backward Elimination" .

É interessante comparar a seleção de subconjuntos de RBF's em uma RBFN com o modelo padrão para otimização em redes neurais artificiais, o qual envolve otimização por gradiente descendente de uma superfície não linear em um espaço multidimensional definido pelos parâmetros da rede. Na otimização por subconjuntos ("Forward Selection" ou "Backward Elimination") o algoritmo de otimização encontra em um espaço discreto de subconjuntos, um conjunto de RBF's com centros e raios fixos de modo a minimizar o erro de previsão. Os pesos não são selecionados, mas sim determinados em função do conjunto de RBF's definido pelo modelo de otimização. Por definição, "Forward Selection" é um algoritmo não linear de otimização para RBFN's, porém possui algumas vantagens:

- Não há necessidade de fixar o número máximo de RBF's;
- Os critérios de seleção dos modelos são tratáveis;
- Não há necessidade de poder computacional elevado.

Outra forma de se otimizar uma rede de RBF's é o modelo de "Ridge Regression" [33,35]. Nesse modelo são definidos parâmetros de regularização que restringem a variância modelada por cada RBF, assim sento, após a otimização do modelo, as RBF's que apresentam o parâmetro de regularização com valores muito altos explicam muito pouco a

variância dos dados. Nesse caso essa respectiva RBF pode ser suprimida (removida da rede) e o valor do parâmetro de regularização é então incorporado ao "off-set" (termo constante da rede – bias).

Numa RBFN otimizada por "Ridge Regression" os parâmetros de regularização (λ) são otimizados através da minimização da função custo, também por mínimos quadrados.

$$C = \sum_{i=1}^{p} \left(y_i - f(x_i)\right)^2 + \lambda \sum_{j=1}^{m} w_j^2 \qquad ; \quad f(x_i) = \sum_{j=1}^{m} w_j h_j(x_i) \tag{3.22}$$

Onde p é o número de amostras (análises) utilizadas para treinar a RBFN e m é o número de RBF's da RBFN, y_i são os valores da propriedade de interesse para cada amostra, $f(x_i)$ são os valores estimados pela rede de RBF's para cada amostra e w_j são os pesos (parâmetros de ponderação) para cada RBF. Ao minimizar a função custo, definida pela Equação 3.22, também pelo método dos mínimos quadrados, encontra-se uma expressão que relaciona os pesos e os parâmetros de regularização, como mostra a Equação 3.23.

$$\hat{W} = \left(H^t H + \Lambda\right)^{-1} H^t y \tag{3.23}$$

onde H é a matriz design definida pela Equação 3.16 .

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_m \end{bmatrix} \tag{3.24}$$

Desse modo os dois (pesos - w e parâmetros de regularização - λ) são encontrados assumindo-se valores iniciais para os parâmetros de regularização e tomando-se um

processo iterativo de otimização de modo a escolher o melhor conjunto de parâmetros de regularização que minimizem a função custo.

É interessante observar que se os parâmetros de regularização forem nulos, a função custo (Equação 3.22) tente ao somatório quadrático dos erros (Equação 3.8) e os pesos encontrados pelo modelo de "Ridge Regression" (Equação 3.23) tentem a se igualar aos pesos encontrados para a RBFN normal (Equação 3.21).

Todos os modelos anteriormente mencionados quando utilizados para a calibração multivariada tendem a encontrar redes de RBF's com muitos parâmetros de regularização. Isso acontece porque eles não otimizam os parâmetros de cada RBF, mas sim a arquitetura da rede de um modo geral.

Para resolver esse problema, utilizou-se um algoritmo genético. A exemplo do que já foi feito para as redes neurais convencionais (multilayer perceptron – MLP) [19,20], em que se pode otimizar as redes com poda de conexões [23-26], utilizou-se o mesmo algoritmo genético que otimiza as RBF's para otimizar também as conexões da RBFN. Para tanto, utilizou-se mais alguns cromossomos para codificar as conexões relevantes da RBFN. Desse modo o AG otimiza tanto os parâmetros internos de cada RBF (que definem sua posição e raio) como as conexões da RBFN otimizando assim a arquitetura da rede.

3.2 – Otimização simultânea dos parâmetros das RBF's e da arquitetura da RBFN por algoritmo genético

Cada RBF tem um raio e um centro para cada variável de entrada. Considerando que uma análise multivariada por espectroscopia no infravermelho próximo (por exemplo)

apresenta cerca de 500 a 1000 comprimentos de onda distintos e que cada RBFN, geralmente apresenta cerca de 10 RBF's, ter-se-ia que otimizar um número muito grande de parâmetros, o que inviabilizaria o processo. Além desse problema, devido às inversões matriciais necessárias ao cálculo dos pesos (vetor w), se as variáveis de entrada forem razoavelmente correlacionadas, haverá problemas numéricos. Para resolver ambos os problemas simultaneamente, utiliza-se um método de ortogonalização, como por exemplo o PCA.

O PCA além de ortogonalizar (descorrelacionar) as variáveis de entrada, compacta a informação contida nas variáveis iniciais (originais) em poucos componentes principais. Desse modo ao invés de se utilizar entre 500 e 1000 variáveis originalmente, utilizam-se poucas componentes principais, tipicamente de 2 a 10. O número de componentes principais utilizadas é baseado na quantidade (relativa) da variância total existentes nas variáveis originais ao longo de todas as amostras utilizadas para treinar a RBFN. Geralmente as componentes que explicam além de 99,5 % da variância total são desprezadas.

Desse modo o problema passa a ser calcular os centros e raios de cada RBF em relação a cada componente principal (que agora são as novas variáveis de entrada de RBFN), bem como quais as conexões são necessárias (relevantes), ou seja, quais componentes principais serão necessárias em cada RBF, e quais as RBF's serão excluídas, caso não haja nenhuma entrada para ela.

A codificação dos cromossomos, e conseqüentemente dos indivíduos do AG, é a etapa que determina a forma da otimização que se pretende realizar. Para encontrar os valores de raios (dispersão) de cada RBF, deve-se definir a faixa de variação em que está cada raio, ou seja, o valor mínimo e o valor máximo dentre os quais pode estar o valor do

raio, para cada RBF. Deve-se também definir a precisão com que se deseja encontrar esses valores, ou, em outras palavras, o número de "bits" para cada cromossomo. Analogamente, o mesmo deve-se realizar em relação aos centros das RBF's para cada componente principal.

Tipicamente, em uma distribuições normais para variáveis aleatórias, os valores das mesmas com 99% de confiança, se encontram dentro de uma faixa que vai até três vezes o desvio padrão (σ) de cada variável. Assim, a faixa de otimização utilizada para os raios das RBF's é obtida pela média ± três vezes o desvio padrão das componentes principais, dentre todas as amostras utilizadas para o treinamento da rede. Foram utilizados 10 bits para codificar cada cromossomo que definem os raios. Desse modo obtém-se uma precisão de $2x3x\sigma/(2^{10})$, ou seja, a menor variação que se pode ter entre cromossomos que otimizam os raios é de $6x10^{-3}$ vezes o desvio padrão do respectivo componente principal. Para definir a faixa dos centros das RBF's para cada componente principal foram tomados os valores máximos e mínimos das componentes principais para as amostras de treinamento da rede. Novamente foram utilizados 10 bits para definir a precisão dos centros. Desse modo, a menor diferença entre cromossomos que otimizam os centros é de (Max(CP) – Min(CP))/(2^{10}) para cada componente principal (CP). Para montar o indivíduo com os cromossomos codificados, utilizou-se o seqüenciamento por RBF, ou seja, o indivíduo é formado pela adição lateral de cromossomos que seqüenciam raios e centros das RBF's.

Desse modo tem-se na seqüência o cromossomo que codifica o raio da 1ª RBF em relação à 1ª CP, seguido do cromossomo que codifica o raio da 1ª RBF em relação à 2ª CP, seguido pelo cromossomo que codifica o raio da 1ª RBF em relação à 3ª componente principal e assim sucessivamente até codificar os raios da 1ª RBF em relação a todos os

componentes principais. Em seguida seguem os cromossomos que codificam os centros da 1ª RBF em relação a todos as CP's. Posteriormente o processo é repetido para a 2ª RBF e assim sucessivamente até codificar os parâmetros de todas as RBF's utilizadas.

Como exemplo, tomemos uma rede de RBF's simplificada, com apenas três RBF's, com duas variáveis de entrada (componentes principais). Essa RBFN é mostrada na Figura 3.4.

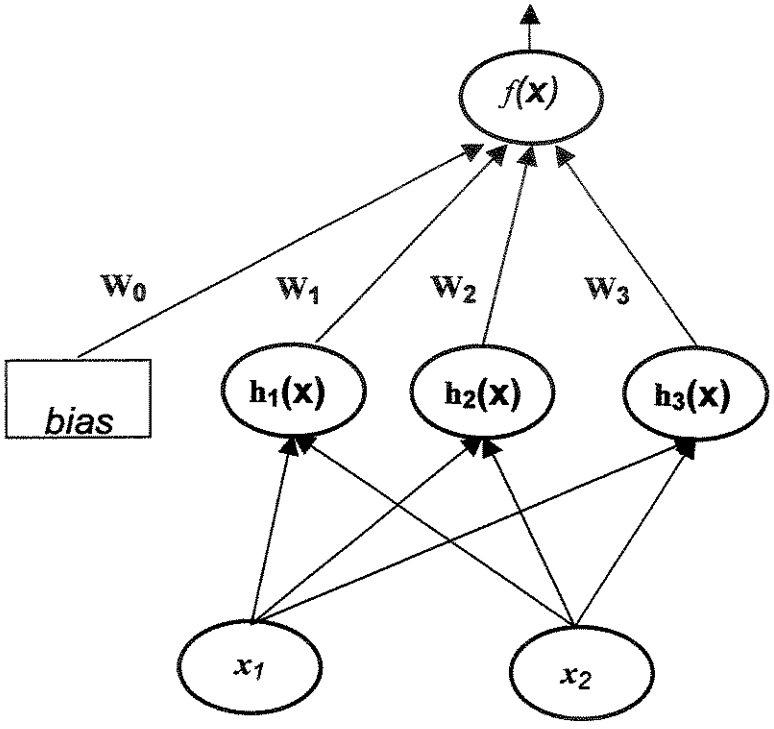


Figura 3.4 – Exemplo de uma RBFN com três RBF's, com duas variáveis de entrada para cada RBF.

Uma vez definido o número máximo de RBF's da rede, e a codificação de cada cromossomo para a otimização dos centros e raios, falta definir o cromossomo para a otimização das conexões. Para otimizar as conexões, deve-se acrescentar um cromossomo adicional que determina o conjunto de conexões selecionadas. Assim, para uma rede com "A" componentes principais e número máximo de RBF's igual a "B", o cromossomo que determina o conjunto de conexões selecionadas deve possuir AxB "bits" (elementos

binários). Cada "bit" determina a seleção de uma conexão (0 → a conexão não é selecionada e 1→ a conexão é selecionada). Esse cromossomo é então adicionado ao fim da codificação do indivíduo para a otimização pelo AG.

No exemplo da RBFN mostrado na Figura 3.4, a codificação de um possível indivíduo é mostrada na Figura 3.5.

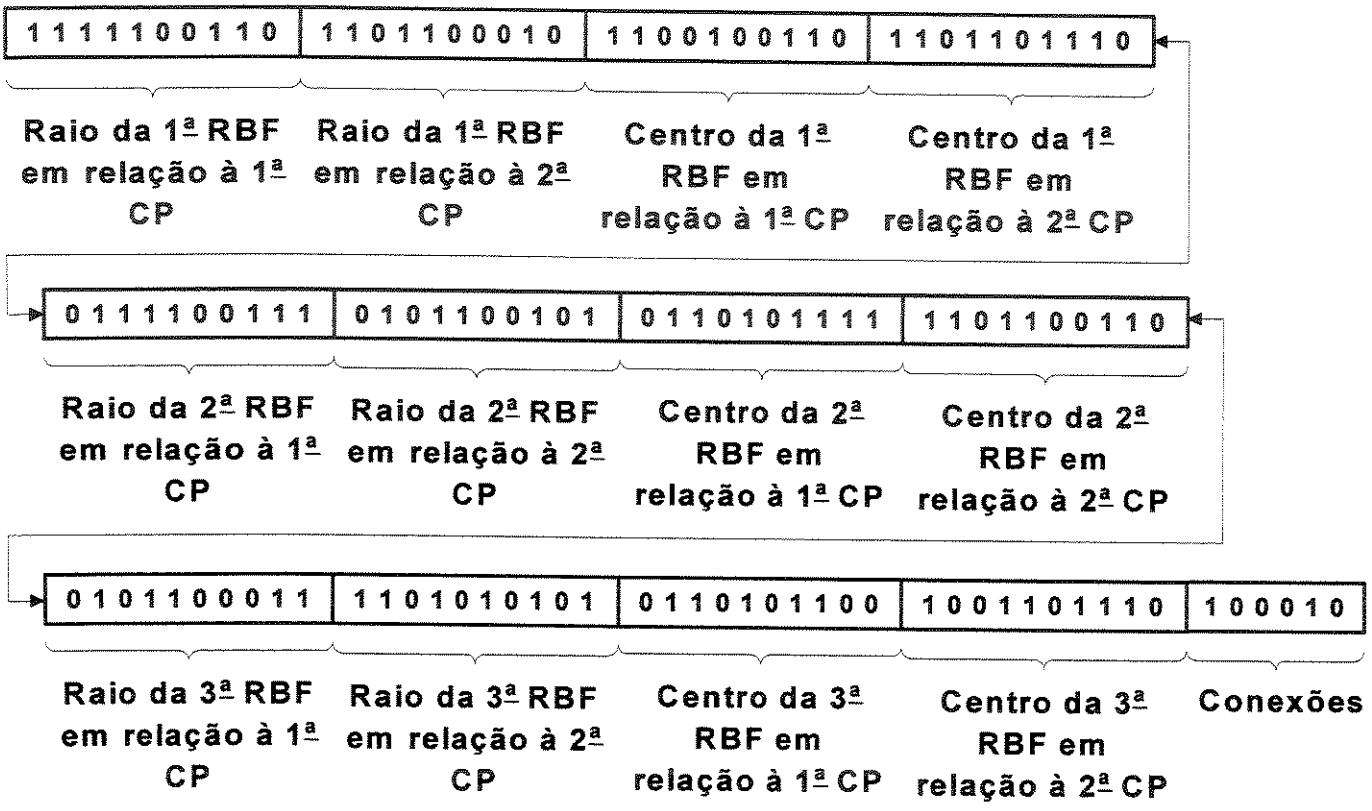


Figura 3.5 – Exemplo da codificação de um indivíduo para a otimização simultânea dos parâmetros das RBF's com precisão de 10 bits e das conexões de uma RBFN com 3 RBF's e duas variáveis de entrada (CP) por AG.

No exemplo do indivíduo codificado na Figura 3.5, de acordo com o último cromossomo, foram selecionadas as conexões do 1ª CP na 1ª RBF e na 3ª RBF. Se esse fosse o melhor indivíduo da população final, ou seja, o indivíduo ótimo, a 2ª RBF seria excluída de rede, bem como só seria utilizado a 1ª CP. A arquitetura da RBFN otimizada seria então a mostrada na Figura 3.6, a qual mostra as conexões eliminadas em pontilhado.

Devido à eliminação de todas as conexões da 2ª RBF, a mesma também foi eliminada da RBFN. Raciocínio análogo pode ser feito em relação à 2ª CP e ao parâmetro de ponderação W_2, ambos mostrados também em pontilhado na Figura 3.6.

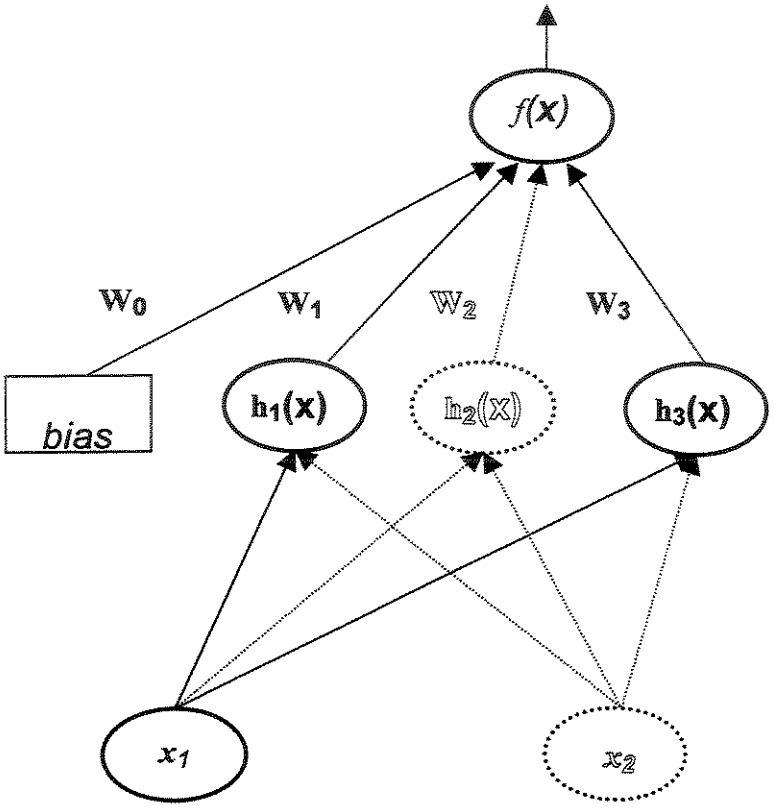


Figura 3.6 – Exemplo da arquitetura da RBFN codificada pelo indivíduo mostrado na Figura 3.5.

Adicionalmente deve-se definir os tipos de funções radiais a serem utilizadas (dentre os quatro tipos mais comuns: gaussiana, multiquádrica, multiquádrica inversa e cauchy).

Ao excluir algumas RBF's da rede, o AG está ao mesmo tempo otimizando os tipos de RBF's da RBFN. Rigorosamente, o que diferencia uma RBF de outra da rede são os parâmetros (raios e centros) e a forma (função) das mesmas, visto que inicialmente todas as RBF's têm as mesmas entradas (as CP's). Como os parâmetros são otimizados pelo AG, a forma (tipo da função) é o verdadeiro diferencial entre elas. Após a otimização, a

permanência de algumas RBF's define quais são as RBF's que melhor modelam os dados em questão.

Capítulo 4

Determinação de parâmetros de qualidade utilizados na indústria sucroalcooleira por espectroscopia no infravermelho próximo utilizando os modelos propostos

Tendo em vista o dispêndio de mão de obra, reagentes e tempo, necessários para a determinação dos parâmetros utilizados para o controle da produção industrial do açúcar e seu controle de qualidade, atualmente tem-se dado ênfase na utilização de métodos espectroscópicos, baseados na absorção de luz na região do infravermelho próximo (Near Infrared – NIR). Esses métodos são não-destrutivos, e podem ser automatizados diretamente na linha de produção, o que permite um controle mais rápido e eficiente do processo industrial.

Os métodos espectroscópicos, baseados no infravermelho próximo, para a determinação dos parâmetros de qualidade devem ser aliados aos métodos quimiométricos de calibração multivariada para se obter informação relevante e permitir a automação industrial. O maior problema é que, para se poder substituir os métodos convencionais, os métodos espectroscópicos (com calibração multivariada) não devem apresentar erros médios de previsão acima de 2% (em relação aos métodos convencionais), conforme o corpo técnico da Usina alta Mogiana.

Como não se pode garantir que as análises espectroscópicas na região no infravermelho próximo sejam lineares (obedeçam à lei de beer), nem sempre é possível utilizar os métodos lineares de calibração multivariada (PLS, PCA, MLR, etc). Ademais, deve-se ter em vista que os métodos convencionais de medida da pol, brix, AR e ART são usados como padrões de calibração dos métodos de calibração multivariada. Por isso os métodos espectroscópicos, com determinação por calibração multivariada, não podem apresentar erros médios inferiores aos obtidos pelos métodos de laboratório (tradicionais).

4.1 – Parâmetros de qualidade

A produção industrial de álcool (etanol) e açúcar (sacarose) no Brasil utiliza a cana de açúcar como matéria-prima básica. Ao longo do processo industrial [36], o teor de açúcar deve ser analisado desde a aquisição da cana, até o produto final. Ademais, o valor pago por essa matéria-prima basicamente depende do teor de sacarose contido na mesma.

No processo industrial de produção de açúcar, a cana é inicialmente lavada, para remover terra e detritos. Depois é picada, desfibrada e esmagada em moendas para a obtenção do caldo da cana. As moendas são formadas por seqüências de ternos (conjuntos com três rolos canelados) com aberturas (distâncias entre os rolos) gradativamente reduzidas, que exercem forte pressão e extraem seqüencialmente o caldo. Extraem-se cerca de 93% do caldo. O caldo é então coado, para remover impurezas grossas, e tratado com cal, para coagular parte da matéria coloidal, precipitar impurezas e modificar o pH. Adiciona-se então ácido fosfórico para ajudar a clarificação do caldo. A mistura é aquecida com vapor de água a alta pressão e decantada em grandes tanques (decantadores de caldo) ou em espessadores contínuos.

Para recuperar o açúcar dos lodos decantados, usam-se filtros a vácuo, contínuos, a tambor rotatório, ou prensas de quadro. O material retido no filtro (torta) atinge cerca de 1 a 4% do peso da cana utilizada e é usado como adubo.

O filtrado obtido é um caldo clarificado com elevado teor de cal e contém cerca de 85% de água. É então evaporado até aproximadamente 40 % de água, tornando-se um xarope grosso, amarelado.

O xarope resultante é lançado no 1° estágio de um evaporador a vácuo de três efeitos, onde atinge um determinado grau de supersaturação. Nesse ponto adicionam-se núcleos de açúcar cristal (semeadura). Pela adição de xarope grosso e evaporação controlada, os cristais crescem até o tamanho desejado. No ponto ótimo, o cristalizador fica quase cheio de cristais de açúcar, com cerca de 10% de água. A mistura de xarope e cristais (massa cozida) é lançada em outro cristalizador, que é um tanque horizontal com agitação, dispondo de serpentinas de arrefecimento. Neste cristalizador há uma deposição adicional de sacarose sobre os cristais já formados, e a cristalização está completa.

A massa cozida é então centrifugada, para remover-se o xarope. Os cristais obtidos formam o açúcar demerara, de boa qualidade, e o xarope é reciclado para dar uma ou mais cristalizações. O líquido residual, depois da reciclagem, é conhecido como melaço e apresenta altos teores de monossacarídeos (glicose e frutose), porém baixos teores de sacarose.

O açúcar demerara (ligeiramente amarelado) possui aproximadamente 97,8% de sacarose e é enviado à refinaria de açúcar.

Para o controle da produção industrial e controle de qualidade do açúcar, são analisados vários parâmetros ao longo de várias etapas do processo [37,38]. A Tabela 4.1 mostra as análises mais freqüentes ao longo das diversas etapas do processo industrial. As principais análises utilizadas para o controle do processo são: brix, polarimetria (pol), teor percentual de açúcares redutores (AR), teor percentual de açúcares redutores totais (ART), pH, teor percentual de umidade e pureza.

Dependendo do ponto onde é coletado o material para a análise, devido à variação do estado físico, as amostras devem ser condicionadas apropriadamente no que diz respeito

à densidade, viscosidade e faixa de concentração. Para tanto podem ser diluídas, aquecidas, etc. Ademais, deve-se adicionar conservantes (reagentes químicos), ou resfriar as amostras, para que não haja alterações dos teores de açúcar em função da ação bacteriana (fermentação).

Tabela 4.1 – Análises necessárias para o controle de qualidade da produção industrial de açúcar.

Material coletado	**Local de coleta**	**Freqüência de análise**	**Tipos de análises**
Água de lavagem da cana	Esteira de transporte da cana	pH a cada 3 horas ART a cada 24 horas	pH e ART
Caldo primário	1º terno da moenda	A cada 3 horas	Brix, pol, pureza, AR e ART.
Caldo misto	Após as peneiras de cada terno de moenda	A cada 3 horas	Brix, pol, pureza, ART e pH.
Caldo residual	Último terno da moenda	A cada 3 horas	Brix, pol e pureza.
Caldo filtrado	Saída dos filtros	A cada 3 horas	Brix, pol, pureza, Ar e pH.
Caldo caleado	Saída do tanque de caleagem	A cada 1 hora	pH.
Caldo clarificado	Bandejas do decantador	A cada 3 horas	Brix, pol, pureza, pH e turbidez.
Torta	União de todos os filtros	A cada 3 horas	% umidade, pol e ART.
Xarope	Saída do último evaporador	A cada 1 hora	Brix, pol, pureza e ART.
Méis	Saídas das centrífugas	A cada 1 hora	Brix, pol, AR, pH e ART.

4.1.1 – Métodos convencionais de análise

4.1.1.1 - Brix

Geralmente, o teor de açúcares é avaliado com um densímetro (ou aerômetro) e expressa-se em °Brix [37-39] (grau brix). Este parâmetro pode ser definido como a percentagem em peso ou em volume (brix-volume) de sólidos solúveis expressos como sacarose, mas na realidade, o °Brix é apenas uma medida dos sólidos solúveis totais em soluções puras de sacarose. Para além disso, esta é uma medida quantitativa dos solutos totais (incluindo os açúcares), não dando qualquer informação qualitativa dos açúcares presentes no produto final.

A concentração de açúcares em meio aquoso pode ser determinada pela medida do grau brix. O grau brix é determinado pela variação da densidade da solução (em relação à água destilada) produzida pela alteração da concentração de sacarose dissolvida (considerando que o único soluto dissolvido seja a sacarose) em função da temperatura. Existem tabelas de correlação entre a densidade, medidas em grau brix, e o teor de sacarose dissolvida a uma dada temperatura (20°C) [37]. Adicionalmente deve-se fazer a correção de temperatura para se obter o grau brix à temperatura ambiente através da utilização de tabelas de correção do brix em função da temperatura [37].

4.1.1.2 – Pol

A luz proveniente de uma fonte de radiação propaga-se em todas as direções. Considerando-se uma determinada direção, a radiação eletromagnética vibra em um número infinito de planos, mas ao atravessar um polarizador, ela vibra em apenas um único plano. Tem-se então uma fonte de luz polarizada. As substâncias que apresentam a propriedade de girar o plano de luz polarizada são denominadas opticamente ativas, como por exemplo os açúcares. O ângulo de rotação de polarização pode ser medido em polarímetros. Sua magnitude é afetada pela temperatura, comprimento de onda da radiação polarizada e pelas características do material opticamente ativo. Mantendo-se constantes essas condições, o ângulo de rotação é diretamente proporcional à concentração do componente ativo e do caminho óptico percorrido pela luz no interior da solução. Uma escala sacarimétrica é definida pelo seguinte princípio: Uma solução pura de sacarose contendo 26 g em 100 mL de solução colocada em um tubo de 20 cm de comprimento proporciona um desvio do plano de luz polarizada ao qual se atribui o valor 100. O valor zero corresponde à água destilada. A partir de uma regra de três direta, pode-se determinar nessa escala sacarimétrica o percentual em volume de açúcares (como sendo sacarose). Como a pol [37,38] é definida como percentagem em peso, o valor na escala sacarimétrica deve ser corrigido com a densidade da solução em questão. A solução deve proporcionar um percurso livre de interferências, ou seja, não pode haver bolhas de ar ou material particulado no caminho óptico.

4.1.1.3 – Açúcares redutores (AR) e açúcares redutores totais (ART)

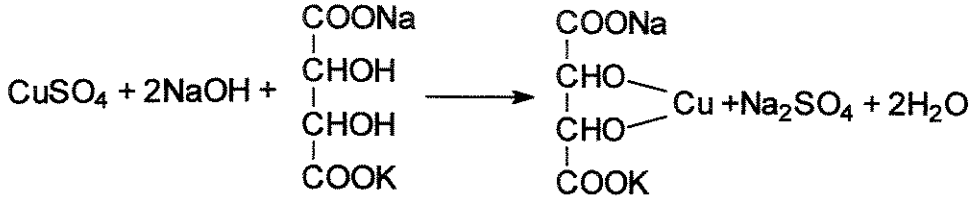
Qualquer sacarídeo que suporta um átomo de carbono anomérico que não formou um anél glicosídico é chamado açúcar redutor [37,38]. Essa denominação ocorre por causa da facilidade com que os grupos aldeído ou cetonas reduzem agentes oxidantes moderados. A maioria dos métodos analíticos para a identificação e quantificação de açúcares, (por exemplo, com o reativo de Fehling) está baseada no aldeído ou cetona presentes nas estruturas de açúcar. Geralmente os testes fornecem uma mudança de coloração decorrente da variação do estado de oxidação de um metal, e conseqüente complexação do mesmo. Os açúcares formam anéis que envolvem os grupos aldeído ou cetona. A formação e reabertura do anel são reversíveis, a menos que o grupo hemiacetal ou hidroxiacetal estejam envolvidos em outra ligação. Anéis fechados não apresentam nenhum aldeído ou cetona livres para reagir (a menos que haja vários anéis, e um deles possa abrir) e formam açúcares não redutores. Há um anel de glicose ao término de cada cadeia em goma e celulose, mas seu efeito é muito pequeno para produzir um teste positivo. A presença de açúcares redutores pode ser determinada pela solução de Fehling [37, 38], ou pelo método de Somogyi e Nelson [37, 38]. Açúcares simples como glicose, dextrose e frutose reduzem a solução de Fehling e dão um precipitado vermelho da solução azul inicial, porém a sacarose não apresenta essa propriedade.

Todos os monossacarídeos, sejam aldoses ou cetoses, são açúcares redutores. A maioria dos dissacarídeos também é. A sacarose (açúcar de mesa comum) é uma exceção notável, porque é um açúcar não redutor.

Para a determinação dos açúcares redutores totais (ART) [37,38], inicialmente procede-se a inversão (hidrólise ácida) da sacarose (para convertê-la em glicose e frutose) e posteriormente determinam-se os açúcares redutores.

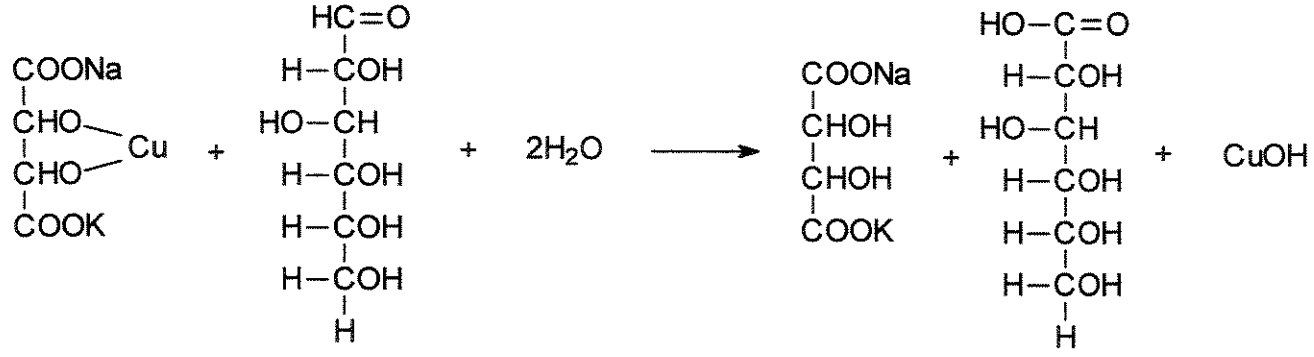
O princípio do método de Fehling é a redução do cobre por açúcares redutores. Normalmente, o cobre precipita em meio alcalino.

Em presença de compostos orgânicos, ricos em oxigênio tais como ácido cítrico, tartárico, etc., ou seus sais, não ocorre a precipitação do cobre devido à formação de um complexo com o cobre. Para evitar sua precipitação, no licor de Fehling o cobre é complexado com o Tartarato duplo de sódio e potássio.



4.1

O complexo formado pelos reagentes do licor de Fehling oxida o açúcar em questão a ácido carboxílico.

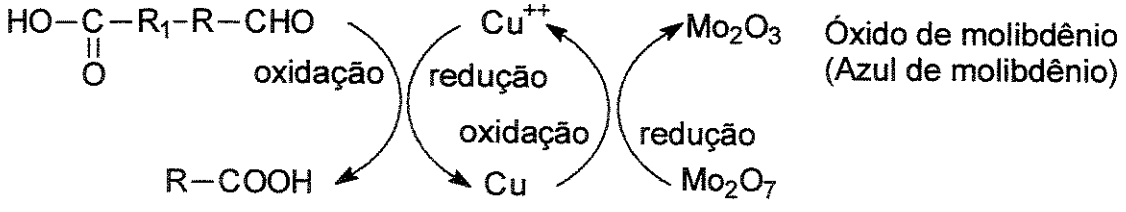


4.2

A determinação da quantidade de açúcar pela titulação do mesmo com o licor de Fehling não é considerada uma medida quantitativa (é apenas semi-quantitativa), pois o complexo formado pelos reagentes do licor (Equação 4.1) não apresenta uma composição muito bem definida. A quantidade de açúcar pode então ser determinada por volumetria de neutralização do ácido carboxílico formado na Equação 4.2.

Segundo o método de Somogyi e Nelson, sob determinadas condições, molibdatos reagem com fosfatos, arseniatos, silicatos, etc., para formar compostos heteropolares tais como, difosfomolibdatos de amônio $(NH_4)_3[P(Mo_3O_{10})_4]$ ou ácido molibdosilísico $H_4[Si(Mo_3O_{10})_4]$. Esses complexos, após redução controlada para dar azul de molibdênio, servem para a determinação colorimétrica de vários agentes redutores ou de elementos que funcionam como átomo central do complexo do ânion, tais como fósforo, arsênio ou silício.

Para a determinação de açúcares redutores pelo método de Somogyi e Nelson, o óxido cuproso, formado pela reação de Somogyi (Equação 4.3) tem o cobre na forma reduzida Cu_2O (essa reação é a mesma explicada pelo método de Fehling). Quando o difosfomolibdato de amônio, contido no reativo de Nelson, reage com o óxido cuproso, oxida-o a Cu^{++} e reduz o molibdato, formando óxido de molibdênio (Mo_2O_3).



4.3

4.1.1.4 – Pureza

A pureza [37, 38] do caldo de cana é definida pela divisão da pol pelo brix e multiplicado por 100.

O pol é uma medida do teor de sacarose na mistura de açúcares, visto que, dos componentes do caldo de cana, apenas a sacarose desvia o plano da luz polarizada. O brix é uma medida indireta de sólidos totais dissolvidos (como se fosse sacarose), logo a razão entre eles indica uma medida de pureza do caldo como se fosse uma solução de sacarose. O fator 100 serve para ponderar percentualmente.

4.2 - Espectroscopia no infravermelho próximo

A região composta pelo infravermelho próximo (NIR – Near Infrared) corresponde à faixa do espectro eletromagnético desde o fim do visível (por volta de 780 nm) até 3000 nm, ou seja, de 12821 cm^{-1} a 3333 cm^{-1}. Bandas de absorção nessa região são compostas de harmônicos, sobretons ou "overtones" (780 a 1800 nm) e combinações de bandas de estiramentos vibracionais fundamentais (1800 a 2500nm). As bandas de absorção envolvidas nessa região (infravermelho próximo) são geralmente devido às ligações C-H, N-H e O-H [40].

Moléculas que absorvem energia na região do infravermelho próximo basicamente vibram em dois modos fundamentais: estiramento e deformação (ou flexão). Estiramento é definido como uma contínua mudança na distância interatômica no eixo entre os átomos, e deformação é definida como a variação do ângulo de ligação entre os átomos. Estiramentos ocorrem em freqüências mais elevadas (comprimentos de onda menores) que as deformações. Estiramentos podem ser simétricos ou assimétricos e deformações podem ser no plano ou fora do plano da molécula.

Sobretons ocorrem quando ligações são excitadas a partir do estado fundamental para níveis excitados mais energéticos (além do 1º nível excitado).

Espectros na região do infravermelho próximo contém informações relativas a diferenças na força de ligação, espécies químicas, eletronegatividade e ligações de hidrogênio. Assim, a alteração de qualquer um desses parâmetros no sistema em análise acarreta em deslocamentos do espectro de absorção. Para amostras sólidas, o espectro na região do NIR apresenta informações a respeito de espalhamento, reflectância difusa, reflectância especular, polimento da superfície, e polarização da luz refletida.

Dentre as vantagens oferecidas pela espectroscopia na região do infravermelho próximo estão a velocidade, simplicidade na preparação de amostras, natureza não-destrutiva da técnica e a instrumentação ser praticamente a mesma da utilizada para análise de luz na região do visível. A maior desvantagem é provavelmente a baixa sensibilidade a constituintes menores.

A espectroscopia no infravermelho próximo tem sido uma poderosa ferramenta para a pesquisa em agricultura, e nas indústrias de alimentos [40, 41], farmacêutica [42], química [43], de polímeros [44] e petróleo [45]. A espectroscopia na região do infravermelho próximo inclui várias aplicações que abrangem o controle de qualidade de produtos farmacêuticos [46], medidas de hemoglobina em sangue [47], determinação da octanagem em gasolina [48], caracterização de propriedades em polímeros [49, 50], tais como configuração, conformação, cristalinidade, orientação e miscibilidade, etc. ou a descrição da cinética e reações de mecanismos de cura (endurecimento) em diferentes sistemas epoxi-amina [51 - 53].

Devido à sua habilidade em analisar rápida e não-destrutivamente várias propriedades químicas e físicas de uma grande variedade de substâncias, as técnicas espectroscopicas na região do infravermelho próximo são bem adaptáveis para testes em controle de processos ou aplicações "on-line" e "at-line".

Um espectro no infravermelho próximo não pode ser interpretado de maneira direta como um espectro na região do infravermelho médio. Espectros na região do infravermelho médio apresentam bandas estreitas e picos, essencialmente relacionados a modos fundamentais de vibração, enquanto espectros na região do infravermelho próximo apresentam bandas largas como resultado da sobreposição de vários picos individuais. Adicionalmente, a presença de ressonâncias de Fermi pode aumentar a complexidade dos espectros no infravermelho próximo. A ressonância de Fermi é a interação ou acoplamento de dois estados de energia vibracional resultando na separação de estados onde um dos estados é um "overtone" ou um "sumtone". Um "overtone" é um caso particular de "sumtone" em que as freqüências dos níveis adicionados são iguais. A próxima seção provê alguns conceitos básicos sobre a espectroscopia vibracional, que torna a espectroscopia NIR um pouco mais compreensível.

4.2.1 – Princípios da espectroscopia vibracional

A luz apresenta comportamento ondulatório (radiação eletromagnética) e corpuscular (partícula). A teoria quântica determina que a energia de um fóton (E_f) ("partícula de luz") é dada por

$$E_f = h\,\nu \qquad 4.4$$

Em que h é a constante de Plank ($6,6256x10^{-27}$ erg seg) e ν é a freqüência da luz. Assim a energia de um fóton pode ser quantificada e é ela que interage com os níveis de energia das ligações químicas.

A radiação infravermelha absorvida por uma molécula excita vibrações nas ligações individuais como a um oscilador diatômico. Logo é conveniente partir de moléculas diatômicas como o mais simples sistema de vibração e depois estender o conceito às moléculas poliatômicas.

4.2.1.1 – Oscilador harmônico.

A freqüência que um dipolo vibra (estiramento ou deformação) é dependente da força da ligação e das massas envolvidas nas mesmas. Quando uma ligação química vibra, a energia é continuamente convertida entre potencial e cinética. Uma aproximação imediata é a modelagem da ligação química por um sistema massa-mola como um oscilador harmônico. No caso de um oscilador harmônico ideal, a energia potencial (V) apresenta apenas um termo quadrático:

$$V = 0,5\ k\ (r-r_e)^2 = 0,5\ k\ x^2 \qquad 4.5$$

Onde k é a constante de força da ligação, r e a distância internuclear, r_e é a posição de equilíbrio internuclear e $x = (r-r_e)$ é o deslocamento em relação à posição de equilíbrio. A energia potencial apresenta forma parabólica e simétrica em relação à posição de equilíbrio (r_e).

O modelo de vibração mecânica para uma molécula diatômica apresenta freqüência de vibração dada por:

$$\nu = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}} \qquad 4.6$$

onde $\mu = m_1 m_2/(m_1+m_2)$ é a massa reduzida do sistema diatômico. Um tratamento mecânico-quântico mostra que o oscilador harmônico apresenta níveis de energia dados por:

$$E_v = \hbar\omega\left(v+\frac{1}{2}\right); \quad \omega = \left(\frac{k}{\mu}\right)^{\frac{1}{2}} = 2\pi\nu; \quad \hbar = \frac{h}{2\pi}; \quad v = 1,2,3,... \qquad 4.7$$

onde h é a constante de Plank, ν é a freqüência clássica de vibração (definida na Equação 4.6) e v é o número quântico de vibração.

Os níveis de energia expressos em unidades de número de onda (cm^{-1}) são dados por:

$$G(v)(cm^{-1}) = \frac{E_{vib}}{\hbar c} = \overline{\nu}\left(v+\frac{1}{2}\right); \quad \overline{\nu} = \frac{\omega}{2\pi c} \qquad 4.8$$

sendo $\overline{\nu}$ o número de onda da transição vibracional.

4.2.1.2 – Oscilador anarmônico.

O modelo harmônico gera erros razoavelmente altos para níveis energéticos mais excitados, de modo que é apenas utilizado para se entender o conceito de vibração molecular. Os níveis de energia para os harmônicos (níveis excitados com maior energia) não são exatamente múltiplos inteiros da energia do nível fundamental ($v=0 \rightarrow E_{vib} = h\nu/2$). Uma boa aproximação para o cálculo dos níveis de energia de uma ligação química é o uso do potencial de Morse, dado pela Equação 4.9.

$$V = hcD_e\left(1 + e^{-\alpha(r-r_e)}\right)^2; \quad \alpha = \left(\frac{\mu}{2hcD_e}\right)^{\frac{1}{2}}\omega; \quad \omega = 2\pi\nu \qquad 4.9$$

Onde D_e é a profundidade do poço de potencial (energia de dissociação), c é a velocidade da luz no vácuo e ω é a freqüência angular.

Perto do mínimo de energia (posição de equilíbrio r_e), o potencial de Morse se comporta como uma função parabólica. Porém em níveis energéticos excitados (harmônicos) ele contempla a possibilidade de dissociação da ligação (quando a energia é igual a D_e).

A equação de Schrödinger pode ser resolvida para o potencial de Morse e obtém-se como resposta os níveis energéticos dados pela equação 4.10.

$$G(v)(cm^{-1}) = \overline{\nu}\left(v+\frac{1}{2}\right) - \overline{\nu}\left(v+\frac{1}{2}\right)^2 Xe + \overline{\nu}\left(v+\frac{1}{2}\right)^3 Ye + \ldots \qquad 4.10$$

onde Xe, Ye, ... , são as constantes empíricas que dependem dos átomos envolvidos na ligação química. Essas constantes podem ser ajustadas à energia de dissociação experimentalmente. Na prática, para o cálculo dos harmônicos superiores (níveis excitados), a equação 4.10 é truncada no 2º termo.

A presença de anarmonicidade apenas é relevante para o cálculo de absorções fracas correspondentes a transições entre os níveis 2 ← 0, 3 ← 0, etc. Pelas regras de seleção da mecânica quântica para o movimento vibracional ($\Delta v = \pm 1$), o 1º "overtone" (transição para o segundo harmônico) é proibido. A explicação para a existência de tais transições é que a regra de seleção é baseada no oscilador harmônico. A presença de anarmonicidade relaxa essa regra de seleção, porém a intensidade com que os "overtones" ocorrem é muito menor que a intensidade dos modos vibracionais fundamentais. Os níveis de energia vibracional são mostrados na Figura 4.1 como linhas horizontais.

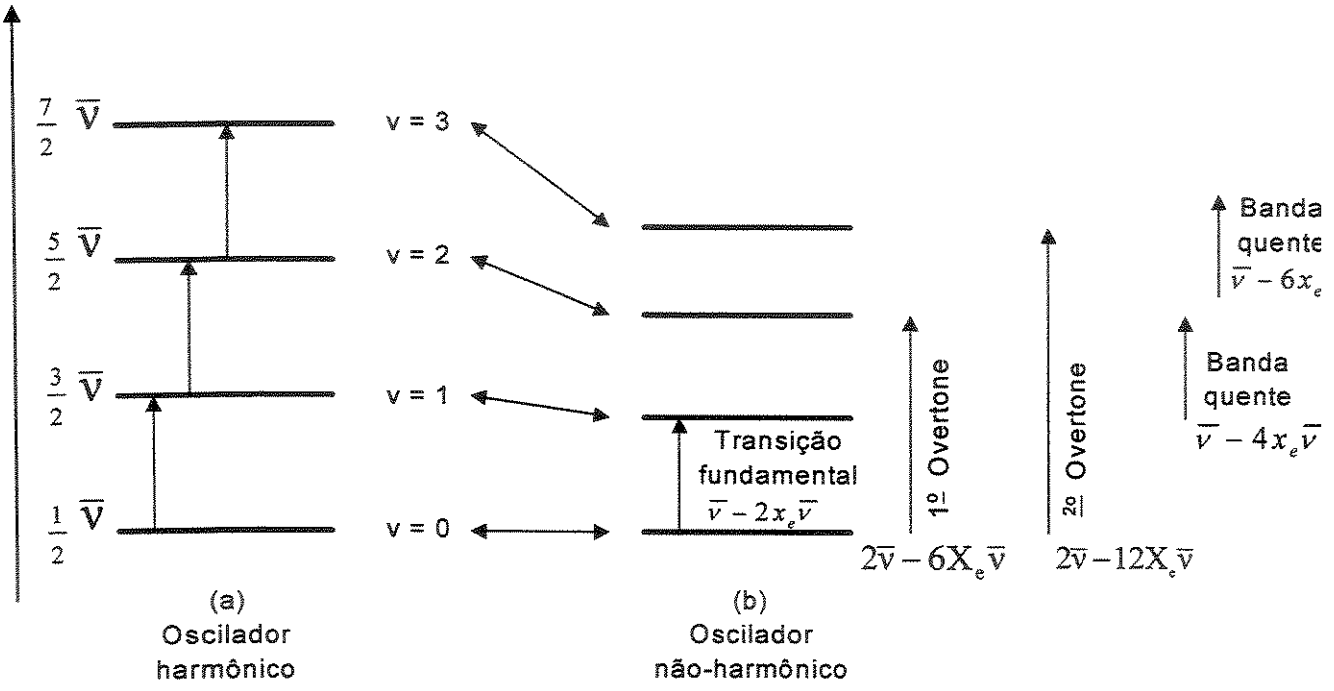


Figura 4.1 – Níveis vibracionais de energia e transições associadas para moléculas diatômicas.

4.2.2 - Tratamento quimiométrico de dados

Fortunadamente, a utilização de métodos estatísticos tem permitido a obtenção de informações qualitativas e quantitativas a partir dos complexos espectros na região do infravermelho próximo. O desenvolvimento de várias técnicas matemáticas e suas disponibilidades em programas de computador ("software") tem popularizado a utilização da espectroscopia no infravermelho próximo.

Através de técnicas quimiométricas, mais especificamente, técnicas de calibração multivariada, a determinação de propriedades químicas e físicas de amostras que absorvem na região do infravermelho próximo tem sido possível. Isso tem sido feito por métodos com modelagem local, como a regressão linear múltipla [9-12], que utiliza a absorbância em

comprimentos de onda selecionados, ou métodos de modelagem global, tais como a regressão por componentes principais (PCR) [9-13] ou mínimos quadrados parciais (PLS) [9-13].

O primeiro passo é realizar os experimentos de calibração, cruciais para determinações quantitativas. Esta etapa envolve a aquisição de um conjunto de espectros de referência (ou de calibração), abrangendo amostras em toda a faixa da propriedade de interesse a ser monitorada, ou seja, qualquer amostra que se pretenda estimar (conjunto de previsão) deve ser bem representada pelas amostras do conjunto de calibração. O propósito dos experimentos de calibração é estabelecer um modelo matemático que relacione os espectros NIR à propriedade de interesse, previamente determinada para todas as amostras do conjunto de calibração por uma técnica independente (método de referência).

Uma vez estabelecido o modelo matemático de calibração, a(s) propriedade(s) de interesse para amostras desconhecidas pode(m) ser determinada(s) a partir dos seus espectros (etapa de previsão).

A espectroscopia no infravermelho próximo não poderia ter alcançado o nível atual de sucesso sem o uso de rotinas quimiométricas aplicadas como pré-processamento nos dados brutos, e como análise de dados para a obtenção de parâmetros relevantes ou para a previsão de propriedades de interesse. Dados brutos, obtidos a partir de espectros NIR, precisam ser matematicamente processados para eliminar defeitos geralmente observados nos espectros.

A derivação numérica [54, 55] pode ser utilizada para remover deslocamentos de linha base ou para resolver picos sobrepostos. A segunda derivação é mais comumente

utilizada. Como essas técnicas aumentam consideravelmente os ruídos, é necessário também aplicar técnicas de alisamento ou remoção de ruídos [16, 54].

A correção de espalhamento multiplicativo (MSC – Multiplicative scatter correction) [56] é outra técnica de correção espectral. Ela enfoca o espalhamento de luz devido à variação no tamanho das partículas. A correção consiste em estabelecer um termo aditivo e um multiplicativo para corrigir cada espectro em relação a um espectro de referência.

Várias técnicas de calibração multivariada têm sido utilizadas para a calibração dos espectros na região do infravermelho próximo. O PCR e o PLS são os métodos mais básicos e mais utilizados para essa finalidade. Regressão por Fourier [57] e, mais recentemente, por redes neurais [58] também têm sido utilizados.

4.3 – Aplicações dos modelos propostos

Os modelos não lineares de calibração multivariada propostos foram aplicados na determinação do pol, ART e brix a partir de espectros na região do infravermelho próximo de amostras de xarope de caldo de cana (caldo de cana concentrado).

Foram utilizados espectros de absorção no infravermelho próximo de xarope de caldo de cana obtidos em um espectrofotômetro FEMTO PLS-PLUS na faixa de 1200 a 2500 nm com amostragem de 626 pontos por espectro, ou seja, com resolução de 2,08 nm. As medidas de absorbância foram feitas a partir da luz transmitida em uma cela com 1 mm de caminho óptico.

Os dados foram obtidos na usina Alta Mogiana em São Joaquim da Barra – SP. Os valores de pol, brix e ART, padrões de referência para calibração multivariada, também foram obtidos pelo corpo técnico da usina Alta Mogiana.

Para a execução da seleção de freqüências de Fourier e da otimização da RBFN, ambos por algoritmo genético, foram construídos programas em ambiente computacional matricial Matlab. Esses programas foram conectados em uma interface gráfica, também construída em Matlab. A Figura 4.2 mostra a janela principal dessa interface gráfica construída. Cada botão acessa comandos ou outras janelas com comandos para executar as operações anteriormente discutidas.

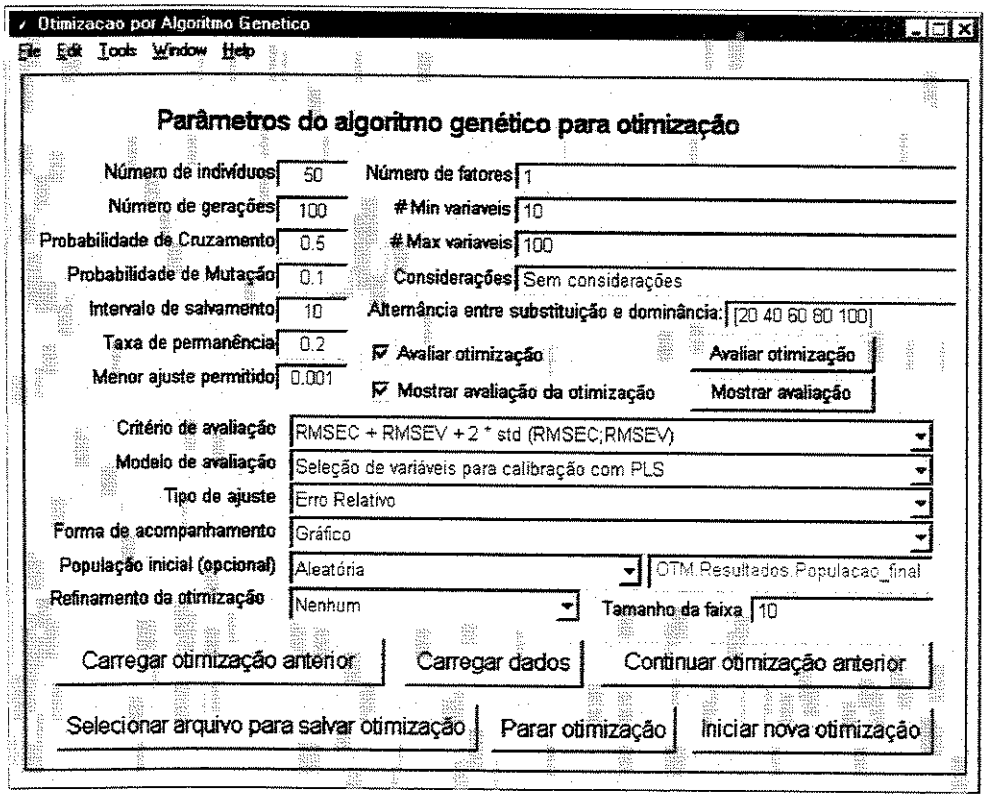


Figura 4.2 - Janela principal da interface gráfica do programa que executa os modelos com otimização por AG propostos.

Através da interface gráfica, pode-se facilmente inserir os parâmetros do algoritmo genético (já discutidos no capítulo 1), bem como abrir outras janelas da interface gráfica para a aquisição (carregamento) de dados e execução de pré-tratamento nos mesmos (se necessário). Os botões são intuitivos e permitem uma fácil utilização dos programas de otimização por AG. A otimização pode ser executada partindo de uma população inicial (inclusive aleatória) ou continuando uma otimização anterior do ponto (número de gerações) em que foi interrompida. Adicionalmente, após a otimização, se houverem dados de previsão, pode ser realizada uma avaliação do modelo otimizado por AG, de modo a verificar os erros de previsão comparando-os com os erros de validação e de calibração. Uma estimativa comparativa desses erros se faz necessária para avaliar o grau de sobre-ajuste do modelo.

Uma documentação de apoio com manuais e instruções de utilização dos programas construídos e da interface gráfica, estão disponíveis gratuitamente no "site" (URL) do Laboratório de Quimiometria em Química Analítica – LAQQA, cujo endereço na Internet é http:\\laqqa.iqm.unicamp.br

4.3.1 – Determinação do brix de amostras de caldo de cana concentrado (xarope) a partir de espectros na região do infravermelho próximo, com o modelo de seleção de freqüências de Fourier por algoritmo genético para calibração com PLS (FFT-PLS-AG).

O modelo de seleção de freqüências de Fourier por algoritmo genético proposto foi aplicado em um conjunto de 467 espectros NIR de caldo de cana concentrado (xarope) para

a determinação do brix. Foram utilizadas amostras com valores de brix dentro da faixa entre 64 e 74 °brix.

As amostras foram divididas em três grupos, sendo 267 amostras para a calibração, 100 para validação e 100 para previsão, escolhidas de modo representativo, isto é, as amostras de cada grupo (calibração, validação e previsão) foram escolhidas de modo que cada grupo possuísse valores de brix em toda a faixa monitorada, sem extrapolação nos conjuntos de validação e previsão. Os espectros NIR das amostras de xarope de caldo de cana são mostrados na Figura 4.3A, e os espectros de freqüência, aplicando a transformada de Fourier dessas amostras são mostrados na Figura 4.3B.

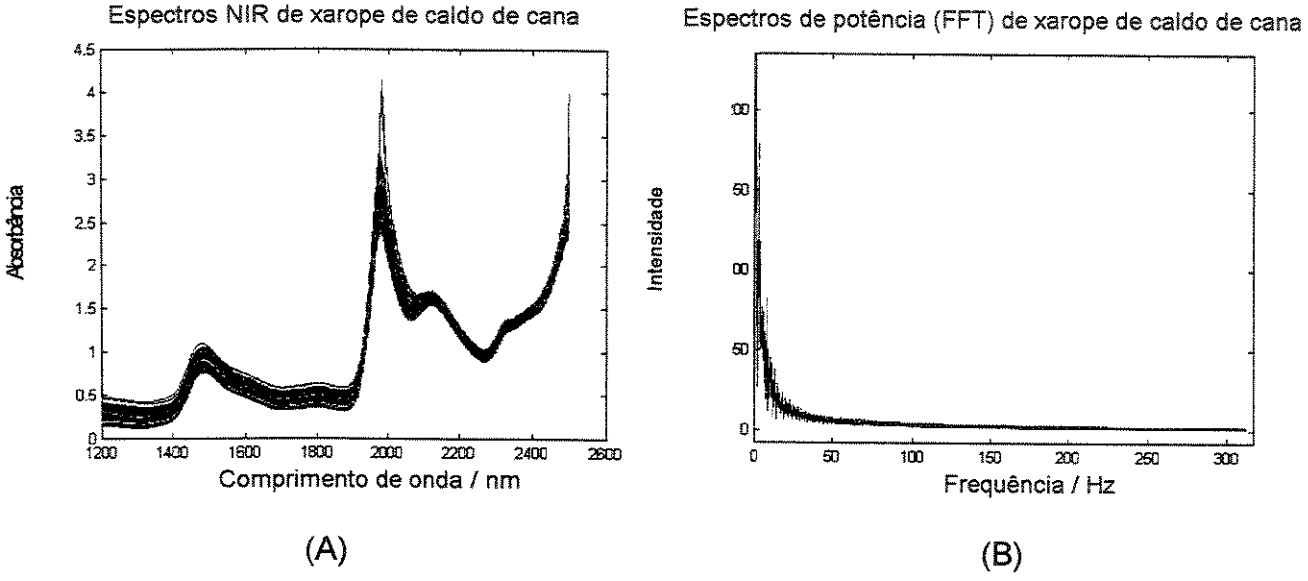


Figura 4.3- A - Espectros NIR de caldo de cana concentrado (xarope); B- Espectros de potência dos dados de xarope de caldo de cana.

Não foram encontradas amostras anômalas ("outliers") dentre as amostras analizadas.

Os espectros da Figura 4.3A mostram claramente um pico em torno de 2000 nm quase saturado correspondente à absorção da água. Essa informação não deve ser

transformada de Fourier) para um espaço com 92 dimensões (número de freqüências selecionadas) e em seguida, através de um modelo linear (PLS com 3 fatores) realizou-se a calibração multivariada desejada.

As freqüências que apresentam maior correlação com o brix foram selecionadas pelo modelo. Pode-se observar claramente que o modelo rejeitou as freqüências próximas a 60 Hz. Embora não seja perceptível pela observação direta dos espectros NIR, possivelmente deve haver uma interferência da rede elétrica (que oscila a 60 Hz) nessas freqüências de modo a diminuir (ou eliminar) a correlação entre as mesmas e o brix. Desempenho análogo é observado em torno de 120 Hz, porém em menor intensidade.

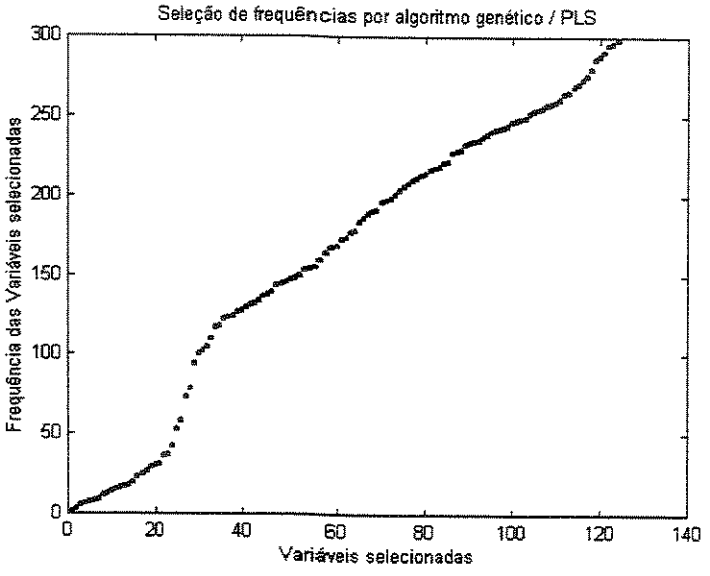


Figura 4.4 - Freqüências selecionadas pelo modelo proposto para a determinação do brix.

É interessante observar que a freqüência 0 Hz foi selecionada pelo modelo proposto, ou seja, um pré-processamento baseado em centrar os dados na média iria acarretar em uma calibração menos efetiva, visto que, conforme o modelo indica, a freqüência 0 Hz possui correlação com o brix.

O desempenho do modelo proposto foi avaliado através da raiz quadrada do erro quadrático médio percentual para calibração (%RMSEC), validação (%RMSEV) e previsão (%RMSEP), calculados pelas Equações 4.11, 4.12 e 4.13 respectivamente.

$$\%RMSEC = \frac{100}{C_m}\left[\frac{\sum_{i=1}^{n}\left(C_i - \hat{C}_i\right)^2}{n-k-1}\right]^{\frac{1}{2}} \qquad 4.11$$

$$\%RMSEV = \frac{100}{C_m}\left[\frac{\sum_{i=1}^{nv}\left(C_i - \hat{C}_i\right)^2}{nv}\right]^{\frac{1}{2}} \qquad 4.12$$

$$\%RMSEP = \frac{100}{C_m}\left[\frac{\sum_{i=1}^{np}\left(C_i - \hat{C}_i\right)^2}{np}\right]^{\frac{1}{2}} \qquad 4.13$$

onde C_m é a concentração média das amostras de calibração, C_i é a concentração real de cada amostra do respectivo conjunto, $\hat{C}_i$ é a estimativa do modelo proposto para a i-ésima concentração, n é o número de amostras para calibração, k é o número de variáveis latentes utilizadas pelo PLS, nv é o número de amostras para a validação e np é o número de amostras para a previsão.

Para efeito de comparação, foram calculados os erros de previsão dos dados com o PLS convencional utilizando 2 variáveis latentes. O número de variáveis latentes foi determinado pelo erro de previsão por validação cruzada tipo "leave one out" [59, 60]. Como os dados de espectro de caldo de cana apresentam apenas uma espécie (por hipótese seria

apenas a sacarose), a segunda variável latente provavelmente deve ser utilizada para modelar não-linearidades inerentes aos ruídos experimentais nos dados.

Tabela 4.2 - Desempenho do PLS direto e do PLS com freqüências de Fourier selecionadas pelo modelo proposto, para os dados de xarope de caldo de cana.

	%RMSEC	%RMSEV	%RMSEP
Determinação do brix diretamente com o PLS	2,77	2,85	2,90
Determinação do brix através do PLS com as freqüências de Fourier selecionadas	1,30	1,26	1,25

Ambos os modelos se mostraram robustos, porém enquanto o PLS direto apresenta erros médios em torno de 2,84 % (considerando os três conjuntos de dados), o modelo de seleção de freqüências apresenta menores erros médios em torno de 1,27%.

Esses dados mostram claramente uma dependência quase linear entre os espectros e o brix, através dos erros obtidos com o PLS. Apesar disso o modelo de FFT-PLS-AG se mostrou eficiente, pois apesar de não melhorar os resultados significativamente, pôde reduzir os erros para a faixa de 1%.

Pode-se usar as amostras de previsão para verificar se os dois conjuntos de erros são significativamente distintos (através do teste F). Tomando os RMSEPs como estimativas dos desvios padrão, temos:

$$\frac{s^2_{PLS}}{s^2_{FFT-AG-PLS}} = \frac{RMSEP^{\,2}_{PLS}}{RMSEP^{\,2}_{FFT-AG-PLS}} = \frac{(0{,}029)^2}{(0{,}0125)^2} = 5{,}38 \quad ; \quad F(99\%,100,100) = 1{,}60$$

Logo, pode-se afirmar com 99% de certeza que as distribuições de erros são estatisticamente distintas.

4.3.2 – Determinação do pol de amostras de caldo de cana concentrado (xarope) a partir de espectros na região do infravermelho próximo, com o modelo de seleção de freqüências de Fourier por algoritmo genético para calibração com PLS (FFT-PLS-AG).

O modelo de seleção de freqüências de Fourier por algoritmo genético proposto também foi aplicado em um conjunto de 202 espectros NIR de caldo de cana concentrado (xarope) para a determinação do pol. Foram utilizadas amostras com valores de pol dentro da faixa entre 4,1994 e 7,1703.

As amostras foram divididas em três grupos, sendo 103 amostras para a calibração, 50 para validação e 49 para previsão, escolhidas de modo representativo. As amostras de cada grupo (calibração, validação e previsão) foram escolhidas de modo que cada grupo possuísse valores de pol em toda a faixa monitorada, sem extrapolação dos conjuntos de validação e previsão. Não foram encontradas amostras anômalas ("outliers") dentre as amostras analisadas. Os espectros NIR das amostras de xarope de caldo de cana são mostrados na Figura 4.5A, e os espectros de potência, após a aplicação da transformada de Fourier, dessas amostras são mostrados na Figura 4.5B.

Para a otimização por algoritmo genético foram utilizados os seguintes parâmetros:

Tamanho da população = 40 indivíduos;

Número de gerações = 400;

Probabilidade de cruzamento = 0,9;

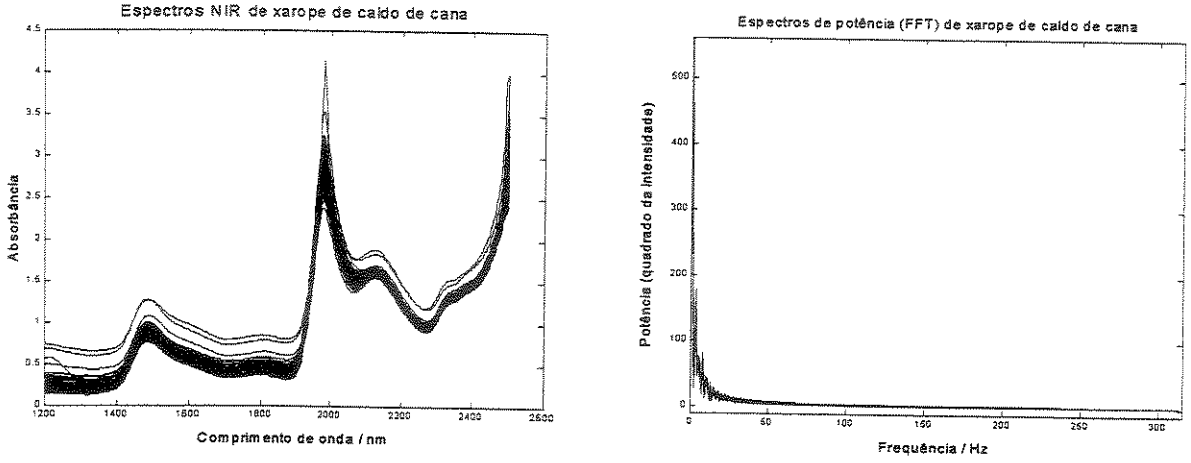
Probabilidade de mutação = 0,1;

Número mínimo de variáveis selecionadas = 10;

Número máximo de variáveis selecionadas = 300;

Número de fatores para a modelagem com PLS = 7;

Critério de avaliação = RMSEC + RMSEV + módulo da diferença entre RMSEC e RMSEV.



(A) (B)

Figura 4.5 - A - Espectros NIR de caldo de cana concentrado (xarope); B- Espectros de potência dos dados de xarope de caldo de cana.

Novamente, os espectros da Figura 4.5A mostram claramente um pico em torno de 2000 nm quase saturado correspondente à absorção da água. Essa informação não deve ser relevante para a previsão do brix, e a(s) freqüência(s) relativas a esse pico provavelmente devem ser excluídas pelo algoritmo genético ao longo da otimização (seleção) de freqüências para gerar um modelo de calibração que melhor explique a variação do brix em função das medidas de espectro NIR.

Também utilizou-se o critério de dominância para o melhor indivíduo.

As variáveis (freqüências de Fourier) selecionadas pelo modelo proposto são mostradas na Figura 4.6. As freqüências que apresentam maior correlação com o pol foram

selecionadas pelo modelo. Pode-se observar, mais uma vez, que o modelo rejeitou as freqüências próximas a 60 Hz e 120 Hz. Embora não seja perceptível pela observação direta dos espectros NIR, possivelmente deve haver uma interferência da rede elétrica (que oscila a 60 Hz) nessas freqüências de modo a diminuir (ou eliminar) a correlação entre as mesmas e o pol. Desempenho análogo é observado em torno de 120 Hz, porém em menor intensidade.

É interessante observar que também a freqüência 0 Hz foi selecionada pelo modelo proposto, ou seja, um pré-processamento baseado em centrar os dados na média iria acarretar em uma calibração menos eficiente, visto que, conforme o modelo indica, a freqüência 0 Hz possui correlação com o POL. Esse fato é reforçado pela comparação com o modelo PLS utilizando os dados de espectro centrados na média (o que é equivalente a remover a freqüência nula). O modelo com os dados centrados na média mostrou-se menos eficiente que o modelo com os dados originais (sem pré-tratamento de variáveis).

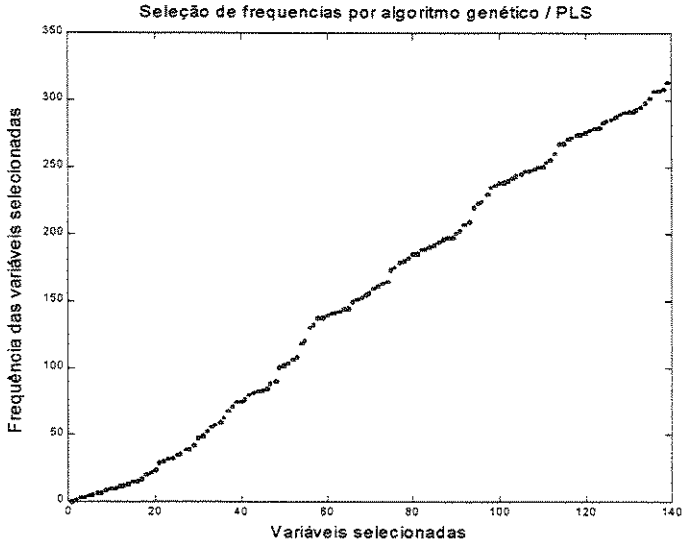


Figura 4.6. Freqüências selecionadas pelo modelo proposto para a determinação da pol.

As variáveis selecionadas pelo modelo proposto são mostradas na Figura 4.4. Foram selecionadas 79 freqüências.

Pelo modelo proposto, partiu-se de um espaço com 626 variáveis correlacionadas, cuja dimensão real é algo menor que 10, Expandiu-se de forma não linear (pela transformada de Fourier) para um espaço com 79 dimensões (número de freqüências selecionadas) e em seguida, através de um modelo linear (PLS com 7 fatores) realizou-se a calibração multivariada desejada.

O desempenho do modelo proposto foi avaliado através da raiz quadrada do erro quadrático médio percentual para calibração (%RMSEC), validação (%RMSEV) e previsão (%RMSEP), calculados pelas Equações 4.11, 4.12 e 4.13 respectivamente. Esse desempenho é avaliado através da discrepância entre o %RMSEC, %RMSEV e %RMSEP. Idealmente eles devem ser iguais, indicando que não há sobreajuste em nenhum conjunto, e baixos indicando que o modelo possui uma boa capacidade de previsão de amostras desconhecidas. A Tabela 4.3 mostra esses critérios de avaliação, comparando o modelo proposto com o modelo PLS (padrão de calibração multivariada).

Tabela 4.3 - Desempenho do PLS direto e do PLS com freqüências de Fourier selecionadas pelo modelo proposto, para os dados de xarope de caldo de cana.

	%RMSEC	%RMSEV	%RMSEP
Determinação da pol diretamente com o PLS	5,14	4,12	4,31
Determinação da pol através do PLS com freqüências de Fourier selecionadas por AG	1,39	1,42	1,50

Para efeito de comparação, foram calculados os erros de previsão dos dados com o PLS convencional utilizando 7 variáveis latentes. O número de variáveis latentes foi determinado pelo erro de previsão para as amostras de validação.

Ambos os modelos são robustos, porém enquanto o PLS direto apresenta erros relativos médios em torno de 4,52%, o modelo de seleção de freqüências apresenta menores erros relativos médios em torno de 1,44%. Essa diminuição no erro é bastante significativa, pois o preço da cana paga pela indústria é dado por esse parâmetro. Para a substituição do método convencional pelo NIR exige-se que esses erros estejam em torno de 2%, conforme informações colhidas na usina Alta Mogiana.

Tomando o RMSEP como estimativa do desvio padrão das distribuições de erro dos modelos, pelo teste F, tem-se:

$$\frac{s^2_{PLS}}{s^2_{FFT-AG-PLS}} = \frac{RMSEP^{\,2}_{PLS}}{RMSEP^{\,2}_{FFT-AG-PLS}} = \frac{(0{,}0431)^2}{(0{,}0150)^2} = 8{,}26 \quad ; \quad F(99\%,49,49) = 1{,}96$$

Logo, pode-se afirmar com 99% de certeza que as a distribuição de erros obtida para o modelo FFT-AG-PLS é estatisticamente distinta.da distribuição de erros obtida pelo PLS.

As distribuições dos erros relativos de previsão para os modelos FFT-PLS-AG e PLS, são mostrados nas Figuras 4.7A e 4.7B respectivamente.

Observa-se, pela Figura 4.7, que ambos os modelos apresentam erros relativos percentuais para as amostras de previsão (ERPP) distribuídos aleatoriamente em torno do zero. Porém o modelo utilizando-se apenas o PLS mostra erros muito maiores e muito altos para as primeiras amostras (que possuem baixo valor de pol).

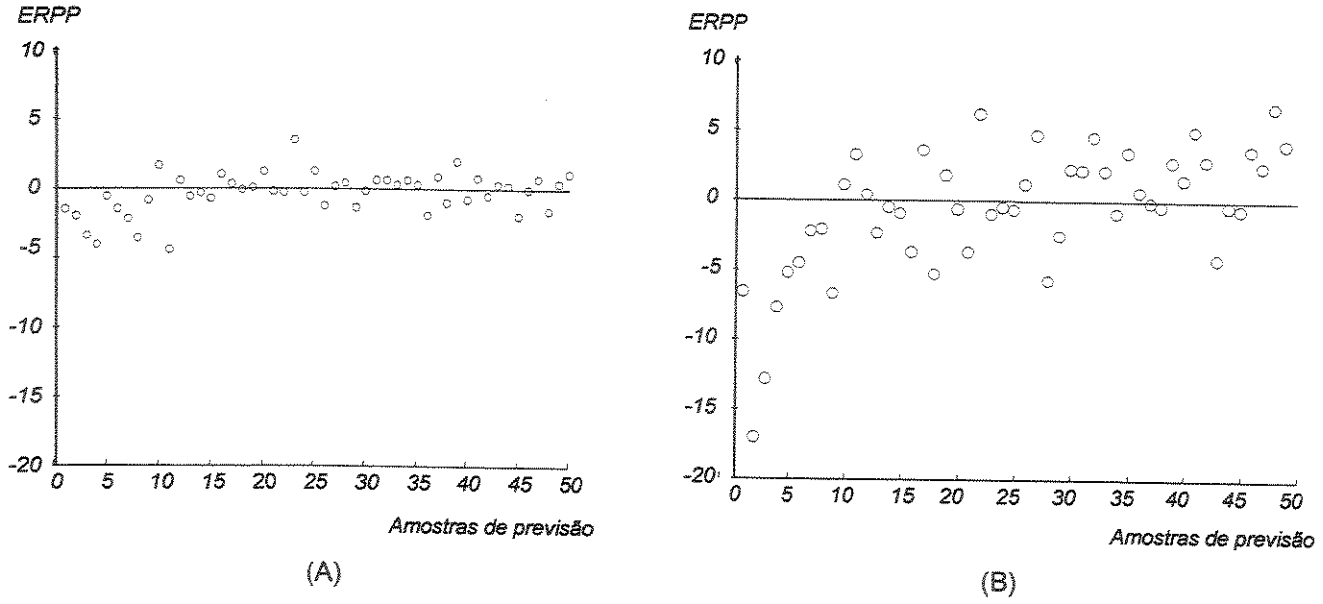
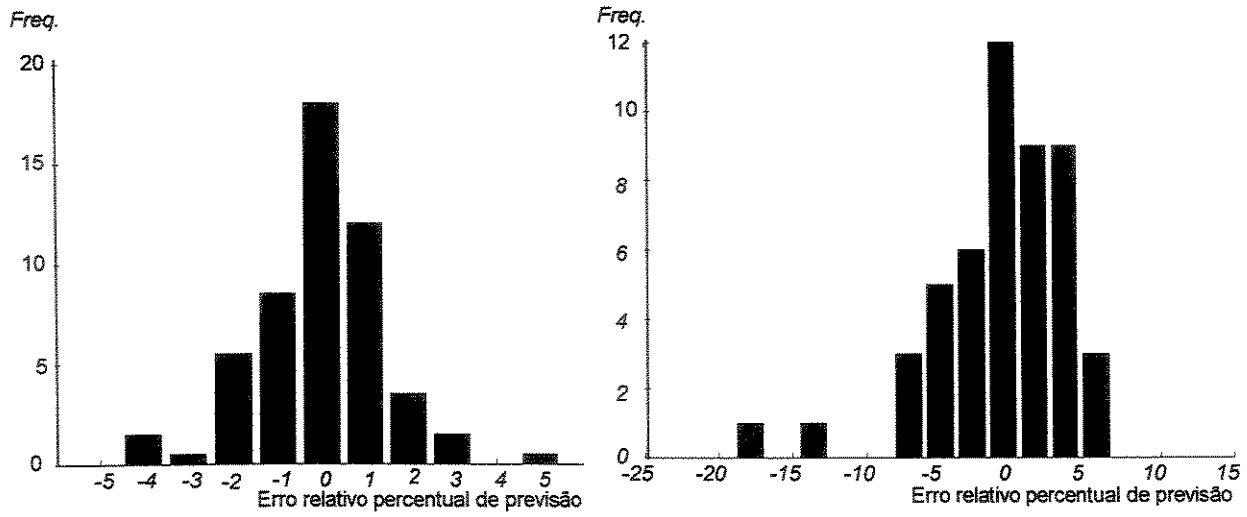


Figura 4.7 – Distribuição dos erros relativos percentuais de previsão (ERPP) obtidos para determinação de pol em amostras de xarope de caldo de cana. A - Pelo modelo de FFT-PLS-AG. B - Pelo modelo de PLS.

As amostras foram ordenadas de modo crescente, assim as primeiras amostras (à esquerda) referem-se aos menores valores de pol.

Os histogramas das distribuições de erro mostrados na Figura 4.7 são mostrados na Figura 4.8. Nela observa-se que a distribuição de erros do modelo proposto se assemelha mais a uma gaussiana que a distribuição de erros obtida pelo PLS. Ou seja, a distribuição de erros do modelo FFT-PLS-AG é realmente aleatória, enquanto que para o PLS parece haver uma certa tendência nos erros, indicando que o modelo desenvolvido (FFT-PLS-AG) é o mais adequado para os dados em questão.



(A) (B)

Figura 4.8 – Histograma da distribuição dos erros relativos percentuais de previsão (ERPP) obtidos para determinação de pol em amostras de xarope de caldo de cana. A - Pelo modelo de FFT-PLS-AG. B - Pelo modelo de PLS.

4.3.3 – Determinação do ART de amostras de caldo de cana concentrado (xarope) a partir de espectros na região do infravermelho próximo, com o modelo de calibração multivariada com RBFN otimizada por algoritmo genético (RBFN-AG)

O modelo de otimização de raios, centros, tipos e número de RBF's com otimização das conexões (poda) em uma RBFN foi aplicado em um conjunto de 192 espectros NIR de caldo de cana concentrado (xarope) para a determinação do ART (açúcares redutores totais). Foram utilizadas amostras com valores de ART dentro da faixa entre 59,82% e 70,32%.

As amostras também foram divididas em três grupos, sendo 97 amostras para calibração, 47 amostras para validação e 47 amostras para previsão, também escolhidas de

modo representativo. Não foram encontradas amostras anômalas ("outliers") dentre as amostras analisadas.

As amostras foram pré-tratadas com utilização de PCA com 6 componentes principais (99,99% de variância explicada) para ortogonalização dos dados de entrada sem perda das informações contidas nos espectros.

Foram utilizadas para otimização um conjunto com 4 RBF's gaussianóides (h_1 a h_4), 4 RBF's cauchys (h_5 a h_8), 4 RBF's hiperquádricas (h_9 a h_{12}) e 4 RBF's hiperquádricas inversas (h_{13} a h_{16}), ou seja, uma possibilidade de construir uma rede com até 16 RBFs.

Para a otimização por algoritmo genético foram utilizados os seguintes parâmetros:

Tamanho da população = 50 indivíduos;

Número de gerações = 800;

Probabilidade de cruzamento = 0,5;

Probabilidade de mutação = 0,1;

Precisão utilizada em cada parâmetro contínuo = 10 bits

Critério de avaliação = RMSEC + RMSEV + módulo da diferença entre RMSEC e RMSEV.

Após a otimização pelo algoritmo genético, foram selecionadas apenas 1 RBF gaussianóide (h_1), 3 RBF's cauchys (h_2, h_3 e h_4), 2 RBF's hiperquádricas (h_5 e h_6) e 2 RBF's hiperquádricas inversas (h_7 e h_8), totalizando 8 RBF's dentre as 16 RBF's disponíveis, como mostrado na Figura 4.9. A otimização por AG eliminou completamente as conexões da 1ª, 2ª e 4ª RBF's (gaussianas); da 5ª RBF (cauchy); da 9ª e 10ª RBF's (hiperquádricas) e da 15ª e 16ª RBF's (hiperquádricas inversas), e desse modo removeu essas RBF's da RBFN. Simultaneamente, a otimização por AG também restringiu as entradas (componentes

principais) das RBF's que permaneceram na rede, eliminando algumas conexões, conforme mostrado na Figura 4.9. Esse processo garante uma baixa probabilidade de sobreajuste, conforme indicam os parâmetros de avaliação da otimização mostrados adiante.

Pelo modelo proposto, partiu-se de um espaço com 626 variáveis correlacionadas, cuja dimensão real é algo menor que 6 (o PCA 6 com variáveis latentes já explica 99,99% da variância total dos dados), expandiu-se de forma não linear (pela transformada de Fourier) para um espaço com 8 dimensões (número de RBFs utilizadas pela rede otimizada) e em seguida, através de um modelo linear (MLR da própria RBFN) realizou-se a calibração multivariada desejada.

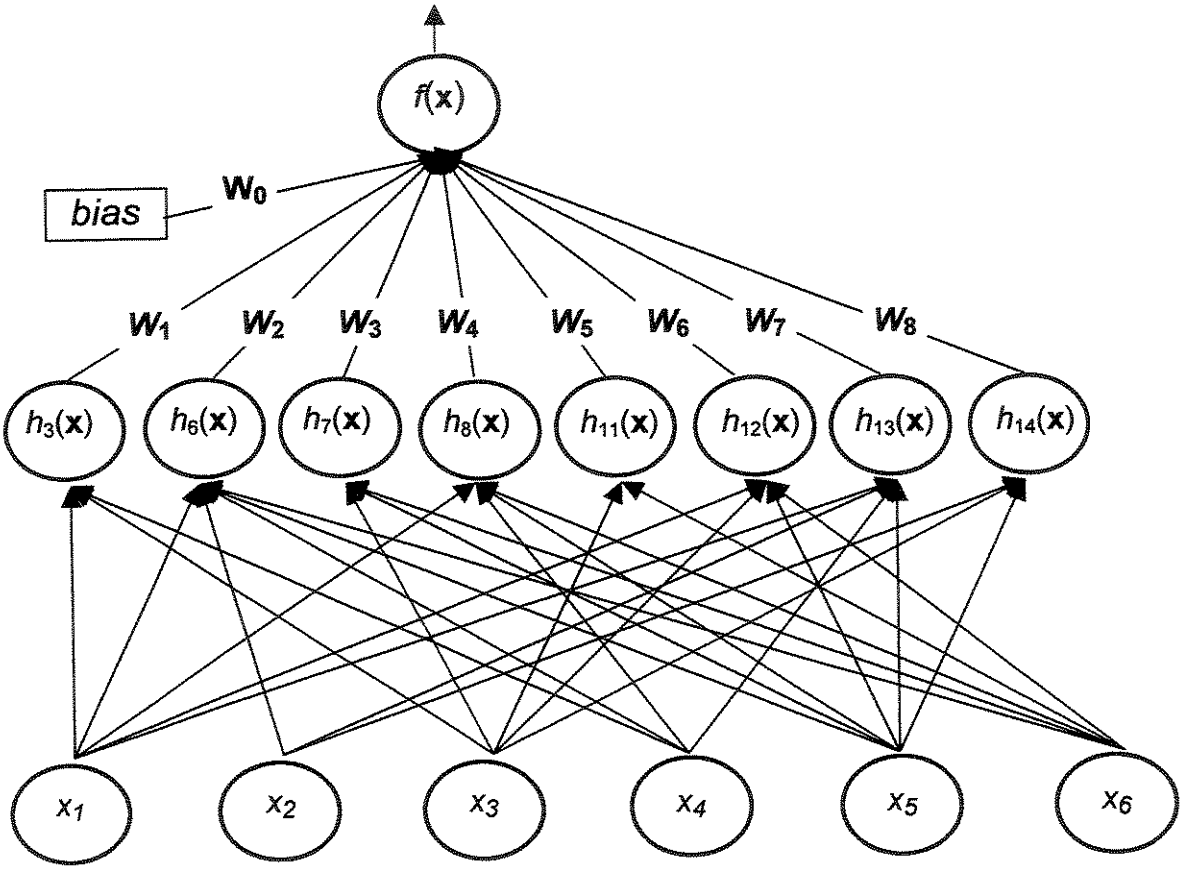


Figura 4.9 - Representação da RBFN otimizada por AG.

A partir da análise da RBFN otimizada pelo AG, pode-se extrair informações a cerca do sistema em questão. Baseado na variância explicada por cada RBF, aliada com as informações a cerca dos raios (dispersão) e centros (posição), pode-se identificar quais as variáveis mais importantes na modelagem, ou seja quais as componentes principais apresentam maiores informações. Posteriormente, pela análise dos "loadings", pode-se identificar quais variáveis do espectro original (absorbâncias em comprimentos de onda específicos) são mais importantes para as componentes principais (variáveis de entrada da RBFN) mais influentes. A Tabela 4.4 mostra a variância explicada pela RBFN otimizada por AG, bem como a parcela de variância explicada por cada RBF da rede.

Tabela 4.4 – Variância explicada pelas RBF's da RBFN para os dados de ART.

	bias	H_3	h_6	h_7	h_8	h_{11}	h_{12}	h_{13}	h_{14}
W	69,24	6,284	15,14	4,936	5,087	-0,3076	0,3232	-7,200	-23,43
Variância explicada	92,19	0,138	0,0173	0,0064	0,00129	0,0346	0,0259	0,181	1,34

Variância total explicada = 93,93%

A Tabela 4.5 mostra os centros e a Tabela 4.6 mostra os raios das diferentes RBF's otimizados para cada entrada (Xi) pelo algoritmo genético. No caso em questão há apenas 6 entradas porque foram utilizadas 6 componentes principais para ortogonalizar os dados.

Os valores de raios e centros marcados com asterisco assinalam conexões que não foram selecionadas (foram cortadas) da RBFN.

Tabela 4.5 - Centros selecionados e otimizados pelo AG para cada entrada de cada RBF da RBFN.

Centros	h_3	h_6	h_7	h_8	h_{11}	h_{12}	h_{13}	h_{14}
X_1	27,356	28,328	*	28,404	*	34,894	31,835	*
X_2	*	1,2034	*	*	*	*	-0,0645	2,6442
X_3	0,2413	*	-0,2549	*	-1,1232	-0,4547	*	2,2259
X_4	-0,6323	-0,4545	*	-0,9374	*	*	0,4662	*
X_5	*	0,3228	-0,2979	0,1906	*	-0,2793	0,3063	2,2259
X_6	*	0,1873	-0,3081	0,1689	0,3476	-0,2445	*	*

Adicionalmente as RBF's que não apresentaram nenhuma conexão foram removidas da rede. Desse modo obteve-se uma rede com o menor número de RBF's e conexões necessárias, de modo a construir um modelo de calibração multivariada com maior número de graus de liberdade. Esse procedimento diminui a possibilidade de sobreajuste.

Tabela 4.6 - Raios selecionados e otimizados pelo AG para cada entrada de cada RBF da RBFN.

Raios	h_3	h_6	h_7	h_8	h_{11}	h_{12}	h_{13}	h_{14}
X_1	1,0568	1,9456	*	3,1742	*	1,6776	1,3149	*
X_2	*	0,4449	*	*	*	*	0,8460	2,4960
X_3	0,6463	*	0,1447	*	0,7014	0,1146	*	0,6776
X_4	0,6181	0,6181	*	0,5508	*	*	0,8030	*
X_5	*	0,2582	0,3965	0,1576	*	0,2508	0,5822	0,3792
X_6	*	0,4248	0,3540	0,1110	0,1088	0,3682	*	*

Para a RBFN otimizada por AG em questão, a RBF h_{14} é responsável pela maior parte da variância explicada (além do "bias"). Logo, pode-se concluir que a 1ª componente principal (excluída pelo AG para a RBF h_{14}) é menos importante para a calibração que as 2ª, 3ª e 5ª componentes principais. Isso acaba com a idéia de que uma rede neural seja uma "caixa preta" e não apresente informações sobre o sistema modelado.

Os avaliadores (%RMSEC, %RMSEV e %RMSEP) são calculados de maneira análoga aos calculados para o PLS, exceto que o número de graus de liberdade (denominador) para os dados de calibração é igual ao número de amostras menos o número de RBF's e bias (termo linear) da RBFN menos 1. O número de graus de graus de liberdade é diferente porque cada parâmetro da rede restringe um grau de liberdade, e como os pesos (w) são determinados por mínimos quadrados, o somatório dos erros quadráticos é nulo, o que restringe mais um grau de liberdade.

Novamente o modelo construído foi comparado com o PLS (modelo padrão de calibração multivariada) e os avaliadores são mostrados na Tabela 4.7.

Para efeito de comparação, foram calculados os erros de previsão dos dados com o PLS convencional utilizando 6 variáveis latentes, sem pré-tratamento (melhor caso). O número de variáveis latentes foi determinado pelo erro de previsão para as amostras de validação.

Ambos os modelos são robustos, porém enquanto o PLS direto apresenta erros médios em torno de 2,45 %, o modelo de RBFN otimizada por AG apresenta erros médios em torno de 1,48% para os dados em questão. Adicionalmente, a discrepância entre os avaliadores de ambos modelos é comparável, diferindo apenas na 2ª casa decimal (desvio

padrão de 0.202 para a RBFN e 0.213 para o PLS). Também nesse caso, foi possível diminuir os erros relativos médios para a faixa de 1%.

Tabela 4.7 - Desempenho do PLS direto e da RBFN otimizada por AG.

	%RMSEC	%RMSEV	%RMSEP
Determinação de açúcares redutores totais diretamente com o PLS	2,34	2,32	2,70
Determinação de açúcares redutores totais através da RBFN otimizada por AG	1,38	1,36	1,72

Tomando o RMSEP como estimativa do desvio padrão das distribuições de erro dos modelos, pelo teste F, tem-se:

$$\frac{s^2_{PLS}}{s^2_{RBFN-AG}} = \frac{RMSEP^{\,2}_{PLS}}{RMSEP^{\,2}_{RBFN-AG}} = \frac{(0{,}0270)^2}{(0{,}0172)^2} = 2{,}46 \quad ; \quad F(99\%,47,47) = 1{,}99$$

Logo, pode-se afirmar com 99% de certeza que as a distribuição de erros obtida para o modelo RBFN-AG é estatisticamente distinta.da distribuição de erros obtida pelo PLS.

Pela distribuição dos erros relativos de previsão para os modelos RBFN-AG e PLS, mostradas nas Figuras 4.10A e 4.10B respectivamente, observa-se que o modelo com RBFN-AG apresenta erros relativos percentuais para as amostras de previsão (ERPP) distribuídos aleatoriamente em torno do zero, o que parece não ser o caso quando se utiliza o PLS. As amostras foram ordenadas de modo crescente, assim as primeiras amostras (à esquerda) referem-se aos menores valores de ART.

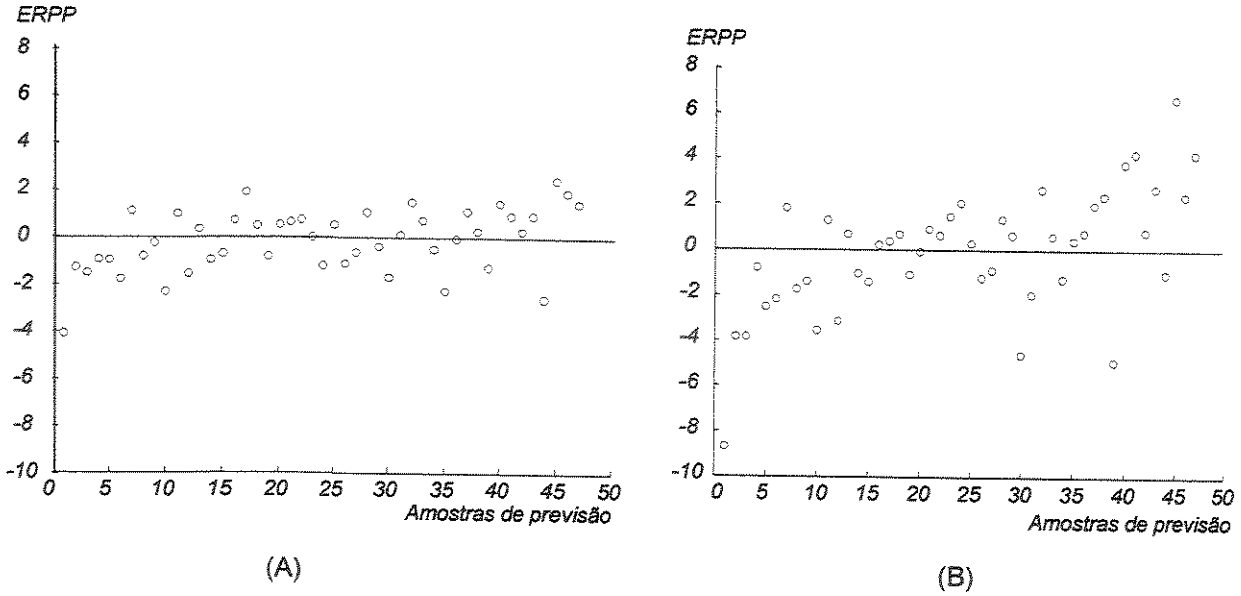
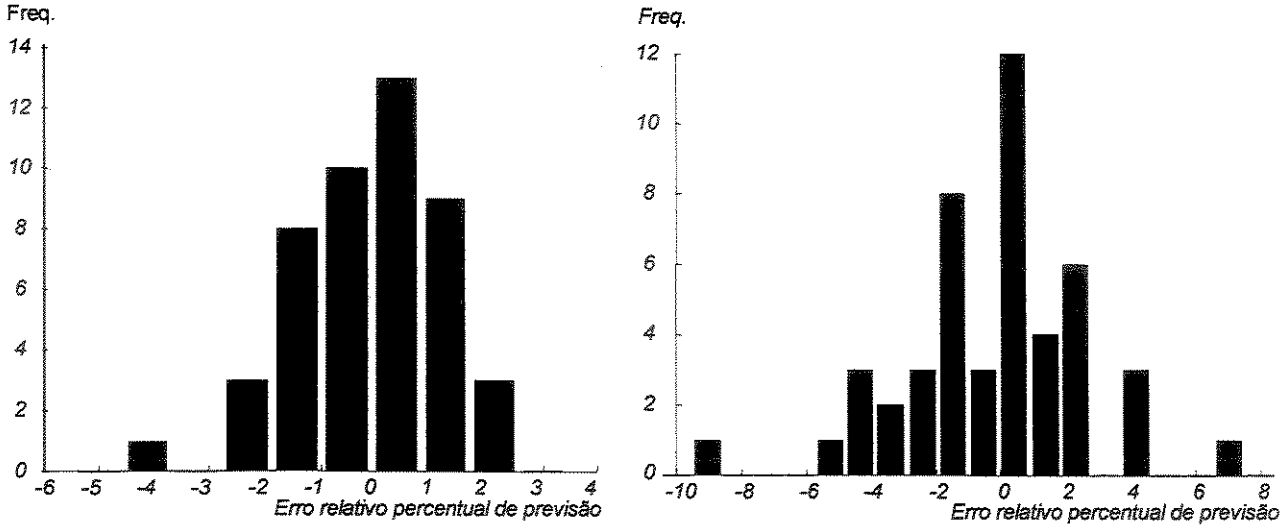


Figura 4.10 – Distribuição dos erros relativos percentuais de previsão (ERPP) obtidos para determinação de ART em amostras de xarope de caldo de cana. A - Pelo modelo de RBFN-AG. B - Pelo modelo de PLS.

Os histogramas das distribuições de erro mostrados na Figura 4.10 são mostrados na Figura 4.11. Nela observa-se que a distribuição de erros do modelo proposto com RBFN-AG novamente se assemelha mais a uma gaussiana que a distribuição de erros obtida pelo PLS. Em outras palavras, o modelo proposto parece ser mais adequado na modelagem de ART.

Apesar de, absolutamente, a diferença entre os erros relativos de previsão com o modelo PLS e RBFN-AG ser pequena, relativamente, a diferença é bastante significativa (quase o dobro). Ademais o PLS obtém erros de previsão que não possibilitam a sua utilização como alternativa viável em substituição aos métodos tradicionais de laboratório para a determinação do ART. Idealmente os métodos espectroscópicos aliados aos modelos quimiométricos não podem apresentar erros relativos médios superiores a 2% para poderem substituir os métodos tradicionais, segundo informações do pessoal técnico da usina Alta Mogiana.



(A) (B)

Figura 4.11 – Histograma da distribuição dos erros relativos percentuais de previsão (ERPP) obtidos para determinação de ART em amostras de xarope de caldo de cana. A - Pelo modelo de RBFN-AG. B - Pelo modelo de PLS.

Capítulo 5

Construção de um potenciostato monocanal

Para a aquisição de dados eletroquímicos, optou-se por construir um potenciostato monocanal interfaceado a um microcomputador. Desse modo, tem-se controle total do potencial aplicado, permitindo a manipulação da rampa de potencial para que se possa aplicar diferentes técnicas voltamétricas. Desenvolveu-se uma formatação própria para o armazenamento dos dados provenientes das análises eletroquímicas. Isso permitiu o total controle dos dados monitorados (corrente elétrica) de modo a desenvolver rotinas quimiométricas para tratá-los.

5.1 – Voltametria

Voltametria compreende um grupo de técnicas eletroanalíticas onde a informação sobre o analito provém de medidas de corrente elétrica como função de um potencial aplicado sobre um eletrodo indicador (ou de trabalho) sob condições de completa polarização, ou seja, os analitos são atraídos ao eletrodo de trabalho devido ao potencial aplicado sobre o mesmo, e ao encontrar o eletrodo, se houver potencial suficiente, sofrem o processo redox, desenvolvendo uma corrente elétrica. A migração de material eltroquimicamente ativo (íons) em direção aos eletrodos gera um gradiente de concentração entre a superfície do eletrodo e o meio da solução.

Na voltametria um potencial variável é aplicado a uma cela eletroquímica e a corrente elétrica de resposta é característica do sinal de excitação aplicado. Os sinais de excitação mais usados em voltametria são mostrados na Figura 5.1. Na Figura 5.1a o potencial cresce linearmente com o tempo, a corrente desenvolvida pela cela é então registrada em função do tempo. As Figuras 5.1b e a Figura 5.1c mostram dois tipos de sinais de excitação na

forma de pulsos. A corrente é então monitorada várias vezes durante o período desses pulsos (antes e após a aplicação de cada pulso). Na Figura 5.1-d o potencial oscila entre dois valores limite, crescendo e decrescendo linearmente com mesma taxa de variação. Este processo pode se repetir várias vezes dando origem a vários ciclos [61, 62].

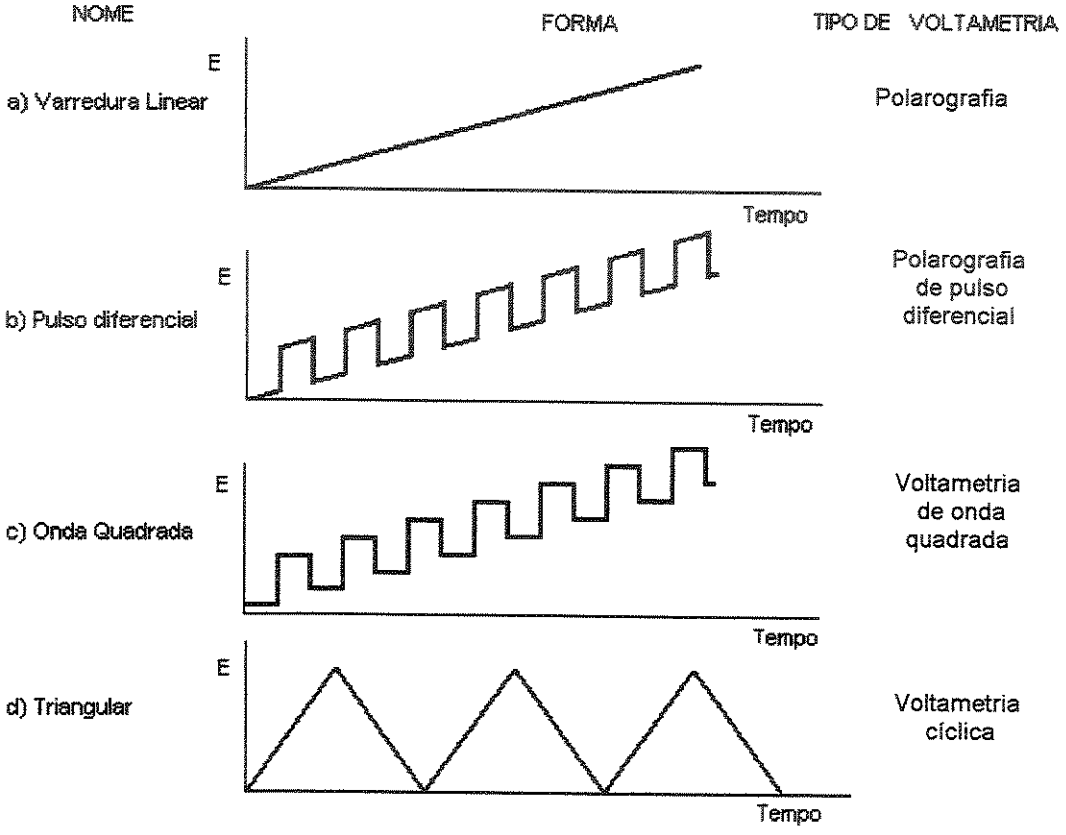


Figura 5.1 – Potenciais elétricos aplicados em voltametria.

5.2 – Sistema voltamétrico

A cela eletroquímica utilizada para a voltametria é geralmente constituída de três eletrodos, sendo um eletrodo de trabalho, um eletrodo de referência cujo potencial é constante e um contra-eletrodo ou eletrodo auxiliar. Aplica-se um potencial (ou diferença de

potencial) entre o eletrodo de trabalho e o eletrodo de referência e monitora-se a corrente entre o eletrodo de trabalho e o eletrodo auxiliar. Isso é feito para minimizar a queda ôhmica causada pela passagem da corrente, caso só houvesse dois eletrodos, o que alteraria o potencial aplicado. Existem, porém, casos em que se utilizam microeletrodos (eletrodos com área superficial pequena) onde a corrente monitorada é tão pequena que a queda ôhmica é desprezível, podendo-se utilizar celas eletroquímicas com apenas dois eletrodos.

Basicamente há três tipos de movimento das espécies eletroquimicamente ativas na voltametria:

1. Convecção – Quando a solução é agitada. Geralmente esse processo é prejudicial à análise, porém há casos em que se trabalha com eletrodos rotatórios em que a convecção é explorada para levar espécies ativas de um eletrodo a outro. Na prática, esse processo é minimizado pelo repouso da solução em estudo;

2. Migração – Devido à diferença de potencial estabelecida entre os eletrodos, as partículas carregadas (íons) migram devido a atração eletrostática. Esse processo é minimizado pela adição de um eletrólito suporte (substância iônica) com concentração muito superior à do analito de interesse;

3. Difusão – Devido ao gradiente de concentração estabelecido entre o meio da solução e as proximidades do eletrodo. Esse processo é o mais explorado em voltametria, e é um dos limitantes da corrente do processo redox.

Em voltametria, após a aplicação de cada ponto do potencial de varredura monitora-se a corrente elétrica resultante de reações de oxi-redução desenvolvidas pelo sistema. Devido à variação do potencial aplicado, há o aparecimento de uma corrente capacitiva resultante do rearranjo das cargas nas proximidades dos eletrodos. Essa corrente decresce

exponencialmente com o tempo de modo que após um certo intervalo de tempo a corrente resultante reflete basicamente o processo redox governado por difusão (no caso de se trabalhar com eletrólito suporte e sem agitação).

5.3 - Montagem do Potenciostato Monocanal

O potenciostato utilizado foi montado conforme o circuito elétrico mostrado na Figura 5.2.

Um potencial de excitação é criado via "software" (V_{in}) e através de uma interface paralela PCL-711S ADVANTECH é aplicado no circuito somador (vide Figura 5.2). O conversor analógico/digital da placa de interface PCL-711S ADVANTECH somente permite a geração de potenciais positivos, daí a necessidade de um circuito somador. A tensão de saída (V_{out}) do circuito somador é dada pela Equação 5.1.

$$V_{out} = - (R4 \times V_{in} / R1 + R4 \times V_{offset} / R2) \qquad 5.1$$

No circuito somador construído todos os resistores utilizados são iguais, logo o ganho do mesmo é unitário, e o potencial final é tão somente a soma dos potenciais de entrada e de "offset" com o sinal invertido. O resistor (trim-pot) que controla o potencial de "offset" foi ajustado de modo que tal potencial foi de –2,50V. Desse modo, quando o potencial de entrada (V_{in}) for zero volt o potencial de saída será igual a +2,5 V, e quando o potencial de entrada for 5V, o potencial de saída será -2,5 V.

Os circuitos integrados (CI) 1, 2, 3, 4 e 5 são amplificadores operacionais. Após o circuito somador o sinal é aplicado ao eletrodo auxiliar através de um amplificador operacional (CI3) pela entrada inversora, desse modo o potencial é novamente invertido.

Assim, ao aplicar um potencial de 0V na entrada do circuito o mesmo converte para -2,5 V entre os eletrodos, e um potencial de entrada de 5 V equivale a 2,5 V entre os eletrodos.

O circuito integrado CI6 é um conjunto de quatro chaves analógicas que permitem a permutação dos resistores ligados entre a saída e a entrada inversora do amplificador operacional CI4. O chaveamento é feito através de um nível lógico (TTL) baixo em um (e somente um) dos pinos 1, 8, 9 e 16 do circuito integrado CI6. Assim, quando o pino 1 tiver um nível lógico baixo, os pinos 2 e 3 entram em curto-circuito e ligam o resistor R12 aos terminais do amplificador operacional CI4. Analogamente, um nível baixo no pino 8, 9, ou 16 provoca a ligação elétrica do resistor R13, R14, ou R15 aos terminais do amplificador operacional CI4 respectivamente. Desse modo ajusta-se o ganho nesse conversor corrente - voltagem. Além disso esse conversor inverte o potencial gerado (a realimentação de potencial é feita pela entrada inversora). Assim os possíveis ganhos neste circuito são: $-4,7x10^2$; $-4,7x10^3$; $-4,7x10^4$ e $-4,7x10^5$ vezes.

A constante de tempo de amostragem do sinal é dada pelo produto da resistência pela capacitância dos elementos chaveados ao conversor corrente-voltagem. Os componentes foram dimensionados para terem uma constante temporal de aproximadamente $4,7\ x10^{-4}$ s.

Os níveis lógicos que controlam o ganho do potenciostato são gerados pelos bits 9, 10, 11 e 12 da saída digital da placa de interface Advantech [63], ou seja são gerados pelo programa que gerencia o potenciostato.

O circuito integrado CI5 é apenas um "buffer" inversor, de modo a inverter e manter o valor de tensão proveniente do conversor corrente – voltagem.

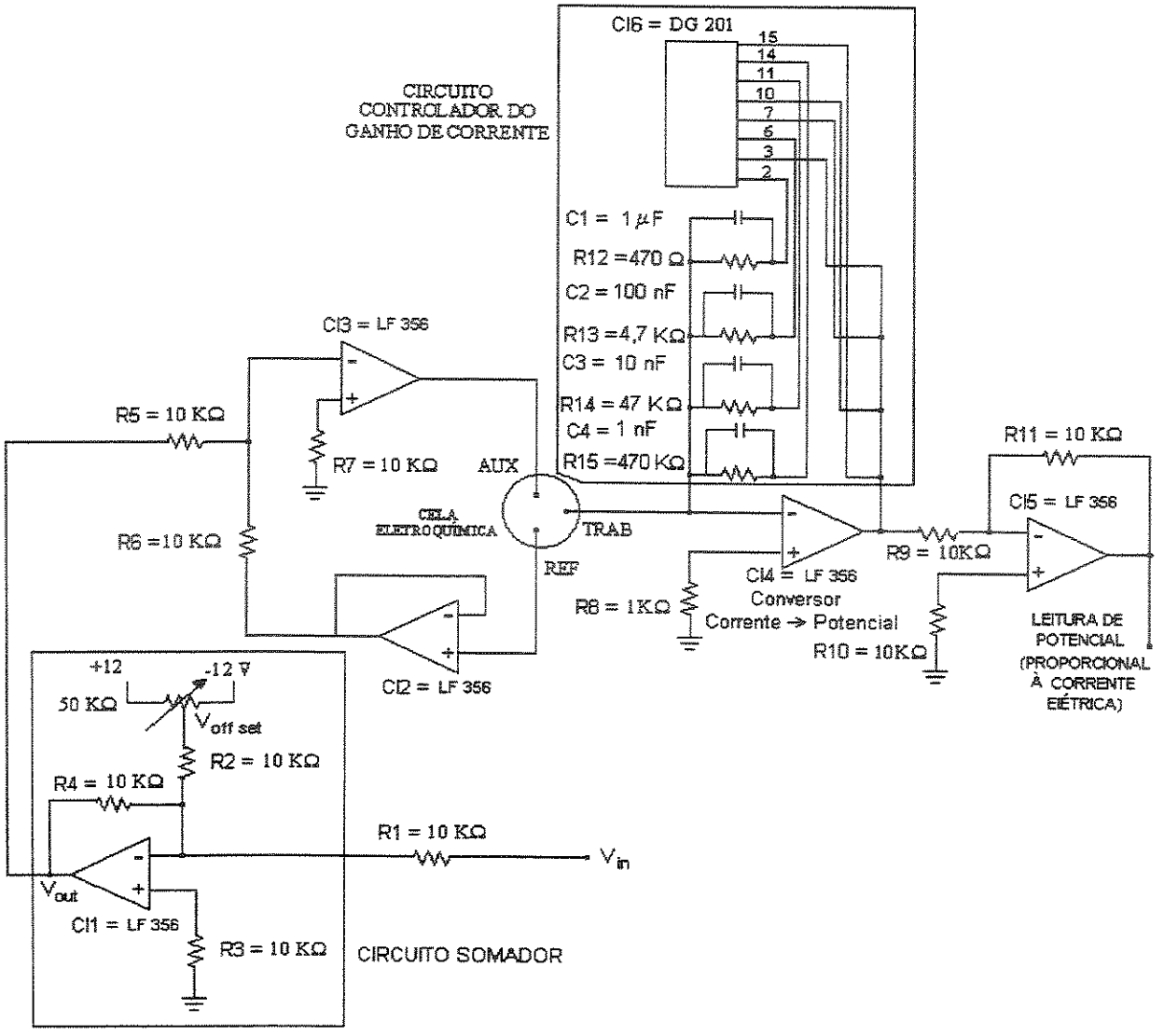


Figura 5.2 - Circuito elétrico do potenciostato construído.

Para operar o potenciostato, o programa controlador aplica um potencial entre 0 e 5 V com resolução de 12 bits (dada pela interface D/A da placa Advantech) na entrada do circuito somador, o qual transforma essa faixa de potencial para 2,50 e –2,50 volts (respectivamente) e aplica essa faixa de potencial entre o eletrodo de referência e o eletrodo de trabalho.

Esse potencial promove reações de oxi-redução na cela eletroquímica, e a corrente elétrica gerada é convertida em potencial e amplificada. A detecção da corrente elétrica gerada é feita utilizando um canal de leitura (entrada analógica) da interface A/D da placa Advantech.

Após a construção do potenciostato faz-se necessário uma etapa de calibração do mesmo, para se obter uma equação que relacione a corrente monitorada com o potencial gerado na conversão (em unidades de A / D). Essa etapa foi feita distintamente para cada ganho (posição da chave analógica). A Figura 5.3 mostra um diagrama do esquema usado para a calibração do potenciostato.

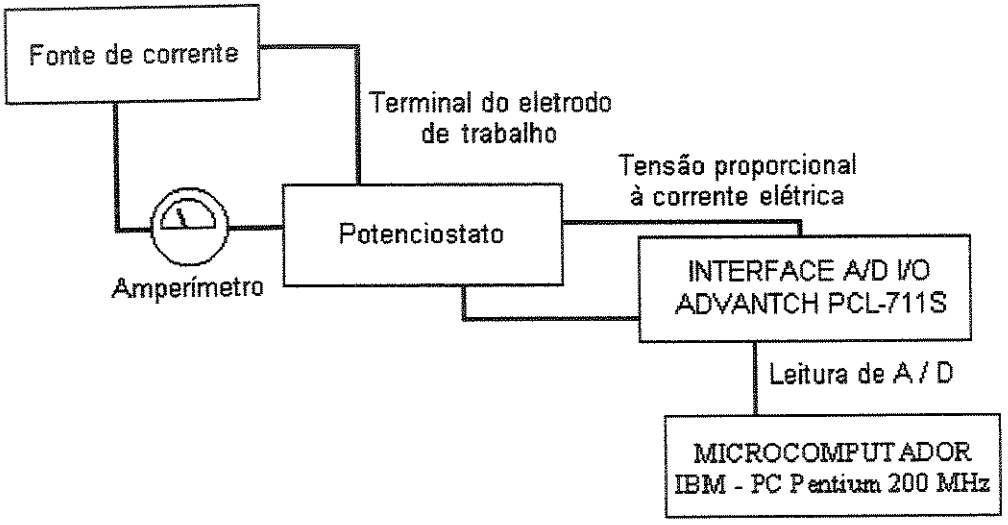


Figura 5.3 – Esquema das conexões do potenciostato com o computador para calibração das correntes elétricas monitoradas.

A Tabela 5.1 mostra as equações lineares obtidas por regressão linear, aos dados da calibração do potenciostato (entre a corrente aplicada e o valor de tensão lido em unidades de A / D), utilizando o método dos mínimos quadrados.

Tabela 5.1 – Resultados da calibração do potenciostato.

Ganho do Potenciostato	Equação linear: I / mA	Coeficiente de correlação linear
$4,7x10^2$	$-8,8482+4,3221x10^{-3}$ x unid de A/D	0,9999468
$4,7x10^3$	$-1,0246+5,0045x10^{-4}$ x unid de A/D	0,9999994
$4,7x10^4$	$-0,10248+5,0215x10^{-5}$ x unid de A/D	0,9999945
$4,7x10^5$	$-0,010257+5,013x10^{-6}$ x unid de A/D	0,9999933

Para cada ganho aplicou-se uma série de correntes (positivas e negativas) com intensidade adequada de modo a não saturar o A/D na conversão analógico / digital e ao mesmo tempo ter uma sensibilidade razoável na faixa em análise. Desse modo obteve-se as seguintes faixas:

Para o ganho de $4,7x10^5$, a faixa monitorada foi de -0,010 a 0,010 mA;

Para o ganho de $4,7x10^4$, as faixas monitoradas foram de -0,100 a -0.010 e de 0,010 a 0,100 mA;

Para o ganho de $4,7x10^3$, as faixas monitoradas foram de - 0,800 a - 0,100 mA e de 0,100 a 0,800 mA;

Para o ganho de $4,7x10^2$, as faixas monitoradas foram de - 8,00 a - 0,800 mA e de 0,800 a 8,00 mA;

5.4 - Desenvolvimento do programa para o controle do potenciostato

O programa para o controle do potenciostato foi desenvolvido em linguagem Visual Basic 3.0 da Microsoft® [64] para o ambiente Windows. Optou-se por utilizar essa linguagem devido à sua maior interatividade com o usuário, uma vez que trabalhando-se no ambiente gráfico do Windows existe a possibilidade da utilização de janelas, barras de menus, textos em diferentes fontes e cores assim como gráficos mais bem elaborados, além da possibilidade de importar dados, gráficos e figuras criados por outros programas para o Windows. A linguagem Visual Basic 3.0 é orientada por eventos (tais como o acionamento dos botões do mouse, entre outros) o que facilita a tomada de decisões por parte do usuário durante a execução do programa e, desse modo, a opção entre sub-rotinas dentro do mesmo.

A principal dificuldade a ser enfrentada na programação com o Visual Basic é que não existem funções para o controle de baixo nível, como acesso a portas de comunicação e endereços de memória. Essa dificuldade foi contornada utilizando-se bibliotecas de acesso dinâmico (dinamic link library - DLL) do repositório de programas para Windows da Universidade de Indiana - Estados Unidos (ftp.cica.indiana.edu), acessado via Internet, denominada INPOUT.DLL. Esta biblioteca tem as funções INP (para ler dados em um endereço de memória) e OUT (para enviar dados a um endereço de memória). Também foi necessário o desenvolvimento de uma DLL em linguagem Visual C++ 2.0 [65] para desabilitar dispositivos periféricos (que é utilizada em casos onde se requer precisão temporal). Essa rotina em Visual C++ é necessária porque o ambiente Windows por ser multi-tarefa fica susceptível à interrupções produzidas pelo mouse, teclado e demais

periféricos. Durante uma estrutura de repetição do tipo FOR - NEXT essas interrupções não são consideradas, porém o computador fica atualizando a posição do cursor na tela, e isso causa uma perda do sincronismo temporal.

Na tela de abertura do programa gerenciador do potenciostato desenvolvido, intitulado "FenixII Mananger", encontra-se uma barra de menus, com as seguintes opções: "Exit", "I /O Calibration" e "Help". A Figura 5.4 mostra a tela de abertura do programa que controla o potenciostato.

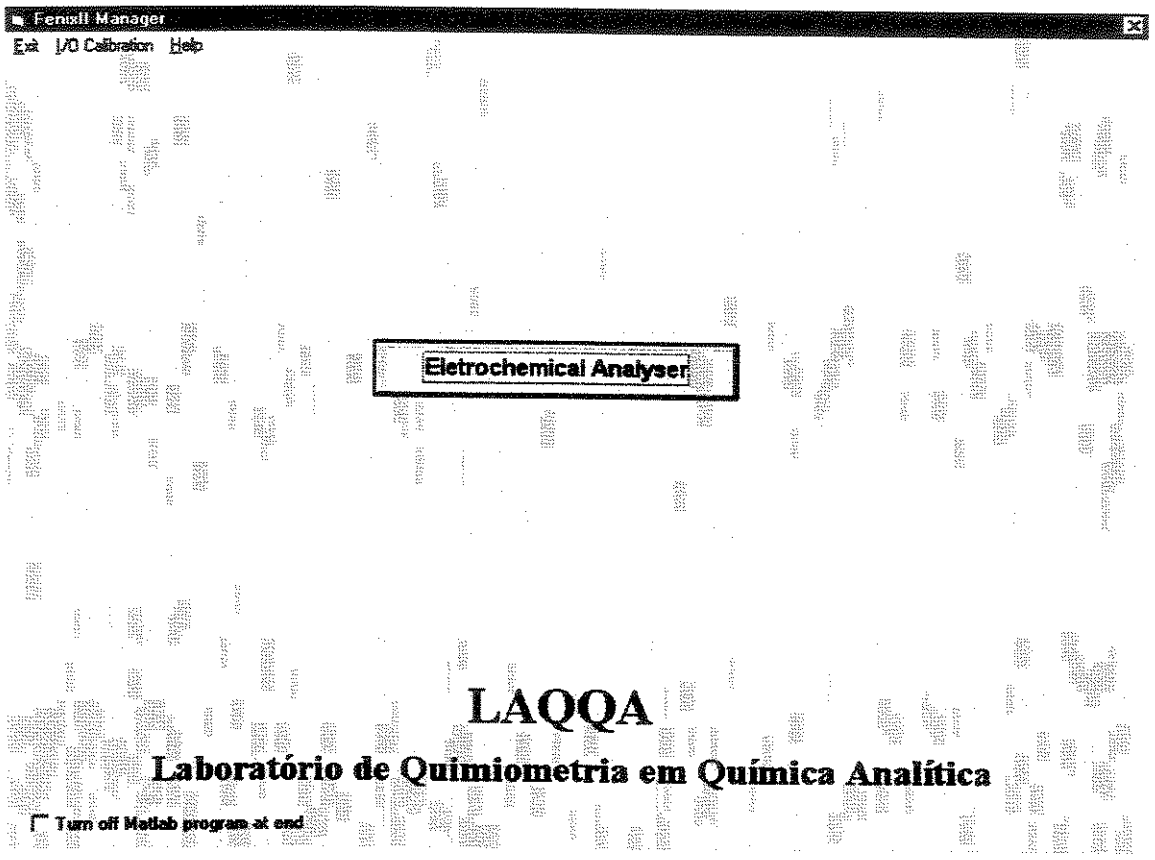


Figura 5.4 - Tela de abertura do programa "FenixII Mananger", que controla o potenciostato.

5.4.1- Saída e Ajuda

A opção "Exit" acessa uma sub-rotina que desliga o programa, e a opção "Help" mostra as funções básicas do equipamento.

5.4.2 – Calibração de sinais de entrada e saída

Essa opção de menu mostra uma janela para testar e calibrar a comunicação de dados entre a interface Advantech e o equipamento construído. A Figura 5.5 mostra a tela dessa janela.

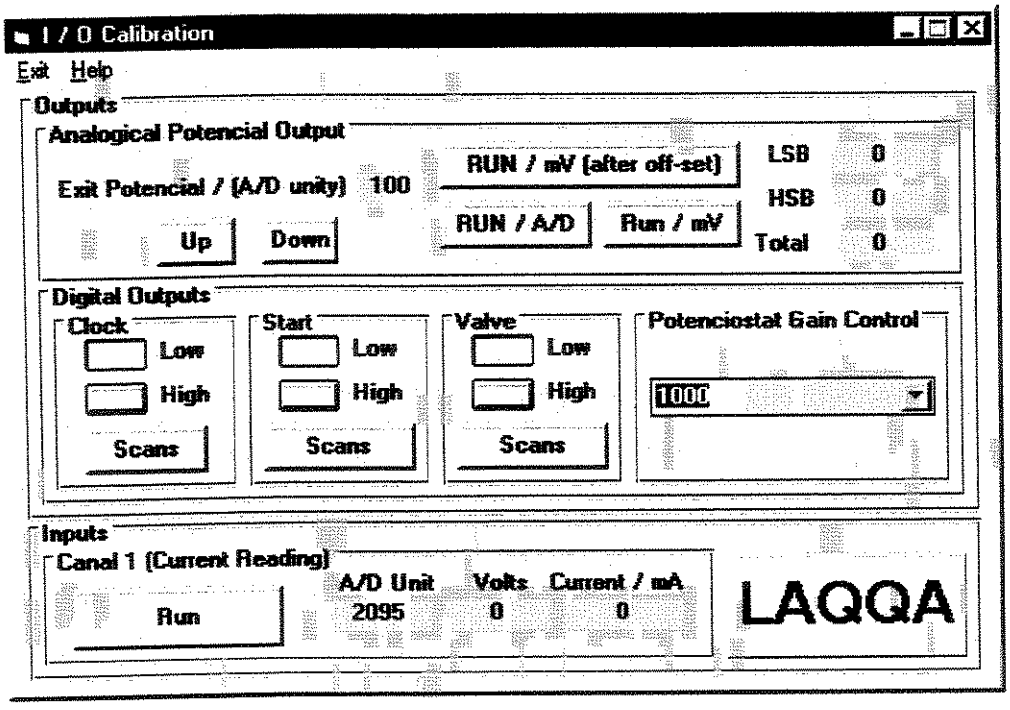


Figura 5.5 - Janela da rotina "I / O Calibration".

Para as opções de saída, nessa janela existem comandos para verificar a saída de potencial pelo conversor digital / Analógico (D/A) em mV e em unidades de A/D, e comandos para alternar os níveis lógicos uma digitais, denominada "valve" que foi utilizada para permutar uma válvula solenóide. Existem também comandos que gerenciam o ganho do potenciostato, através do controle de quatro "bits" ligados à chave analógica DG 201 no circuito do potenciostato (ver Figura 5.2).

Para as opções de entrada de potencial, há comandos para gerenciar o canal de entrada do conversor analógico/digital da interface, utilizado pelo potenciostato. Nessa entrada analógica simplesmente monitora-se o potencial da mesma. Essa rotina é usada na calibração de correntes do potenciostato, para monitorar o potencial gerado pelo conversor corrente/voltagem.

5.4.3 – Gerenciamento do potenciostato

O potenciostato construído apresenta opções para a execução de medidas de voltametria, onde se aplica um potencial e monitora-se a corrente. O gerenciador do potenciostato possui rotinas que permitem ao usuário estabelecer a faixa de potencial de trabalho, o número de varreduras, o intervalo de tempo entre cada potencial aplicado, o ganho do potenciostato, a utilização de filtros digitais e correção de linha de base. Após estabelecer as condições de trabalho, o usuário deve executar o comando "RUN". O gerenciador do potenciostato irá então executar a voltametria e retornará a velocidade de varredura.

O gerenciador do potenciostato permite a utilização de quatro tipos de potenciais de excitação: varredura linear, onda quadrada, pulso diferencial e voltametria cíclica. O controle de ganho permite monitorar correntes elétricas entre 10 μA e 8 mA.

A Figura 5.6 mostra a janela aberta para o gerenciador do potenciostato.

Além das rotinas de aquisição e controle de dados, o gerenciador do potenciostato permite também a manipulação de arquivos, com funções para armazenar os dados coletados em arquivos, carregar arquivos anteriormente armazenados e imprimir dados.

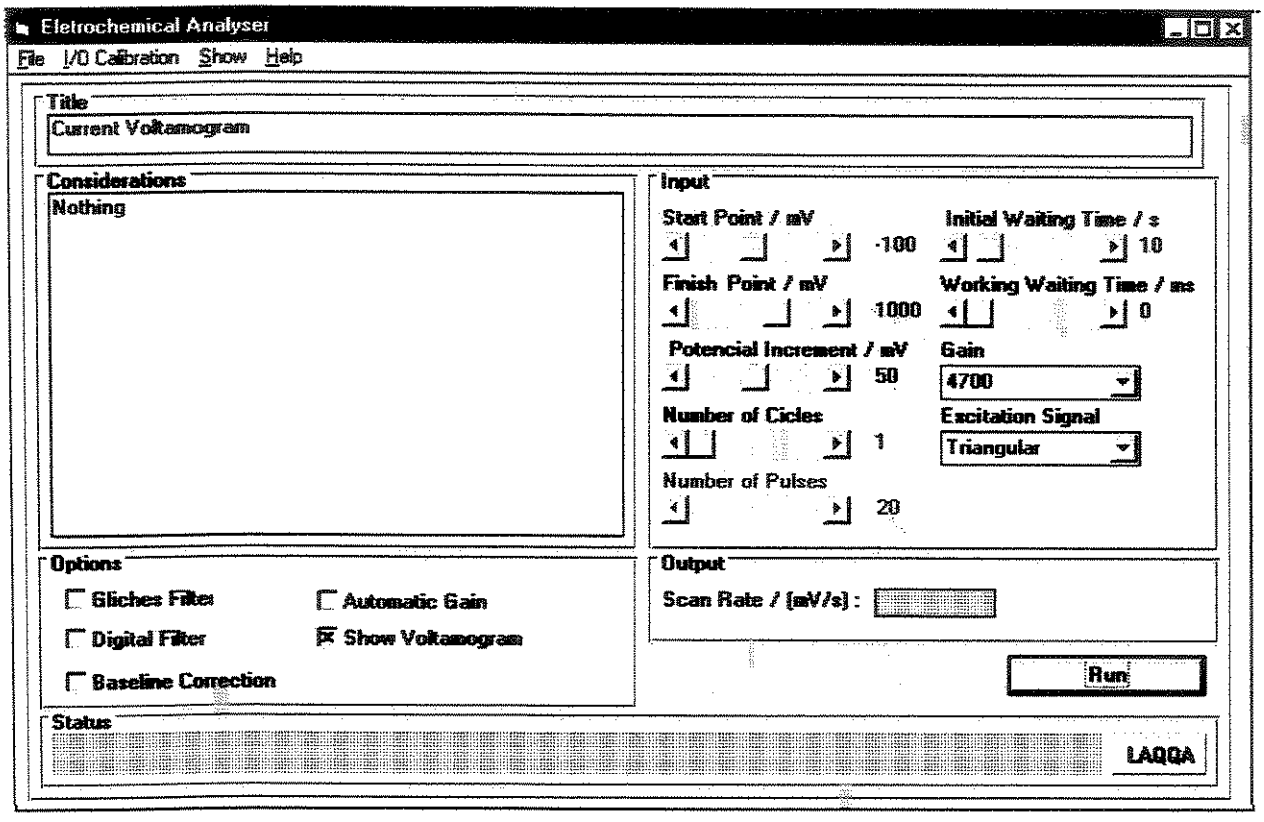


Figura 5.6 - Janela da rotina que controla o potenciostato.

5.4.4 - Utilização da Troca dinâmica de dados (DDE) do Windows para realizar a conversação entre o Visual Basic e o Matlab

Para a visualização e impressão dos espectros e voltamogramas foi utilizada uma das propriedades existentes no ambiente Windows, que é a possibilidade de realizar a chamada "troca dinâmica de dados" (Dynamic Data Exchange - DDE) [66] entre o Visual Basic e o Matlab.

O ambiente Windows é caracterizado por ser multi-tarefa, ou seja, vários aplicativos podem ser executados simultaneamente. A "troca dinâmica de dados" (DDE) é um mecanismo existente no sistema operacional Windows, que torna possível que duas aplicações que estejam sendo executadas simultaneamente "conversem" uma com a outra continuamente, trocando dados e executando funções. A DDE automatiza operações manuais, por isso, sua utilização é muito interessante em casos onde possam ser utilizados recursos disponíveis num aplicativo que não se encontram em outro. Desta forma, desenvolveu-se o programa para o controle do potenciostato utilizando-se o Visual Basic, mas para a apresentação dos gráficos, impressão e realização de cálculos avançados, foi utilizado o Matlab que possui muitos recursos gráficos e de cálculo que dificilmente poderiam ser implementados pelo Visual Basic.

O Matlab (Matrix Laboratory) [18] é um ambiente computacional de alta performance em computação numérica e visualização de dados, onde pode-se fazer cálculos matemáticos com matrizes de qualquer grandeza (em princípio, limitado apenas pela memória do computador), de modo que o mesmo é facilmente utilizável para cálculos quimiométricos. O Matlab integra análise numérica, computação matricial, processamento de

sinais e visualização gráfica em um ambiente que possui comandos similares aos utilizados em álgebra linear, inclusive no que se refere à notação dos comandos.

O sistema de representação gráfica do Matlab possui uma variedade de sofisticadas técnicas de apresentação e visualização de dados, inclusive objetos gráficos, tais como, gráficos de linhas e de superfície, cuja aparência pode ser controlada ajustando os valores das propriedades desses objetos.

O Matlab apresenta também a capacidade de executar programas, desenvolvidos em linguagem própria, que permitem a utilização seqüencial de qualquer função do Matlab (funções internas), ou de outras subrotinas geradas anteriormente. A versão 4.2 do Matlab (ou superior) também possui a capacidade de realizar a troca dinâmica de dados (DDE), podendo assim comunicar-se internamente com outros aplicativos que possuam essa habilidade. A visualização de voltamogramas, por exemplo, é na verdade a execução de um sub-programa feito em linguagem própria do Matlab

No programa desenvolvido para o controle do potenciostato, o Visual Basic carrega (executa) o Matlab, e este fica em segundo plano, ou seja, minimizado pelo Windows. A aquisição de dados, como já mostrado, é realizada por rotinas desenvolvidas para o Visual Basic. Após esta aquisição, os dados são gravados em um RAM-drive, e carregados automaticamente para o Matlab, que gera o gráfico para a apresentação do voltamograma. A figura correspondente ao gráfico criado pelo Matlab é então transferida para o Visual Basic (via DDE), onde uma nova janela é aberta para mostrá-la.

A Figura 5.7 mostra o gráfico de um voltamograma de uma solução 0,001 mol L^{-1} de ferrocianeto de potássio adquirido pelo gerenciador do potenciostato e visualizado através de uma DDE entre o Matlab e o Visual Basic. Foram utilizados eletrodos de trabalho e

auxiliar confeccionados em platina, e o potencial de excitação foi estabelecido em relação a um eletrodo de referência de Ag / AgCl, utilizando como eletrólito suporte uma solução 0,1 mol L^{-1} de $NaNO_2$.

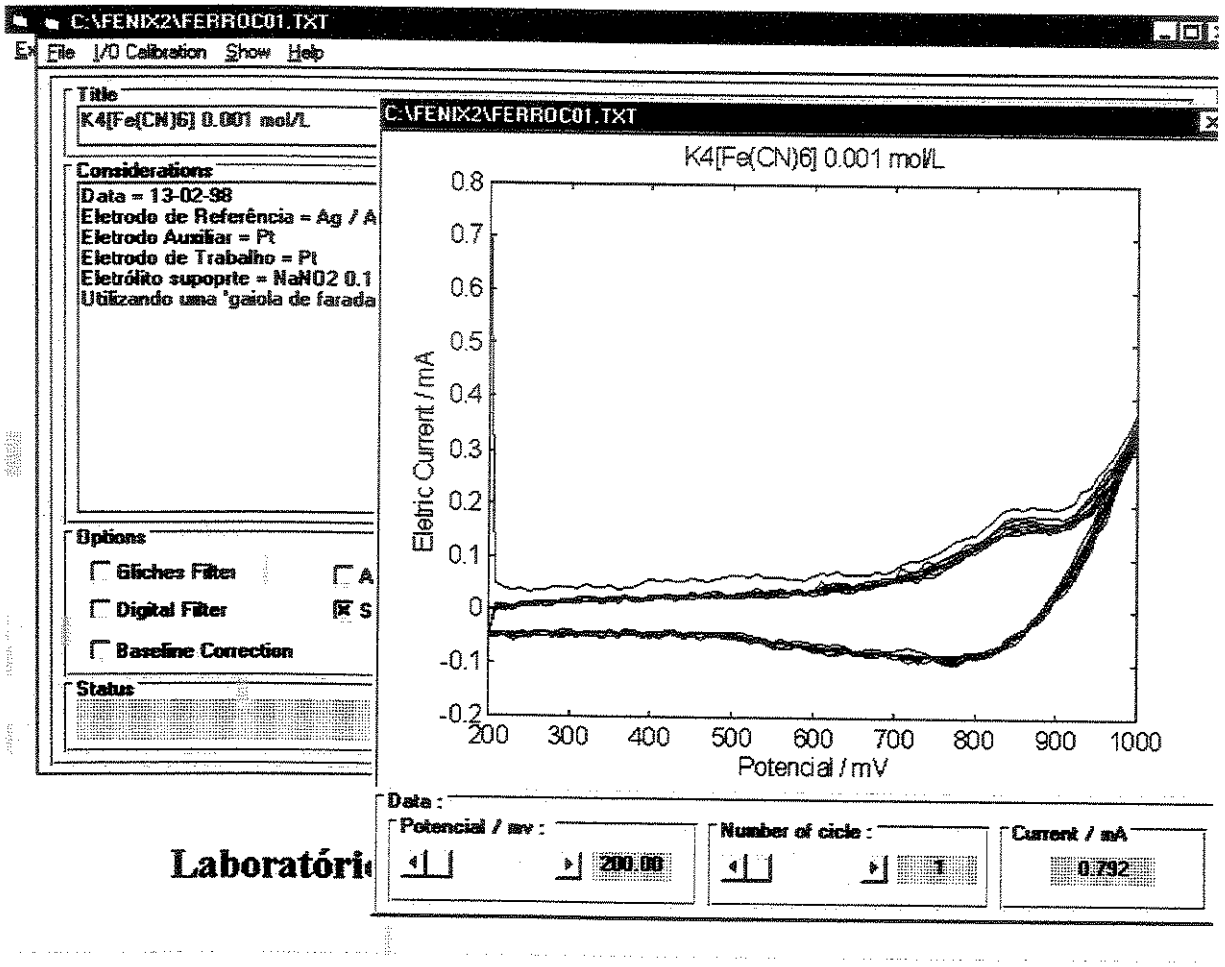


Figura 5.7 - Voltamograma de uma solução 10^{-3} mol/L de ferrocianeto de potássio exibida através de uma DDE entre o Matlab e o Visual Basic.

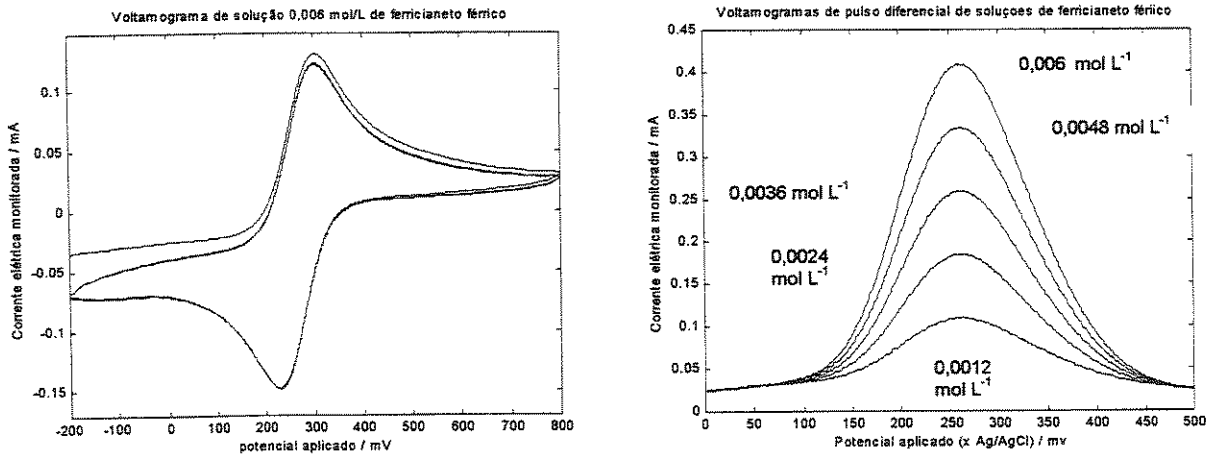
A utilização de DDE também é realizada para a execução de filtros digitais mais elaborados, como por exemplo a utilização de filtro digital com transformada de Fourrier, pois o Matlab já possui sub-rotinas prontas para a execução desse tipo de cálculo. Nesse caso, o gerenciador do equipamento coloca os dados no RAM-drive e o programa do Matlab

carrega esses dados e processa o filtro. Novamente o programa do Matlab salva os dados filtrados no RAM-drive e o programa gerenciador do equipamento carrega esses dados filtrados para depois serem devidamente processados (exibidos ou armazenados).

5.5 - Avaliação do potenciostato construído

O potenciostato construído foi avaliado através da medida de voltametria de soluções de ferricianeto férrico. Essa espécie química possui um comportamento eletroquímico bem estabelecido [61 - 62], e por comparação com a literatura, pode-se avaliar a performance do equipamento em questão.

A Figura 5.8 mostra os voltamogramas obtidos para o ferricianeto férrico.



A B

Figura 5.8. Voltamogramas de solução de ferricianeto férrico com eletrodos de trabalho e auxiliar de platina e eletrodo de referência Ag/AgCl, utilizando como eletrólito suporte uma solução de KNO_3 1,0 mol L^{-1} em velocidade de varredura de 50 mV / s. A – Varredura cíclica; B – Varredura de pulso diferencial em vários níveis distintos de concentração (com 50 mV de amplitude de pulso).

Obteve-se uma diferença de potencial de pico (anódico – catódico) de 73 mV em média, e a razão entre as correntes de pico (catódica / anódica) de 0,95.

Idealmente, em um processo eletroquímico reversível, a diferença entre o potencial de pico para um sistema eletroquímico de uma reação redox envolvendo um elétron deve ser aproximadamente igual a 59 mV (RT/mF), e a razão entre as correntes de pico deve ser unitária ($I_{anódica} = I_{catódica}$). Segundo dados da literatura [61,62], admite-se até um valor de 100 mV para a diferença entre os potenciais de pico (em processos redox que envolvem a transferência de apenas um elétron) e uma razão entre as correntes de pico entre 0,8 e 1,2.

A dependência entre a concentração da espécie eletroativa e a corrente elétrica (de pico) foi avaliada através dos voltamogramas, com pulso diferencial, de uma série de soluções de ferricianeto férrico entre $1,2 \times 10^{-3}$ e $6,0 \times 10^{-3}$ mol L^{-1}.

Observou-se um comportamento linear entre a corrente de pico e a concentração da espécie:

$$\text{Concentração / mMol } L^{-1} = 0,632 + 13,5 \times \text{Corrente de pico / } \mu A \; ; \; R = 0,997 \qquad 5.2$$

A quantidade de material que sofre oxi-redução em um sistema eletroquímico é definida pela carga elétrica que passa pelo circuito da cela eletroquímica em um intervalo de tempo definido. Para um intervalo de tempo definido (qualquer) a corrente elétrica é proporcional à carga. Logo, espera-se que a corrente elétrica monitorada pelo sistema eletroquímico seja proporcional à concentração da espécie química em questão (no caso ideal).

Pela análise do coeficiente de correlação linear (R), do ajuste entre a corrente de pico e a concentração do reagente (mostrado na Equação 5.2) pôde-se observar uma dependência linear entre eles. A corrente de pico pode ser utilizada como parâmetro para

estimar a dependência entre a corrente elétrica monitorada em um potencial qualquer e a concentração do reagente.

Capítulo 6

Adição de padrão multivariada para a determinação simultânea de ácido úrico e ácido ascórbico por voltametria de pulso diferencial

No preparo dos padrões para o desenvolvimento do modelo de calibração para mais de um analito, geralmente prepara-se uma série de amostras com concentrações variáveis (independentemente) para todas as espécies utilizadas de modo a seguir um planejamento fatorial [67, 68]. Isso demanda muito trabalho experimental além de gerar uma grande quantidade de resíduos. Adicionalmente, existem alguns casos onde não se sabe a faixa de possíveis concentrações das amostras reais. Para solucionar esses problemas, elaborou-se um planejamento experimental baseado em adição de padrão simultânea para calibração multivariada com duas espécies químicas.

6.1 - Modelo de adição de padrão para análise simultânea multivariada.

A idéia básica do modelo de adição de padrão é minimizar o efeito de matriz. Isso pode ser realizado pela medida da amostra e pela medida do sistema com adições de quantidades conhecidas das espécies químicas de interesse sobre essa amostra. Por diferença das medidas experimentais, pode-se determinar a resposta experimental do efeito de cada adição. O problema é fazer isso simultaneamente para mais de uma espécie química, visto que a adição consecutiva de uma das espécies químicas não permite a livre adição dos diversos níveis das demais espécies químicas.

Normalmente fixa-se o nível (concentração) de uma das espécies químicas e adicionam-se quantidades conhecidas de uma segunda espécie química nos diversos níveis do planejamento. Ao se adicionar mais de uma quantidade conhecida da primeira espécie química para elevar sua concentração ao segundo nível do planejamento fica impossível utilizar os mesmos níveis anteriores para a segunda espécie química, visto que, a mesma já

está em seu nível maior (pelas adições de quantidades conhecidas realizadas anteriormente).

Para contornar o problema apresentado no parágrafo anterior, pode-se realizar um planejamento em que se divide a amostra em um número de alíquotas igual ao número de níveis da 1ª espécie. Para cada alíquota, fixa-se a concentração da primeira espécie química pela adição de uma quantidade conhecida da mesma. Em seguida adicionam-se quantidades conhecidas da segunda espécie para fazê-la variar conforme os níveis propostos no planejamento experimental. Caso o primeiro nível da primeira espécie química seja zero (espécie ausente), procede-se apenas a adição da segunda espécie para esse nível da primeira espécie. O processo é repetido para a segunda alíquota, sendo que nela a quantidade adicionada da primeira espécie química deve ser necessária para fixar sua concentração no segundo nível do planejamento (para essa espécie). O processo se repete para as demais alíquotas até atingir todos os níveis da primeira espécie.

No caso de medidas eletroquímicas, antes da análise de cada alíquota da amostra, deve-se realizar a medida de referência, ou seja, o voltamograma do eletrólito suporte puro. Esse voltamograma é realizado apenas para verificar a reprodutibilidade do sistema eletroquímico a cada nova alíquota (nova solução) no que diz respeito à remoção de oxigênio e condicionamento de temperatura.

Por exemplo, no caso em questão, em que se pretende quantificar uma mistura de ácido úrico e ácido ascórbico, por voltametria de pulso diferencial, utilizando adição de padrão multivariada, utilizou-se 6 níveis distintos para cada uma das duas espécies (além do nível nulo), totalizando-se 36 amostras com diferentes concentrações das duas espécies que serão quantificadas. Desse modo, segundo o planejamento proposto, dividiu-se a

amostra (onde será realizada a adição de padrão) em 6 alíquotas, e em cada uma delas adicionou-se uma solução padrão da primeira espécie que se deseja determinar (ácido úrico) de modo a atingir seu respectivo primeiro nível no planejamento. Sobre essa solução, foi adicionada a segunda espécie em diferentes níveis, de modo a obter seis pontos do planejamento, sendo que a cada adição da segunda espécie foi tomado um voltamograma. Obteve-se assim a 1ª parte (1/6) do planejamento, tendo o primeiro analito em seu 1º nível e todos os níveis do segundo analito simultaneamente. O processo foi repetido executando-se cada nível do 1º analito até obter os seis níveis.

A Figura 6.1 mostra a ordem de execução do primeiro 1/6 do planejamento proposto para um sistema com duas espécies a serem determinadas por adição de padrão com seis níveis em cada uma delas. Nesse caso, a primeira espécie que se deseja determinar apresenta nível fixo zero (não se adiciona essa espécie) e a segunda espécie que se deseja determinar é variada nos seis níveis do planejamento (além do nível 0).

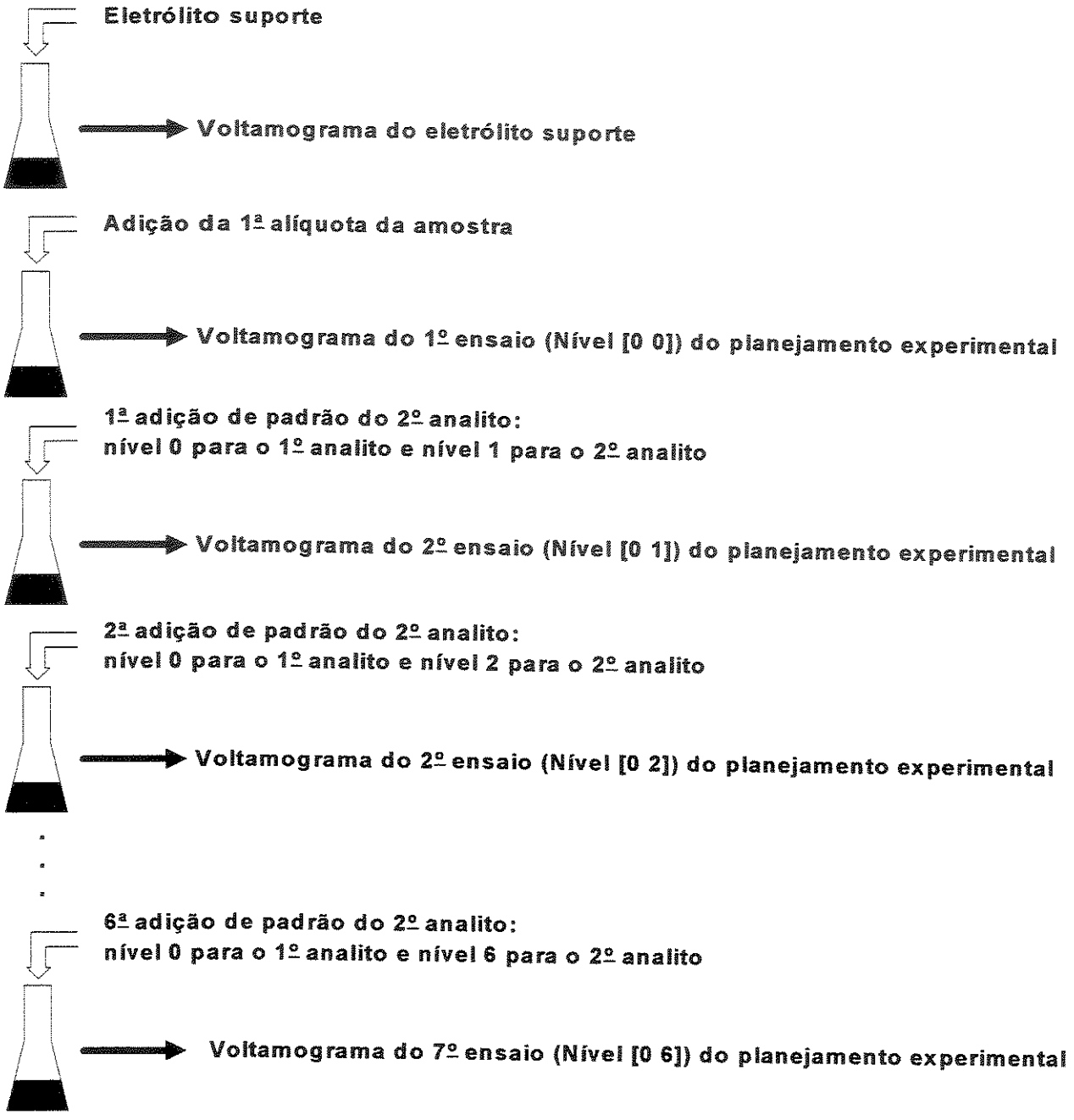


Figura 6.1 – Seqüência do planejamento experimental para a 1ª alíquota da amostra (1/6 do planejamento).

A Figura 6.2 mostra a ordem de execução da segunda parte do planejamento, com a segunda alíquota da amostra. Nesse caso, o primeiro analito apresenta o primeiro nível fixo

(apenas uma adição de padrão) e o segundo analito é variado nos seis níveis do planejamento (além do nível 0).

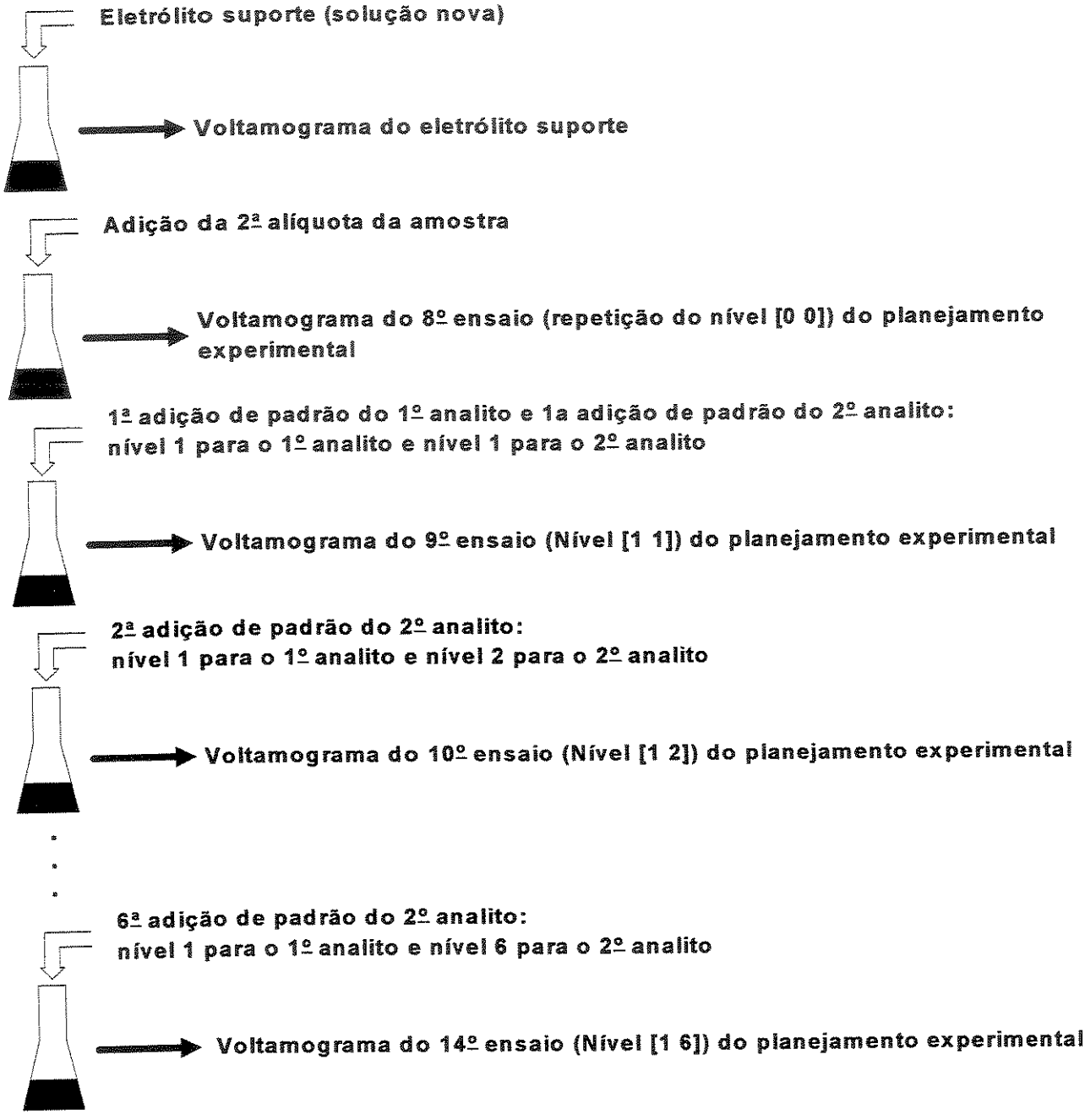


Figura 6.2 – Seqüência do planejamento experimental para a 2ª alíquota da amostra (1/6 do planejamento).

O procedimento experimental continua até atingir o nível [6 6] do planejamento. É interessante observar que a solução padrão da primeira espécie é adicionada apenas cinco

vezes (o primeiro nível é nulo) e a segunda espécie é adicionada seis vezes. Desse modo, o planejamento experimental apresenta ao todo seis níveis distintos para cada espécie química. O voltamograma da amostra pura (nível [0 0] do planejamento) é tomado seis vezes, uma vez por alíquota de amostra, para ter uma referência sempre atualizada.

Posteriormente as respostas do sistema em questão são modeladas com um modelo linear de calibração multivariada e a medida da amostra pura é determinada pela extrapolação do modelo.

6.1 - O Modelo de calibração multivariada por adição de padrão

Em um modelo univariado de adição de padrão ajusta-se uma reta, pelo método dos mínimos quadrados, entre os valores de concentração produzidos pelas adições do padrão e suas respectivas medidas analíticas (sinal analítico). Então o valor da amostra é tomado pelo cruzamento da extrapolação da reta com o eixo das abscissas.

No modelo multivariado de adição de padrão, ajusta-se a um modelo linear a variação das medidas experimentais (voltamogramas) contra a variação da concentração ocasionada pala adição do padrão. Assim, para o modelo multivariado tem-se:

$$\Delta \mathbf{y} = \Delta \mathbf{X}.\ \mathbf{B}^t \qquad 6.1$$

onde $\Delta \mathbf{X}$ é a matriz formada pelos voltamogramas das adições de padrão, devidamente subtraídos do voltamograma da amostra, $\Delta \mathbf{y}$ é o vetor formado pelo incremento de concentração ocasionado pela adição do padrão, e **B** é a matriz dos coeficientes de regressão obtida pelo modelo linear de calibração multivariada.

Para a etapa de previsão tem-se:

$$\mathbf{y}_{prev} = \mathbf{X}_{amostra}.\mathbf{B}^{t} \qquad 6.2$$

onde **B** é o mesmo obtido na etapa de calibração.

O mérito desse modelo é que tomando a variação dos dados (**ΔX**) na parte de calibração, elimina-se o efeito de matriz e o modelo de calibração multivariada ajusta apenas a variação dos dados que apresenta dependência com as respostas (**y**), ou seja a matriz do sistema não é modelada (por construção). Quando executa-se a previsão, a matriz dos coeficientes de regressão (**B**) elimina o efeito de matriz dos dados da amostra, visto que o voltamograma da amostra (com efeito de matriz) é projetado em um espaço vetorial construído pela modelagem de dados sem efeito de matriz (**B**).

Para a calibração multivariada pode-se utilizar vários métodos lineares distintos (PLS, PCR, MLR, etc.), de modo que optou-se pelo modelo de regressão contínua de potências (CPR - continuum power regression) [69 - 72], visto que o mesmo otimiza o modelo linear que melhor se adapta aos dados em questão, discernindo entre otimizar a variância (PCR), a correlação (MLR), a covariância (PLS) ou qualquer combinação intermediária entre esses modelos.

6.2 - Calibração multivariada com o modelo de continuum power regression (CPR)

Seja uma matriz de dados **X** (n x p) com n amostras e p variáveis, de posto r ≤ min(n,p) e uma matriz de respostas **Y** (n x m) com m respostas para cada uma das n amostras. No caso em questão **X** é a matriz obtida pala digitalização dos voltamogramas das adições de padrão, subtraindo o voltamograma da amostra e **Y** a matriz com a concentração dos respectivos analitos adicionados. Geralmente ambos **X** e **Y** são centrados na média

Tomando a decomposição em valores singulares (singular value decomposition – SVD) [54, 73] de **X** temos:

$$\mathbf{X} = \mathbf{U}\,\mathbf{S}^{1/2}\,\mathbf{V}^t \qquad 6.3$$

Sendo **U** (n x r) contendo os vetores singulares à esquerda e **V** (p x r) os vetores singulares à direita. Os conjuntos de vetores singulares são ortonormais entre si. Assim $\mathbf{U}^t\mathbf{U} = \mathbf{V}^t\mathbf{V} = \mathbf{I}$ = matriz identidade de ordem r.

$\mathbf{S}^{1/2}$ é uma matriz diagonal com r valores singulares não nulos $\lambda_a^{1/2}$ (a = 1,2,...,r) em ordem decrescente na diagonal. As dimensões associadas com valores singulares nulos são desconsideradas (truncadas).

As colunas de **U** correspondem aos escores normalizados dos componentes principais de **X**, enquanto que as colunas de **V** correspondem aos seus loadings . Desse modo temos:

$\mathbf{X}^t\mathbf{X} = \mathbf{V}\mathbf{S}\mathbf{V}^t$ e $\mathbf{X}\mathbf{X}^t = \mathbf{U}\mathbf{S}\mathbf{U}^t$,onde **S** (r x r) é uma matriz diagonal de autovalores (quadrado dos valores singulares).

Segundo Stone e Brooks [68], um parâmetro α é indicado para selecionar a posição no caminho a partir da regressão linear múltipla (multiple linear regression - MLR) com $\alpha = 0$ até o PCR ($\alpha = 1$), o qual pode passar pelo PLS ($\alpha = ½$). Uma parametrização alternativa envolve um parâmetro de potência (γ) e pode ser feita com $\gamma = \alpha/(1-\alpha)$. Desse modo quando $\gamma = 0$ tem-se a MLR, quando $\gamma = 1$ tem-se o PLS e quando $\gamma = \infty$ tem-se o PCR. Aplica-se então o algoritmo padrão do PLS [9 - 13] entre $\mathbf{X}^\gamma$ e **Y** para obter uma matriz de regressão **B**, de modo que $\mathbf{Y} = \mathbf{X}^\gamma\mathbf{B}^t$; sendo que $\mathbf{X}^\gamma = \mathbf{U}\,\mathbf{S}^{\gamma/2}\,\mathbf{V}^t$.

A Figura 6.3 ilustra a ponderação contínua feita pelo parâmetro de potência (γ) da regressão contínua, entre os modelos lineares. Variando o parâmetro de potência em intervalos tão pequenos quanto se queira (continuamente), a regressão contínua gera modelos "híbridos" entre os modelos lineares clássicos (MLR, PLS e PCR).

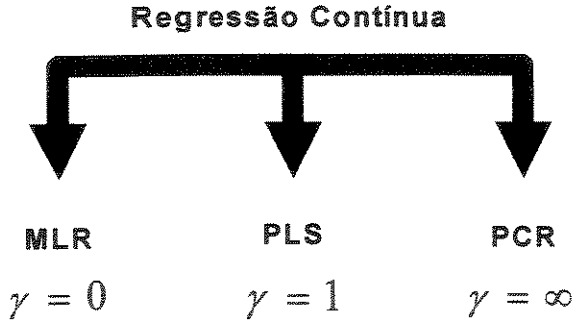


Figura 6.3 – Relação entre a regressão contínua e os modelos lineares de calibração multivariada.

Pode-se mostrar que a regressão contínua converge numericamente à solução do PCR para valores altos de γ, e para a MLR para $\gamma = 0$.

O algoritmo da regressão contínua pode ser entendido como a expansão do espaço das variáveis independentes (**X**) pelo parâmetro de potência.

Quando o valor de γ tende a infinito (ou um número grande), a matriz resultante $\mathbf{X}^{\gamma}$ é expandida em relação a **X** no sentido de explicar a variância de **X**, dessa forma as variáveis que possuem maior influência na determinação dos "loadings" de **X** (variáveis que possuem mais informação) têm sua importância relativa ampliada em $\mathbf{X}^{\gamma}$. Dessa forma as variáveis independentes que têm maior importância ficam mais importantes, e as que têm menor importância ficam menos importantes. Como o PLS captura a variância ao longo da correlação, fazendo a matrix **X** mais direcional, o algoritmo do PLS (embutido na CPR) produz variáveis latentes que são mais finamente alinhadas aos componentes principais da

matriz **X** original (o modelo contempla a "variância canônica"). Na prática isso é numericamente equivalente ao PCR.

Quando o valor de γ tende a zero, os valores singulares da matriz **S** tendem a se igualar (qualquer número elevado a zero é igual à unidade). Desse modo $\mathbf{X}^{\gamma}$ torna-se igualmente ponderado em relação a todas as variáveis (menos direcional). Nesse caso, o algoritmo do PLS (embutido na CPR) capturará apenas a correlação ("correlação canônica"). Numericamente, isso é equivalente a fazer uma regressão linear múltipla, porque a regressão é feita sem ponderar a informação contida nas diversas variáveis.

Como o algoritmo utilizado no CPR é o do PLS, quando γ =1, a matriz de valores singulares é igual a ela própria (todo número elevado à unidade é igual a ele próprio), o modelo se iguala ao próprio PLS.

Para a determinação do valor do γ ótimo, constrói-se uma superfície do valor do PRESS (Erro quadrático médio de previsão), obtido por validação cruzada "leave one out" [59], em função do número de fatores e de uma série (discreta) de possíveis valores de γ. O valor do γ ótimo e do número ótimo de fatores são obtidos pelas coordenadas do ponto em que se obtém o menor valor do PRESS.

Capítulo 7

Determinação voltamétrica simultânea de ácido ascórbico e ácido úrico por adição de padrão multivariada

7.1 – Problemas em determinações voltamétricas simultâneas

Geralmente as medidas de voltametria são governadas por difusão. Quando o potencial elétrico aplicado ao eletrodo de trabalho é suficiente para promover uma reação de oxi-redução, as espécies eletroquimicamente ativas sofrem oxi-redução e migram do eletrodo para a solução e/ou da solução para o eletrodo. Em ambos os casos cria-se um gradiente de concentração das espécies formadas e/ou consumidas. Esse gradiente depende da geometria do eletrodo, mas geralmente é proporcional à distância em relação ao eletrodo. Devido à formação do gradiente de concentração, cria-se um movimento de difusão dos íons no sentido de homogeneizar a solução, ou seja, a difusão iônica tende a eliminar o gradiente gerado pela aplicação do potencial elétrico ao eletrodo.

Devido à formação do gradiente de concentração e da difusão dos íons em solução, a resposta eletroquímica se apresenta como um sinal de baixa freqüência, pois a difusão iônica gera um retardamento na resposta do sistema, alargando o sinal analítico. Dessa forma os voltamogramas de espécies eletroquimicamente ativas se apresentam na forma de bandas largas.

Se existe mais de uma espécie com propriedades eletroquímicas semelhantes em solução, é muito provável que as mesmas apresentem um certo grau de sobreposição no voltamograma. Métodos clássicos de determinação quantitativa possivelmente falham para sistemas com sobreposição de respostas, visto que são baseados em calibração univariada. Além disso, sistemas com espécies eletroquimicamente ativas com grupos funcionais semelhantes apresentam potenciais de pico relativamente próximos, aumentando assim a sobreposição e suas respostas. Outro fator importante na diferenciação da resposta

eletroquímica é o tamanho da molécula e sua afinidade pelo solvente e/ou íons presentes em solução. Esses parâmetros afetam a velocidade de difusão dessas espécies ao longo do gradiente de concentração.

Por exemplo: misturas de ácidos orgânicos com mesma ordem de grandeza na cadeia orgânica. Devido à estrutura da cadeia orgânica, existe maior ou menor estabilização das espécies oxidadas e reduzidas o que causa certa diferenciação no potencial de oxi-redução. As diferentes velocidades de difusão das espécies geram voltamogramas com perfis de resposta diferentes. Se as medidas forem razoavelmente sobrepostas não se pode resolver quantitativamente as espécies eletroquimicamente ativas baseado em calibração univariada, pois a presença de uma espécie interfere na corrente de pico da outra, e vice-versa. Nesse caso pode-se utilizar calibração multivariada para modelar as diferenças na posição (potencial de pico) e perfil de oxi-redução (forma do voltamograma). Dessa forma pode-se determinar quantitativamente as espécies eletroquímicas simultaneamente.

7.2 – Determinação de condições experimentais de varredura

Através de um planejamento fatorial prévio, otimizou-se as condições experimentais para velocidade de varredura de 10 mV/s; com altura do pulso diferencial de 50 mV. Utilizou-se 5,00 mL de solução 0,1 mol L^{-1} de PIPES pH = 7 como eletrólito suporte. Foram utilizados dois eletrodos de platina (eletrodo de trabalho e auxiliar), e um eletrodo de referência de Ag/AgCl.

Como o eletrólito suporte (solução de PIPES, pH = 7) apresenta certa atividade eletroquímica na faixa de potenciais em que ocorre a oxidação do ácido ascórbico e do ácido úrico, optou-se por restringir a faixa de potenciais aplicados entre 0 e 400 mV.

A Figura 7.1 mostra o voltamograma de uma solução 0.1 mol L^{-1} do eletrólito suporte PIPES (pH=7) puro.

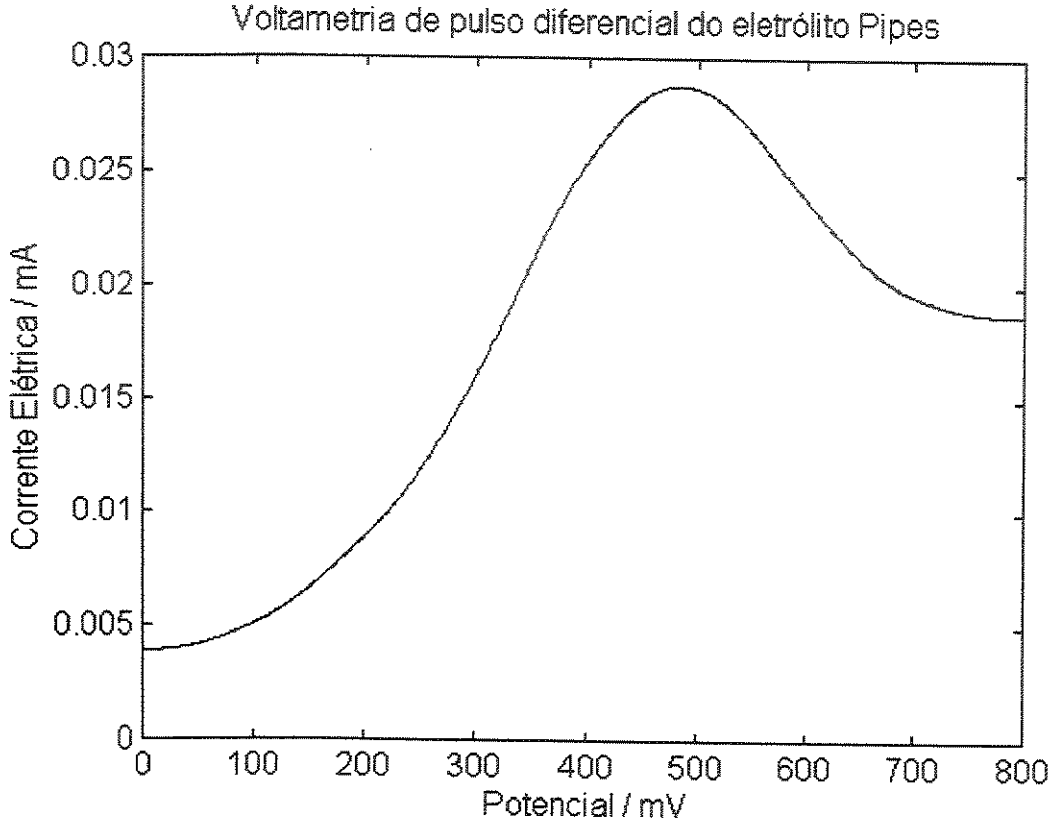


Figura 7.1 – Voltamograma de uma solução 0,1 mol L^{-1} de eletrólito suporte PIPES pH = 7.

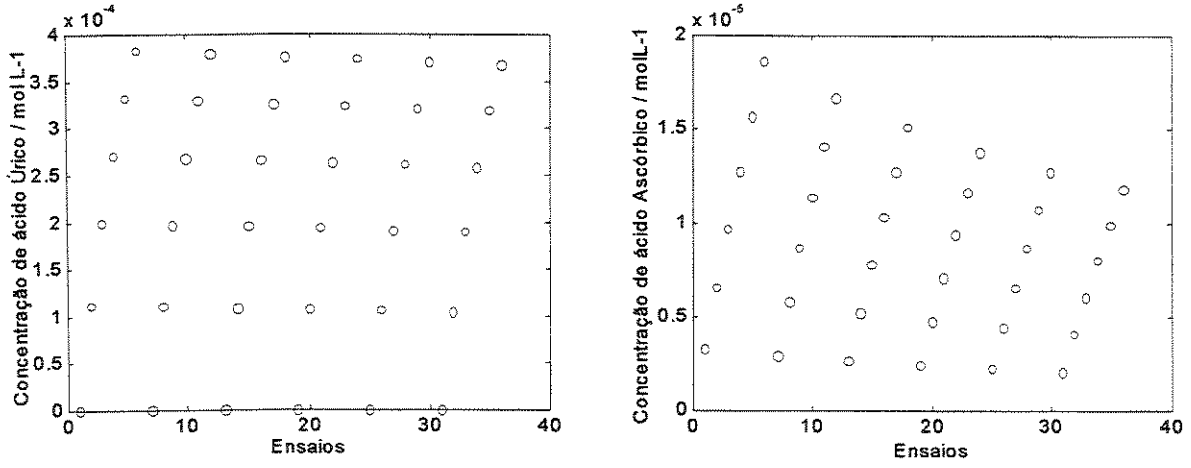
7.3 – Planejamento experimental

O modelo de adição de padrão em questão foi aplicado na determinação de ácido úrico e ácido ascórbico sobre uma matriz de urina humana.

Pelo modelo de adição de padrão proposto, para cada alíquota da amostra tomou-se inicialmente o voltamograma de 10,00 mL de uma solução de eletrólito suporte puro para servir de referência.

Em seguida, adicionou-se sobre a solução do eletrólito suporte uma alíquota de 5,00 mL da amostra e tomou-se seu voltamograma. Na seqüência, realizaram-se as adições de padrão sobre o sistema 10,00 mL de eletrólito suporte + 5,00 mL de alíquota da amostra. A cada ponto do planejamento tomou-se o voltamograma da solução obtida no respectivo ensaio.

A Figura 7.2 mostra as concentrações finais, com as diluições e aumentos de concentração corrigidos, obtidas em função das adições de padrão em cada ensaio do planejamento



A B

Figura 7.2 – Concentrações de (A) ácido úrico e (B) ácido ascórbico, obtidos pelo planejamento do modelo de adição de padrão.

Utilizaram-se 6 alíquotas de 100 μL ácido ascórbico 5 x 10^{-2} mol L^{-1} (para cada nível de ácido úrico) e 5 alíquotas de 1000 μL de ácido úrico 3 x 10^{-3} mol L^{-1} além do nível nulo,

totalizando 6 níveis, (para cada nível de ácido ascórbico). Desse modo obteve-se um planejamento com duas variáveis em 6 níveis cada, totalizando 36 ensaios obtidos por adição de padrão.

Para um nível fixo do ácido úrico (primeiro analito), a adição de 100 μL do ácido ascórbico (segundo analito) provoca uma pequena diluição na concentração do ácido úrico. O contrário também é observado (ver Figura 7.2), sendo que devido ao fato da solução de ácido úrico ser mais diluída (devido à sua menor solubilidade foi necessário diluí-la), a diluição produzida na concentração de ácido ascórbico em função da adição de ácido úrico é mais pronunciada.

7.4 – Resultados experimentais

O modelo de adição de padrão proposto foi aplicado na determinação de ácido úrico e ácido ascórbico sobre uma matriz de urina humana. Nessa amostra de urina, conhecia-se que a concentração de ácido úrico era 8,0 $x10^{-4}$ mol L^{-1} e que a de ácido ascórbico era de 1,0 $x10^{-4}$ mol L^{-1}.

Tomaram-se os voltamogramas das alíquotas da amostra, e das soluções obtidas pelas adições de padrão, e subtraiu-se dos mesmos as respectivas medidas voltamétricas da solução de eletrólito puro. Subtraindo-se as medidas voltamétricas relativas às alíquotas de amostra das medidas voltamétricas das soluções após as adições de padrão, obteve-se os voltamogramas mostrados na Figura 7.3. Esta figura mostra as medidas voltamétricas relativas a apenas as adições de padrão (previamente descontados o efeito do eletrólito suporte e da amostra).

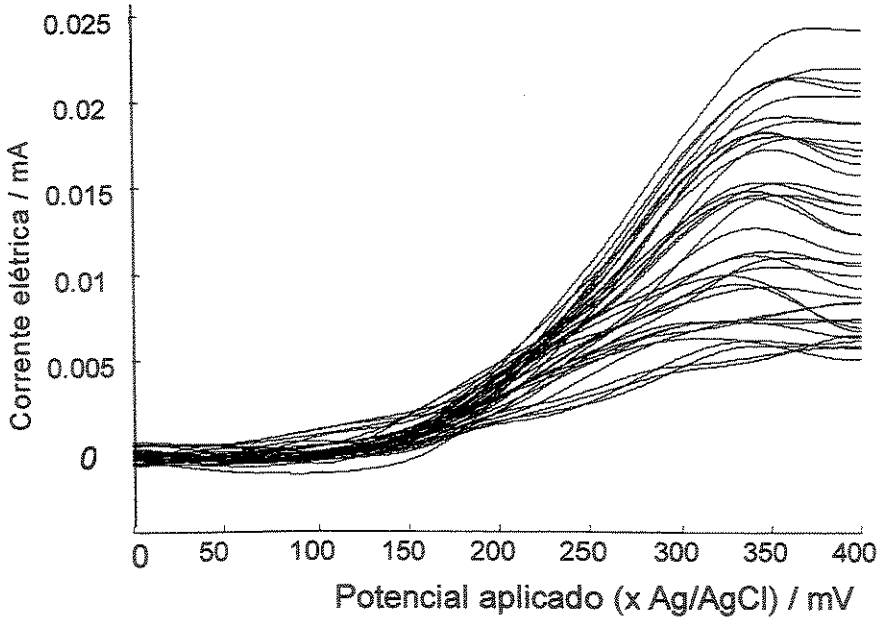


Figura 7.3 – Respostas voltamétricas das adições de padrão.

As medidas voltamétricas relativas a apenas às alíquotas da amostras (já descontadas as medidas voltamétricas relativas ao eletrólito suporte) são mostradas na Figura 7.4.

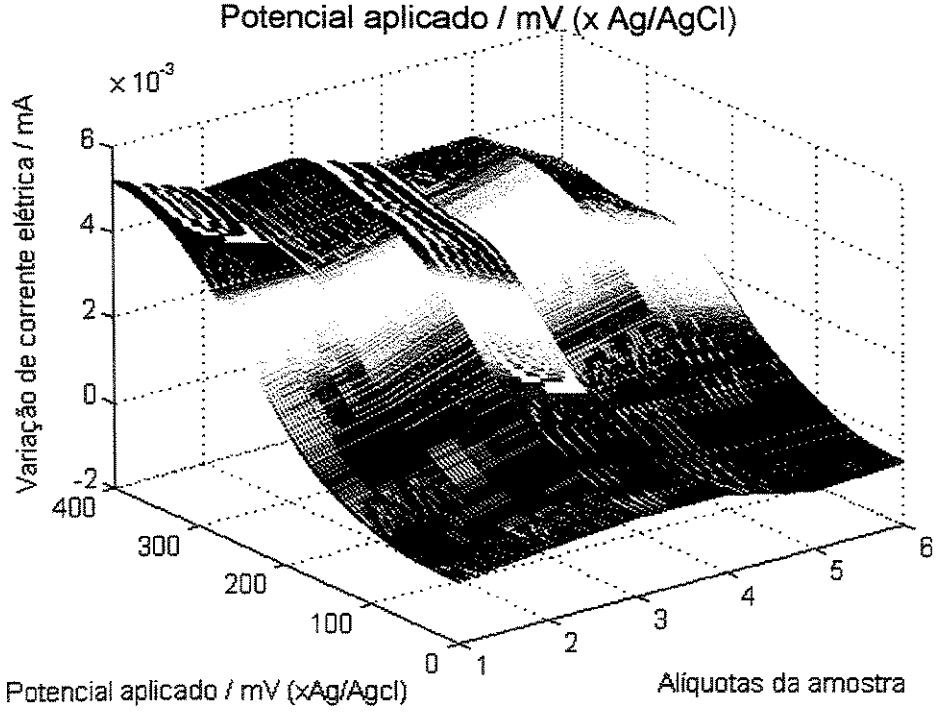


Figura 7.4 – Respostas voltamétricas relativas às alíquotas da amostra.

Para construir o modelo de CPR, inicialmente precisa-se obter os valores ótimos do parâmetro de continuidade e do número de fatores utilizados. Esses parâmetros foram obtidos a partir da análise do PRESS (erro quadrático médio para amostras de previsão) obtido por validação em blocos [59]. A Figura 7.5 mostra a superfície de PRESS (em função do número de fatores utilizados e do parâmetro de continuidade) para os dados obtidos pelos voltamogramas da adição de padrões para o ácido úrico e a Figura 7.6 mostra a superfície do PRESS para o ácido ascórbico.

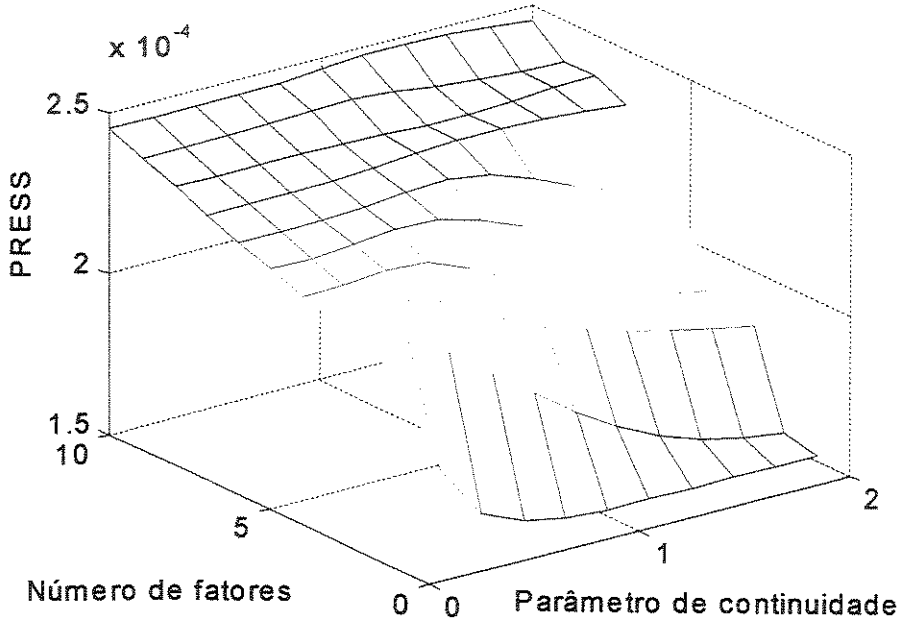


Figura 7.5 – Superfície de PRESS, obtida por validação cruzada em blocos, para o ácido úrico, por calibração multivariada com regressão contínua.

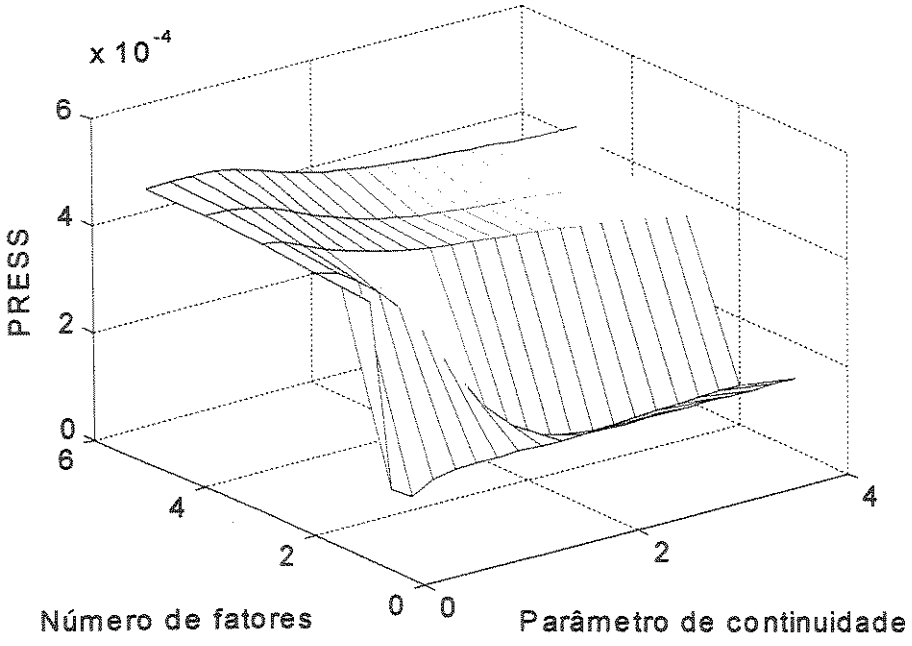


Figura 7.6 – Superfície de PRESS, obtida por validação cruzada em blocos, para o ácido ascórbico, por calibração multivariada com regressão contínua.

Tomaram-se as coordenadas do ponto mínimo das duas superfícies de PRESS e determinaram-se os parâmetros de continuidade e o número de fatores ótimos para os dados em questão.

Os valores desses parâmetros são mostrados na Tabela 7.1.

Tabela 7.1 – Determinação do parâmetro de continuidade e do número de fatores ótimos para calibração multivariada por regressão contínua.

Analito	Parâmetro de continuidade	Número de fatores
Ácido úrico	**2**	**1**
Ácido ascórbico	**2,6**	**2**

De acordo com os dados obtidos na Tabela 1 ambos os modelos são híbridos entre o PLS e o PCR. Extrapolando os modelos construídos para realizar a previsão da amostras obteve-se a Tabela 2.

Tabela 7.2 - Desempenho do modelo de adição de padrão proposto.

Analito	Alíquota	Concentração real da amostra / mol.L^{-1}	Concentração estimada para a amostra / mol.L^{-1}	Erro % relativo
Ácido Úrico	1	$8,0 \times 10^{-4}$	$6,31 \times 10^{-4}$	-21,1
Ácido Úrico	2	$8,0 \times 10^{-4}$	$8,56 \times 10^{-4}$	7,0
Ácido Úrico	3	$8,0 \times 10^{-4}$	$7,64 \times 10^{-4}$	-4,5
Ácido Úrico	4	$8,0 \times 10^{-4}$	$6,98 \times 10^{-4}$	-12,7
Ácido Úrico	5	$8,0 \times 10^{-4}$	$8,44 \times 10^{-4}$	5,4
Ácido Úrico	6	$8,0 \times 10^{-4}$	$7,86 \times 10^{-4}$	-1,7
Ácido Úrico	**média**	**$8,0 \times 10^{-4}$**	**$7,6 \times 10^{-4}$**	**4,6**
Ácido ascórbico	1	$1,0 \times 10^{-4}$	$1,19 \times 10^{-4}$	18,7
Ácido ascórbico	2	$1,0 \times 10^{-4}$	$1,03 \times 10^{-4}$	2,7
Ácido ascórbico	3	$1,0 \times 10^{-4}$	$1,05 \times 10^{-4}$	5,2
Ácido ascórbico	4	$1,0 \times 10^{-4}$	$8,62 \times 10^{-5}$	-13,8
Ácido ascórbico	5	$1,0 \times 10^{-4}$	$9,12 \times 10^{-4}$	-8,8
Ácido ascórbico	6	$1,0 \times 10^{-4}$	$7,82 \times 10^{-5}$	-21,9
Ácido ascórbico	**média**	**$1,0 \times 10^{-4}$**	**$9,7 \times 10^{-5}$**	**3,0**

Pode-se obter uma previsão para cada alíquota, visto que, foi obtido um voltamograma para cada alíquota, mas o valor real de concentração do ácido úrico e do ácido ascórbico são melhor representados pela média das estimativas das alíquotas. Com esse procedimento pode-se estimar o erro de previsão para a amostra em questão.

Desse modo pode-se dizer que as concentrações estimadas pelo modelo foram de $7,63 \times 10^{-4}$ mol L^{-1} para o ácido úrico e $9,7 \times 10^{-5}$ mol L^{-1} para o ácido ascórbico, com erros percentuais relativos de determinação de 4,56% e 2,96% respectivamente.

Considerando a faixa de concentração monitorada, os erros relativos são razoáveis. Apesar de se obter erros individuais de previsão relativamente altos em algumas alíquotas

da amostra, esses erros apresentam distribuição normal em torno da média. Dessa forma quando se considera o valor médio de previsão, o erro relativo cai consideravelmente.

Conclusões

O modelo de algoritmo genético desenvolvido mostrou-se bastante robusto e flexível de modo a poder otimizar tanto variáveis discretas (presença ou ausência de freqüências de Fourier, RBF's específicas e/ou conexões específicas das RBFs), quanto variáveis contínuas (raios e centros das RBF's).

O primeiro modelo de calibração multivariada desenvolvido, baseado na seleção de freqüências de Fourier por AG, mostrou-se bastante promissor, uma vez que conseguiu reduzir o número de variáveis do modelo, assim como o erro médio de previsão. Esse modelo proposto mostrou-se satisfatório tanto em sistemas químicos relativamente lineares, como é o caso da determinação do brix em amostras de xarope de caldo de cana, a partir de seus espectros de absorbância na região do infravermelho próximo, quanto em sistemas claramente não-lineares, como o caso da determinação de pol em amostras de caldo de cana, a partir de seus espectros de absorbância na região do infravermelho próximo. Em todos os casos, o modelo proposto (FFT-PLS-AG) apresentou erros quadráticos médios de previsão (RMSEP) menores que o PLS, (modelo considerado padrão em calibração multivariada).

O segundo modelo de calibração multivariada proposto, baseado na otimização de uma RBFN por AG, também conseguiu modelar satisfatoriamente os espectros de absorbância na região do infravermelho próximo das amostras de xarope de cana de açúcar. O AG utilizado conseguiu otimizar todos os parâmetros da RBFN, não necessitando conhecimento prévio do sistema em questão. Esse modelo de calibração multivariada também apresentou erros quadráticos médios de previsão menores que os obtidos com o PLS.

A aplicação de modelos não lineares de calibração multivariada, em todos os casos estudados, mostrou-se mais eficaz para resolver problemas em sistemas em que não existe uma relação linear entre os dados analíticos (espectros) e a concentração das espécies químicas. Isso demonstra que a estratégia de uma transformação não-linear aumenta a capacidade de separação dos padrões de informação, para uma posterior regressão linear.

Tendo em vista a dificuldade de disponibilidade de verbas, a construção de um potenciostato controlado por computador mostrou-se uma alternativa viável e promissora para a aplicação de métodos quimiométricos a baixo custo. Além do preço, a versatilidade é outra vantagem do mesmo. O desenvolvimento do programa gerenciador do equipamento construído permitiu o desenvolvimento de uma formatação específica que possibilita o controle total sobre os dados adquiridos com o potencistato construído.

Por fim, o modelo multivariado de adição de padrão mostrou-se satisfatório, permitindo a determinação simultânea de duas espécies químicas sem a necessidade de se preparar uma grande quantidade de padrões.

Perspectivas futuras

Dentro do escopo da modelagem não linear, pode-se utilizar os modelos de calibração multivariada não-linear desenvolvidos (FFT-PLS-AG e RBFN-AG) em outros tipos de medidas multivariadas. Em eletroanalítica existe um grande número de medidas não lineares possíveis de serem utilizadas. As medidas de corrente elétrica em função do potencial aplicado (voltamograma) basicamente são lineares (a corrente elétrica é proporcional à concentração da espécie eletroquimicamente ativa), porém em potenciometria as medidas são proporcionais ao logaritmo da concentração (lei de Nerst). O perfil multivariado pode ser obtido pelo monitoramento do potencial gerado em um sistema de injeção em fluxo (FIA) [74, 75].

Medidas eletroquímicas são muito suceptíveis a ruídos na rede elétrica, de modo que o modelo FFT-PLS-AG possivelmente deve obter bons resultados para esses tipos de dados.

Em medidas eletroquímicas freqüentemente há interação entre espécies quimicas distintas, devido à efeito estérico e/ou por adsorção de espécies na superfície dos eletrodos. Adicionalmente, a aplicação de um potencial elétrico, pode induzir reações entre as espécies químicas. Enfim, há um número razoável de medidas eletroquímicas que podem não ser linearmente correlacionadas com a concentração das espécies. Sobre esses sistemas os modelos de calibração multivariada propostos possivelmente se aplicam satisfatoriamente.

Adicionalmente, há um grande número de substâncias que podem ser monitorados por espectroscopia vibracional na região do infravermelho próximo em que hajam desvios de linearidade. Possíveis aplicações desses modelos poderiam ser realizadas em sistemas

onde são realizadas medidas de reflectância de sólidos, onde são conhecidos problemas com não linearidades e ruídos [76, 77].

Referências Bibliográficas

1 - T. M. Cover, "Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition", IEEE Transactions on eletronic Computers, EC14, 326, (1965).

2- C. Darwin, "On the Origin of Species by Means of Natural Selection", Murray, London, (1859).

3 - J. H. Holland, "Adaptation in Natural and Artificial Systems", University of Michigan Press, Ann Arbor, (1975).

4 - D. E. Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning", Addison-Wesley, Reading, (1989).

5 - L. Davis, (ed.), "Handbook of Genetic Algorithms", Van Nostrand Reinhold, New York, (1991).

6 – Z. Michalewicz, "Genetic Algorithms + Data Structures = Evolution Programs", Springer-Verlag, Berlin Heidelberg, (1992).

7 – C. Reeves, "Modern Heuristic Techniques for Combinatorial Problems", Blackwell Scientific Publications, (1993).

8 – H. Martens e T. Naes, "Multivariate Calibration", Wiley, Chichester, (1989).

9 - K. R. Beebe e B. R. Kowalski, "An Introduction to Multivariate Calibration and Analysis", Anal. Chem., 59, 1007A, (1987).

10 – B. G. M. Vandeginste; D. L. Massart; L. M. C. Buydens; S de Jing; P. J. Lewi e J. Smeyers-Verbeke, "Handbook of Chemometrics and Qualimetrics: Part B", Elsevier, Amsterdan, (1998).

11 – M. Otto, "Chemometrics. Statistical and Computer Application in Analytical Chemistry", Whiley-VCH,Weinheim,(1999).

12 – K. R. Beebe; R. J. Pell e M. B. Seasholtz, "Chemometrics: A practical Guide", Wiley, (1998).

13 - P. Geladi e B. R. Kowalski, "Partial Least Squares Regression: A Tutorial", Anal. Chim. Acta, 185, 1, (1986).

14 – A. V. Oppenheim e R. W. Schafer, "Discrete-Time Signal Processing", Prentice Hall, New Jersey, (1989).

15 – G. B. Arfken e H. J. Weber, "Mathematical Methods for Physicists", Academic Press, San Diego, (1995).

16 – E. O. Cerqueira, R. J. Poppi e L. T. Kubota, "Utilização de Filtro de transformada de Fourier Para a Minimização de Ruídos em Sinais Analíticos", Quím. Nova, 23, 690, (2000).

17 - A. Savitzky and M. J. E. Golay, Anal. Chem., 36, 1627, (1964).

18 - Matlab for Windows Ver. 5.2,The Math Works Inc.,Natick, (1998).

19 – S. Haykin, "Redes Neurais", Artmed Editora LTDA, Porto Alegre,(2001).

20 – J. Zupan e J. Gasteiger,"Neural Networks in Chemistry and Drug Design", Whiley-VCH, Weinheim, (1999).

21 - Wythoff, B.; Chemom. Intell. Lab. Sys., 18, 115, (1993).

22 - J. R. S. Jang; C. T. Sun e E. Mizutani; "Neuro Fuzzy e soft computing"; Prentice Hall; Upper Saddle River, (1997).

23 - B. Hassibi e D. G. Stork; "In Advances in neural information processing systems 5", Ed. Morgan Kaufmann, San Mateo,(1993).

24 -Y. L. Cun; B. Boser; J. S. Denker e S. A. Solla, "Optimal brain damage: Advances in Neural information processing systems, vol.2"; Morgan Kaufman; San Mateo, (1990).

25 - R. J. Poppi; D. L. Massart; "The Optimal Brain Surgeon for Pruning Neural Network Achitecture Applied to Multivariate Calibration", Anal. Chim. Acta, 375, 187, (1998).

26 - C. Mello; R. J. Poppi; J. C. De Andrade e H. Cantarella, "Pruning Neural Network for Architecture Optimization Applied to Near Infrared Reflectance Spectroscopic Measurements. Determination of the nitrogen contents in wheat Leaves", Analyst, 124, 1669, (1999).

27 - I. T. Nabney, "Netlab Algorithms for Patern Recognition", Springer, London, (2002).

28 - F. Giraud e Z. M. Salameh, "Analysis of the effects of a passing cloud on a grid-interactive photovoltaic system with battery storage using neural networks", IEEE Transactions on Energy Conversion, 14, 1572, (1999).

29 - M.F. Wilkins; L. Boddy, C. W. Morris e R. R. Jonker, "Identification of phytoplankton from flow cytometry data by using radial basis function neural networks", Applied and Environmental Microbiology, 65, 4404, (1999).

30 – B. Walczak e D. L. Massart, "The radial basis functions - Partial least squares approach as a flexible non-linear regression technique", Anal. Chim. Acta, 331, 177, (1996).

31 - B. Walczak e D. L. Massart, "Application of Radial Basis Functions - Partial Leas Squares to non-linear pattern recognition problems: Diagnosis of process faults", Anal. Chim Acta, 331, 187, (1996).

32 – M. J. L. Orr, "Introduction to Radial Basis Function Networks", Centre for Cognitive Science, University of Edinburg,2 Buccleuce Place, Edinburg EH8 9LW, Scotland, 1996. Site (URL): http://anc.ed.ac.uk/~mjo/rbf.html

33 - M. J. L. Orr, "Recent Advances in Radial Basis Function Network", Institute for Adaptative and Neural computation Division of Informatics, Edinburgh University, Edinburg EH8 9LW, Scotland, 1999. Site (URL): http://anc.ed.ac.uk/~mjo/rbf.html

34 – T. Stubbings e H. Hutter, "Classification of analytical images with radial basis function networks and forward selection", Chemom. Intel. Labs. Syst., 49, 163. (1999).

35 - M. J. L. Orr, "Regularisation in the Selection of RBF Centres", Institute for Adaptative and Neural computation Division of Informatics, Edinburgh University, Edinburg EH8 9LW, Scotland, 1999. site (URL): http://anc.ed.ac.uk/~mjo/rbf.html

36 – R. N. Shreve e J. A. Brink Jr., "Indústria de Processos Químicos", Ed. Guanabara Dois, Rio de Janeiro, (1980).

37 – H. V. Amorim et all, "Métodos analíticos para o controle da produção de álcool e açúcar", ESALQ, Piracicaba, (1996).

38 – Conselho dos produtores de cana-de-açúcar, açúcar e álcool do estado do Paraná: CONSECANA-PR, "Normas Operacionais de Avaliação da Qualidade da cana-de-açúcar", FAEP - Federação da Agricultura do Estado do Paraná, Curituba, (2002). Site (URL): http://www.faep.com.br/consecana/normasop.htm

39 – B. M. Silva; R. M. Seabra ; P. B. Andrade ; M. B. Oliveira e M. A. Ferreira, "Adulteração por Adição de Açúcares a Sumos de Frutos: Ima revisão", Cienc. Tecnol. Aliment. 2, 184, (1999).

40 – B. G. Osborne; T. Fearn e P. H. Hindle, "Practical NIR Spectroscopy with Applications in Food and Beverage analysis". Longman Scientific & Technical, Essex, (1993).

41 – A. M. C. Watson, "Near IR Reflectance Spectrophotometric Analysis of Agricultural Products",Anal. Chem., 49, A835, (1977).

42 – J. D. Kirrsch e J. K. Drennen, "Near-Infrared Spectroscopy – Applications in the Analisis of Tablets and Solid Pharmaceutical Dosage Forms", Appl. Spectrosc. Rev., 30, 139, (1995).

43 – L. G. Weyer, "Near-Infrared Spectroscopy of Orgânic Substances", Appl. Spectrosc. Rev., 21,1,(1985).

44 – K. A. B. LEE, "Comparison of MID-IR with NIR in Polymer Analysis", Appl Spectrosc. Rev., 28, 231, (1993).

45 – B. Descales; I. Cermelli; J. R. Llinas; G. Margail e A. Martens, "Analysis online of Petrochemicals Units Using Near-infrared Spectrophotometry", Analusis, 21, M25, (1993).

46 – A. E. Carlsson e K. L. R. Janne, 'Near-Infrared Spectroscopy as an alternative to biological Testing for Quality-Control of hyaluronan – Comparison of Data Preprocessing Methods for Classification", Appl. Spctrosc., 49, 1037, (1995).

47 – J. T. Kuenstner; K. H. Norris e W. F. Mccarthy, "Measurement of Hemoglobin in unlysed Blood by Near-Infrared Spectroscopy", Appl. Spectrosc., 48, 484, (1994).

48 – H. Prufer e D. Mamma, "Near-Infrared Online Analysis of Motor Octane number ir Gasoline with an Acoustooptic Tunable transmission Spectrophotometer", Analusis, 23, M14 (1995).

49 – T. Buffeteau; B. Desbat e L. Bokobza, "The Use of Near-Infrared Spectroscopy Couplec to the Molecular-Orientation in Uniaxially Stretched Polymers", Polymer,

36, 4339, (1995).

50 – Y. Ghebremeskel; J. fields e A. Garton, " The use of Near Infrared (NIR) spectroscopy to Study specific Interactions in Polymer Blends", J. Polymer Science Part B – Pplymer Physics, 32, 383, (1994).

51 – N. A. Stjon e G. A. George, "Cure Kinetics and Mechanisms of a Tetraglicidyl-4,4'Diaminodiphenilmethane Diaminophenylsulphone Epoxi-Resin Using Near IR Spectroscopy", Polymer, 33, 2679, (1992).

52 – G. A. George; P. Coleclarke; N. A. Stjon e G. Friend, "Real-time Monitoring of the Cure Reaction of a TGDD/DDS Epoxy-resin Using Fiber Optic TF-IR", J. Appl Polymer Science, 42, 643, (1991).

53 – B. G. Min; Z. H. Stachurski; J. H. Hodgkin e G. R. Heat, "Quantitative-Analysis of the Cure Reaction of DGEBA DDS Epoxy-resins witout and with Thermoplastic Polysulfone Modifier Using Near-Infrared spectroscopy", Polymer, 34, 3620, (1993).

54 - W. H. Press et all, "Numerical Recipes in C: The Art of scientific Computing", Cambridge University Press, Cambridge, (1997).

55 – M. A. G. Ruggiero e V. L. R.Lopes, "Cálculo Numérico – Aspectos Teóricos e Computacionais", McGraw – Hill, Rio de Janeiro, (1988).

56 – B. M. Wise e N. B. Gallagher, "PLS_Toolbox 2.0 for use with MATLAB™", Eigenvector Reasearch Ink, (1998).

57 – L. Pasti; D. Jouan-Rimbaud; D. L. Massart e O. E. de Noord, "Application of Fourier Transform to Multivariate Calibration of Near-Infrared Data", Anal. Chim. Acta, 364, 253, (1998).

58 – F. Despagne e D. L. Massart, "Neural Networks in Multivariate Calibration", Analyst, 123, 157R, (1998).

59 – M. Stone, "Cross-Validatory Choice and Assessment of Statistical Prediction", J. Roy Stat. Soc., B, 111, (1974).

60 – R. D. Snee, "Validation of Regression Models: Methods and Examples"Technometrics, 19, 415, (1976).

61 – P. T. Kissinger e W. R. Helneman, "Cyclic Voltametry", J. Chem. Educ., 60, 702, (1983).

62 – G. A. Mabbott, "An Introduction to cyclic voltametry", J. Chem. Educ., 60, 697, (1983).

63 – PC-Labcard, PCL 711B, Advantech Co., (1993).

64 - Microsoft Visual Basic V. 3.0, Microsoft Co.,(1993).

65 – Microsoft Visual C++ V. 2.0, Microsoft Co.,(1995).

66 – E. O. Cerqueira e R. J. Poppi, "Using Dinamic Data Exange to Exange Information Between Visual Basic and Matlab", Trends in Analytical Chemistry", 15 500, (1996).

67 – B. de Barros Neto; I. S. Scarminio e R. E. Bruns, "Planejamento e Otimização de Experimentos", Editora da UNICAMP, Campinas, (1996).

68 – G. E. P. Box; W. G. Hunter e J. S. Hunter, "Statistics for Experimenters", John Wiley & Sons, New York, (1978).

69 - M. Stone e R. J. Brooks, "Continuum Regression: Cross-Validated Sequentially Constructed Prediction Embracing Ordinary Leart Squares, Partial Least squares and Principal Components Regression", J. R. Statist. Soc. B, 52, 237, (1990).; Corrigendum; 54 906, (1992).

70 – S. de Jong; B. M. Wise e N. L. Ricker, "Canonical Partial Least Squares and Continuum Power Regression", J. Chemom., 15, 85, (2001).

71 – M. C. Ortiz, J. Arcos e L. Sarabia, "Using Continuum Power Regression for Quantitative Analysis with Overlapping Signal Obtained by Differencial Pulse Polarography", Chemom. Int. Lab. Syst., 34, 245, (1996).

72 – B. M. Wise e N. L. Ricker, "Indentification of Finite Inpulse Response Models with Continuum Regression", J. Chemom., 7, 1, (1993).

73

– E. R. Malinowski, "Factor Analysis in Chemistry", Wiley, New York, (1991).

74 – J. Ruzika e E. H. Hansen, "Flow Injection Analysis", Wiley, New York, (1968).

75 - C. Ridder e L. Norgaard, "Simultaneous determination of Cobalt and Nickel by Flow Injection Analysis and Partial Least Squares Regression with Outlier Detection", Chemom. Intell. Lab. Syst., 14, 297, (1992).

76 – P. H. Fidêncio; R. J. Poppi e J. C. de andrade, "Determination of Organic Matter in Soils Using Radial Basis Function Network and Near Infrared Spectroscopy", Anal. Chim. Acta, 453, 125, (2002).

77 – W. F. McClure; A. Hamid; F. G. giesbrecht e W. W. Weeks, "Fourier Analysis Enhances NIR Difuse Reflectance Spectroscopy", Appl. Spectrosc., 38, 322, (1984).