UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE QUÍMICA
LAQQA – Laboratório de Quimiometria em Química Analítica

ALGORITMOS GENÉTICOS PARA SELEÇÃO DE VARIÁVEIS EM MÉTODOS DE CALIBRAÇÃO DE SEGUNDA ORDEM

Dissertação de mestrado

RENATO LAJARIM CARNEIRO
Orientador: Prof. Dr. Ronei Jesus Poppi

Campinas
2007

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO INSTITUTO DE QUÍMICA DA UNICAMP

C215a

Carneiro, Renato Lajarim.
Algoritmos genéticos para seleção de variáveis em métodos de calibração de segunda ordem / Renato Lajarim Carneiro. -- Campinas, SP: [s.n], 2007.

Orientador: Ronei Jesus Poppi.

Dissertação - Universidade Estadual de Campinas, Instituto de Química.

1. Seleção de variáveis. 2. Algoritmos genéticos. 3. Métodos de calibração de segunda ordem. I. Poppi, Ronei Jesus. II. Universidade Estadual de Campinas. Instituto de Química. III. Título.

Título em inglês: Genetic algorithm for selection of variables in second-order calibration methods

Palavras-chaves em inglês: Selection of variables, Genetic algorithm, second-order calibration methods

Área de concentração: Química Analítica

Titulação: Mestre em Química na Área de Química Analítica

Banca examinadora: Prof. Dr. Ronei Jesus Poppi (orientador), Profa. Dra. Adriana Vitorino Rossi (UNICAMP-IQ) e Dionísio Borsato (UEL, DQ)

Data de defesa: 10/07/2007

Dedico este trabalho a todos aqueles que
participaram desta minha caminhada;

A eles e a Ele que me deram a vida;

Àqueles que caminharam comigo desde o início, às vezes
deixando-me cair, sempre me ensinando a levantar;

Àqueles e a ela, os quais nossos caminhos se cruzaram,
e fizeram parte desta minha prazerosa jornada.

"A distância entre a loucura e a genialidade
é a distância entre o fracasso ou sucesso."

"Polvo, barro, sol y lluvia
Es Camino de Santiago
Millares de peregrinos
Y más de un millar de años,

Peregrino ¿Quién te llama?
¿Qué fuerza oculta te atrae?
Ni el Campo de las Estrellas
Ni as grandes catedrales.

No es la bravura navarra,
Ni el vino de los riojanos,
Ni los mariscos gallegos,
Ni los campos castellanos,

Peregrino, ¿Quién te llama?
¿Qué fuerza oculta te atrae?

Ni las gentes del Camino,
Ni las costumbres rurales,

No es la historia y la cultura,
Ni el gallo de La Calzada,
Ni el palacio de Gaudí,
Ni el Castillo Ponteferrada,

Todo lo veo al pasar,
Y es un gozo verlo todo,
Mas la voz que a mi me llama
La siento mucho más hondo

La fuerza que a mi me empuja
La fuerza que a mi me atrae
No sé explicarla ni yo,
¡Sólo el de Arriba lo sabe!"

E.G.B

AGRADECIMENTOS

- Ao meu pai Uri, minha mãe Marlene, e meu irmão Gustavo, pelo apoio e compreensão durante estes anos;

- À Alessandra Borin, por seu amor, confiança e apoio nos momentos em que mais necessitei;

- Ao prof. Dr. Ronei Jesus Poppi, pela orientação e oportunidade de realizar este trabalho;

- Ao prof. Dr. Romà Tauler Ferré do *Consejo Superior de Investigaciones Científicas (CSIC)*, pela aceitação e apoio na minha permanência de seis meses na Espanha;

- Aos prof. Dionísio Borsato, Rui Sérgio e Evandro Bona, pela amizade e por minha iniciação na quimiometria;

- Aos amigos da república "Toca dos gatos": Marcelo P2, Haroldão, Lomba, Pereira e Pedrão, pela amizade, cervejas, e churrascos;

- Aos amigos do LAQQA: Jez (em especial, sem o qual a realização deste trabalho seria muito mais difícil), Paulo Henrique, Danilo, Trevisan, Genésio, Werickson, Gilmare, Patrícia e claro, Alessandra;

- Aos amigos da república em Barcelona: Dominique (EUA), Cathy (Filipinas), Tony (Espanha), Shain (Turquia), Efrain, Dennis e Raymundo (México), os quais me acolheram, e fizeram da minha estadia um período muito agradável;

- Aos amigos do CSIC: Stephan, Silvia (Cochi), Marta, Leonel e Débora;

- Aos amigos do Caminho de Santiago que percorreram comigo 860 quilômetros, durante 28 dias de caminhada pelo norte da Espanha: José, Miguel, Jorge, Israel, Frank, Marina, Nicolas e muitos outros, com os quais vivi e aprendi muitas lições que levarei comigo durante toda minha vida;

- Aos amigos da minha querida cidade de Londrina, ao pessoal do Parigot e ao pessoal da turma de Bacharelado em Química 2004 da UEL;

- À UNICAMP pela infra-estrutura que possibilitou o desenvolvimento deste trabalho. À FAPESP (Processo 05/53280-4) pelo financiamento e ao Banco Banespa - Santander pela bolsa de mobilidade internacional;

- Finalmente, à todos aqueles que contribuíram de alguma forma com minha formação pessoal e profissional, e que (por esquecimento) não foram citados anteriormente.

SÚMULA CURRICULAR

1. Dados pessoais

- Estado civil: Solteiro;
- Nacionalidade: Brasileiro;
- Data de nascimento 27/11/1982.
- Naturalidade: Assis Chateaubriand – PR
- E-mail: renatolajarim@pop.com.br

2. Formação Acadêmica

2.1 Mestrado (08/2005 a 07/2007)

- Mestrado em Química Analítica na Universidade Estadual de Campinas com permanência de 6 meses no Consejo Superior de Investigaciones Científicas (IIQAB-CSIC), Barcelona, Espanha.

2.2 Graduação (02/2001 a 12/2004)

- Graduação em Bacharelado em Química com atribuições de Bacharel em Química e Química Tecnológica na Universidade Estadual de Londrina, UEL.

3. Produção científica

3.1 Iniciação científica (02/2002 a 08/2004)

- Projeto: Desenvolvimento de aplicativo para microinformática que permita otimizaçao simultânea em sistemas alimentares com respostas múltiplas: generalização da função de Derringer-Suich; Instituição financiadora: de 02/2002 a 07/2003 Universidade Estadual de Londrina (IC-UEL) e de 08/2003 a 08/2004 CNPq; Universidade Estadual de Londrina.

3.2 Principais resumos de trabalhos científicos apresentados em congressos

- CARNEIRO, R. L.; BRAGA, J. W. B; BOTOLLI, C. B. G.; POPPI, R. J. . Application of Genetic Algorithm for variables selection in BLLS method for pesticides and metabolites determinations in wine. In: 10th International Conference on Chemometrics in Analytical Chemistry. Águas de Lindóia, 2006.
- CARNEIRO, R. L.; BRAGA, J. W. B; BOTOLLI, C. B. G.; POPPI, R. J. . Aplicação de Algoritmo Genético para seleção de variáveis em calibração de segunda ordem na determinação de pesticidas em vinho. In: 29ª Reunião da Sociedade Brasileira de Química, 2006, Águas de Lindóia.
- BONA, E. ; BORSATO, D. ; SILVA, R. S. S. F. ; SILVA, L. H. M. ; FIDELIS, D. A. S. ; ARAUJO, A. ; CARNEIRO, R. L. . Modelagem e simulação da difusão do NaCl e KCl em queijo prato através do método de elementos finitos: Salga sem agitação.. In: 28ª Reunião da Sociedade Brasileira de Química, 2005, Poços de Caldas.
- BORSATO, D. ; CARNEIRO, R. L. ; ARAUJO, A. ; BONA, E. ; SILVA, R. S. S. F. ; FIDELIS, D. A. S. ; SILVA, L. H. M. . Modelagem e simulação da difusão do NaCl e KCl em queijo prato através do método de elementos finitos:Salga com agitação. In: 28ª Reunião da Sociedade Brasileira de Química, 2005, Poços de Caldas.
- CARNEIRO, R. L. ; SILVA, R. S. S. F. ; BORSATO, D. . Métodos de gradiente para otimização simultânea em sistemas alimentares.. In: XIX Congresso Brasileiro de Ciência e Tecnologia de Alimentos, 2004, Recife. Anais do XIX CBCTA em CD. Recife: CEJEM, 2004., 2004.
- FUCHS, R. H. B. ; HAULY, M. C. O. ; OLIVEIRA, A. S. ; LIUTTI, G. C. ; BONA, E. ; CARNEIRO, R. L. ; BORSATO, D. . Aplicação do método Complex na otimização da formulação do iogurte de soja suplementado com inulina e oligofrutose.. In: XXVI Congresso Latinoamericano de Química e 27ª Reunião Anual da Sociedade Brasileira de Química, 2004, Salvador.

3.3 Artigos publicados

- RENATO L. CARNEIRO, JEZ W.B. BRAGA, CARLA B.G. BOTTOLI, RONEI J. POPPI. Application of genetic algorithm for selection of variables for the BLLS method applied to determination of pesticides and metabolites in wine. Analytica Chimica Acta. 595, p. 51-58. **2007**.

- EVANDRO BONA; RENATO L. CARNEIRO; DIONISIO BORSATO; RUI S.S.F. SILVA; DAYANNE A. S. FIDELIS; LUIZ H. M. SILVA. Simulation of NaCl and KCl mass transfer during salting of prato cheese in brine with agitation: a numerical solution. Brazilian Journal of Chemical Engineering. 24 (03) p. 337 – 349. **2007**.

- RENATO L. CARNEIRO; RUI S. S. F. SILVA; DIONISIO BORSATO; EVANDRO BONA. Métodos de gradiente para otimização simultânea: estudo de casos de sistemas alimentares. Semina: Ciências Agrárias, Londrina, v. 26, n. 3, p. 353-362, jul./set. **2005**.

4. Outros

4.1 Monitoria

- de 03/2002 a 12/2002 na disciplina Química Inorgânica na Universidade Estadual de Londrina, Departamento de Química;

4.2 Estágio

- de 08/2004 a 02/2005 na Cia. Iguaçu de Café Solúvel nos departamentos de Controle de Qualidade e Pesquisa e Desenvolvimento, onde foi realizada a validação de métodos analíticos empregados;

4.3 Participações em congressos

- 1ER Encuentro de Jóvenes Investigadores en Quimiometría. Tarragona, Espanha. 2006.
- 29ª Reunião Anual da Sociedade Brasileira de Química. Águas de Lindóia - SP. 2006.

- XXVI Congresso Latinoamericano de Química e 27ª Reunião Anual da Sociedade Brasileira de Química..XXVI Congresso Latinoamericano de Química e 27ª Reunião Anual da Sociedade Brasileira de Química. Salvador - BA. 2004. (Participações em eventos/Congresso).
- XIII Encontro Anual de Iniciação Científica e I Seminário Estadual de Políticas de Pesquisa e Pós-graduação (SEPG).XIII Encontro Anual de Iniciação Científica e I Seminário Estadual de Políticas de Pesquisa e Pós-graduação (SEPG). Londrina - PR. 2004.
- 26ª Reunião Anual da Sociedade Brasileira de Química.26ª Reunião Anual da Sociedade Brasileira de Química. Poços de Caldas - MG. 2003.
- XII EAIC - PIBIC/CNPq - Encontro Anual de Iniciação Científica.XII EAIC - PIBIC/CNPq - Encontro Anual de Iniciação Científica. Foz do Iguaçú - PR. 2003.

RESUMO

Titulo: ALGORITMOS GENÉTICOS PARA SELEÇÃO DE VARIÁVEIS EM MÉTODOS CALIBRAÇÃO DE SEGUNDA ORDEM

Autor: Renato Lajarim Carneiro

Orientador: Ronei Jesus Poppi

Esse trabalho teve por objetivo desenvolver um programa em MatLab baseado no Algoritmo Genético (GA) para aplicar e verificar as principais vantagens deste na seleção de variáveis para métodos de calibração de segunda ordem (BLLS-RBL, PARAFAC e N-PLS). Para esta finalidade foram utilizados três conjuntos de dados:

1. Determinação de pesticidas e um metabólito em vinho tinto por HPLC-DAD em três situações distintas. Nestas três situações foram observadas sobreposições dos interferentes sobre os compostos de interesse. Estes compostos eram os pesticidas carbaril (CBL), tiofanato metílico (TIO), simazina (SIM) e dimetoato (DMT) e o metabólito ftalimida (PTA).
2. Quantificação das vitaminas B2 (riboflavina) e B6 (piridoxina) por espectrofluorimetria de excitação/emissão em formulações infantis comerciais, sendo três leites em pó e dois suplementos alimentares.
3. Análise dos fármacos ácido ascórbico (AA) e ácido acetilsalicílico (AAS) em formulações farmacêuticas por FIA com gradiente de pH e detecção por arranjo de diodos, onde a variação de pH causa alteração na estrutura das moléculas dos fármacos mudando seus espectros na região do ultravioleta.

A performance dos modelos, com e sem seleção de variáveis, foi comparada através de seus erros, expressados como a raiz quadrada da média dos quadrados dos erros de previsão (RMSEP), e os erros relativos de previsão (REP). Resultados melhores foram claramente observados quando o GA foi utilizado para a seleção de variáveis nos métodos de calibração de segunda ordem.

ABSTRACT

Title: GENETIC ALGORITHM FOR SELECTION OF VARIABLES FOR SECOND-ORDER CALIBRATION METHODS

Author: Renato Lajarim Carneiro
Adviser: Ronei Jesus Poppi

The aim of this work was to develop a program in MatLab using Genetic Algorithm (GA) to apply and to verify the main advantages of variables selection for second-order calibration methods (BLLS-RBL, PARAFAC and N-PLS). For this purpose three data sets had been used:

1. Determination of pesticides and a metabolite in red wines using HPLC-DAD in three distinct situations, where overlappings of the interferentes on interest compounds are observed. These composites were the pesticides carbaryl (CBL), methyl thiophanate (TIO), simazine (SIM) and dimethoate (DMT) and the metabolite phthalimide (PTA).
2. Quantification of the B2 (riboflavine) and (pyridoxine) B6 vitamins for spectrofluorimetry of excitation-emission in commercial infantile products, being three powder milk and two supplement foods.
3. Analysis of ascorbic acid (AA) and acetylsalicylic acid (AAS) in pharmaceutical tablets by FIA with pH gradient and detection for diode array, where the variation of pH causes alterations in the structure of molecules of analites shifting its spectra in the region of the ultraviolet.

The performance of the models, with and without selection of variable, was compared through its errors, expressed as the root mean square error of prediction (RMSEP), and the relative errors of prediction (REP). The best results were obtained when the GA was used for the selection of variable in second-order calibration methods.

ÍNDICE

LISTA DE ABREVIATURAS

GA	Algoritmo Genético
NPLS	Quadrados Mínimos Parciais N-Dimensional
PARAFAC	Análise de Fatores Paralelos
BLLS-RBL	Quadrados Mínimos Bilineares – Bilinearização residual
RMSEP	Raiz quadrada da média dos quadrados dos erros de previsão
REP	Erros relativos de previsão
MLR	Regressão Linear Múltipla
PCR	Regressão por Componentes Principais
PLS	Regressão por Mínimos Quadrados Parciais
ALS	Algoritmo de mínimos quadrados alternados
SVD	Decomposição de valores singulares
EEM	Matrizes de fluorescência de excitação-emissão
SPE	Extração em fase sólida
UV-Vis	Radiação eletromagnética ultravioleta e visível
FIA-DAD	Análise por Injeção em Fluxo e detecção por arranjo de diodos
NIR	Infravermelho próximo
HPLC-DAD	Cromatografia Líquida de Alta Eficiência com detecção por arranjo de fotodiodos
HPLC-MS	Cromatografia Líquida de Alta Eficiência com detecção por Espectrômetro de Massas
CG-MS	Cromatografia Gasosa com detecção por Espectrômetro de Massas
RMN-2D	Ressonância Nuclear Magnética em duas dimensões
CBL	Pesticida carbaril
TIO	Pesticida tiofanato metílico
SIM	Pesticida simazina
DMT	Pesticida dimetoato
PTA	Metabólito ftalimida
Vitamina B2	Riboflavina

Vitamina B6	Piridoxina
APL, **BPL**, **CPL**, **DPL** e **EPL**	Representam cada uma das preparações lácteas comerciais utilizadas no capítulo 5
AA	Ácido ascórbico, podendo representar também a forma protonada (meio ácido)
AA^+	Espécie desprotonada do AA (meio básico)
AAS	Ácido acetilsalicílico, podendo representar também a forma protonada (meio ácido)
AAS^+	Espécie desprotonada do AAS (meio básico)
JMD, **KMD**, **LMD** e **MMD**	Representam cada um dos medicamentos comerciais utilizados no capítulo 6

LISTA DE TABELAS

Capítulo 4

Capítulo 5

Capítulo 6

LISTA DE FIGURAS

Capítulo 2

Capítulo 3

Capítulo 4

Capítulo 5

Capítulo 6

CAPÍTULO 1

INTRODUÇÃO E OBJETIVOS GERAIS

Nos processos de calibração empregados na determinação de propriedades de interesse em sistemas químicos, normalmente o uso de somente algumas variáveis que contêm mais informação pode favorecer a interpretação do modelo, pois elimina ruídos e não-linearidades. Os algoritmos genéticos (GA) são métodos de otimização numérica que simulam o processo de evolução biológica baseada na teoria de Darwin, sendo aplicados com sucesso em muitas situações para seleção de variáveis utilizando uma função de otimização adequada [1]. A seleção é realizada através de uma seqüência de dígitos binários (cromossomos), armazenados na memória do computador, que evoluem a todo tempo, de maneira muito semelhante ao modo com o que os indivíduos de uma população natural evoluem. Embora os procedimentos computacionais sejam muito simplificados comparados com os processos naturais, os GAs são capazes de otimizar problemas complexos e interessantes [2].

Em química analítica, o GA tem sido aplicado em vários trabalhos [3-11] com métodos de calibração de primeira ordem, tais como regressão por mínimos quadrados parciais (PLS) e regressão linear múltipla (MLR), com o propósito de selecionar as variáveis mais relevantes a fim de conseguir uma melhor estimativa da concentração de alguns analitos numa amostra.

A aplicação do GA com calibração de segunda ordem é recente, com poucos artigos publicados. São dados alguns exemplos de artigos publicados [12-14] utilizando o GA e métodos de calibração de segunda ordem:

- selecionar os melhores subconjuntos nos N-modos mantendo a informação estrutural do conjunto de dados do método PARAFAC [12];
- selecionar melhores conjuntos para o modelo de calibração no N-PLS [13] e;
- selecionar os melhores comprimentos de onda na região NIR para se propor uma rota no estudo de uma reação de polimerização [14].

No anexo A é apresentado um artigo, que avalia a otimização por GA objetivando uma melhora da previsão dos modelos construídos utilizando somente soluções padrão, a fim de analisar amostras que contenham componentes

desconhecidos, que é uma das principais características dos métodos de calibração de segunda ordem. Esse artigo foi publicado no desenrolar desta dissertação.

O objetivo desta dissertação foi desenvolver um programa baseado no Algoritmo Genético para seleção de variáveis em métodos de calibração de segunda ordem, aplicando e verificando as principais vantagens deste quando utilizado conjuntamente com: BLLS-RBL (quadrados mínimos bilineares – bilinearização residual) e PARAFAC (análise de fatores paralelos), que apresentam vantagem de segunda ordem e ainda N-PLS (regressão por mínimos quadrados parciais n-dimensional). Para esta finalidade foram utilizados três conjuntos de dados:

1. Determinação de pesticidas em vinho tinto por Cromatografia Líquida de Alta Eficiência com arranjo de fotodiodos (HPLC-DAD) em três situações distintas, onde foram observadas sobreposições dos picos dos interferentes com os picos dos compostos de interesse. Estes compostos eram os pesticidas carbaril (CBL), tiofanato metílico (TIO), simazina (SIM) e dimetoato (DMT) e o metabólito de pesticida ftalimida (PTA);
2. Determinação de riboflavina (vitamina B2) e piridoxina (vitamina B6) em formulações lácteas comerciais através de espectrofluorimetria de emissão/excitação;
3. Análise dos fármacos ácido ascórbico (AA) e ácido acetilsalicílico (AAS) em formulações farmacêuticas por FIA com gradiente de pH e detecção por arranjo de fotodiodos, onde a variação de pH causa alteração na estrutura das moléculas dos fármacos mudando seus espectros na região do ultravioleta.

A performance dos modelos, com e sem seleção de variáveis, foi comparada através de seus erros, expressos como a raiz quadrada da média dos quadrados dos erros de previsão (RMSEP), e dos erros relativos de previsão (REP). Quando os métodos de segunda ordem foram utilizados com o GA, pôde-se observar a diminuição dos RMSEPs e a melhora dos níveis de recuperação, nos três casos analisados.

CAPÍTULO 2

MÉTODOS QUIMIOMÉTRICOS UTILIZADOS

2.1. Métodos de Calibração

O processo de calibração pode ser definido como uma série de operações que estabelecem, sob condições específicas, uma relação matemática entre medidas instrumentais e valores da propriedade de interesse correspondente, realizada em padrões. Os métodos de calibração existentes podem ser divididos, quanto à complexidade, ou dimensionalidade dos dados, em calibração de ordem zero, primeira e segunda ordem [15]. A Figura 1 mostra uma representação esquemática dos três tipos de dados utilizados nos três níveis de calibração citados, obtidos através de um equipamento HPLC-DAD.

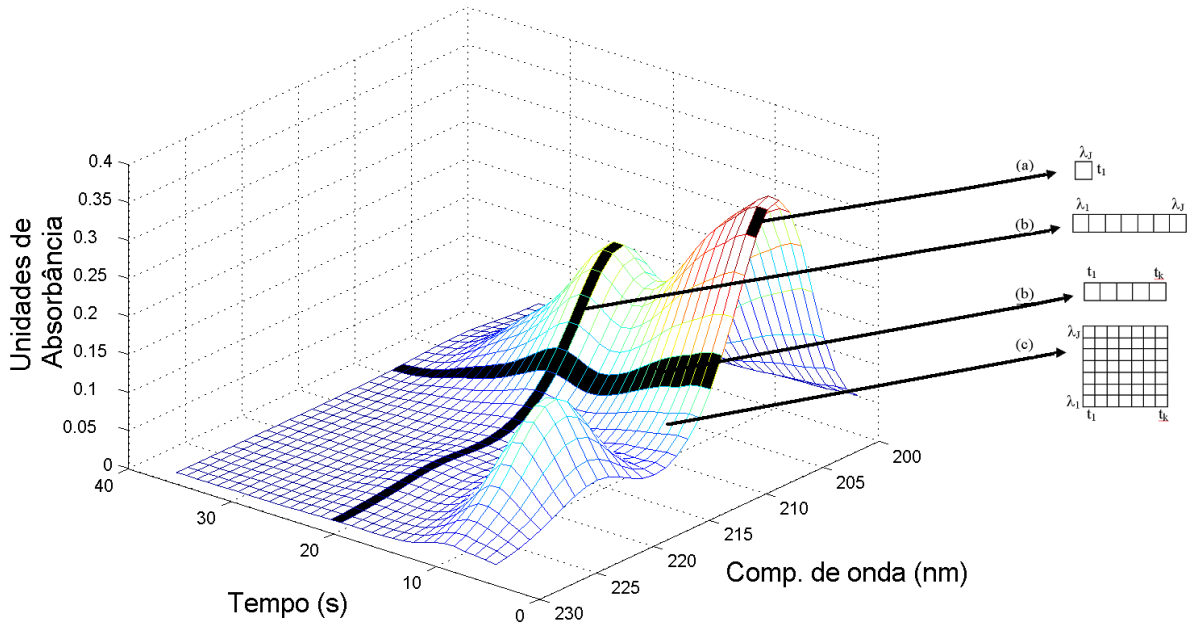


Figura 1 – Representação esquemática dos diferentes tipos de dados em um sistema HPLC-DAD: (a) escalar (uma única medida instrumental), (b) vetor de dados, (c) uma matriz de dados.

Em modelos de calibração de ordem zero apenas uma medida é feita por amostra, isto é, um único valor escalar é obtido por amostra, sendo então uma calibração univariada. Segundo o exemplo mostrado na Figura 1 ele é representado por um ponto da superfície, que corresponde à leitura da

absorbância em um comprimento de onda λ_1 em um tempo t_1. Exemplos típicos de instrumentos que geram esse tipo de dados incluem medidas eletroquímicas, espectrofotométricas ou cromatográficas, com monitoramento de um único ponto ou variável. Calibrações de ordem zero são as mais aplicadas em análises de rotina, contudo sua aplicação requer que a grandeza que é medida diretamente no sistema (ex.: absorbância de um composto) seja livre de interferentes que possam provocar desvios entre sua relação com a propriedade de interesse.

Métodos de calibração de primeira ordem fazem uso de um vetor de medidas instrumentais para cada amostra. Pelo exemplo da Figura 1, são possíveis dois tipos de dados, monitorar um comprimento de onda λ_i nos tempos de t_1 a t_k, ou monitorar os comprimentos de onda λ_1 a λ_J em um determinado tempo. Exemplos de dados multivariados incluem medidas espectrométricas e eletroquímicas em que se monitora uma determinada faixa espectral ou de variação de potencial. Esses métodos possibilitam análises mesmo na presença de interferentes (desde que esses interferentes estejam presentes nas amostras de calibração), determinações simultâneas e análises sem resolução. Diversos modelos de calibração multivariada vêm sendo utilizados, tais como: Regressão Linear Múltipla (MLR), Regressão por Componentes Principais (PCR) e Regressão por Mínimos Quadrados Parciais (PLS). Esses modelos têm apresentado ótimos resultados, com aplicações em diversas áreas, como no tratamento de dados de infravermelho próximo em análise de bebidas, madeira, polímeros, produtos farmacêuticos, agrícolas, etc [16]. Porém sua utilização requer um número grande de amostras de calibração.

Calibrações de segunda ordem são construídas para métodos que geram uma matriz de dados por amostra. Como é mostrada na Figura 1, uma matriz de dimensões j x k, que define a superfície mostrada na figura é obtida para cada amostra. Caso forem analisadas "i" amostras, teremos um tensor de dados **X** de dimensões **i x j x k**. Esse tipo de dados pode ser gerado através de diversas técnicas, como: Cromatografia Líquida de Alta Eficiência ou Gasosa com detecção por Espectrômetro de Massas (HPLC-MS e CG-MS respectivamente), Cromatografia Líquida com detecção por Arranjo de Diodos (HPLC-DAD),

Ressonância Nuclear Magnética em duas dimensões (RMN-2D), Fluorescência Molecular de Excitação e Emissão, Análise por Injeção em Fluxo com gradiente de pH e detecção por Arranjo de Diodos, dados de Microscopias e imagens em geral, etc [15]. Esses métodos têm a grande vantagem de permitir a determinação de espécies de interesse na presença de interferentes, mesmo que estes interferentes não tenham sido incluídos nas amostras de calibração. Essa característica é a chamada "vantagem da segunda ordem". Além disso, o perfil de cada composto linearmente independente, presente na amostra, pode ser estimado com dados de segunda ordem. O número de amostras requerido para a construção de modelos de calibração de segunda ordem é sensivelmente menor que aquele necessário para modelos de primeira ordem [15].

2.2. Modelos de Calibração para dados de segunda ordem

A maior parte dos modelos de calibração de segunda ordem é construída com base na decomposição do tensor de dados em um grupo de vetores base (tríades), como é mostrado na Figura 2. Essas tríades podem ser denominadas também de "fatores". Nos modelos PARAFAC e BLLS cada fator corresponde à uma fonte de sinal analítico, ou seja, está associado com uma espécie química que responda à medição instrumental ou com um sinal de origem física como o efeito Schilieren, que será explicado mais adiante. No modelo NPLS, cada fator corresponde a uma variável latente. A Figura 2 ilustra um exemplo com "k" fatores.

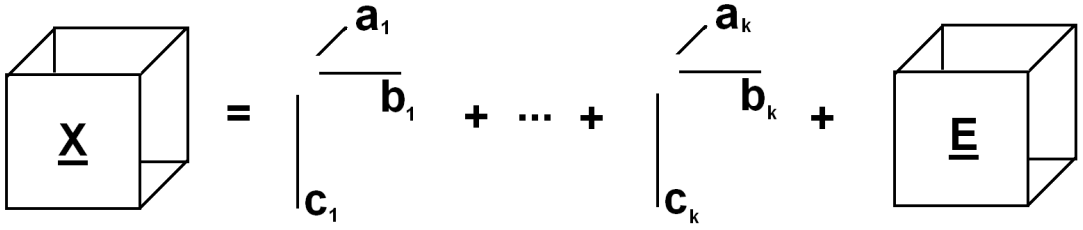


Figura 2 – Decomposição de dados tridimensionais em tríades.

2.2.1. Análise de Fatores Paralelos (PARAFAC):

Do inglês *"Parallel Factor Analysis",* desenvolvido no começo dos anos 70, este método apresenta grandes vantagens na modelagem de dados espectroscópicos, principalmente de fluorescência de excitação e emissão e dados cromatográficos, devido à estrutura desses dados que parecem se ajustar perfeitamente ao modelo. Na análise espectral de misturas, os espectros reais das espécies puras (cromatogramas, perfis cinéticos, etc.) podem ser recuperados se os dados forem de fato trilineares, o número correto de fatores for usado e a razão sinal/ruído for adequada [17].

O modelo PARAFAC para um conjunto de dados trilinear é dado por três matrizes de pesos (em inglês "loadings"), **A**, **B** e **C**. O modelo trilinear é encontrado minimizando a soma dos quadrados dos resíduos e_{ijk} utilizando o algoritmo de mínimos quadrados alternados (ALS) [17]. O modelo de decomposição pode ser representado pela figura 2 ou por :

$$\underline{\mathbf{X}}_{i,j,k} = \sum_{f=1}^{F} \mathbf{a}_{i,f}\mathbf{b}_{j,f}\mathbf{c}_{k,f} + \underline{\mathbf{E}}_{i,j,k} \qquad [1]$$

onde "F" é o número de fatores e "$\underline{\mathbf{E}}$" um arranjo cúbico que contém os erros de decomposição.

Através do modelo, o perfil de cada espécie da mistura é armazenado em um fator do modelo PARAFAC. O modelo de regressão é obtido por mínimos quadrados entre o vetor peso relacionado com a concentração e o vetor com as concentrações de referência das amostras de calibração.

2.2.2. Regressão por Mínimos Quadrados Parciais N-dimensional (N-PLS):

N-PLS, do inglês *"N-way Partial Least Squares",* é uma extensão do modelo de Mínimos Quadrados Parciais (PLS) utilizado para dados de primeira ordem [18]. O algoritmo do N-PLS decompõe os arranjos tridimensionais $\underline{\mathbf{X}}_{(i\ x\ j\ x\ k)}$ juntamente com o vetor de concentrações de referência, $\mathbf{y}_{(i\ x\ 1)}$, em um conjunto de tríades. Cada tríade é equivalente a uma variável latente em um modelo PLS e

consiste de um vetor de escores, **t**, e dois vetores de pesos, $\mathbf{w}^{j}$ e $\mathbf{w}^{k}$. A decomposição de **X** pode ser representada por:

$$\mathbf{X}_{i,j,k} = \sum_{f=1}^{F} \mathbf{t}_{i,j}\mathbf{w}^{j}_{j,f}\mathbf{w}^{k}_{k,f} + \mathbf{E}_{i,j,k} \qquad [2]$$

onde: $\mathbf{E}_{(i,j,k)}$ contém os resíduos e "f" é o número de variáveis latentes. O vetor **y** de concentrações é decomposto segundo a equação:

$$y = Tb^{T} \Rightarrow b = \left(T^{T}T\right)^{-1}T^{T}y \qquad [3]$$

onde: **T** é uma matriz de escores, cujas colunas consistem dos vetores escores individuais de cada componente e **b** são os coeficientes de regressão. A concentração de novas amostras y*, pode ser estimada a partir de novos escores, **T***, de acordo com a equação: $y^{*} = T^{*}b^{T}$.

2.2.3. Quadrados Mínimos Bilineares – Bilinearização Residual (BLLS-RBL):

Em comparação com outros modelos de calibração de segunda ordem a introdução do BLLS-RBL, do inglês "*Bilinear Least Squares"*, pode ser considerada relativamente recente [19,20], e vem demonstrando ser capaz de obter resultados analíticos comparáveis a modelos mais conhecidos, como o PARAFAC, em amostras mais complexas [21]. Foi também observado que ele possibilita modelar adequadamente sistemas com compostos de interesse que se apresentam como espécies em equilíbrio (sistemas linearmente dependentes), onde o surgimento de uma espécie é decorrente do desaparecimento de outra [22,23]. No BLLS, a concentração do analito é introduzida na etapa de decomposição, onde somente as matrizes dos padrões estão presentes, a fim de obter aproximações das matrizes dos analitos puros em concentração unitária ($\mathbf{S_n}$). Para cada amostra medida num sistema HPLC-DAD, por exemplo, é obtida uma matriz **X** de informações formada por J tempos e K comprimentos de onda. Quando todos I padrões de calibração são colocados um sobre o outro, um arranjo de três dimensões é formado, de dimensões I x J x K. Para estimar $\mathbf{S_n}$, os dados de

calibração são primeiro vetorizados e unidos numa matriz $\mathbf{V_x}$ de dimensão JK x I [24,25]:

$$\mathbf{V_x} = [\text{vec}(\mathbf{X_1}) \mid \text{vec}(\mathbf{X_2}) \mid ... \mid \text{vec}(\mathbf{X_I})] \quad [4]$$

onde "vec" indica a operação de "desdobramento", onde a matriz é vetorizada. Então é aplicado diretamente quadrados mínimos para obter a informação do analito puro, $\mathbf{V_S}$ [24,25]:

$$\mathbf{V_s} = \mathbf{V_x}\mathbf{y}^{T+} \quad [5]$$

onde os subscritos "T" e "+" correspondem à transposta e pseudoinversa, respectivamente, e **y** é o vetor de concentração de referência. Se mais que um analito está presente, o vetor **y** será uma matriz **Y** com dimensões I x N_c, onde N_c é o numero de analitos calibrados. $\mathbf{V_S}$ então contém as matrizes de interesse $\mathbf{S_n}$ na forma vetorizada:

$$\mathbf{V}_s = [vec(\mathbf{S_1}) \mid vec(\mathbf{S_2}) \mid ... \mid vec(\mathbf{S_{Nc}})] \quad [6]$$

Para obter os perfis de tempo e espectrais presentes nas matrizes $\mathbf{S_n}$, é empregado a decomposição de valores singulares (SVD, do inglês *singular value decomposition*) [19,20]. Os perfis dos componentes são obtidos através da decomposição de valores singulares dos componentes isolados de cada matriz $\mathbf{S_n}$:

$$(\mathbf{b_n}, g_n, \mathbf{c_n}) = \text{SVD}(\mathbf{S_n}) \quad [7]$$

onde g_n é o primeiro valor singular, e $\mathbf{b_n}$ e $\mathbf{c_n}$ são os primeiros vetores singulares esquerdo e direito de $\mathbf{S_n}$, respectivamente. As concentrações numa amostra desconhecida (cuja matriz de dados é $\mathbf{X_u}$) são estimadas, contanto que nenhuma interferência ocorra, pela aplicação direta de quadrados mínimos [19,20,24,25]:

$$\mathbf{y_u} = \mathbf{S}_{cal}^{+}\text{vec}(\mathbf{X_u}) \quad [8]$$

onde $\mathbf{y_u}$ é o vetor de concentração estimada 1 x N_c de N_c analitos em $\mathbf{X_u}$, e $\mathbf{S_{cal}}$ é uma matriz de calibração JK x N_c dada por:

$$\mathbf{S_{cal}} = [g_1(\mathbf{c_1} \otimes \mathbf{b_1}) \mid g_2(\mathbf{c}_2 \otimes \mathbf{b}_2) \mid ... \mid g_{Nc}(\mathbf{c}_{Nc} \otimes \mathbf{b}_{Nc})] \quad [9]$$

onde $\otimes$ indica o produto de Kronecker.

Quando os analitos calibrados produzem sinais que estão sobrepostos com o de interferentes presentes em $\mathbf{X_u}$, um processo de separação independente, chamado bilinearização residual (RBL, do inglês *residual bilinearization*) é

empregado para encontrar os perfis dos interferentes, que são incorporados numa versão expandida do $\mathbf{S_{cal}}$:

$$\mathbf{S}_{int} = [\mathbf{S_{cal}} \mid g_{int}(\mathbf{c}_{int} \otimes \mathbf{b}_{int})] \quad [10]$$

onde g_{int}, $\mathbf{b_{int}}$ e $\mathbf{c_{int}}$ são obtidos por SVD da matriz de resíduos ($\mathbf{E_u}$) calculada ao ajustar os dados da soma da contribuição de vários componentes:

$$\mathbf{E_u} = \mathbf{X_u} - \sum_{n=1}^{Nc} g_n \mathbf{b_n}(\mathbf{c_n^T})\mathbf{y_{u,n}} \quad [11]$$

$$(\mathbf{b}_{int}, g_{int}, \mathbf{c}_{int}) = SVD_1(\mathbf{E_u}) \quad [12]$$

O processo RBL pode ser realizado por um método interativo [19,24,26] ou pelo processo de minimização de Gauss-Newton [22,24]. É importante notar que no modelo BLLS nenhum procedimento de inicialização ou restrição é necessário, e que a vantagem de segunda ordem é adquirida pela análise RBL da matriz residual de $\mathbf{X_u}$. O número de interferentes presentes pode ser estimado pela comparação dos resíduos deixados pelo modelo na previsão da amostra com os resíduos nas amostras de calibração ou com o nível de ruído instrumental (obtido através de replicatas do branco).

2.3. Cálculo de erros

2.3.1. Raiz Quadrada da Média Quadrática dos Erros de Previsão (RMSEP):

Do inglês “root mean square error prediction”. O RMSEP é dado pela equação 13 onde $\hat{y}_i$ é o valor previsto, y_i é o valor real e n é o número de amostras:

$$\mathrm{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad [13]$$

2.3.2. Erro Relativo de Previsão (REP):

Do inglês "relative error prediction". O REP é dado pela equação 14 onde $\hat{y}_i$ é o valor previsto, y_i é o valor real e n é o número de amostras:

$$REP = \sqrt{\sum_{i=1}^{n}\left(\frac{(y_i - \hat{y}_i)^2}{y_i^{\,2} \times n}\right)} \times 100 \qquad [14]$$

2.4. Programas utilizados

Os programas NPLS e PARAFAC utilizados fazem parte do pacote **The *N*-way Toolbox for MATLAB** e foram obtidos gratuitamente da internet [27]. O BLLS-RBL, o GA e todos demais subprogramas utilizados, foram desenvolvidos por Jez W. B. Braga e Renato L. Carneiro.

CAPÍTULO 3

DESENVOLVIMENTO DO PROGRAMA DE ALGORITMO GENÉTICO

O GA básico envolve cinco passos: *codificação das variáveis*, *criação da população inicial*, *avaliação da resposta, cruzamento* e *mutação*. A implementação do GA na seleção de variáveis difere-se das aplicações normalmente realizadas no que tange à codificação do problema e a função de resposta, já que as outras etapas permanecem inalteradas. No caso da seleção de variáveis, considera-se que o cromossomo possui "p" genes, onde cada gene representa uma das variáveis do sinal analítico (espectro, por exemplo) sendo então o número de genes igual ao número de variáveis contidas nesse sinal.

Na seleção de variáveis utiliza-se o auxílio do código binário (0,1) para codificar o problema. Cada gene pode assumir o valor um ou zero. Quando a posição referente a uma determinada variável for igual a um, implicará na seleção desta variável, se a posição conter o valor zero, a variável não será selecionada. A figura 3 mostra a codificação de um cromossomo para a seleção de variáveis em um sistema FIA-DAD ou HPLC-DAD.

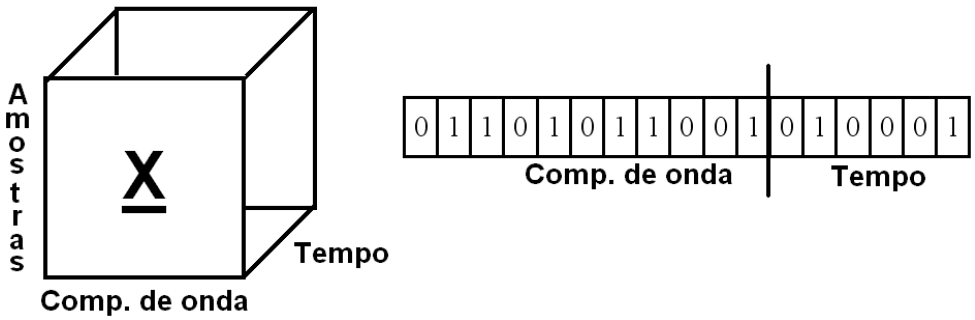


Figura 3 – Codificação de um cromossomo para seleção de variáveis em calibração de segunda ordem.

O programa GA para seleção de variáveis foi construído em ambiente MatLab 6.5, e trabalha em conjunto com as rotinas de calibração dos métodos de segunda ordem. O programa foi construído seguindo o fluxograma da figura 4:

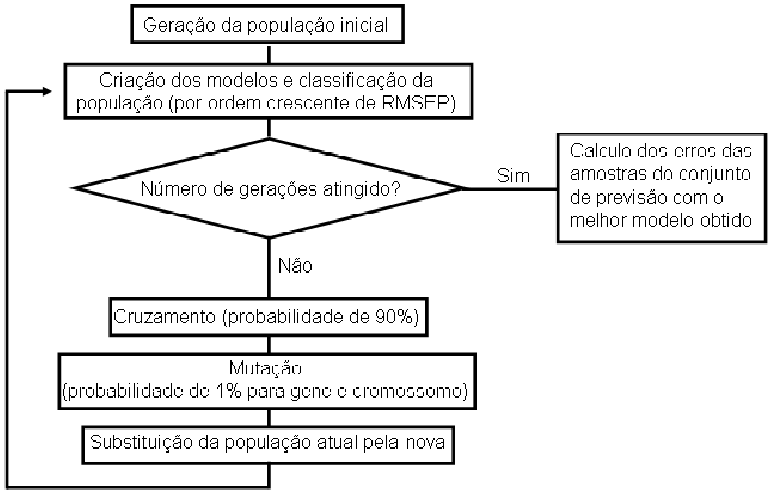


Figura 4 – Fluxograma representativo do programa de GA criado.

3.1. Geração da população inicial

A geração aleatória da população inicial é feita criando uma matriz com valores aleatórios de um ou zeros, onde cada linha é um cromossomo e cada coluna representa uma variável. Como o método é de segunda ordem, há duas espécies de variáveis, uma pertencente a cada modo da nossa calibração (perfil espectral e temporal, por exemplo), deste modo, o número de colunas dessa matriz será a soma de todas as variáveis dos dois modos (dos comprimentos de onda e dos tempos na discretização adequada, como na figura 3). Na rotina construída, podemos escolher o número de cromossomos, e também que o número de variáveis escolhidas gire em torno de um número pré-determinado. Resumindo, a primeira etapa consiste na criação de uma matriz com valores aleatórios de um e zero, onde podemos estabelecer a percentagem de “uns” que apareceram nessa matriz, ou seja, o número de variáveis selecionadas.

3.2. Criação do modelo e avaliação da população

Cada cromossomo era avaliado seguindo o procedimento a seguir. O cromossomo é divido em duas partes, a primeira parte correspondente às

variáveis do primeiro modo, e a segunda parte correspondente às variáveis do segundo modo. Desta maneira, as duas espécies de variáveis (dos dois modos) que foram unidas na geração aleatória, são agora separadas. E em todas etapas do GA (geração, cruzamento e mutação) as variáveis de ambos modos nos cromossomos, serão unidas e então separadas novamente na etapa de avaliação. Após o cromossomo ser divido em cromossomo$_{modo1}$ e cromossomo$_{modo2}$, o tensor de dados tem suas dimensões reduzidas eliminando as variáveis de acordo com o valor dos seus respectivos genes (0 ou 1) correspondentes. Supomos que tenhamos a medida de uma amostra qualquer num sistema FIA-DAD que gere gradiente de pH com o tempo, obtendo uma matriz de dados por amostra. Supondo que na matriz de dados desta amostra temos 6 variáveis no modo das linhas e 11 variáveis no modo das colunas (originando uma matriz 6 x 11), temos portanto 17 variáveis, e nosso cromossomo terá 17 genes. Sendo o modo das linhas o tempo (t_i) e o das colunas o espectro (λ_j), o processo de divisão do cromossomo é mostrado na figura 5.

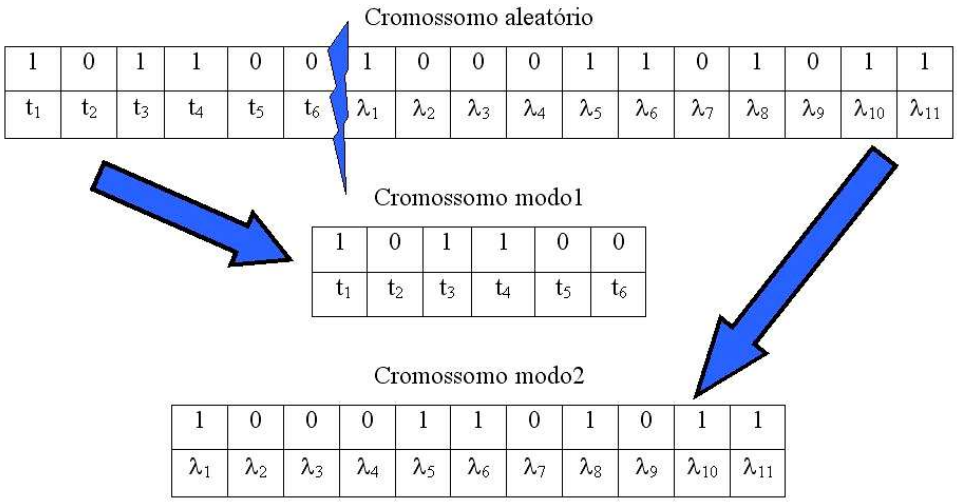


Figura 5 – Ilustração de um exemplo da divisão do cromossomo para seleção das variáveis na matriz de dados gerada por uma amostra.

A figura 6 representa a matriz de dados de uma amostra, com os espectros em função dos tempos. Os quadrados em cinza são variáveis excluídas e os

quadrados em branco são as variáveis conservadas, aplicando o cromossomo ilustrado na figura 5.

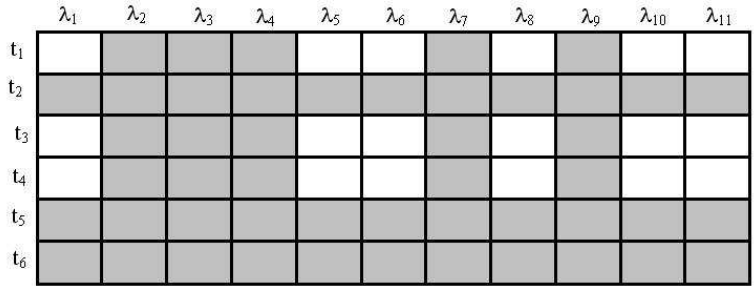


Figura 6 – Matriz de dados obtidos a partir de uma amostra onde os quadrados em cinza são variáveis excluídas e os quadrados em branco são as variáveis conservadas, aplicando o cromossomo ilustrado na figura 5.

A matriz da figura 7 representa a matriz de dados selecionada pelo GA.

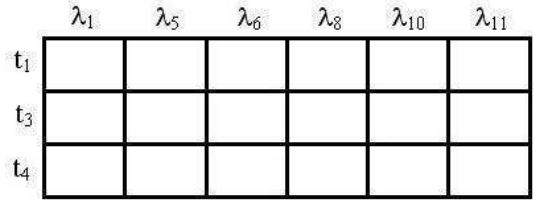


Figura 7 – Matriz com as variáveis selecionadas, obtida a partir da matriz original da figura 6, aplicando o cromossomo da figura 5.

A matriz selecionada apresentada corresponde a apenas uma amostra. Se tivermos 10 amostras, por exemplo, teremos 10 matrizes selecionadas como a apresentada (inclusive com as mesmas variáveis selecionadas, pois o que mudará de uma amostra para outra serão os valores de unidade de absorbância relacionada a cada tempo). A partir daí são construídos os modelos de segunda ordem com as matrizes selecionadas das amostras, e os erros relativos de previsão são calculados através de validação cruzada do conjunto de calibração ou através de um conjunto de validação externo.

Esse passo é repetido com cada cromossomo, e estes são então classificados em ordem crescente de erros para posterior cruzamento.

3.3. Critério de parada

Na rotina criada, a otimização termina quando o algoritmo realiza um certo número de iterações, onde cada iteração significa a substituição da população anterior por uma nova e todos os passos do GA são realizados novamente. No GA essas iterações recebem a denominação de “gerações”, pois uma população antiga dá lugar a uma mais nova, e assim como na teoria da evolução de Darwin, uma população mais nova está naturalmente mais adaptada ao seu meio devido à seleção natural que ocorre no decorrer do tempo. Desta maneira, o critério de parada da otimização é definido pelo usuário, determinando o número de gerações que deseja analisar.

3.4. Cruzamento

A etapa de cruzamento é a etapa mais importante do GA, onde dois cromossomos previamente avaliados se unem para dar origem à dois novos cromossomos. O cruzamento foi realizado seguindo uma ordem fixa, e foi baseada na classificação que estes obtiveram na etapa de avaliação. O cruzamento aleatório é evitado pois pode causar a convergência precoce da população. Quanto mais diferentes forem os cromossomos no cruzamento, melhor é, pois aumenta a heterogeneidade da população, o que, assim como num ecossistema, melhora a adaptação da população ao meio. No GA proposto, o cruzamento ocorreu da seguinte maneira:

a) Toda a programação foi feita em termos de percentagem, então para melhor entendimento suponhamos uma população de 100 cromossomos. Selecionou-se da população inicial, levando em conta a avaliação prévia dos cromossomos, uma população temporária que através do cruzamento daria origem à nova população. Esta população temporária consiste então dos 60 melhores cromossomos e mais 10 cromossomos (selecionados aleatóriamente) que estão entre os 61 e 90 melhores. Desta maneira, os 10 piores não entram na etapa de

cruzamento sendo eliminados. Foram então selecionados 70 cromossomos para o cruzamento.

b) O cruzamento entre dois cromossomos é realizado trocando em média 50% dos seus genes aleatoriamente, como ilustrado na figura 8. O vetor preto é denominado “máscara”, é gerado aleatoriamente e define quais genes serão trocados (genes com valor “1” na máscara).

c) Para dar origem à nova população, os cromossomos foram cruzados da seguinte maneira: o 1º com o 36º, o 2º com o 37º .... até ... o 35º com o 70º. Cada cruzamento dá origem à dois novos cromossomos. A etapa de cruzamento ocorre com uma probabilidade de 90 %, sendo que nos casos em que não ocorrem, os cromossomos cruzantes são preservados na nova população. Desta maneira, até aqui obtemos 70 cromossomos para a nova população, os outros 30 cromossomos são obtidos cruzando os 15 melhores cromossomos com outros 15 cromossomos escolhidos aleatoriamente dentro da população temporária. Totalizando então 100 cromossomos gerados.

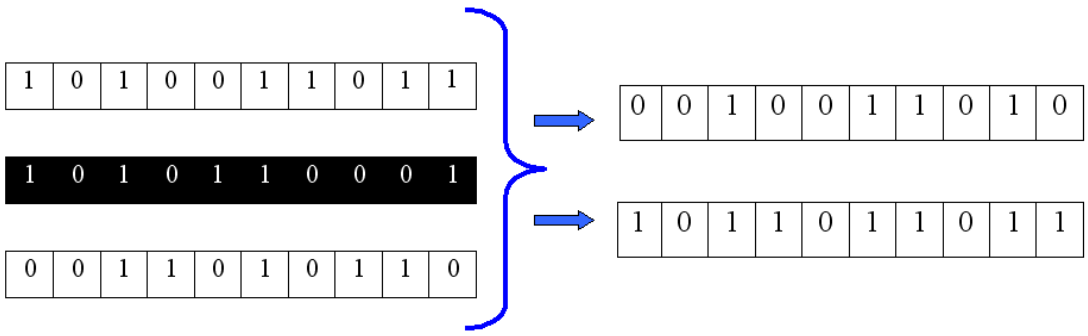


Figura 8 – Cruzamento entre dois cromossomos. Os vetores em branco representam os cromossomos que cruzam entre si, e o vetor em preto representa a máscara do cruzamento.

3.5. Mutação

A etapa de mutação ocorre logo depois dos cruzamentos. Nesta operação, uma pequena percentagem de cromossomos (ao redor de 1%) sofre mutação em alguns dos seus genes (1% dos genes em média). Isso é feito para eliminar a possibilidade de todos cromossomos possuírem um gene de mesmo valor (1 ou 0), pois neste caso, nenhum cruzamento possível poderia alterar este gene, deixando o sistema menos heterogêneo.

3.6. Substituição da população antiga pela nova

Nesta última etapa, a população originada do cruzamento e que passou posteriormente pela mutação, é agora a nova população (nova geração de cromossomos). Essa nova população é avaliada gerando dois gráficos que mostram a evolução da média dos erros e o erro do melhor cromossomo, em cada geração. Assim, todo o processo desde a avaliação do critério de parada (tópico 3.3.) é executado novamente, até que o critério de parada seja atingido.

3.7. Parâmetros do GA para a seleção de variáveis

O programa do GA foi utilizado com cada um dos métodos de segunda ordem estudados, de maneira que a rotina destes métodos era acessada como sub-rotinas do GA. Os parâmetros iniciais utilizados no GA para as otimizações estão ilustrados a seguir, e foram iguais para todos os casos:

- Percentagem de variáveis no modo 1 e no modo 2 = 23 %;
- Probabilidade de cruzamento = 90 %;
- Percentagem de cromossomos mutantes por geração = 1 %;
- Percentagem de genes que mutam nos cromossomos mutantes = 1%;
- Número de gerações = 100;
- Número de cromossomos = 100;

Os parâmetros do GA são estabelecidos empiricamente, através de testes prévios com os conjuntos de dados em questão.

Desta maneira o modelo foi iniciado utilizando 23 % das variáveis para cada modelo, porém durante as etapas do GA esse número pode aumentar ou diminuir, objetivando encontrar o cromossomo com menor erro. Este valor foi escolhido após alguns testes com o GA.

A probabilidade de cruzamento foi de 90 % pois é um número suficientemente grande para dar origem à nova população e dá uma margem de probabilidade para alguns indivíduos permanecerem no sistema sem realizarem o cruzamento.

A probabilidade de mutação foi de 1% dos cromossomos, e 1% dos genes. Essa probabilidade deve ser mínima, pois é apenas uma barreira para impedir a presença de um mesmo gene, em diferentes cromossomos, com apenas um valor em toda população. Após a mutação desse gene, ele será passado à diante nos futuros cruzamentos, diminuindo muito a probabilidade de que todos cromossomos apresentem um gene com mesmo valor. Além disso, se a mutação for muito alta, a busca acaba por se tornar aleatória.

O aumento no número de gerações pode levar a um melhor resultado, porém após certas gerações a queda do erro já não é tão expressiva, e considerando o tempo de processamento computacional gasto, optou-se como critério de parada a análise de 100 gerações.

O número de cromossomos também pode influenciar na otimização. Teoricamente, quanto mais cromossomos, mais rápido se atinge a convergência (modelo ótimo), isso porque, assim como em ecossistemas proposto por Darwin, quanto maior o número de indivíduos em uma população, maior será a probabilidade de um deles apresentar um comportamento satisfatório para sobrevivência no seu ambiente. Porém, um número muito alto de cromossomos poderia levar a um grande consumo de tempo de processamento. A opção de 100 cromossomos mostrou-se mais interessante.

A medida de erro utilizada para otimização do GA foi o RMSEP de duas amostras de previsão que não foram utilizadas na calibração, nas quais se conhecia a concentração dos analitos, obtida pelo método de referência.

Então para que o GA selecione ou não variáveis nas quais a interferência está presente, ele necessita destas amostras de previsão. Importante notar que as amostras de previsão são utilizadas somente no cálculo do RMSEP, e em nenhum momento são inseridas na construção dos modelos de segunda ordem, e que também a concentração ou o número de interferentes não necessitam ser conhecidos nessas amostras.

CAPÍTULO 4

APLICAÇÃO 1: DETERMINAÇÃO DE PESTICIDAS EM VINHO POR HPLC-DAD.

4.1. Aquisição de dados de segunda ordem por HPLC-DAD

Em um sistema HPLC-DAD os dados obtidos são de segunda ordem pois a absorbância varia tanto em função do tempo de eluição (perfil temporal) como em função do comprimento de onda (perfil espectral). A cada intervalo de tempo apropriado após a injeção, o detector DAD registra um espectro na região UV-Vis, desta maneira, cada espectro estará relacionado com um tempo de eluição. Os dados para cada amostra são então bidimensionais (na forma de uma matriz de dados) e um exemplo é mostrado na figura 9:

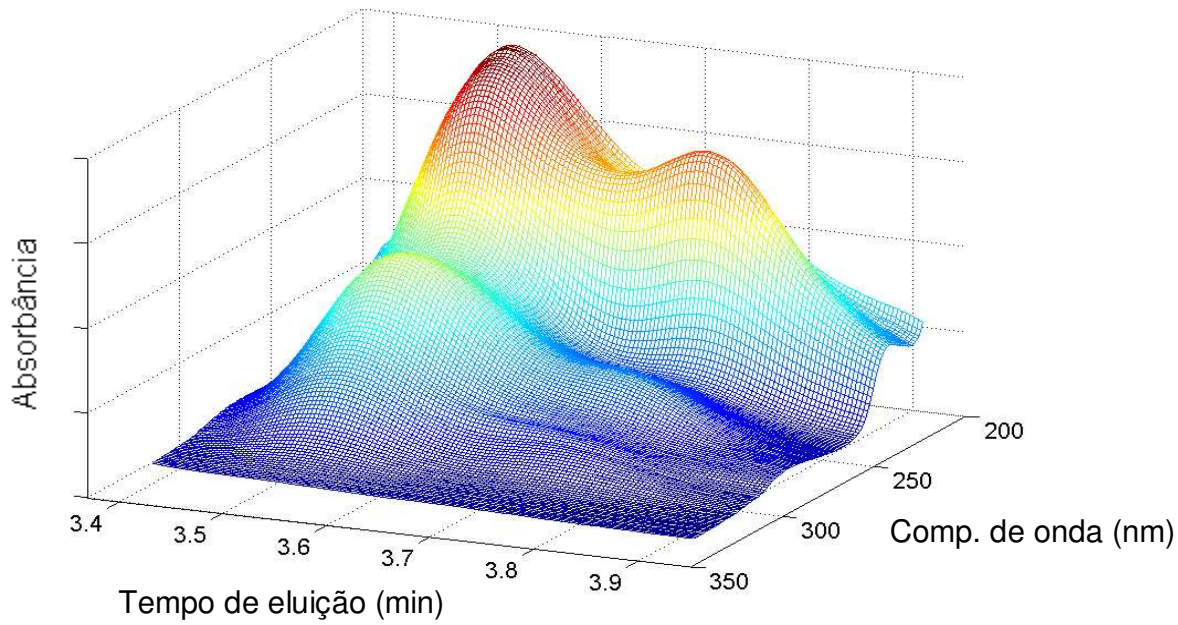


Figura 9 – Superfície de uma amostra de vinho contaminada para o tempo de eluição dos pesticidas tiofanato metílico e simazina.

Na figura 9, pela análise do perfil temporal, pode-se observar a existência de uma sobreposição de compostos. O fato do detector ser de arranjo de diodos possibilita a separação quimiométrica através da utilização de modelos de calibração de segunda ordem.

4.2. Procedimento experimental

Os procedimentos experimentais foram realizados por Braga *et al.* [28].

4.2.1. Amostras de calibração

Os solventes utilizados para a preparação da análise cromatográfica foram acetonitrila (grau HPLC,Tedia), água (Milli-Q, Millipore), ácido fosfórico (Merck), acetato de etila (Tedia), metanol (grau HPLC,Tedia) e isopropanol (Merck). Eles foram filtrados usando uma membrana 0,45 μm de fluoreto de poli(vinilideno) (PVDF) (Millipore). Os padrões dos pesticidas foram simazina (SIM) (98,3%) da Novartis, carbaril (CBL)(99,8%) da Supelco, tiofanato metílico (TIO) (98,5%) e dimetoato (DMT) da Riedel-de-Häen. O metabolito ftalimida (PTA) (99,9%) era da Riedel-de-Häen. O metabólito PTA é o produto gerado da degradação do pesticida Folpet.

As soluções estoque de cada analito foram preparadas com acetonitrila nas seguintes concentrações: 1046 μg mL^{-1} para PTA, 1077 μg mL^{-1} para CBL, 1028 μg mL^{-1} para TIO, 1011 μg mL^{-1} para DMT e 402 μg mL^{-1} para SIM. Soluções intermediárias para cada analito foram obtidas a partir da diluição apropriada das soluções estoque de cada analito utilizando 50:50 (v/v) acetonitrila:água, obtendo assim as seguintes concentrações para as soluções intermediárias: 20,15 μg mL^{-1} para PTA, 105,7 μg mL^{-1} para CBL, 39,63 μg mL^{-1} para TIO, 99,21 μg mL^{-1} para DMT e 19,40 μg mL^{-1} para SIM. Para os analitos PTA, TIO e SIM foram realizadas duas diluições. Estas soluções foram guardadas a 4 ºC na ausência de luz. Na construção do modelo, seis padrões de calibração contendo todas as misturas de interesse foram preparadas cobrindo a faixa analítica apresentada na tabela 1 e com concentrações distribuídas igualmente. As faixas de concentração foram utilizadas baseado em injeções preliminares no HPLC, obtendo a área do pico cromatográfico usando o método isocrático (IM) com detecção a 220 nm, ou seguindo as recomendações do *Codex Alimentarius* para limites máximos de residuos permitidos [29]

Tabela 1. Faixa de concentração, em μg mL^{-1}, usada para os analitos de interesse.

Analito	Faixa de concentração μg mL^{-1}
DMT	1,00 – 7,50
PTA	0,10 – 1,40
TIO	0,50 – 5,37
SIM	0,10 – 1,24
CBL	1,00 – 6,00

4.2.2. Amostras de vinho

As amostras utilizadas foram de um vinho tinto comercial fabricado no Rio Grande do Sul. Este vinho foi obtido de uvas provenientes de vinhas onde não houve a utilização de pesticidas sintéticos (processo "orgânico" de produção). Cada amostra foi submetida a um procedimento de extração em fase sólida (SPE) para limpeza da amostra. O método SPE empregou cartuchos 1,00 mL Oasis HLB que foram primeiramente condicionados com 2,50 mL de metanol e 2,50 mL de água. Então 2,50 mL de vinho foram adicionados ao cartucho, permitindo assim o escoamento. Os cartuchos eram lavados com 1,50 mL de uma solução de isopropanol 2% (v/v) e secos por 20 minutos. O extrato que havia sido retido no cartucho, foi eluído diretamente com 3,00 mL de acetato de etila para um cartucho com Florisil construído no próprio laboratório, permitindo o escoamento através dos cartuchos sob pressão, e o extrato era então coletado num tubo de ensaio. O solvente foi evaporado por secagem sobre um fluxo de nitrogênio à temperatura ambiente. A amostra seca foi redissolvida com 1,00 mL de acetonitrila, obtendo um fator de concentração de 2,5 vezes. Finalmente, a solução foi transferida para um *vial* para análise. Foram realizadas adições em seis extratos com os compostos de interesse, e estes foram analisados por um processo isocrático (sem utilização de gradiente na fase móvel). Para melhor avaliação da previsão dos erros dos modelos de segunda ordem, os analitos foram adicionados nos extratos após o SPE, desta maneira nenhuma perda no SPE foi considerada nas seis amostras. Um volume de 100 μL foi usado para adição dos analitos nos

extratos, assim os interferentes nestas amostras foram diluidos por 10 %. Todas amostras e padrões foram analisados em duplicata.

4.3. Obtenção dos dados por HPLC – DAD

O sistema HPLC utilizado foi um Cromatógrafo Shimadzu VP Series Liquid equipado com um amostrador automático SIL-10AXL, uma bomba de solvente modelo LC-10ATVP e um detector SPD 10AVP DAD. Para a separação foi utilizada uma coluna Anovapack C18 (4μm) (150mm x 4.6 mm i.d.) da Waters conjuntamente com uma coluna de guarda similar. No método isocrático (IM), a separação foi conduzida utilizando 50:50 (v/v) acetonitrila:água como fáse móvel, a água foi acidificada para pH 3.0 com ácido fosfórico antes da mistura, o fluxo usado foi de 0,60 mL min^{-1}. Foi utilizado um tempo total de 12 minutos nos padrões e 25 minutos nas amostras dos extratos (tempo necessário para limpeza da coluna devido a outros componentes presentes no extrato).

4.4. Resultados e discussão

A figura 10 (cromatograma) foi obtida monitorando os tempos de eluição entre 2 e 5 minutos no comprimento de onda 220 nm. Nos seis padrões de calibração DMT e PTA estão muito sobrepostos apresentando apenas um pico, em aproximadamente 2,7 minutos. Para TIO e SIM uma média sobreposição é observada, enquanto que para o CBL o pico está resolvido. Pode ser observado também que quando uma amostra de vinho é analisada, existe sobreposição de interferentes sobre DMT, PTA, TIO, SIM e CBL, gerando três situações distintas onde a utilização de um método de calibração de segunda ordem foi necessária. Pode ser observado também na figura 10 que a amostra de vinho apresenta um deslocamento no tempo quando comparado com os padrões de calibração. Este tipo de desvio deve ser corrigido antes de desenvolver o modelo, pois produz desvios da estrutura bilinear da matriz de dados. Os deslocamentos de tempo foram corrigidos baseados no procedimento proposto por Prazen *et al* [30],

baseado na minimização dos erros obtidos por SVD das matrizes deslocadas e das matrizes dos padrões, onde não ocorre o deslocamento.

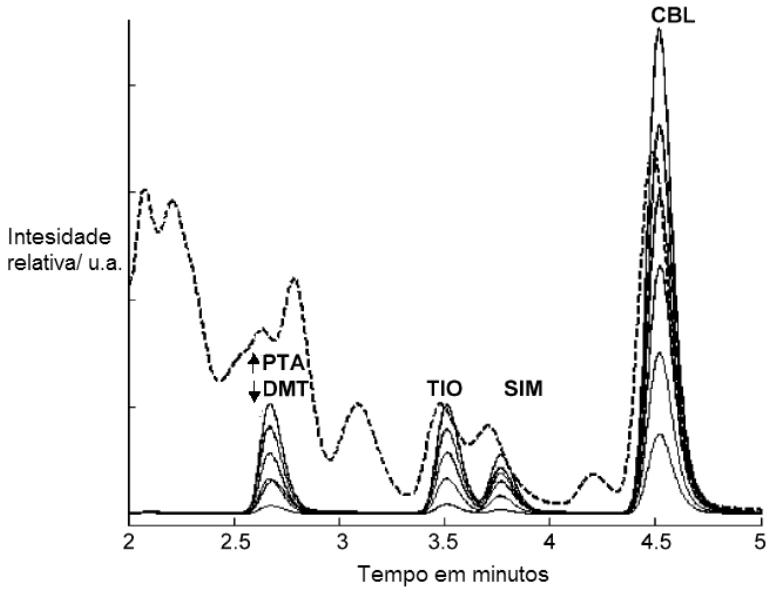


Figura 10 – Cromatograma na região dos compostos de interesse. As linhas sólidas são os padrões de calibração, e a linha tracejada é o perfil de uma amostra de vinho contaminada com os pesticidas. O cromatograma foi obtido monitorando o comprimento de onda de 220 nm.

Cada um dos casos descritos foi analisado independentemente, resultando então em três situações distintas de sobreposição, sendo elas, SIM e TIO com interferente, DMT e PTA com interferente, e CBL com interferente. A figura 11 apresenta os perfis espectrais dos analitos, recuperados pelo modelo PARAFAC a partir dos conjuntos de calibração das três distintas situações.

A tabela 2 traz uma comparação entre os resultados obtidos com o BLLS-RBL e com a utilização conjunta com o GA para os compostos CBL, TIO, SIM, DMT e PTA.

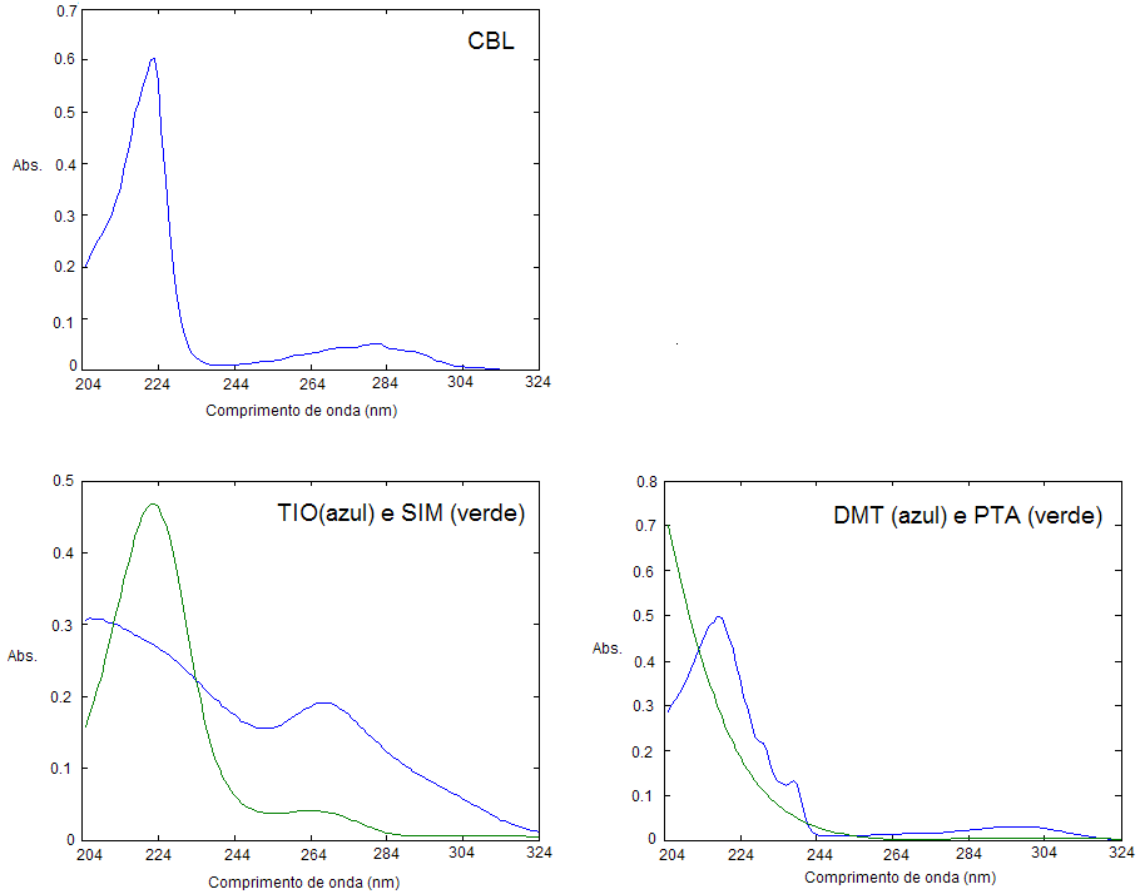


Figura 11 – Perfis espectrais dos analitos recuperados do conjunto de calibração pelo modelo PARAFAC.

Somente o BLLS-RBL foi utilizado neste caso, pois os erros obtidos pelo PARAFAC e NPLS foram muito elevados, mesmo com a utilização do GA. Uma possível explicação é a não total trilinearidade dos dados, decorrente do deslocamento do tempo de retenção dos analitos nas amostras problema em relação ao conjunto de calibração, fato que pode ser contornado pelo BLLS-RBL.

Na tabela 2, o RMSEP, o REP e os valores de recuperação são apresentados comparando o modelo com todas as variáveis, e o modelo construído utilizando as variáveis selecionadas pelo GA. Para SIM, TIO, PTA e DMT, foi observado que as amostras com baixas concentrações apresentam um erro maior que das outras amostras. Uma comparação dos cromatogramas a comprimentos de ondas fixos, destas amostras com as outras, mostrou que o

deslocamento de tempo para as amostras de baixa concentração não foi corrigido devido às baixas concentrações e à presença de interferentes. Estas amostras foram consideradas *outliers*. Assim os valores apresentados na tabela 2 foram calculados utilizando as seis amostras somente para o CBL, e para os outros casos foram utilizadas as cinco amostras de maior concentração, rejeitando a de menor concentração. Pode ser observado que os erros, dados por RMSEP e REP, assim como o desvio padrão, foram reduzidos quando o GA foi utilizado no desenvolvimento do modelo.

Tabela 2. Comparação das recuperações e erros obtidos utilizando ou não o GA.

Pesticida	BLLS-RBL		BLLS-RBL-GA	
	Recuperações[a]	RMSEP[b]	Recuperações[a]	RMSEP[b]
CBL	98 (±2)	0,09 (2)	101 (±1)	0,09 (2)
SIM	92(±2)	0,08 (7)	99 (±2)	0,02 (2)
TIO	98 (±2)	0,13 (3)	100 (±2)	0,07 (1)
PTA	105 (±6)	0,08 (7)	102 (±3)	0,03 (3)
DMT	101 (±5)	0,27 (5)	102 (±3)	0,19 (3)

[a] Valores de recuperação em porcentagem com estimativas de desvio padrão entre parênteses.

[b] RMSEP em $\mu g\ mL^{-1}$ e REP entre parênteses.

Para estes dados, o GA foi executado de maneira que ao invés de se trabalhar com as variáveis isoladas, elas foram utilizadas em blocos de cinco (um gene para cada cinco variáveis), como pode ser visualizado na seleção das variáveis mostrada na figura 12.

A figura 12 apresenta as superfícies de uma amostra de vinho, em cada uma das regiões de eluição de interesse, na forma de linhas de contorno, onde as variáveis selecionadas pelo GA são mostradas. Para estas cinco otimizações, o GA excluiu do modelo aproximadamente 80, 90, 85, 95 e 80% das variáveis, para CBL, SIM, TIO, PTA e DMT, respectivamente.

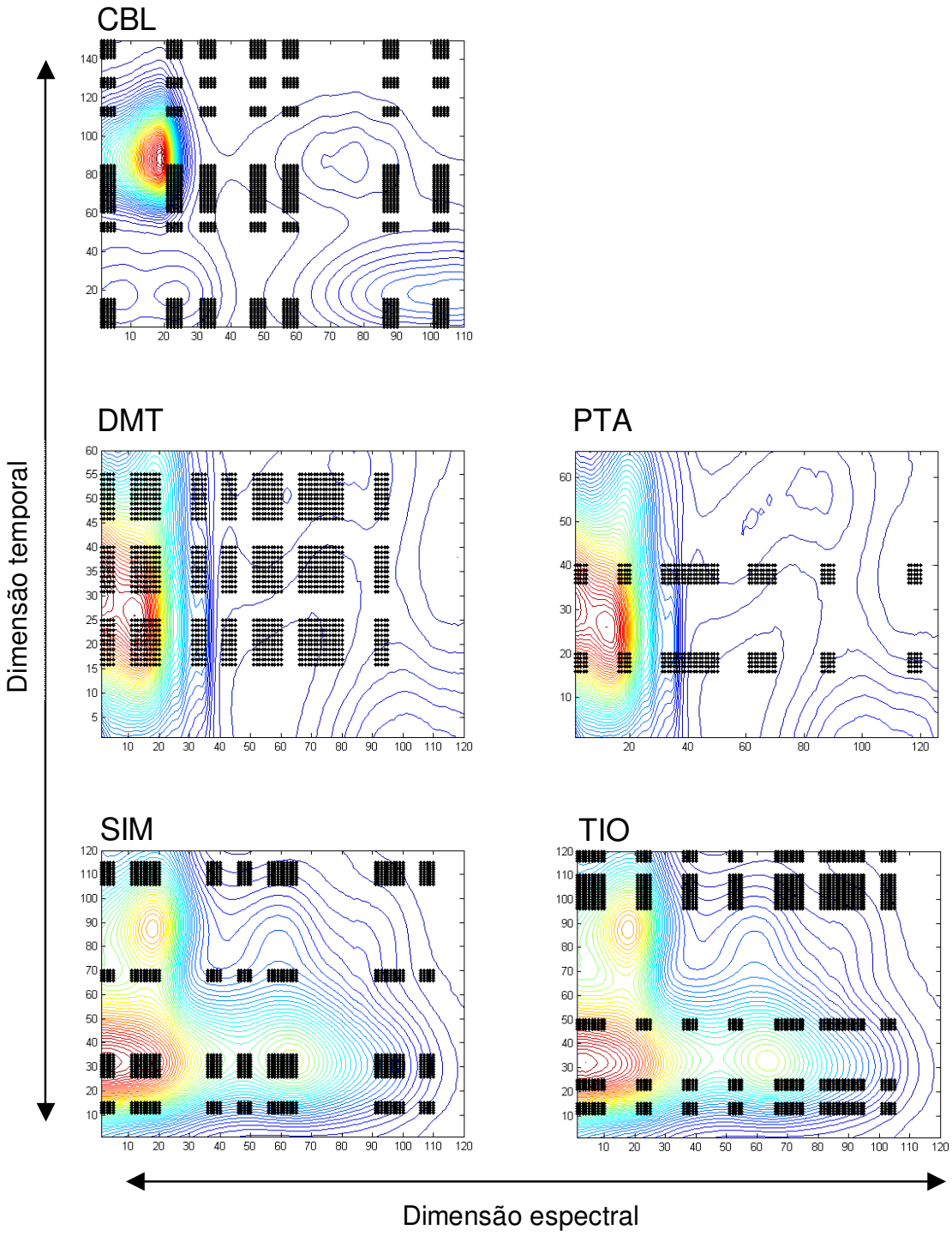


Figura 12 – Gráficos de curvas de nível dos casos estudados onde os pontos em preto indicam as variáveis escolhidas pelos melhores cromossomos para a determinação de cada analito, nas três situações de sobreposição em questão.

A interpretação das variáveis selecionadas nos casos apresentados é muito complexa, pois na busca de um modelo de previsão eficiente, o GA também tem que selecionar variáveis representativas do interferente, no caso da utilização da vantagem de segunda ordem. Além disso, o processo de calibração é um algoritmo matemático que, em busca do melhor modelo, pode selecionar variáveis sem um sentido químico relevante, mas que na construção do modelo possibilitam a obtenção de melhores resultados.

Ainda assim, ao se analisar a seleção para SIM, TIO, DMT e PTA, pode-se visualizar a tendência da seleção de variáveis sobre os máximos das curvas. Além disso também é possível observar uma baixa densidade de variáveis selecionadas no final do espectro ultravioleta, para CBL, DMT e PTA, pois o espectro destas substâncias nesta região contém pouca informação química, como mostram os espectros na figura 11.

A figura 13 mostra um exemplo do gráfico de saída do programa GA, apresentando a evolução do RMSEP a cada geração para o caso da SIM utilizando o BLLS-RBL.

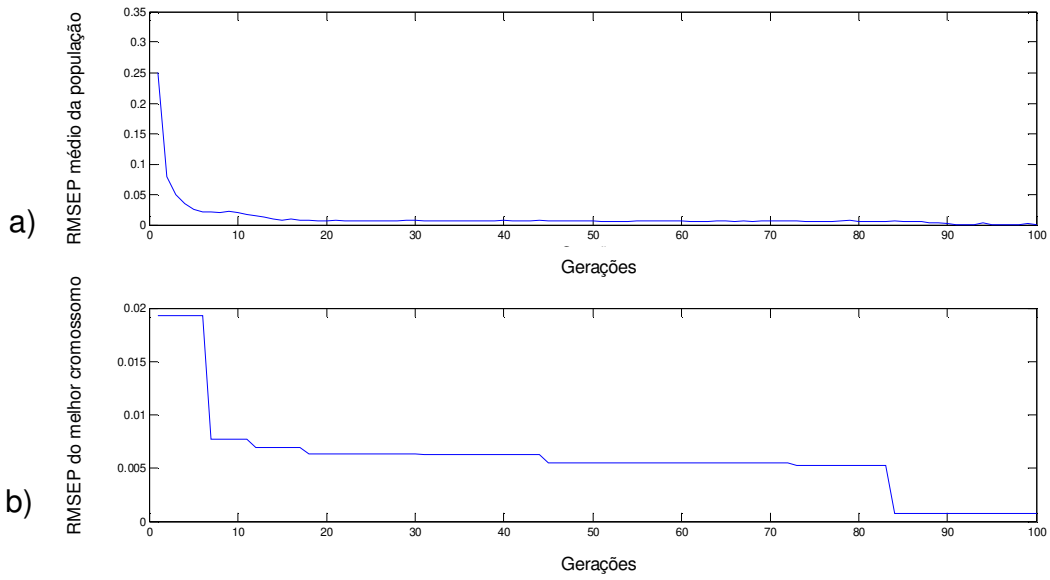


Figura 13 – Evolução do RMSEP utilizando o GA na determinação de SIM utilizando o BLLS-RBL: a) média de todos os cromossomos da população a cada geração e; b) melhor cromossomo a cada geração.

CAPÍTULO 5

APLICAÇÃO 2: DETERMINAÇÃO DE RIBOFLAVINA E PIRIDOXINA EM PREPARAÇÕES LÁCTEAS POR FLUORESCÊNCIA MOLECULAR DE EXCITAÇÃO/EMISSÃO

5.1. Aquisição de dados de segunda ordem por Fluorescência molecular de Excitação-Emissão

A fluorimetria é uma técnica analítica consagrada que fornece uma grande sensibilidade e seletividade analítica, muitas vezes empregada na análise de sistemas em que os analitos estão presentes em concentração nano ou picomolares [31].

Matrizes de fluorescência de excitação-emissão ou do inglês *Excitation-Emission Matrix* (EEM), são geradas pela obtenção de espectros de excitação e espectros de emissão de uma amostra. Os espectros de excitação são produzidos pela medida da intensidade de luminescência mantendo-se constante o comprimento de onda de emissão e varrendo-se o de excitação. Espectros de emissão são obtidos de forma contrária, mas através do mesmo princípio, mantendo-se a excitação constante e varrendo-se o modo de emissão. Desta maneira, matrizes de excitação-emissão são geradas por duas dimensões independentes de comprimentos de onda, onde uma destas dimensões caracteriza-se pelos perfis de excitação e outra, possuindo informação referente ao espectro de emissão. A projeção destes espectros gera uma superfície tridimensional de fluorescência total [32]. É apresentado na figura 14 um espectro de emissão-excitação para uma amostra de vitamina B2:

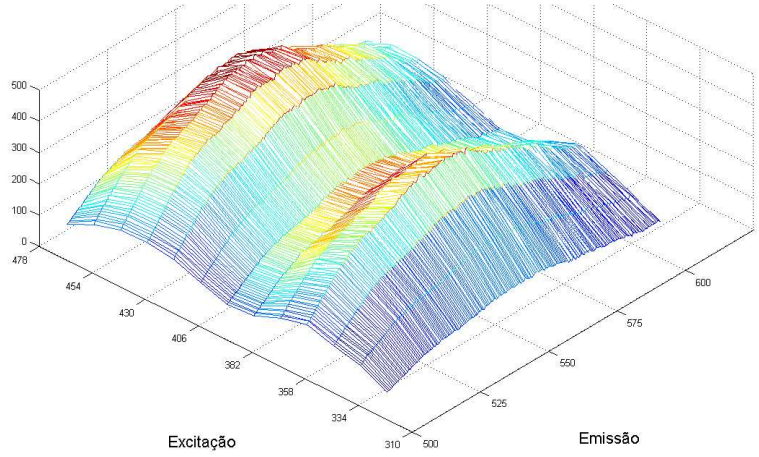


Figura 14 – Superfície de excitação/emissão de fluorescência obtida para uma solução padrão de riboflavina.

5.2. Procedimento experimental

A escolha das amostras de preparações lácteas foi feita verificando no rótulo se havia tanto vitamina B2 (riboflavina) como B6 (piridoxina), e também se o produto era solúvel em água, a fim de facilitar as análises.

O conjunto de calibração foi construído com cinco pontos medidos em triplicata, totalizando 15 medidas. Foram realizadas seis medidas (três do produto sem adição, mais três com diferentes níveis de adição das vitaminas B2 e B6) para cada produto, em triplicata, sendo no total 18 medidas para cada produto.

O ácido tricloroacético foi utilizado para a extração das vitaminas e precipitação das proteínas nos produtos analisados [33].

5.2.1. Padrões

1. Dissolveu-se a riboflavina (98,4%, doação) com 5,00 mL de HCl (Merck, P.A.) concentrado e a piridoxina (99,5%, doação) com água deionizada. As soluções foram transferidas para um mesmo balão volumétrico e o volume foi completado obtendo-se uma solução estoque de 50,0 mg/L para cada um dos analitos.

2. Foi preparada uma solução intermediária de concentração 2,50 mg/L transferindo-se 5,00 mL da solução estoque para um balão de 100,0 mL e o volume completado com água deionizada.

3. A solução de trabalho de 0,250 mg/L foi preparada transferindo-se 10,00 mL da solução intermediária para um balão de 100,0 mL e completando o volume com tampão $HPO_4^{-2}/H_2PO_4^-$ de pH =7,0, preparado com Na_2HPO_4 (Synth, 99,0%) e HCl (Merck, P.A.).

4. Os padrões de calibração foram obtidos transferindo os volumes da tabela 3, correspondentes à solução de trabalho para balões de 100,0 mL, adicionando 1 mL de uma solução de ácido tricloro ácetico 33%(m/m),

(Synth, 99,0%), 0,925 mL de solução de NaOH de aproximadamente 5 mol/L (Synth, 97,0%) e completando com solução tampão pH=7,0 ($H_2PO_4^-$ /HPO_4^{2-}).

A tabela 3 ilustra as concentrações dos pontos utilizados na curva de calibração, tanto para a riboflavina quanto para piridoxina.

Tabela 3. Concentrações dos pontos utilizados na curva de calibração, tanto para a riboflavina quanto para piridoxina.

Padrão	Vol. sol. Trabalho (mL)	Concentração(µg/L) de B2 e B6
1	2	5,00
2	8	20,0
3	15	37,5
4	20	49,5
5	28	70,0

5.2.2. Amostras

1. Foram preparadas soluções de 100 mL dos produtos APL, BPL, CPL, DPL e EPL em triplicata, com as massas ilustradas na tabela 4, que foram escolhidas baseadas no teor nominal de riboflavina e piridoxina presentes em cada produto:

Tabela 4. Produtos utilizados, e massas utilizadas para preparar 100 ml de solução.

Massas médias (g), para preparar 100mL de solução de cada produto				
APL	BPL	CPL	DPL	EPL
12	7	11	3,5	4

2. Para cada produto, foram transferidos 40 mL (produtos APL, BPL e CPL) ou 45 mL (produtos DPL e EPL) das soluções preparadas anteriormente, para balões de 50 mL, e o volume destes foi completado com a solução de TCA.

3. Centrifugou-se por 5 minutos (2400 rpm).

4. Filtraram-se os precipitados com papel filtro comum.

5. Diluiu-se 10 vezes a solução filtrada, levando 5 mL desta a um balão de 50 mL completando com solução Na_2HPO_4 0,12 mol/L e corrigindo o pH para aproximadamente 7 utilizando uma soluçao de NaOH 5 mol/L.

6. 25 mL das soluções de APL, BPL, CPL, DPL e EPL, foram transferidas para um balão de 50mL e o volume completado com tampão $HPO_4^{-2}/H_2PO_4^-$ (pH=7,0). Para amostras com adição das vitaminas, foram adicionados também 1mL, 2mL ou 3mL de uma solução 0,250 µg/L de ambos analitos para as adições 1, 2 e 3, respectivamente antes de completar o volume com tampão $HPO_4^{-2}/H_2PO_4^-$ (pH=7,0).

A tabela 5 ilustra a concentração final de cada uma das soluções analisadas para cada produto. O valor baseia-se no teor encontrado pelo método de HPLC descrito a seguir, para cada produto, mais as adições. As adições foram realizadas em 3 níveis, acrescentando 5, 10 e 15 µg/L, em cada uma das concentrações abaixo.

Tabela 5. Concentração final em cada uma das soluções, corrigida pelo valor obtido pelo método de comparação.

Concentração das soluções analisadas em µg/L									
APL		BPL		CPL		DPL		EPL	
B2	B6	B2	B6	B2	B2	B2	B6	B2	B6
48,0	15,0	41,1	10,1	45,8	45,8	41,1	20,8	45,8	39,2

5.3. Obtenção dos dados

Foi utilizado um espectrofluorímetro PerkinElmer LS 55 para obtenção dos dados com a varredura iniciando a 250 nm, com mais 19 passos de 12 nm chegando à 478 nm na última varredura. O espectro de emissão foi monitorado de 300 a 600 nm com resolução de 0,5 nm. As fendas de emissão e de excitação foram mantidas em 10 nm. Utilizou-se um banho termostatizado para manter as amostras à aproximadamente 24ºC, que era a temperatura mais próxima à ambiente.

5.4. Método de comparação

Foi utilizado o cromatógrafo Shimadzo Prominence High Performance Liquid Chromatography equipado com um detector APD-M20A de arranjo de diodos, uma amostrador automático e uma coluna Microsorb MV C18 5 µm (250 mm x 4.6 mm) da Varian. A fase móvel utilizada foi metanol:agua (40:60) (v/v) com dodecilsulfato de sódio na concentração 5 x 10^{-3} mol/L, o pH desta solução foi ajustado para 3. Foi utilizada uma vazão de 1mL min^{-1} e um tempo de análise de 25 minutos.

5.4.1. Preparação dos padrões

Foi preparada uma solução padrão estoque de 50 mg/L de ambas as vitaminas, através da dissolução de 25,00 mg destas em água deionizada, num balão de 500 mL. Antes de completar a dissolução, foram adicionados 2,0 mL de HCl concentrado para auxiliar na solubilização da vitamina B2. A curva de calibração para ambas as vitaminas consistiu de seis pontos com concentrações variando de 0,250 a 7,500 mg/L. Os frascos foram envoltos em papel alumínio e armazenados no refrigerador até a análise.

5.4.2. Preparação das amostras dos produtos

Foram medidos 20,00g de cada produto e adicionados em balões volumétricos de 100 mL com um pouco de água deionizada (sem completar o volume do balão), estes balões foram deixados no ultrassom por 30 minutos. Nos balões foram acrescentados 20 mL de uma solução de ácido tricloroacético 33,3%, e o volume dos balões foram então completados com água deionizada e deixados no ultrassom por mais 30 minutos. As soluções foram então centrifugadas por 5 minutos a 2000 rpm e filtradas em papel comum. Os filtrados foram recolhidos em frascos escuros e foram armazenados no refrigerador até a análise.

5.5. Resultados e Discussão

Como pode ser verificado nas figuras 15 e 16, os espectros de emissão da piridoxina e da riboflavina não se sobrepõem e desta maneira, pode-se construir um conjunto de calibração com os dois analitos e obter suas superfícies de emissão separadas. Antes do tratamento dos dados, foi selecionada uma faixa de emissão onde não havia espalhamento de radiação proveniente da fonte do espectrofluorímetro. As faixas de excitação e emissão utilizados para tratar os dados obtidos para a riboflavina e piridoxina são os que ilustram as figuras 15 e 16 respectivamente. A seguir, nas figuras 17 a 20, são apresentados os gráficos de valor real contra previsto, utilizando o conjunto de calibração, modelado com o PARAFAC.

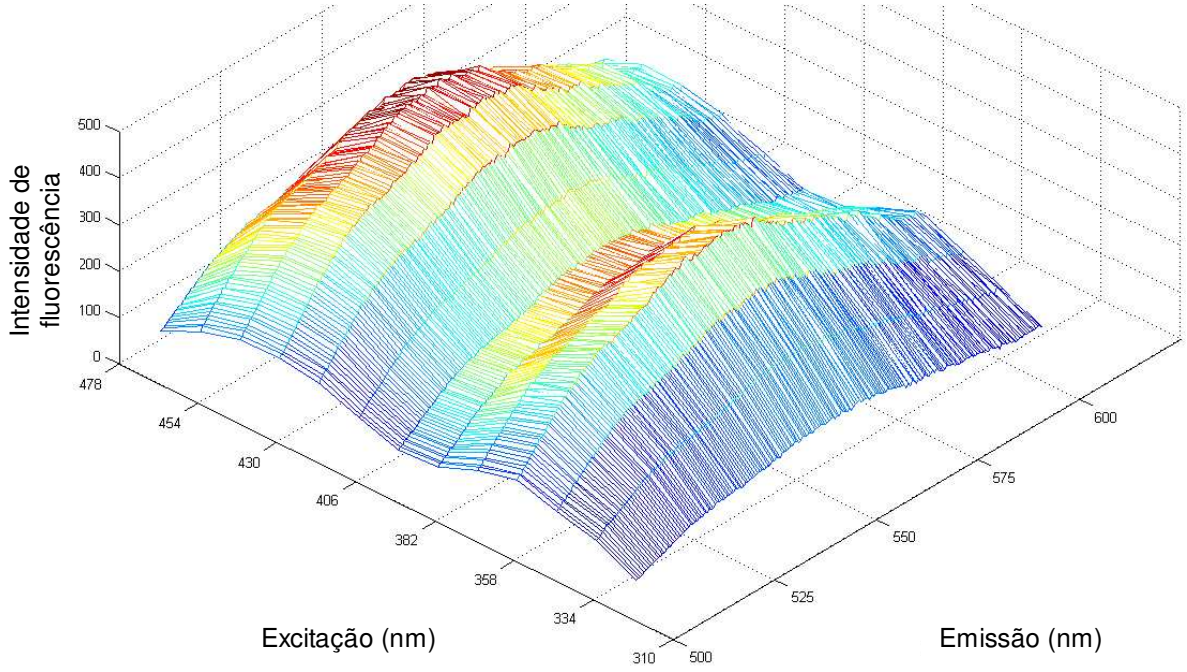


Figura 15 – Superfície obtida de um padrão do conjunto de calibração para a riboflavina na concentração de 70 μg/L.

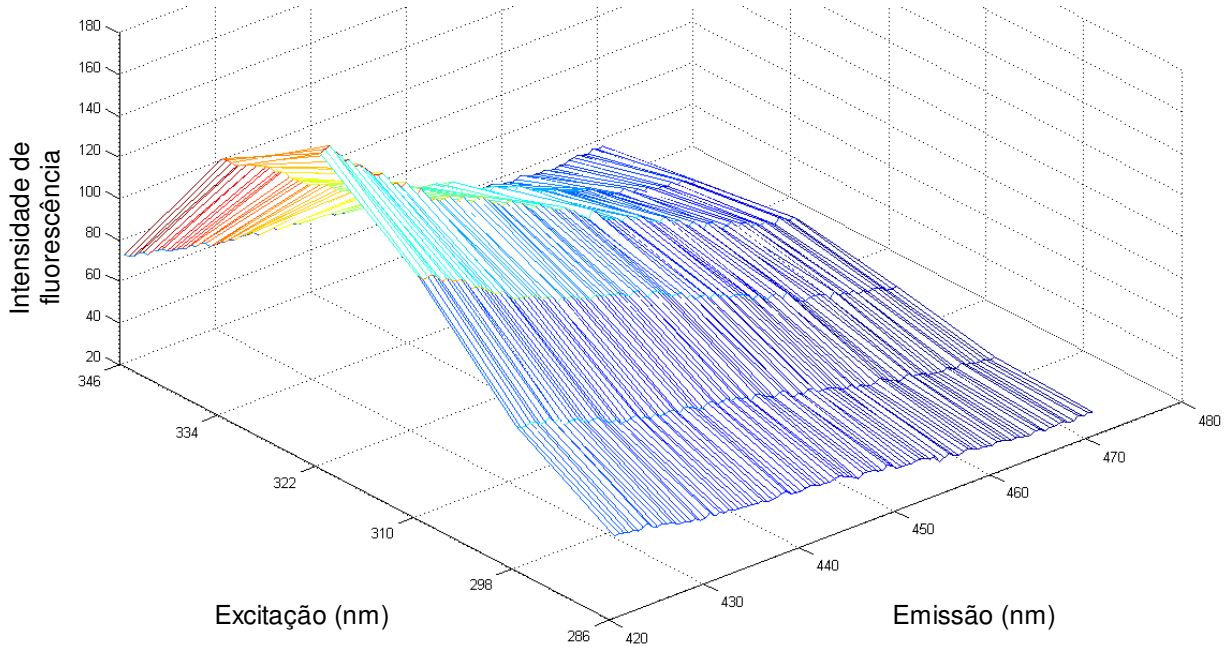


Figura 16 – Superfície obtida de um padrão do conjunto de calibração para a piridoxina na concentração de 70 μg/L.

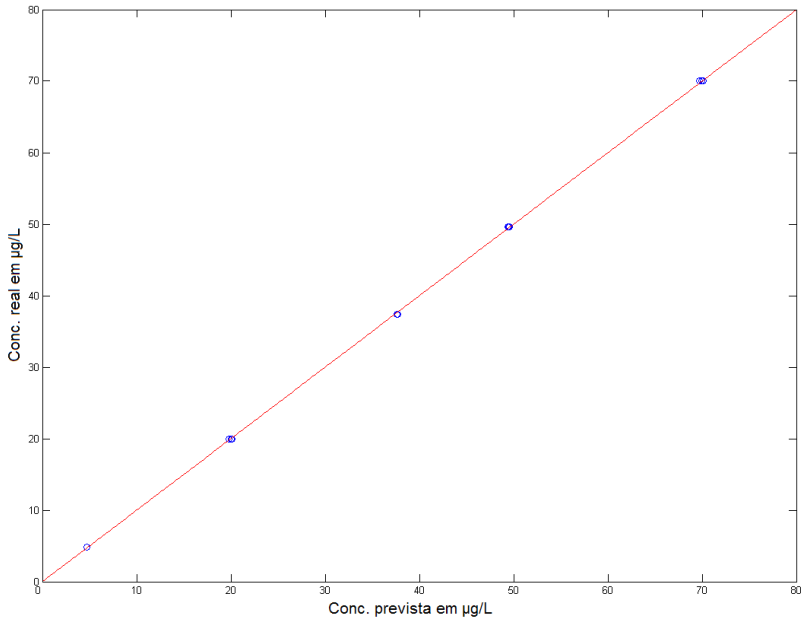


Figura 17 – Valor real contra previsto, utilizando o conjunto de calibração, modelado com o PARAFAC e um fator para riboflavina.

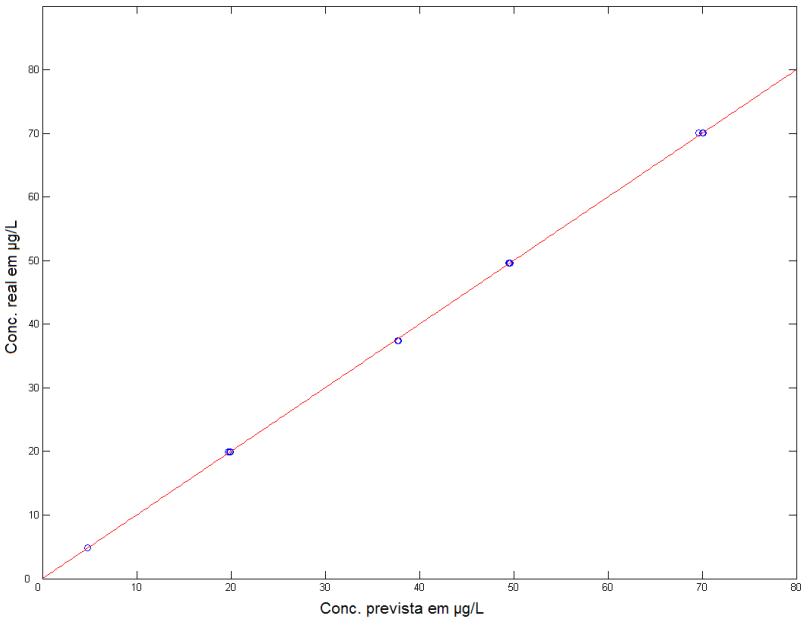


Figura 18 – Valor real contra previsto, utilizando o conjunto de calibração, modelado com o PARAFAC e dois fatores para riboflavina.

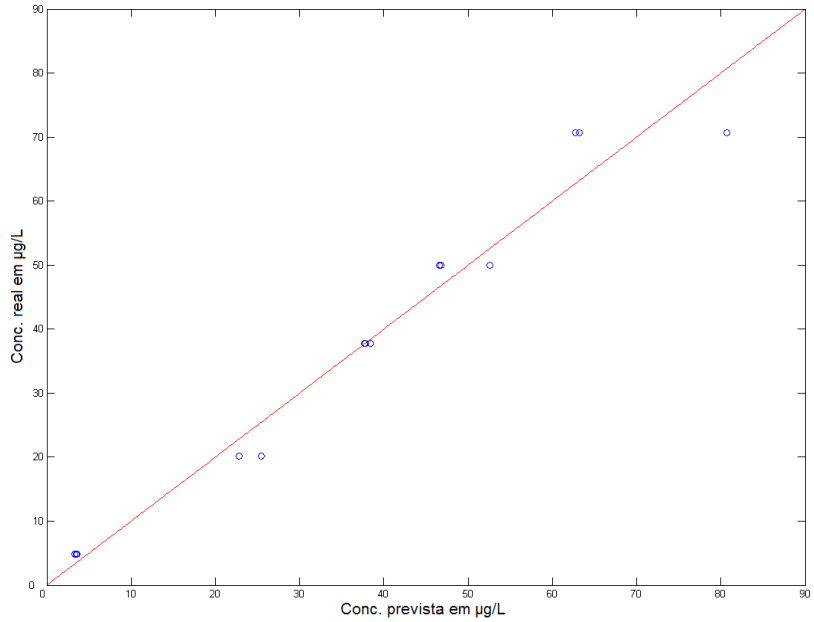


Figura 19 – Valor real contra previsto, utilizando o conjunto de calibração, modelado com o PARAFAC e um fator para piridoxina.

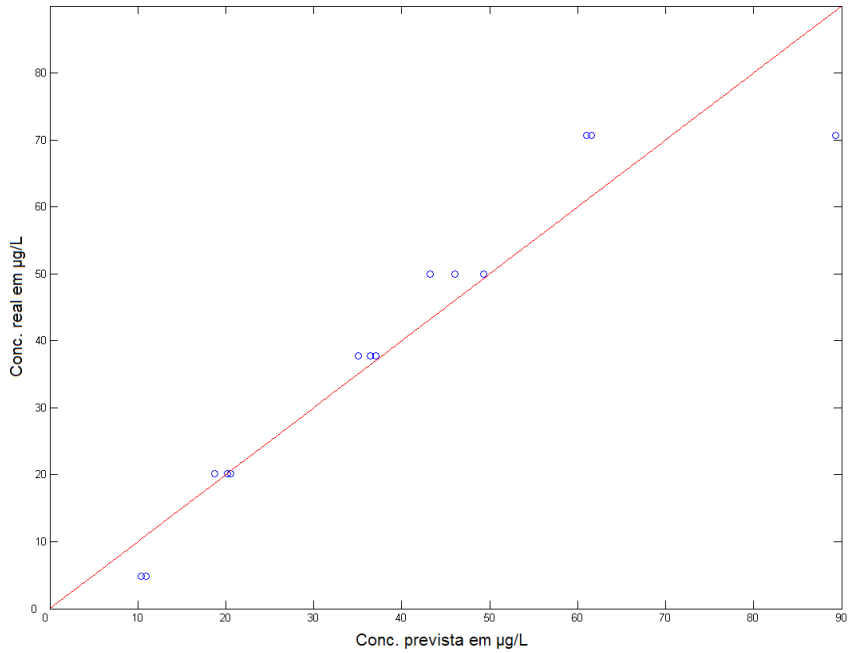


Figura 20 – Valor real contra previsto, utilizando o conjunto de calibração, modelado com o PARAFAC e dois fatores para piridoxina.

Quando se utiliza o modelo PARAFAC no conjunto de calibração, é possível analisar como se comportam os dados, ou seja, se são trilineares (fato que ocorre na ausência de supressão). Pode-se ver nos gráficos 16 e 17 a concordância entre os valores reais e previstos para o conjunto de calibração para a riboflavina, tanto utilizando um como para dois fatores. Nas figuras 19 e 20, para a piridoxina, pode-se verificar que não há uma concordância ótima entre valores reais e previstos, como ocorreu na riboflavina, e isso não era esperado, já que esta curva foi construída ao mesmo tempo em que a curva para riboflavina, eliminando a possibilidade de problemas instrumentais. Além disso, uma triplicata do último nível de concentração parece estar deslocada, sendo considerada posteriormente um *outlier*. É necessário levar em consideração, porém, que a emissão de fluorescência da piridoxina é menor que da riboflavina, e o primeiro ponto da curva é de concentração muito baixa (5 µg/L). Problemas de supressão foram estudados, e nessa faixa de concentração não houve supressão para nenhum dos dois analitos nos padrões. Os produtos analisados também não apresentaram supressão quando foram realizadas medidas mediante consecutivas diluições para cada um dos produtos, sendo a intensidade de fluorescência diretamente proporcional à concentração.

Nas tabelas 6 a 11 são apresentados os resultados obtidos dos produtos analisados, utilizando os modelos PARAFAC, NPLS e BLLS-RBL com e sem a seleção de variáveis por GA.

5.5.1. Resultados do PARAFAC

Nas tabelas 6 e 7 são mostrados os resultados obtidos com o PARAFAC.

Tabela 6 – Erros de previsão do PARAFAC sem GA.

PARAFAC sem GA (RMSEP µg/L)				
	riboflavina		piridoxina	
Produto	1 fator	2 fatores	1 fator	2 fatores
APL	1,26 (2,5)	1,68 (3,4)	56,98 (74,6)	7,95 (29,9)
BPL	3,93 (7,8)	3,52 (7,0)	59,20 (70,0)	11,00 (30,5)
CPL	1,52 (3,4)	1,15 (2,48)	140,03 (90,2)	14,78 (49,9)
DPL	17,55 (37,9)	11,34 (28,9)	176,45 (72,4)	44,52 (39,6)
EPL	9,26 (15,6)	9,29 (15,6)	79,65 (63,6)	65,18 (59,3)

* Erro relativo de previsão entre parênteses (REP).

Tabela 7 – Erros de previsão do PARAFAC com GA.

PARAFAC com GA (RMSEP µg/L)				
	riboflavina		piridoxina	
Produto	1 fator	2 fatores	1 fator	2 fatores
APL	5,05 (10,4)	7,07 (15,5)	101,04 (83,0)	6,90 (21,8)
BPL	1,17 (2,5)	1,23 (2,6)	96,35 (78,8)	18,69 (45,0)
CPL	1,14 (2,4)	1,33 (2,9)	100,85 (87,0)	4,00 (69,6)
DPL	18,67 (40,2)	16,47 (37,2)	48,93 (41,8)	32,51 (101,4)
EPL	3,51 (7,3)	5,72 (12,8)	83,18 (65,0)	22,76 (165,7)

* Erro relativo de previsão entre parênteses (REP).

Pela análise das tabelas 6 e 7, pode-se observar que os erros para a piridoxina foram muito elevados mesmo utilizando o GA com um ou dois fatores. Isso porque o método PARAFAC não encontrou uma solução matemática razoável para criar um modelo para prever a piridoxina.

5.5.2. Resultados do NPLS

Os erros dos modelos NPLS são mostrados nas tabelas 8 e 9.

Tabela 8 – Erros de previsão do NPLS sem GA.

NPLS sem GA (RMSEP µg/L)				
	riboflavina		piridoxina	
Produto	1 fator	2 fatores	1 fator	2 fatores
APL	1,34 (2,6)	1,65 (3,3)	40,00 (67,5)	15,25 (44,7)
BPL	3,93 (7,9)	3,49 (6,9)	38,93 (60,7)	19,02 (43,2)
CPL	1,67 (3,67)	1,34 (3,0)	84,36 (84,8)	53,50 (78,5)
DPL	17,00 (37,3)	19,66 (39,7)	79,72 (54,2)	52,45 (43,6)
EPL	8,90 (15,1)	9,53 (15,9)	41,20 (47,9)	34,6 (73,6)

* Erro relativo de previsão entre parênteses (REP).

Tabela 9 – Erros de previsão do NPLS com GA.

NPLS com GA (RMSEP µg/L)				
	riboflavina		piridoxina	
Produto	1 fator	2 fatores	1 fator	2 fatores
APL	1,26 (2,4)	1,21 (2,4)	34,85 (64,5)	1,60 (9,5)
BPL	3,40 (7,4)	3,31 (6,6)	32,81 (56,6)	5,02 (16,1)
CPL	1,29 (2,8)	1,11 (2,4)	70,61 (82,5)	1,87 (12,5)
DPL	12,37 (30,7)	8,23 (23,2)	73,51 (52,2)	6,96 (10,8)
EPL	8,51 (14,6)	8,62 (14,6)	35,96 (44,9)	21,45 (31,1)

* Erro relativo de previsão entre parênteses (REP).

Pela análise das tabelas 8 e 9, pode-se observar que os erros para a piridoxina foram também elevados, porém a utilização do GA fez com que os erros diminuíssem, principalmente utilizando dois fatores.

5.5.3. Resultados do BLLS

Os erros dos modelos BLLS são mostrados nas tabelas 10 e 11.

Tabela 10 – Erros de previsão do BLLS sem GA.

BLLS sem GA (RMSEP µg/L)				
	riboflavina		piridoxina	
Produto	1 fator	2 fatores	1 fator	2 fatores
APL	1,26 (2,5)	1,62 (3,23)	48,56 (71,5)	7,57 (27,2)
BPL	3,93 (7,8)	3,62 (7,2)	51,71 (67,1)	9,81 (27,9)
CPL	1,49 (3,3)	1,16 (2,52)	120,76 (88,9)	10,53 (42,2)
DPL	16,87 (37,0)	18,78 (39,4)	152,34 (69,3)	13,25 (16,3)
EPL	9,26 (15,6)	9,48 (15,9)	70,63 (61,0)	18,16 (26,6)

* Erro relativo de previsão entre parênteses (REP).

Tabela 11 – Erros de previsão do BLLS com GA.

BLLS com GA (RMSEP µg/L)				
	riboflavina		piridoxina	
Produto	1 fator	2 fatores	1 fator	2 fatores
APL	1,12 (2,2)	1,17 (2,4)	38,38 (66,7)	5,38 (21,4)
BPL	3,80 (7,6)	3,66 (7,3)	40,34 (61,4)	20,06 (205,7)
CPL	1,14 (2,5)	1,47 (3,1)	94,39 (86,3)	4,08 (39,4)
DPL	12,40 (30,5)	1,90 (6,8)	124,14 (64,8)	6,52 (9,2)
EPL	9,07 (15,3)	8,63 (14,8)	59,02 (56,9)	13,72 (23,9)

* Erro relativo de previsão entre parênteses (REP).

Pela análise das tabelas 10 e 11, pode-se observar que o BLLS-RBL foi o modelo de calibração que obteve os melhores resultados gerais para a piridoxina quando comparado com os outros modelos sem a utilização do GA, porém a utilização do GA ainda minimiza esses erros, como mostrado na tabela 11.

5.5.4. Discussão dos resultados

A tabela 12 mostra os modelos que levaram aos menores RMSEP's e seus REP's na determinação das vitaminas B2 e B6 nos produtos analisados.

Tabela 12 – Melhores modelos para cada produto/vitamina:

Produto	Analito	Modelo / fatores	RMSEP em µg/L
APL	B2	BLLS com GA / 1	1,12 (2,2)
	B6	NPLS com GA / 2	1,60 (9,5)
BPL	B2	PARAFAC com GA / 1	1,17 (2,5)
	B6	NPLS com GA / 2	5,02 (16,1)
CPL	B2	NPLS com GA / 2	1,11 (2,4)
	B6	NPLS com GA / 2	1,87 (12,5)
DPL	B2	BLLS com GA / 2	1,90 (6,8)
	B6	BLLS com GA / 2	6,52 (9,2)
EPL	B2	PARAFAC com GA / 1	3,51 (7,3)
	B6	BLLS com GA / 2	13,72 (23,9)

* Erro relativo de previsão entre parênteses (REP).

Uma análise das tabelas 6 à 11 mostra que, apesar de algumas exceções, os melhores resultados foram obtido quando o GA foi utilizado conjuntamente com o método de calibração de segunda ordem. Observando os casos em que o GA não obteve um bom resultado, conclui-se que são casos que possuem erros muito elevados e os modelos, sem a utilização do GA, não explicam de maneira satisfatória a relação entre os dados e a concentração, e talvez por esse fato nem mesmo o GA tenha conseguido encontrar uma relação razoável entre algumas variáveis e o valor esperado para os analitos.

Pela análise da tabela 12 assim como das tabelas 6 à 11, pode-se observar que o GA, quando executado com os modelos de calibração de segunda ordem, apresentou uma melhora considerável nos resultados, quando comparado com os resultados sem a utilização do GA, de maneira que os melhores resultados para todos produtos/analitos foram obtidos quando empregou-se o GA. Uma outra

observação a ser feita é que em todos os modelos, o fato de utilizar dois fatores para a piridoxina foi muito mais significativo que quando utilizado para a riboflavina, em termos da diminuição de RMSEP. Isso ocorre porque a emissão da piridoxina, relativamente à riboflavina, é menor, estando assim mais susceptível a interferência, além de seu espectro estar localizado numa região de menor comprimento de onda (mas próximo ao ultravioleta), onde existe maior probabilidade de ocorrer absorção/emissão por parte de interferentes presentes nas amostras de previsão.

Na tabela 12, pode-se observar também que os melhores valores para cada caso se distribuem com quatro melhores resultados para BLLS-RBL e NPLS, e dois melhores resultados para o PARAFAC, utilizando o GA em todos os casos. O NPLS, que não possui vantagem de segunda ordem e necessita da presença dos interferentes no conjunto de calibração para poder realizar a previsão, conseguiu os melhores resultados em muitos casos quando utilizado conjuntamente com o GA, principalmente para a piridoxina, mostrando que o GA conseguiu contornar a presença dos interferentes nas amostras de previsão. Na tabela 13 são mostrados os índices de recuperação das adições, com os melhores da tabela 12.

Tabela 13 – Índices médios de recuperação obtidos com os melhores modelos de cada caso (Tabela 12). Dados em percentagem de recuperação.

Produto	Analito	Recuperação em %
APL	B2	101,1 (±7,3)
	B6	88,4 (±30,9)
BPL	B2	98,2 (±8,8)
	B6	89,0 (±33,7)
CPL	B2	99,8 (±9,5)
	B6	120,8 (±18,5)
DPL	B2	74,3 (±8,1)
	B6	48,2 (±84,2)
EPL	B2	93,5 (±7,2)
	B6	322,9 (±187,0)

*Estimativa de desvios padrão entre parênteses

Na tabela 13 pode-se observar que a recuperação das adições de piridoxina para os produtos DPL e EPL foram ruins, porém, analisando a tabela 12 pode-se notar que os RMSEP's desses modelos também foram altos. Ao analisar a tabela 13 é necessário ter em mente que as adições foram muito pequenas, e um RMSEP de 1 µg/L na tabela 12 pode gerar um erro grande nas recuperações das adições. Como já foi dito, os níveis de adição foram de 5, 10 e 15 µg/L, portanto um erro de recuperação de 20% na tabela 13 corresponde a um erro de concentração de somente 1, 2 e 3 µg/L, para o primeiro, segundo e terceiro nível de adição, respectivamente. Além disso, observando as figuras 15 e 16, as intensidades máximas de emissão de fluorescência para uma solução com concentração de 70 µg/L é cerca de 140 unidades para piridoxina e 250 unidades para riboflavina. Desta maneira temos uma relação "intensidade de fluorescência/concentração(µg/L)" de cerca de 2 para piridoxina e de 3 para riboflavina. Supondo que o ruído instrumental, verificado pela oscilação da intensidade de fluorescência emitida, está na faixa de ± 2 unidades, e que a relação "intensidade de fluorescência/concentração(µg/L)" é de 2, analisando em função da concentração, o ruído instrumental presente será então de ± 1 µg/L, tornando difícil a quantificação neste nível.

Desta maneira, recuperações entre 80 a 120 % já eram esperadas. Se fossem realizadas adições maiores, com certeza os erros seriam menores, porém não faria sentido realizar adições de 100 %, por exemplo. Não seria possível obter informações relevantes sobre a recuperação, já que a tendência é que estas se aproximassem deste valor. As adições realizadas foram propostas tendo em vista as concentrações dos analitos nos produtos analisados, ilustrados na tabela 5. Apesar das concentrações dos analitos estarem em um nível baixo para a análise, não se pode aumentar a concentração das soluções dos produtos estudados, pois ocorreria auto-supressão da emissão de fluorescência.

As variáveis selecionadas para os casos estudados são mostradas nas figuras 21 e 22. Nota-se nessas figuras que no perfil de excitação, onde havia menos variáveis, o mínimo selecionado foi de duas variáveis, que é o mínimo possível para que seja possível utilizar os métodos de segunda ordem aqui

empregados. A interpretação das variáveis selecionadas é normalmente difícil, porém nesta aplicação parece ser um pouco mais clara.

Analisando primeiramente o caso da riboflavina (figura 21), pode-se observar que para todos os produtos, existem variáveis selecionadas sobre algum dos máximos de emissão da riboflavina. Para o produto DPL, não foram selecionadas variáveis sobre o maior pico de emissão, isso porque havia interferências neste máximo de emissão (isso pode ser observado, pois o máximo deste pico está deslocado para cima em relação à todos outros produtos), e o GA selecionou, desta maneira, as variáveis que estavam no segundo pico de excitação.

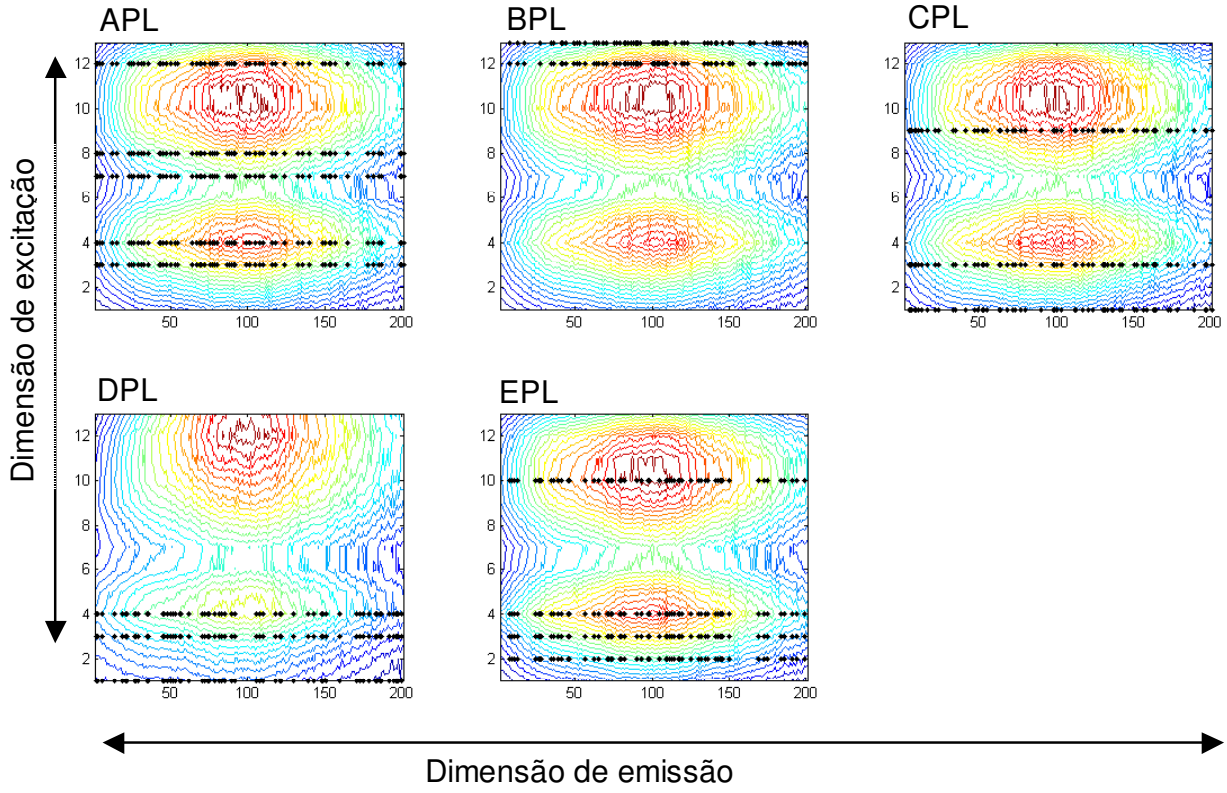


Figura 21 – Variáveis selecionadas nos melhores modelos (tabela 12) para riboflavina nos cinco produtos. As variáveis selecionadas são indicadas por pontos nas superfícies.

Analisando o caso da piridoxina (figura 22), pode-se observar que para nenhum dos produtos foram selecionadas variáveis sobre o máximo de emissão, mas sim em regiões tangentes ao máximo. Em todos os casos foi selecionada a variável 2 do perfil de excitação, sendo esta importante para a quantificação da piridoxina em meio aos interferentes dos produtos analisados.

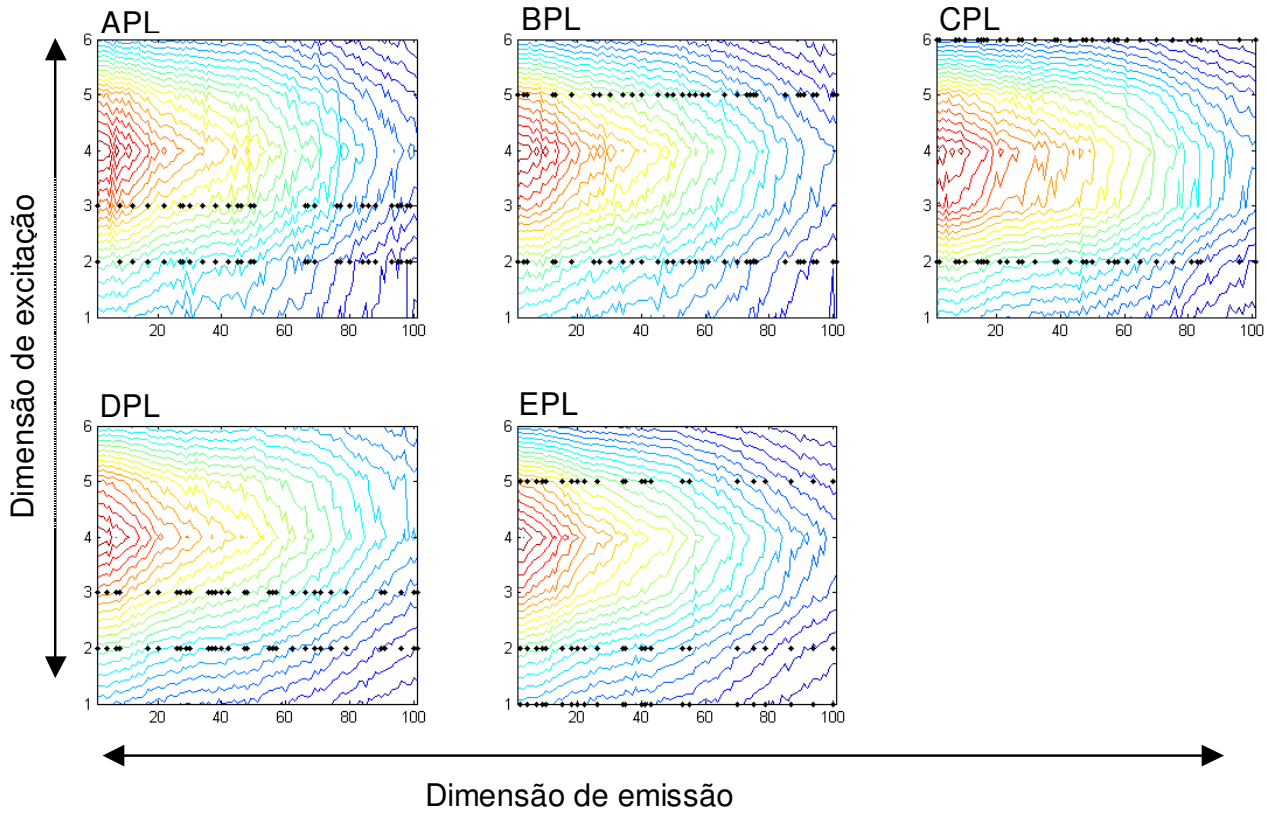


Figura 22 – Variáveis selecionadas nos melhores modelos (tabela 12) para piridoxina nos 5 produtos. As variáveis selecionadas são indicadas por pontos na superfície.

A figura 23 mostra um exemplo do gráfico de saída do programa GA, apresentando a evolução do RMSEP a cada geração para o caso da determinação de piridoxina no produto CPL, utilizando o modelo NPLS com dois fatores.

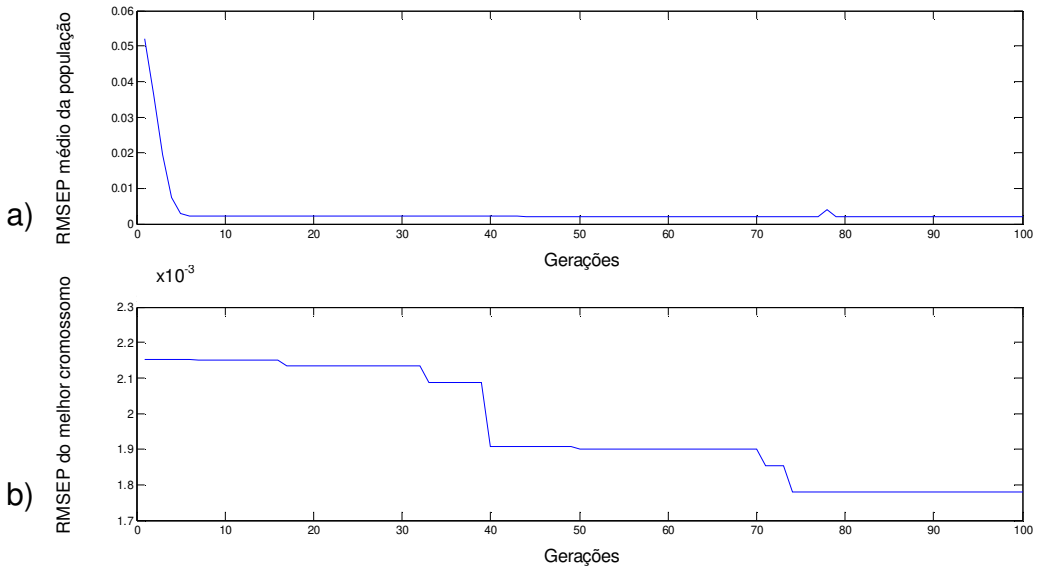


Figura 23 – Evolução do RMSEP utilizando o GA na determinação de piridoxina no produto CPL, utilizando o modelo NPLS com dois fatores: a) média de todos cromossomos da população a cada geração e; b) melhor cromossomo a cada geração.

CAPÍTULO 6

APLICAÇÃO 3: DETERMINAÇÃO SIMULTÂNEA DE AA E AAS EM MEDICAMENTOS POR FIA COM GRADIENTE DE pH

6.1. Aquisição de dados de segunda ordem por FIA com gradiente de pH

Num sistema de análise por injeção em fluxo (FIA) com gradiente de pH e detector de arranjo de diodos, os dados obtidos são de segunda ordem pois a absorbância varia tanto em função do comprimento de onda (perfil espectral) quanto em função do tempo (perfil temporal). O perfil temporal neste caso é resultante da variação do pH durante o tempo, decorrente do gradiente de pH gerado. A cada tempo, o espectrofotômetro registra um espectro na região UV-Vis, desta maneira, cada espectro estará relacionado com um pH. Os dados para cada amostra são então bidimensionais (na forma de uma matriz de dados) e um exemplo é mostrado na figura 24:

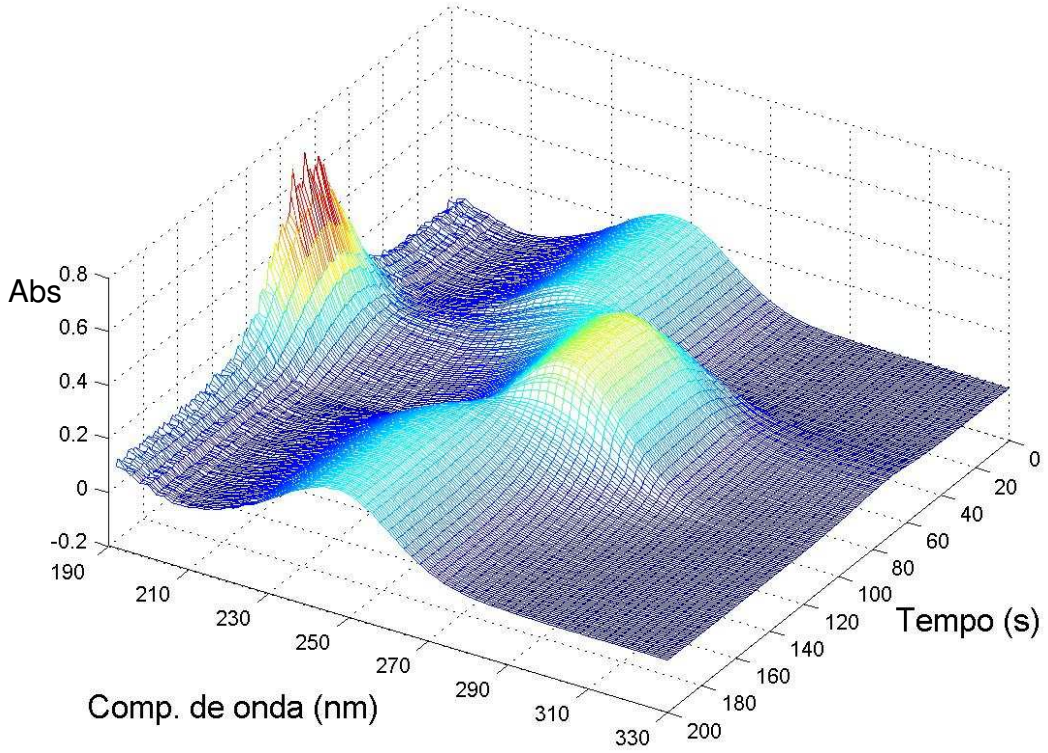


Figura 24 – Superfície gerada para o ácido ascórbico no sistema FIA com gradiente de pH e detecção por arranjo de diodos.

É possível notar na figura 24 o gradiente de pH gerado pelo deslocamento do pico do AA. O perfil espectral do AAS e do AA é sensível ao pH devido ao fato de serem ácidos fracos e estarem susceptíveis a protonação/desprotonação. Os espectros de uma espécie protonada e da mesma espécie desprotonada são geralmente diferentes, sendo essa variação que permite a utilização do FIA com gradiente de pH e detector de arranjo de diodos como um método de segunda ordem.

6.2. Construção do sistema FIA

Houve a necessidade de construir um sistema FIA que gerasse um gradiente de pH no qual estivesse contido o pKa's dos ácidos fracos presentes nos fármacos estudados. Para a construção deste sistema foram utilizadas 3 válvulas solenóides de três vias da *N-Research* que operam a 12 Volts e consomem 3 Watts , uma bomba peristáltica Ismatec IPC com tubos de Tygon e tubos de politetrafluoroetileno (PTFE) com 0,8 mm de diâmetro interno, uma fonte de 12 Volts e potência de 24 Watts e suporte, junções e reator de acrílico, além de um agitador magnético comum. A saída do fluxo deste sistema foi acoplada a uma célula de fluxo para ser medida espectrofotometricamente por um equipamento HP 8452 equipado com detector de arranjo de diodos.

O sistema foi montado de modo que a amostra passava continuamente pela cela de fluxo no espectrofotômetro, sofrendo uma diluição prévia constante. O gradiente de pH foi gerado injetando uma solução aquosa de K_2HPO_4 de concentração 0,05 mol L^{-1} através de uma alça de injeção de 31,5 cm, equivalente à 160 μL. O carregador foi uma solução aquosa de H_3PO_4 de concentração 0,01 mol L^{-1}. Em uma das junções, havia então confluência do ácido com a amostra diluída, e após a injeção, a amostra confluía com o gradiente de pH formado, modificando assim seu meio. Como foram analisados dois fármacos com pKa's dentro da faixa de pH gerado pelo FIA (pKa de 4,2 para acido ascórbico e 3,5 para acido acetilsalicílico), teremos as espécies na forma desprotonada (pH >> pKa) e na forma protonada (pKa >> pH).

Na figura 25 é mostrado o esquema do sistema FIA. O controle das válvulas foi unificado, de forma que são ligadas ou desligadas todas ao mesmo tempo. Inicialmente carregava-se o sistema com os fluidos correspondentes a cada segmento do FIA.

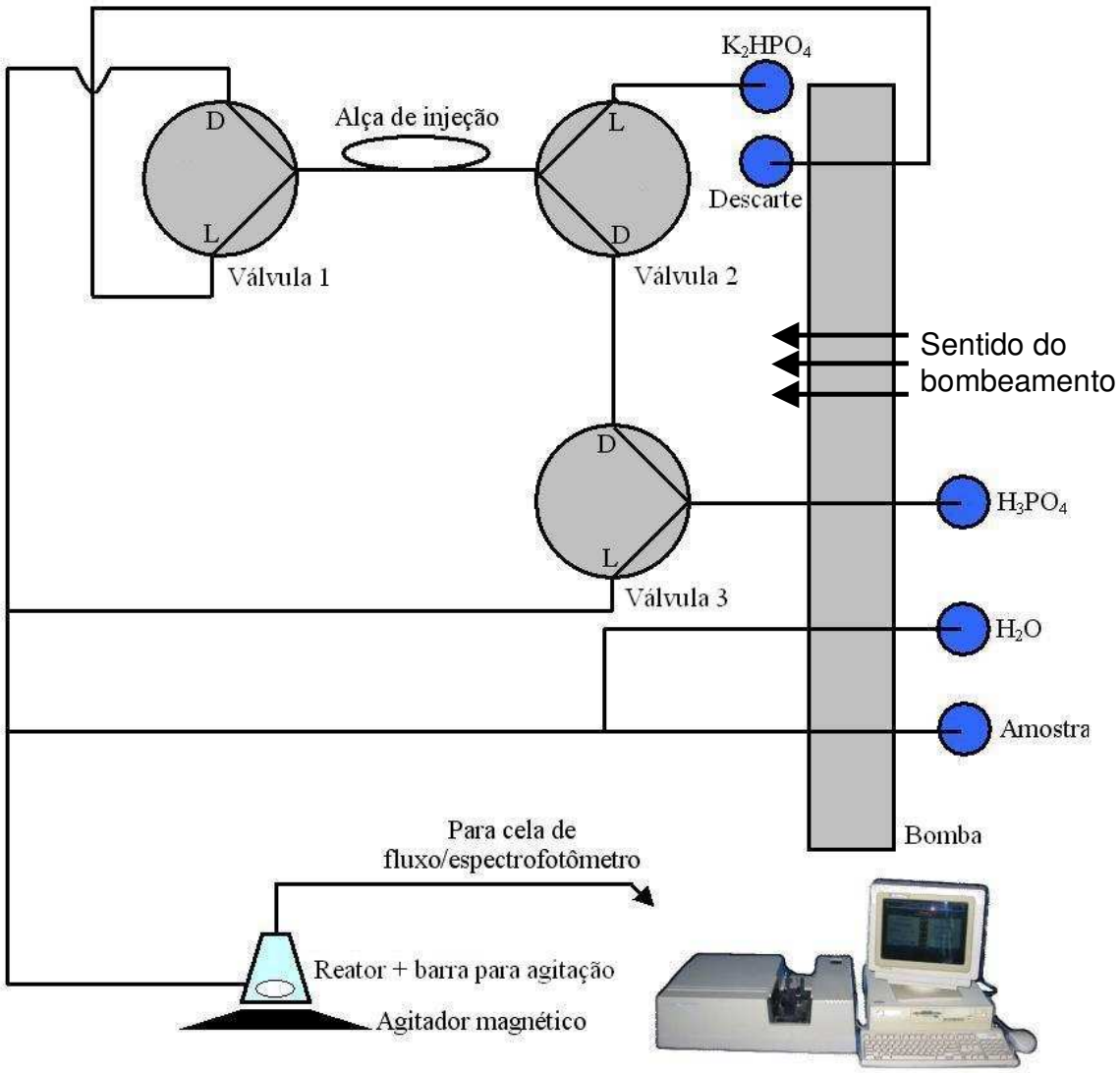


Figura 25 – Esquema do sistema FIA montado. L e D indicam saída do fluxo com as válvulas ligadas ou desligadas, respectivamente.

Quando as válvulas estavam desligadas, a solução de H_3PO_4 passava pela alça de injeção. Assim que as válvulas eram ligadas, o fluxo da solução de H_3PO_4 era desviado pela válvula solenóide 3 e por sucção as válvulas solenóides 1 e 2

permitiam o preenchimento da alça de injeção pela solução de K_2HPO_4. Após as válvulas serem desligadas, o volume de solução da alça era então injetado no sistema. A dispersão do K_2HPO_4 no H_3PO_4 gera um gradiente de pH suave no fluxo. A dispersão é um fenômeno característico do FIA, e ocorre em função do escoamento laminar dos fluidos dentro do sistema e em função do reator [34]. O fenômeno pode ser observado na figura 26 [35].

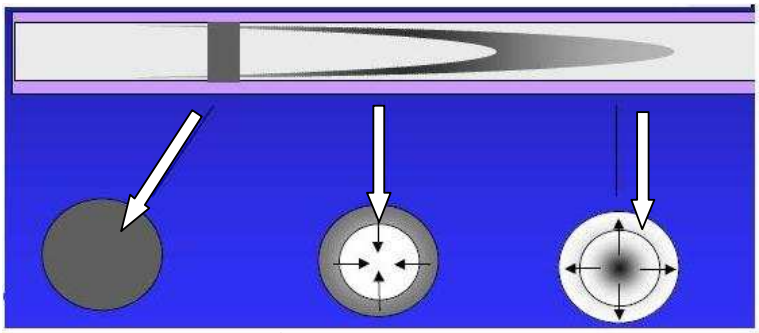


Figura 26 – Dispersão num sistema FIA. Os tons de cinza representam quantativamente a presença do K_2HPO_4 injetado, e a parte branca representa o carregador (solução de H_3PO_4). A dispersão é caracterizada pelo perfil parabólico. Dentro do reator esse perfil é homogeneizado dando origem à um fluxo com concentração de K_2HPO_4 variável, o que leva a uma variação do pH (gradiente) [35].

6.3. Calibração do sistema FIA

O sistema FIA foi calibrado a fim de verificar a faixa de pH alcançado no sistema proposto. Para isso foram preparadas diversas soluções tampão numa faixa de pH que abrangesse os pKa's dos fármacos em estudo e também uma solução de indicadores que respondesse espectrofotometricamente à toda faixa de interesse. Foi então calibrada uma faixa de pH de 2,0 até 6,4. Na tabela 14 é mostrado como as soluções foram preparadas:

Tabela 14. Preparo das soluções tampão e pH das soluções antes e na saída do sistema FIA, medidas por pHmetro.

Volumes para 100 mL de soluções tampão			
Ác. cítrico (0,1 M)	Na_2HPO_4 (0,2 M)	pH da solução	pH na saída do FIA
98,00	2,00	1,79	2,01
90,30	9,70	2,25	2,39
81,10	18,90	2,67	2,77
72,40	27,60	3,10	3,22
65,20	34,80	3,57	3,71
59,00	41,00	4,02	4,19
53,80	46,20	4,46	4,67
49,00	51,00	4,90	5,11
44,80	55,20	5,31	5,52
40,20	59,80	5,71	5,91
34,60	65,40	6,18	6,37

A solução de indicadores era uma solução aquosa de alaranjado de metila numa concentração de $3x10^{-4}$ mol L^{-1} e verde de bromocresol na concentração de $1x10^{-4}$ mol L^{-1}. Para a calibração, as soluções tampão foram alimentadas no caminho com maior fluxo (originalmente o da água para diluição) no sistema FIA, a solução de indicadores no caminho de menor fluxo (originalmente o da amostra) e água nos outros caminhos. Desta maneira, foi construído um modelo de calibração entre o pH e os espectros na região do visível (400 a 650 nm) da solução na saída do FIA. Após isso, o sistema foi carregado como originalmente proposto (com H_3PO_4 e K_2HPO_4), ainda com a solução de indicadores no caminho de menor fluxo (esta solução indicadora foi alimentada da mesma maneira para se obter a mesma concentração de indicadores na calibração e previsão do gradiente) e foi injetada a solução de K_2HPO_4. O espectrofotômetro monitorou o gradiente no fluxo através do seu programa de cinética, capturando 1 espectro por segundo durante 300 segundos.

Para encontrar os pH's no sistema FIA, foi desenvolvido um modelo de calibração multivariada baseado no método dos mínimos quadrados parciais (PLS). O conjunto de calibração do modelo PLS foram os espectros das soluções tampão com seus respectivos pH's na saída do FIA. Desta maneira, os espectros obtidos a cada segundo após a injeção do K_2HPO_4 foram relacionados aos pH's através da calibração feita utilizando PLS. Assim, pode-se verificar a faixa de pH alcançado no gradiente (entre 2,0 e 6,4), assim como a relação "tempo de injeção"/pH. A figura 27 apresenta o equilíbrio entre as espécies protonada e desprotonada do ácido ascórbico recuperado e normalizado pelo PARAFAC a partir de dados obtidos com o sistema FIA proposto.

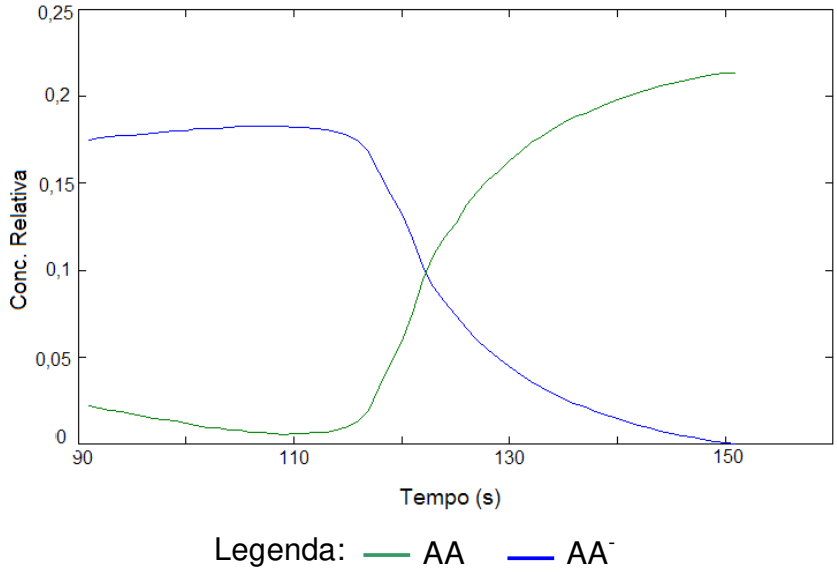


Figura 27 – Apresenta o equilíbrio entre as espécies protonada e desprotonada do ácido ascórbico recuperado e normalizado pelo PARAFAC a partir de dados obtidos com o sistema FIA proposto.

6.4. Procedimento experimental

6.4.1. Amostras de calibração, validação e validação de 2º ordem

O planejamento experimental realizado consistiu de quatro conjuntos: calibração, validação, validação com vantagem de segunda-ordem (utilizando cafeína como interferente) e amostras de medicamentos com e sem adição de

analito (para teste de recuperação). A Figura 28 ilustra o planejamento dos conjuntos de calibração, validação e validação com vantagem de segunda ordem. A concentração nos níveis é referente à amostra antes de ser alimentada no sistema, ou seja, antes da etapa de diluição no FIA. As diluições decorrentes das confluências não alteram a calibração pois a relação de concentração entre as amostras permanece inalterada.

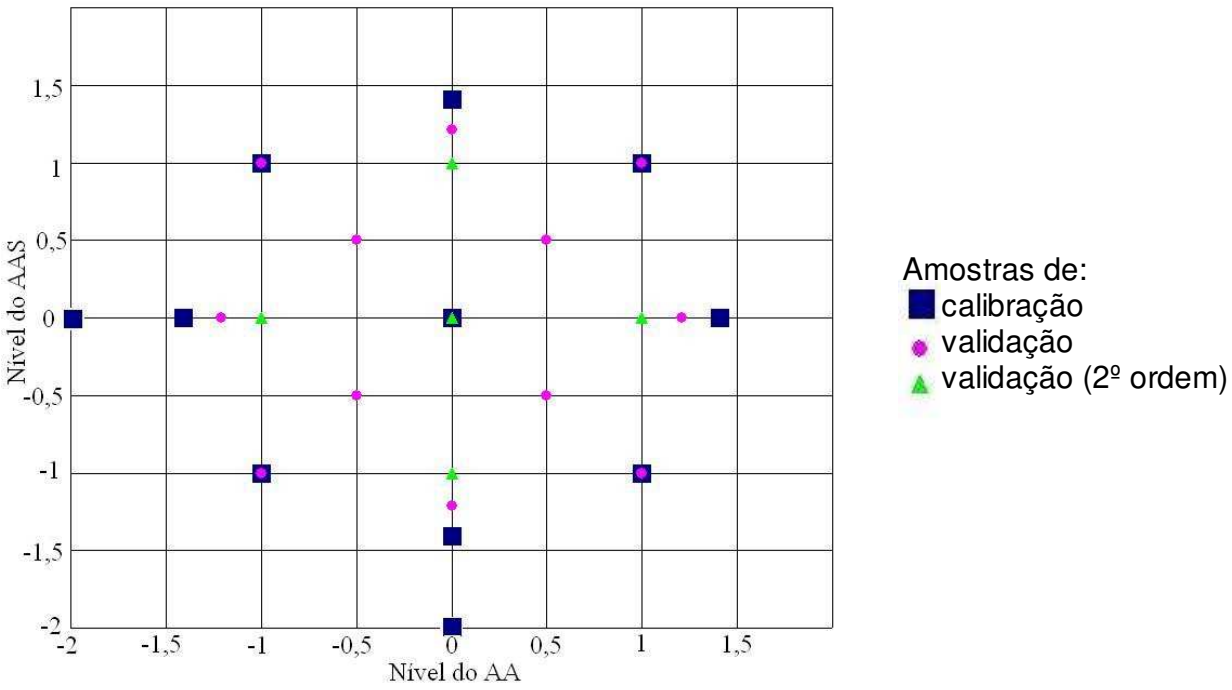


Figura 28 – Gráfico do planejamento experimental utilizado. Foram utilizados oito níveis e dois fatores para construir os conjuntos de calibração, validação e validação de segunda ordem.

Na tabela 15 é apresentada as codificações dos níveis utilizados no planejamento da figura 28.

Tabela 15. Codificação dos níveis.

Níveis	-2	-1,41	-1,21	-1	-0,5	0	0,5	1	1,21	1,41
AAS (mg/L)	0	23,6	31,6	40,0	60,0	80,0	100,0	120,0	128,4	136,4
AA (mg/L)	0	14,2	19,0	24,0	36,0	48,0	60,0	72,0	77,0	81,8

A escolha das concentrações dos níveis foi feita realizando testes prévios no sistema FIA proposto, levando em conta o sinal de absorbância na saída do sistema FIA.

As soluções trabalho foram preparadas à partir de soluções estoque de AAS (Synth, 99,9%) de concentração 400 mg L^{-1} e de AA (ECIBRA, 99,0%) de concentração 240 mg L^{-1}. Nas soluções do conjunto de validação de 2ª ordem, foi utilizada uma solução estoque de cafeína (VETEC, 99%) de 19,5 mg L^{-1}, e esta foi adicionada nas amostras deste conjunto de maneira que a concentração de cafeína em cada amostra fosse de 4 mg L^{-1}. Essa concentração foi escolhida baseando-se na razão entre o AAS e a cafeína presente no medicamento **LMD**.

6.4.2. amostras dos medicamentos

Os medicamentos analisados foram:

- **JMD,** comprimido efervescente sabor laranja. O valor nominal dos princípios ativos em cada comprimido eram de: 400 mg de AAS e 240 mg de AA. O excipiente não era definido na embalagem do produto.
- **KMD,** comprimido efervescente sabor limão. O valor nominal dos princípios ativos em cada comprimido eram de: 400 mg de AAS e 240 mg de AA. O excipiente era composto de bicarbonato de sódio, ácido cítrico, ciclamato de sódio, sacarina sódica e aroma.
- **LMD,** o valor nominal dos princípios ativos em cada comprimido eram de: 500 mg de AAS e 30 mg de cafeína. O excipiente é composto de amido e corante CI nº 45.430 em quantidades não especificadas.
- **MMD,** comprimido efervescente. O valor nominal dos princípios ativos em cada comprimido eram de: 1000 mg de AA e 250 mg de cálcio ionizável na forma de lactobionato de cálcio e carbonato de cálcio. Além de bicarbonato de sódio, ácido cítrico, 1,8 g de sacarose, aromatizante, sacarina e polietilenoglicol.

As soluções de cada medicamento foram preparadas da seguinte maneira:

1. Foram obtidas as massas de 10 comprimidos para verificar a massa média de cada comprimido.
2. Após isso, esses comprimidos foram macerados em cadinho cerâmico e homogeneizados.
3. Foram preparados 500 mL de solução adicionando a massa necessária medicamento para que fosse produzida uma solução com as concentrações dos analitos de interesse (AAS e AA) próximas ao nível zero do conjunto de calibração.
4. Somente para o medicamento **LMD** houve a necessidade de solubilização prévia em 10 mL de etanol (grau HPLC da Merck), pois neste medicamento não foi possível a solubilização direta com água. A quantidade de etanol utilizada não influenciou significativamente nas medidas, pois até a detecção, a amostra foi diluída várias vezes dentro do FIA. Os outros medicamentos, todos efervescentes, não necessitaram de solubilização prévia em etanol.
5. Foram realizadas também, adições dos analitos em 3 diferentes níveis para cada medicamento a fim de verificar índices de recuperação. Essas adições foram feitas adicionando volumes das soluções estoque no preparo das soluções de 500 mL. As adições foram de 10, 25 e 60 %, de modo que mesmo com as adições, as amostras continuassem dentro do conjunto de calibração (60% de adição do analito equivale ao nível 1,21).

Na tabela 16 são apresentados os valores das concentrações finais obtidas levando em conta os valores nominais para cada fármaco. Esses valores foram corrigidos pelos métodos de referência como será mostrado a seguir.

Tabela 16. Estimativa das concentrações de AAS e AA nas soluções.

	Sem adição (mg/L)		Adição 10 % (mg/L)		Adição 25 % (mg/L)		Adição 60 % (mg/L)	
Medicamento	AAS	AA	AAS	AA	AAS	AA	AAS	AA
JMD	80,5	48,3	88,3	53,1	100,7	60,6	128,7	77,6
KMD	80,4	48,2	88,2	52,9	100,6	60,3	128,5	77,1
LMD	80,3	*	88,1	*	100,4	*	128,4	*
MMD	*	48,4	*	53,2	*	60,5	*	77,1

* Ausente.

6.5. Obtenção dos dados por FIA

Antes de realizar as medidas do conjunto de amostras dos medicamentos, foi feita a calibração e tentativa de previsão do conjunto de validação. Houve muita dificuldade de se conseguir bons resultados devido a problemas de degradação do AA e devido às variações de pressão existente no fluxo, provocado pela bomba peristáltica. O problema das variações de pressão no fluxo foi resolvido trocando o reator tubular por uma câmara de mistura, pois essa além de servir como reator atuava como um amortecedor de fluxo, ou seja, o volume de solução dentro da câmara de mistura absorvia em parte a variação de pressão do fluxo de entrada deixando este mais suave em sua saída. Desta maneira o fluxo chegava praticamente contínuo no detector.

O problema de degradação do AA [36] foi resolvido utilizando água deionizada saturada com $N_{2(g)}$ (borbulhou-se $N_{2(g)}$ por 3 minutos para cada litro de água deionizada) já que o processo de degradação do AA para ácido dehidroascórbico (que não absorve no UV) é acelerado pela presença de $O_{2(g)}$. Outro fator controlado no preparo da solução foi a temperatura da água utilizada e do estoque da solução, que permaneceu sempre entre 1 e 5 ºC (banho de gelo). Em um breve estudo cinético da decomposição do AA, realizando monitoramento por absorção molecular UV, verificou-se que a taxa de decomposição do AA nas

condições de temperatura ambiente é da ordem de 50 % por hora. Já nas condições sugeridas (água saturada com $N_{2(g)}$ e temperatura controlada entre 1 e 5 ºC) a decomposição do AA é reduzida a menos de 0,5 % por hora. Ainda assim, por precaução, as soluções estoque eram preparadas a cada 2 horas, de maneira que o erro por decomposição do AA na amostra seria no máximo 1 %.

Além disso, o sistema apresentou o efeito Schlieren [37], um problema causado pela variação do índice de refração no fluxo devido à injeção da solução contida na alça. Como esse efeito era pequeno, além de ser constante, ele não foi motivo de preocupação, já que os modelos de calibração de segunda ordem que seriam utilizados o levariam em consideração.

Depois de resolvidos os problemas acima descritos, as medidas dos conjuntos de calibração, validação, validação de segunda ordem das amostras de medicamentos foram realizadas, todas em triplicata.

6.6. Obtenção dos dados pelo método de referência

Foram realizadas determinações dos fármacos nos medicamentos utilizando HPLC para AAS [38] e iodimetria para o AA [39] e realizada a comparação entre os resultados obtidos pelos métodos de calibração de segunda ordem, com AG e FIA.

No método de referência do AAS foi utilizado o cromatógrafo Shimadzo Prominence High Performance Liquid Chromatography equipado com um detector APD-M20A de arranjo de diodos, um amostrador automático e uma coluna Microsorb MV C18 5 µm (250 mm x 4.6 mm) da Varian. A fase móvel utilizada foi uma solução 15:85 (v/v) acetonitrila:água, a água foi acidificada para pH 3,0 com ácido fosfórico antes da preparação da fase móvel. O fluxo foi de 1,0 mL min^{-1}. A curva analítica foi estabelecida utilizando seis padrões.

A determinação para o AA consistiu da titulação de AA em uma solução acidificada com uma solução padrão de iodo e uma solução de amido como indicador.

Os resultados das análises de referência são mostrados na tabela 17.

Tabela 17. Quantidade em porcentagem relativa entre o valor encontrado e o teor nominal de AAS e o AA para os medicamentos, por HPLC e iodimetria, respectivamente.

Medicamento	AAS (%)	AA (%)
JMD	93,14	109,94
KMD	94,05	101,23
LMD	103,93	*
MMD	*	98,99

* Ausente.

6.7. Resultados e Discussão

Nas figuras a seguir pode-se verificar a relação entre o valor previsto e real do conjunto de validação (figuras 29 e 30) e do conjunto de validação de segunda ordem (figuras 31 e 32) utilizando o modelo PARAFAC. Este tratamento prévio foi realizado sem seleção de variáveis, utilizando a faixa de 210 a 300 nm do espectro UV e o tempo de 90 a 150 segundos após a injeção da alça.

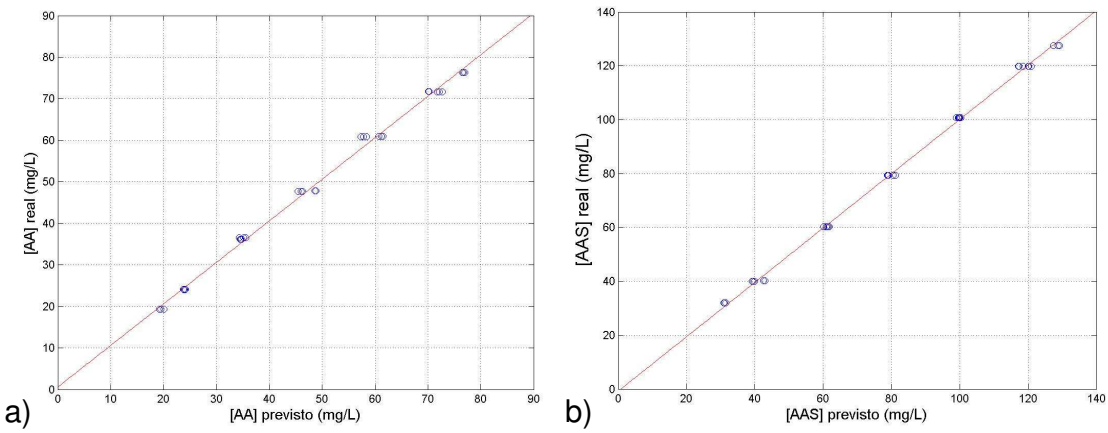


Figura 29 – Valor real versus previsto do conjunto de validação. O modelo de calibração foi construído utilizando o PARAFAC com quatro fatores. a) valores para [AA]; b) valores para [AAS].

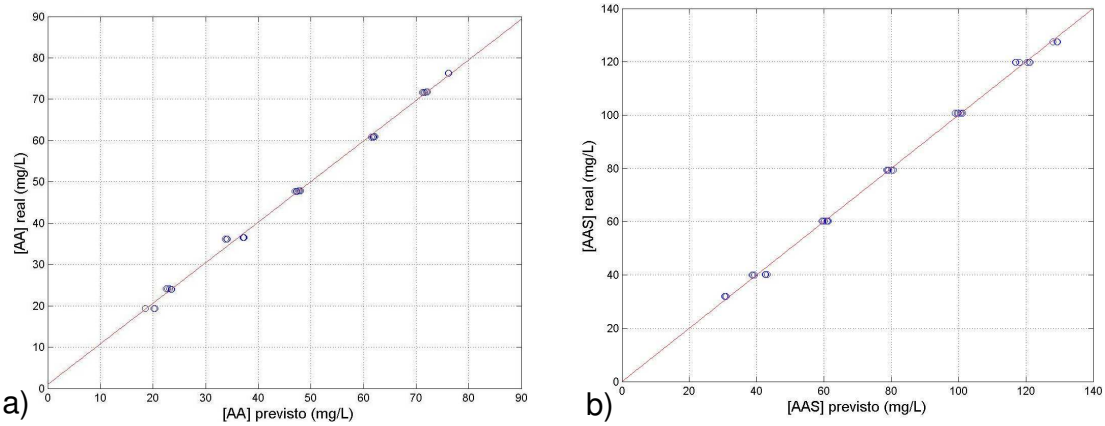


Figura 30 – Valor real versus previsto do conjunto de validação. O modelo de calibração foi construído utilizando o PARAFAC com cinco fatores, para modelar possível interferente ou ruído. a) valores para [AA]; b) valores para [AAS].

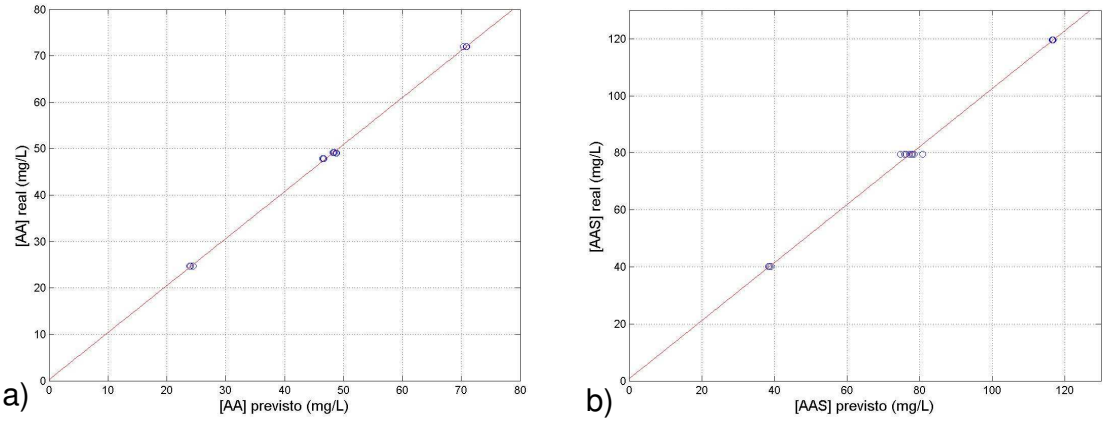
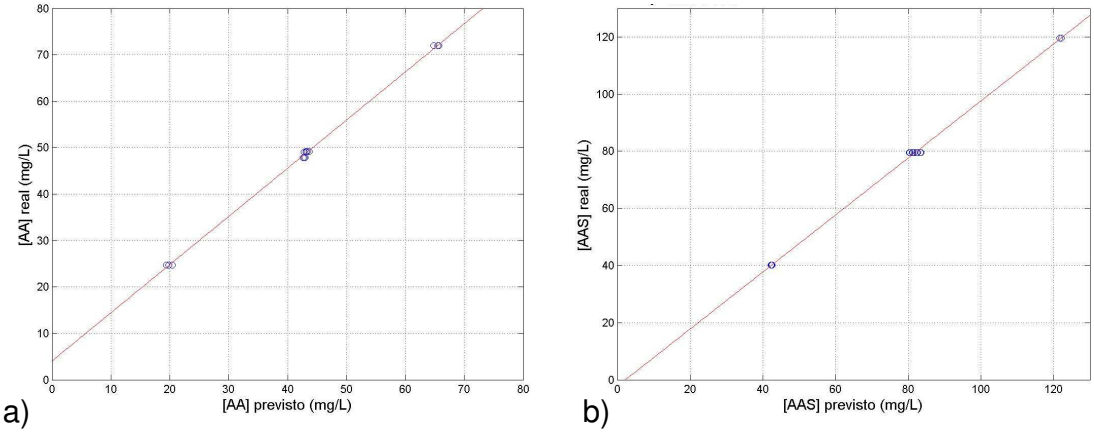


Figura 31 – Valor real versus previsto do conjunto de validação com interferente utilizando PARAFAC com quatro fatores. a) valores para [AA]; b) valores para [AAS].



a) b)

Figura 32 – Valores real versus previsto do conjunto de validação com interferente. O modelo de calibração foi construído utilizando o PARAFAC com cinco fatores. a) valores para [AA]; b) valores para [AAS].

No PARAFAC, os fatores podem ser interpretados como as espécies que possuem perfis em uma das dimensões. Como são quatro espectros possíveis (AAS, AAS^+, AA, AA^+) para o conjunto de calibração e validação, o mais correto é utilizar quatro fatores, porém, pode-se utilizar um quinto fator neste caso para descrever possíveis ruídos, efeito Schlieren ou a cafeína introduzida no conjunto de validação de segunda ordem.

Comparando as figuras 29, 30, 31 e 32, pode-se ver que a inclusão de um fator a mais no PARAFAC não altera significantemente a previsão. Nas figura 31 e 30 pode-se ver que mesmo quando adiciona-se cafeína como interferente, um modelo com quatro fatores obteve previsão tão boa quanto um modelo com cinco fatores. Isso porque a cafeína está em baixa concentração e sua absorbância contribui muito pouco com a absorbância total da amostra, assim o modelo distribui os perfis da cafeína para as outras espécies de maneira que a previsão não melhora significativamente com a adição de um fator a mais.

Nas figuras 33 e 34 são mostrados os perfis espectral e temporal normalizados, recuperados na deconvolução do PARAFAC utilizando quatro e cinco fatores respectivamente.

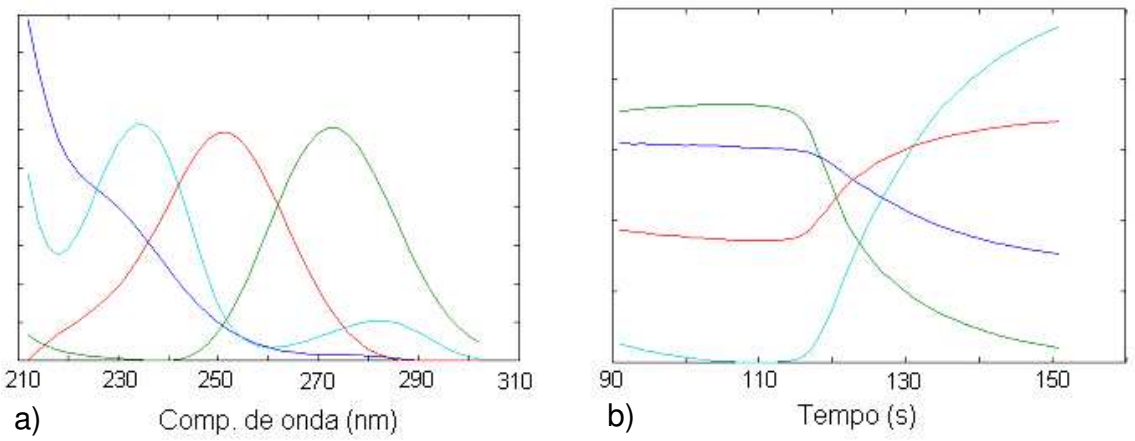


Figura 33 – Perfis espectral (a) e temporal (b), normalizados, recuperados na deconvolução do conjunto de validação utilizando PARAFAC com quatro fatores. Legenda: AAS AAS^- AA AA^-

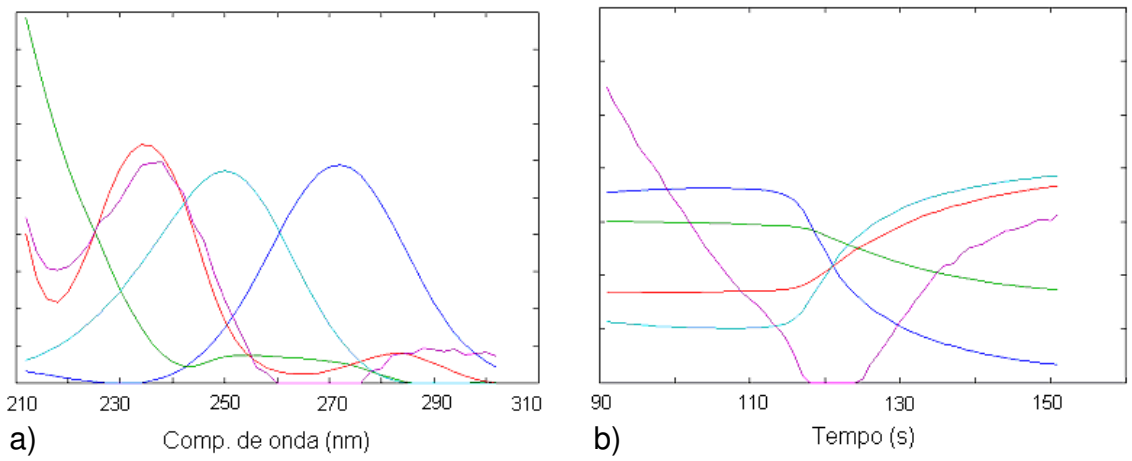


Figura 34 – Perfis espectral (a) e temporal (b), normalizados, recuperados na deconvolução do conjunto de validação utilizando PARAFAC com cinco fatores. Legenda: AAS AAS^+ AA AA^+ Interferência

É possível observar perfis temporais melhor recuperados na figura 34, onde o PARAFAC foi executado utilizando cinco fatores. O perfil espectral do possível interferente se mistura ao perfil do AAS, porém o perfil temporal da interferência é totalmente diferente dos outros perfis. A faixa de tempo escolhida para se trabalhar com o gradiente foi de 90 a 150 segundos, como dito anteriormente, onde existe a transição do meio básico pro meio ácido. O perfil temporal do

interferente na figura 34 está relacionado ao efeito Schlieren. Ao se adicionar mais um fator no conjunto de calibração, obteve-se melhor resolução dos perfis temporais para as espécies de interesse (comparando as figuras 33 e 34), pois este fator a mais, descreve o desvio relacionado a esse efeito.

6.7.1. Resultados do PARAFAC

Nas tabelas 18 e 19 são apresentados os resultados obtidos pelo PARAFAC.

Tabela 18. Erros de previsão do PARAFAC sem AG.

PARAFAC sem AG (RMSEP mg/L)				
	AAS		AA	
Medicamento	4 fatores	5 fatores	4 fatores	5 fatores
JMD	40,6	13,1	1,1	6,9
KMD	20,5	2,3	1,0	6,7
LMD	1,8	4,9	*	*
MMD	*	*	2,4	3,2

* Ausente

Tabela 19. Erros de previsão do PARAFAC com AG.

PARAFAC com AG (RMSEP mg/L)				
	AAS		AA	
Medicamento	4 fatores	5 fatores	4 fatores	5 fatores
JMD	120,8	7,6	2,0	5,1
KMD	4,3	8,7	0,5	1,0
LMD	6,5	59,6	*	*
MMD	*	*	3,0	4,1

* Ausente

Como pode ser visualizado nas tabelas 18 e 19, a utilização do AG com o PARAFAC para esta aplicação não apresentou um ganho significativo, aumentando alguns erros e diminuindo outros.

6.7.2. Resultados do NPLS

Os erros dos modelos NPLS são mostrados nas tabelas 20 e 21.

Tabela 20. Erros de previsão do NPLS sem AG.

NPLS sem AG (RMSEP mg/L)				
	AAS		AA	
Medicamento	4 fatores	5 fatores	4 fatores	5 fatores
JMD	24,3	23,5	1,0	1,0
KMD	10,8	10,5	0,4	0,4
LMD	1,7	2,2	*	*
MMD	*	*	3,5	2,5

* Ausente

Tabela 21. Erros de previsão do NPLS com AG.

NPLS com AG (RMSEP mg/L)				
	AAS		AA	
Medicamento	4 fatores	5 fatores	4 fatores	5 fatores
JMD	3,4	3,5	1,2	0,9
KMD	2,2	1,7	0,5	0,4
LMD	0,9	0,9	*	*
MMD	*	*	1,5	1,4

* Ausente

Através da análise das tabelas 20 e 21, se pode verificar a melhora nos resultados quando o AG foi utilizado conjuntamente com o NPLS, principalmente para a quantificação do AAS nos medicamentos **JMD** e **LMD**. Somente na análise do AA para os medicamentos **JMD** e **KMD**, o modelo com AG apresentou um pequeno aumento do erro.

6.7.3. Resultados do BLLS

Os erros dos modelos BLLS são mostrados nas tabelas 22 e 23.

Tabela 22. Erros de previsão do BLLS sem AG.

BLLS sem AG (RMSEP mg/L)				
	AAS		AA	
Medicamento	4 fatores	5 fatores	4 fatores	5 fatores
JMD	20,8	8,6	3,0	3,5
KMD	12,8	2,7	1,9	2,3
LMD	1,3	11,1	*	*
MMD	*	*	4,5	2,3

* Ausente

Tabela 23. Erros de previsão do BLLS com AG.

BLLS com AG (RMSEP mg/L)				
	AAS		AA	
Medicamento	4 fatores	5 fatores	4 fatores	5 fatores
JMD	2,3	2,2	0,8	1,0
KMD	2,7	2,1	0,4	1,2
LMD	1,2	0,7	*	*
MMD	*	*	1,4	1,9

* Ausente

Através da análise das tabelas 22 e 23, se pode verificar a melhora em todos os casos quando o AG foi utilizado conjuntamente com o BLLS-RBL.

6.7.4. Discussão dos resultados

A tabela 24 mostra os modelos que levaram aos menores RMSEP's na determinação de AA e AAS nos medicamentos analisados.

Tabela 24. Melhores modelos para cada fármaco/medicamento:

Medicamento	Analito	Modelo (fatores)
JMD	AAS	BLLS_AG (5)
	AA	BLLS_AG (4)
KMD	AAS	NPLS_AG (5)
	AA	NPLS (4)
LMD	AAS	BLLS_AG (5)
MMD	AA	PARAFAC_AG (4)

Uma análise das tabelas 18 à 23 mostra que os melhores resultados foram obtidos quando utilizamos o AG, porém no caso do AA a diferença dos RMSEP's com e sem o AG foi pequena como pode ser visto em todos os casos estudados na determinação de AA, já que a quantificação sem a seleção de variáveis resultou em erros baixos.

Pela análise da tabela 24 assim como das tabelas 18 à 23, pode-se ver que o AG realizado com os modelos de segunda ordem apresentou uma melhora nos resultados, principalmente no caso do AAS, quando comparados com os modelos sem a utilização do AG, de maneira que os melhores resultados para todos fármacos/medicamentos foram obtidos quando empregou-se o AG, com exceção do AA para o medicamento **KMD**, no qual se obtiveram resultados iguais utilizando o AG com NPLS e cinco fatores. Um detalhe importante a ser observado e que mostra que o programa AG utilizado atingiu o objetivo proposto, é o fato de os métodos NPLS e BLLS-RBL com quatro fatores apresentarem resultados muito bons com erros relativos de previsão entre 0,4 e 4 % (RMSEP de 0,4 à 3,4 mg/L). Isso ocorreu porque possivelmente o AG está encontrando um conjunto de variáveis que permitem obter bons resultados mesmo na presença dos

interferentes, já que a otimização com quatro fatores não abre espaço para modelar o interferente.

Em especial, o NPLS, que não possui vantagem de segunda ordem e necessita da presença dos interferentes no conjunto de calibração para poder realizar a previsão, conseguiu ótimos resultados, já que na calibração foi utilizado somente soluções de AAS e AA. O fato de não haver os interferentes no modelo de calibração foi contornado pelo AG.

De maneira geral o método que se mostrou mais eficiente foi o BLLS quando utilizado conjuntamente com o AG. Na tabela 25 são mostrados os índices de recuperação das adições feitas nos três níveis com os melhores modelos obtidos com o BLLS-AG.

Tabela 25. Índices de recuperação obtidos com os melhores modelos BLLS-AG. Dados em percentagem de recuperação.

Medicamento	Analito	Nível 1	Nível 2	Nível 3
JMD	AAS	102,6 (±1,0)	100,8 (±2,7)	100,4 (±0,1)
	AA	102,0 (±0,4)	101,2 (±0,4)	99,0 (±0,3)
KMD	AAS	103,2 (±2,0)	99,7 (±1,1)	98,4 (±1,6)
	AA	98,9 (±0,6)	99,5 (±0,1)	100,3 (±0,4)
LMD	AAS	99,2 (±0,4)	99,5 (±0,4)	99,7 (±1,0)
MMD	AA	103,4 (±0,1)	103,1 (±0,7)	99,2 (±0,7)

*Estimativa dos desvios padrão entre parênteses

É possível observar na tabela 25 que as recuperações, se aproximam de 100 %, principalmente para os níveis 2 e 3, pois neles as adições são maiores, gerando erros relativos menores.

Na figura 35 estão ilustradas as variáveis selecionadas sobre as superfícies dos medicamentos, obtidas pelo FIA com gradiente de pH, tanto para o AAS como para o AA.

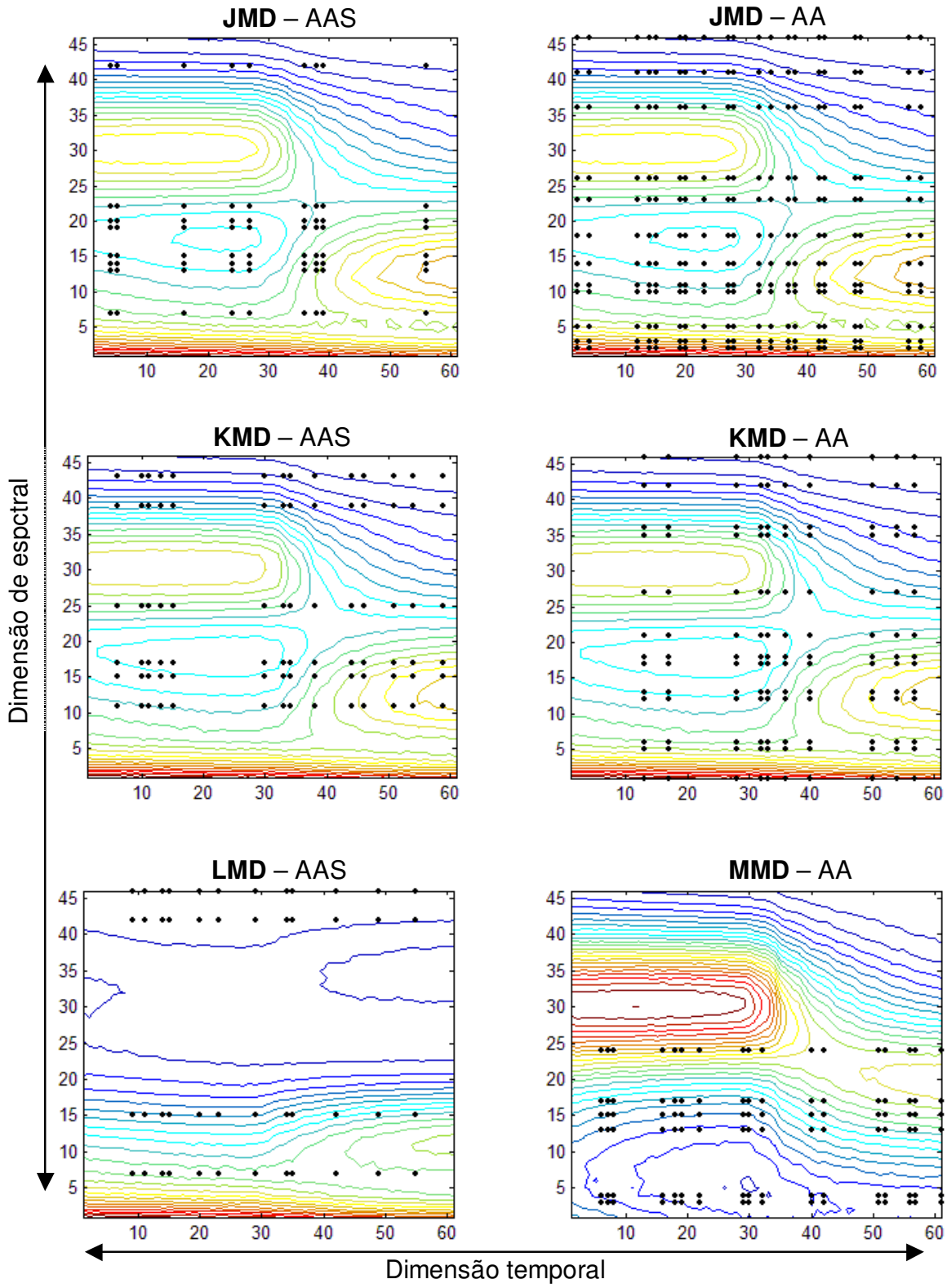


Figura 35 – As variáveis selecionadas pelo AG para os melhores modelos BLLS-RBL são indicadas pelos círculos em preto.

Através das superfícies geradas pelos medicamentos **LMD** (só contém AAS) e **MMD** (só contém AA), pode-se observar as regiões nas quais existem máximos de absorção para o AAS e AA, respectivamente. Sendo assim, era de se esperar que as variáveis selecionadas para o AA estivessem em comprimentos de onda maiores, e para o AAS, comprimentos de onda menores, como mostrado também nos perfis espectrais da figura 34. Porém, antes de se realizar uma análise da seleção das variáveis em função de máximos de absorção, deve-se notar que nesta análise, temos quatro espécies presentes, porém temos somente dois analitos, que são o AAS e o AA, e a quantificação poderia ser realizada tanto através das espécies protonadas (meio ácido do gradiente) quanto através das espécies desprotonadas (meio básico do gradiente). Pela análise dos erros, a quantificação através das espécies ácidas mostrou ser mais confiável, além disso o perfil ácido do AAS é de resolução mais fácil, ao contrário da espécie básica que possui um espectro pouco característico, como se pode visualizar na figura 34. Desta maneira, o fato de existir uma tendência na seleção de variáveis de menores comprimentos de onda no perfil espectral, pode ser justificada pelo fato dos perfis ácidos, que foram usados para a quantificação, possuírem máximos mais próximos ao final da faixa da região ultravioleta utilizada.

Nos perfis temporais mostrados na figura 35, pode-se observar que o fato da amostra estar a todo tempo no sistema, em concentração constante, faz com que a seleção de variáveis neste modo esteja distribuída em todo tempo, para ambos os fármacos e em todos os medicamentos, como era de se esperar.

Na figura 35 pode-se observar também que os medicamentos que só possuiam um dos fármacos, o **LMD** (AAS) e o **MMD** (AA), tiveram o menor número de variáveis selecionadas no perfil espectral, justamente porque havia menos interferência e um número menor de variáveis foi suficiente para ajustar o modelo.

A figura 36 mostra um exemplo do gráfico de saída do programa GA, apresentando a evolução do RMSEP a cada geração para a determinação de AAS no produto **JMD**, utilizando o modelo BLLS com quatro fatores.

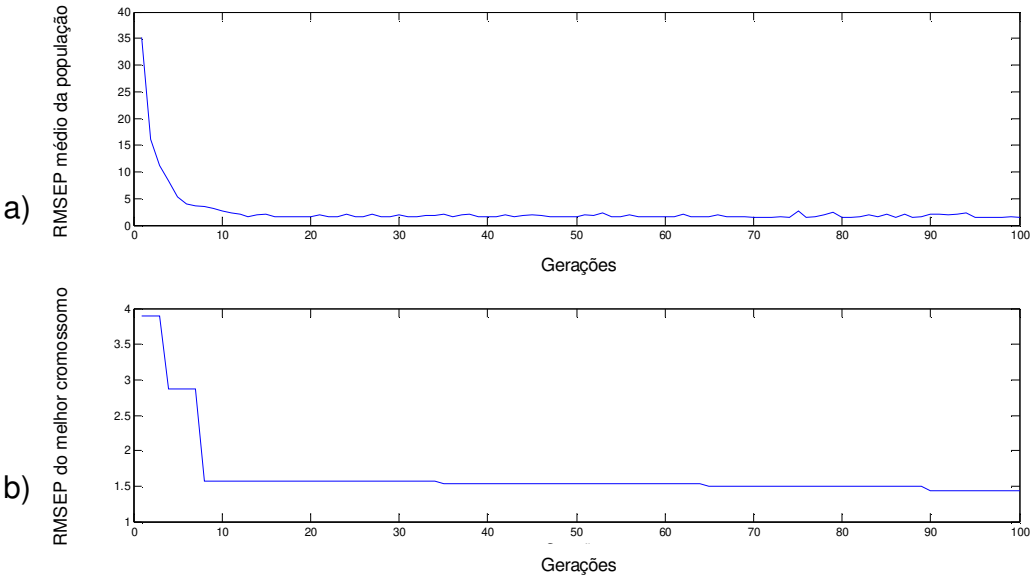


Figura 36 – Evolução do RMSEP utilizando o GA na determinação de AAS no produto **JMD**, utilizando o modelo BLLS com quatro fatores: a) média de todos cromossomos da população a cada geração e; b) melhor cromossomo a cada geração.

CAPÍTULO 7

CONCLUSÕES

Pode-se concluir, de maneira geral, que quando o AG foi utilizado, verificou-se a minimização do RMSEP para a maioria dos modelos, em todas as aplicações. Porém, em qualquer um dos casos apresentados, a otimização do AG foi condicionada à presença de uma amostra onde estivesse presente o analito e o interferente, isso porque o AG necessita saber onde está o interferente, para desta maneira selecionar as variáveis mais adequadas para a minimização dos erros. É importante ressaltar que este interferente, qualquer que seja, não é calibrado.

Na determinação dos pesticidas por HPLC-DAD, a seleção de variáveis pelo AG utilizando o modelo BLLS-RBL encontrou resultados melhores para todos os pesticidas.

Na determinação de riboflavina e piridoxina por espectrofluorimetria de excitação/emissão, os modelos BLLS-RBL e NPLS apresentaram diminuição dos erros quando utilizados com as variáveis selecionadas pelo AG. Já o modelo PARAFAC não apresentou ganhos com o AG, obtendo resultados melhores e piores. Os altos erros apresentados nas tabelas 6 a 11 para a determinação da piridoxina com a utilização de um fator, são decorrentes da presença de forte interferência, onde a utilização de um fator não possibilita a criação de um modelo consistente ainda que o AG seja utilizado. Como se pôde visualizar nas mesmas tabelas, quando se adicionou mais um fator, os erros diminuíram consideravelmente, principalmente quando o AG foi utilizado. Todos os melhores modelos para quantificar a piridoxina e a riboflavina foram obtidos quando o AG foi utilizado.

Também na determinação de AAS e AA em medicamentos, não se obteve melhoras nos erros ao se utilizar o modelo PARAFAC com as variáveis selecionadas pelo AG, porém para o NPLS e o BLLS-RBL houve a diminuição dos erros em quase todos os casos quando o AG foi utilizado, de maneira que também nesta aplicação todos os melhores modelos foram obtidos utilizando o AG.

A eficiência do método proposto está diretamente ligada, aos casos em que não foram adicionados fatores para modelar as interferências e ainda assim obtendo bons resultados, como na aplicação de AG com NPLS e BLLS-RBL e 4 fatores para prever o AAS nos medicamentos JMD e KMD.

A interpretação das variáveis selecionadas é complexa porque estão envolvidos processos onde o erro é minimizado matematicamente, havendo a possibilidade de se perder o sentido químico das variáveis selecionadas, ou seja, a seleção não está condicionada à uma solução trivial como encontrar um máximo de absorção num espectro. Ainda assim foi possível verificar em certos casos um sentido químico das variáveis escolhidas, tal como regiões de maior informação, ou sem interferentes.

Os resultados obtidos nestas três aplicações sugerem que o AG é uma ferramenta útil de seleção de variáveis em métodos de calibração de segunda ordem, tal como BLLS-RBL e NPLS e em alguns casos também com PARAFAC, apresentando uma efetiva diminuição de erros num processo de calibração, mesmo na presença de interferentes.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] D. E. Goldberg. Genetic Algorithm in Search, Optimization, and Machine Learning. Addison-Wesley, New York (1989).

[2] S. Forrest. Science, 261 (1993) 872.

[3] A. S. Bangalore, R. E. Shaffer, G. W. Small, M. A. Arnold. Anal. Chem., 68 (1996) 4200.

[4] Q. Ding, G. W. Smal, M. A. Arnold. Anal. Chem., 70 (1998) 4472.

[5] P. A. Costa Filho, R. J. Poppi. Anal. Chim. Acta, 446 (2001) 39.

[6] M. A. A. Lomillo, O. Renedo, M. J. A. Martinéz. Anal. Chim. Acta, 449 (2001) 167.

[7] J. Ghasemi, A. Niazi, R. Leardi. Talanta, 59 (2003) 311.

[8] A. Abdollahi, L. Bagheri. Anal. Chim. Acta, 514 (2004) 211.

[9] H. A. D. Filho, E. S. O. N. Souza, V. Visani, S. R. R. C. Barros, T. C. B. Saldanha, M. C. U. Araújo, R. K. H. Galvão. J. Braz. Chem. Sos., 16 (2005) 58.

[10] J. Ghasemi, D. M. Ebrahimi, L. Hejazi, R. Leardi, A. Niazi. J. Anal. Chem., 61 (2006) 92.

[11] V. G. Franco, J. C. Perín, V. E. Mantovani, H. C. Goicoechea. Talanta, 68 (2006) 1005.

[12] W. Wu, Q. Guo, D. L. Massart, C. Boucon, S. de Jong. Chemom. Intell Lab. Syst., 65 (2003) 83.

[13] J. A. Lopes, J. C. Menezes. Chemom. Intell Lab. Syst., 68 (2003) 75.

[14] S. Gourvénec, X. Capron, D. L. Massart. Anal. Chim. Acta, 519 (2004) 11.

[15] K. S. Booksh, B. R. Kowalski. Anal. Chem., 66 (1994) 782A.

[16] D. A. Burns, E. W. Ciurczak. Handbook of Near-Infrared Analysis, 2ª ed. Marcel Dekker, Ney York (2001).

[17] R. Bro. Chemom. Intell. Lab. Syst., 38 (1997) 149.

[18] R. Bro. Multi-way Analysis in the Food Industry, Ph.D. thesis. Amsterdan (1998).

[19] M. Linder, R. Sundberg. Chemom. Intell. Lab. Syst., 42 (1998) 159.

[20] M. Linder, R. Sundberg. J. Chemom., 16 (2002) 12.

[21] P. C. Damiani, A. J. Nepote, M. Bearzotti, A. C. Olivieri. Anal.Chem., 76 (2004) 2798.

[22] N. R. Marsili, A. Lista, B. S. F. Band, H. C. Goicoechea, A. C. Olivieri. Analyst, 130 (2005) 1291.

[23] H. C. Goicoechea, A. C. Olivieri. Appl.Spectrosc., 59 (2005) 926.

[24] A. Haimovich, R. Orselli, G. M. Escandar, A. C. Olivieri. Chemom. Intell. Lab. Syst., 80 (2006) 99.

[25] N. M. Faber, J. Ferré, R. Boqué, J. H. Kalivas. Chemom. Intell. Lab. Syst., 63 (2002) 107.

[26] J. Öhman, P. Geladi, S. Wold. J. Chemom., 4 (1990) 79.

[27] http://www.models.kvl.dk/source/nwaytoolbox/index.asp, acessado em 01/2006.

[28] J.W.B. Braga, C.B.G. Bottoli, I.C.S.F. Jardim, H.C. Goicoechea, A.C. Olivieri, R.J. Poppi. J. Chromatogr. A, 1148 (2007) 200.

[29] http://www.codexalimentarius.net/mrls/pestdes/jsp/pest_q-e.jsp acessado em 21/06/2005.

[30] B.J. Prazen, R.E. Synovec, B.R. Kowalski. Anal. Chem., 70 (1998) 218.

[31] G. G. Guilbault. Practical Fluorescence, 3º ed. Marcel Dekker, New York (1990).

[32] M.G. Trevisan. Aplicação de Métodos Quimiométricos de Ordem Superior e Fluorescência Molecular na Análise em Matrizes Biológicas. Dissertação de mestrado. Campinas (2003).

[33] R. Bianchini, M.D.V.C Penteado. Ciênc. Tecnol. Aliment., 20 (3) (2000).

[34] E. A. G. Zagatto, C. C. Oliveira, C. H. Collins. Quím. Nova, 22 (1) (1999) 143.

[35] http://gia.iqm.unicamp.br/, acessado em 05/2006.

[36] J. C. B. Fernandes, G. O. Neto, L. T. Kubota. Anal. Chim. Acta, 366 (1998) 11.

[37] A. C. B. Dias, E. P. Borges, E. A. G. Zagatto, P. J. Worsfold. Talanta 68 (2006) 1076.

[38] J.T. Franeta, D. Agbaba, S. Eric, S. Pavkov, M. Aleksic, S. Vladimirov. Il Farmaco 57 (2002) 709.

[39] United States Pharmacopeia (2005) 178.

ANEXO A

Available online at www.sciencedirect.com
ScienceDirect
Analytica Chimica Acta 595 (2007) 51–58
ANALYTICA CHIMICA ACTA
www.elsevier.com/locate/aca

Application of genetic algorithm for selection of variables for the BLLS method applied to determination of pesticides and metabolites in wine

Renato L. Carneiro, Jez W.B. Braga, Carla B.G. Bottoli, Ronei J. Poppi *

Universidade Estadual de Campinas, Instituto de Quimica, C.P. 6154, 13084-971 Campinas, SP, Brazil

Received 15 October 2006; received in revised form 13 December 2006; accepted 14 December 2006
Available online 19 December 2006

Abstract

A variable selection methodology based on genetic algorithm (GA) was applied in a bilinear least squares model (BLLS) with second-order advantage, in three distinct situations, for determination by HPLC–DAD of the pesticides carbaryl (CBL), methyl thiophanate (TIO), simazin (SIM) and dimethoate (DMT) and the metabolite phthalimide (PTA) in wine. The chromatographic separation was carried out using an isocratic elution with 50:50 (v/v) acetonitrile:water as mobile phase. Preprocessing methods were performed for correcting the chromatographic time shifts, baseline variation and background. The optimization by GA provided a significant reduction of the errors, where for SIM and PTA a decrease of three times the value obtained using all variables, and an improvement in the distribution of them, reducing the observed bias in the results were observed. Comparing the RMSEP of the optimized model with the uncertainty estimates of the reference values it is observed that GA can be a very useful tool in second-order models.

Keywords: BLLS; Genetic algorithm; Pesticides; Wine; HPLC–DAD

1. Introduction

Genetic algorithms (GA) are methods of numerical optimization that simulate biological evolution based on the Darwin theory and are widely used in many situations for variable selection. The selection operates on strings of binary digits stored in the computer memory, and over time, the functionality of these strings evolves in much the same way that natural populations of individuals evolve. Although the computational settings are highly simplified compared with the natural world, GA are capable of evolving surprisingly complex and interesting structures [1]. In calibration processes employed for determination of a property of interest in a system, often the use of a few variables that contain more information can provide enhancement in the interpretation of the model, beyond eliminating noise and nonlinearity. GA constitutes a valuable tool for this purpose by the appropriate use of an optmization function [2].

In analytical chemistry GA have been applied in several papers [3–11] with first-order calibration methods, such as partial least squares regression (PLS) and multiple linear regression (MLR), with the purpose of selecting the most relevant variables to acquire a better estimate of the concentration of some compound of interest in a sample. GA application with second-order calibration is recent. To the best of our knowledge only three papers have been published [12–14] where GA was applied to select the best N-way subset that keeps the structure information of the multiway data set of the PARAFAC method [12], selecting a better set of batches to include in the model calibration in N-PLS [13] and selecting the best NIR wavelengths for resolution purposes in a polymerization reaction study [14]. However, none of these papers evaluated GA optimization of the prediction ability of the models when they were built with only standard solutions and samples analyzed with the occurrence of unknown compounds, which are one of the main characteristics of second-order calibration methods, know as the second-order advantage.

The aim of this work was to apply and verify the main advantages in employing GA for variable selection in a relatively recently proposed bilinear least squares method (BLLS) when the second-order advantage is active. For this purpose, three distinct situations were studied in the determination of pesticides and a metabolite in red wine by HPLC–DAD. In these three

* Corresponding author. Tel.: +55 19 35213126; fax: +55 19 35213023.
E-mail address: ronei@iqm.unicamp.br (R.J. Poppi).

0003-2670/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.aca.2006.12.023

situations overlap of matrix interferences with the compounds of interest was present. These compounds were the pesticides carbaryl (CBL), methyl thiophanate (TIO), simazin (SIM) and dimethoate (DMT) and the metabolite phthalimide (PTA). The performances of the models were compared mainly based on their accuracies, expressed by the root mean square errors of prediction (RMSEP), and relative errors of prediction (REP), as well as, by the sensitivity and bias of the results.

1.1. Bilinear least squares (BLLS)

Bilinear least squares (BLLS) is a second-order calibration methodology that has been recently introduced in the second-order scenario [15,16], and has been demonstrated to provide analytical results comparable with PARAFAC in complex samples [17]. It has also been shown to work adequately with analytes presenting equilibrium species (linear dependent systems) [18,19]. In BLLS, the analyte concentration is introduced into the decomposition step, where only matrices of standards are present, in order to obtain approximations of pure-analyte matrices at unit concentration, ($\mathbf{S}_n$). For each sample measured in a HPLC–DAD system, a data matrix formed by J times and K wavelengths is obtained. When all I calibration standards are stacked on top of each other, a three-way array $\mathbf{X}$, with dimensions $I \times J \times K$, is formed. To estimate $\mathbf{S}_n$, the calibration data are first vectorized and joined into a $JK \times I$ matrix $\mathbf{V}_x$ [20,21]:

$$\boldsymbol{V}_x = [\mathrm{vec}(X_1)|\mathrm{vec}(X_2)|\ldots|\mathrm{vec}(X_I)] \quad (1)$$

where "vec" indicates the unfolding operation. Then a direct least squares procedure is used to obtain the pure-analyte information [20,21]:

$$\boldsymbol{V}_s = \boldsymbol{V}_x\, \boldsymbol{y}^{T+} \quad (2)$$

where "T" and "+" superscript are the transpose and pseudo-inverse operations, respectively, and $\mathbf{y}$ is the vector of the reference concentrations. If more than one analyte is present, $\mathbf{y}$ will be a matrix $\mathbf{Y}$ with dimensions $I \times N_c$, where N_c is the number of calibrated analytes. $\mathbf{V}_s$ then contains the required $\mathbf{S}_n$ matrices in vectorized form:

$$\boldsymbol{V}_s = [\mathrm{vec}(S_1)|\mathrm{vec}(S_2)|\ldots|\mathrm{vec}(S_{N_c})] \quad (3)$$

To obtain the chromatographic and spectral profiles presented in the $\mathbf{S}_n$ matrices, singular value decomposition (SVD) is employed [15,16]. The component profiles are obtained by single component singular value decomposition (SVD_1) of each $\mathbf{S}_n$ matrix, obtained after appropriate reshaping of the unfolded vec($\mathbf{S}_n$) [15,16]:

$$(\boldsymbol{b}_n, \boldsymbol{g}_n, \boldsymbol{c}_n) = \mathrm{SVD}_1(\boldsymbol{S}_n) \quad (4)$$

where $\mathbf{g}_n$ is the first singular value, and $\mathbf{b}_n$ and $\mathbf{c}_n$ are the first left and right singular vectors of $\mathbf{S}_n$, respectively. The concentrations in a unknown sample (whose matrix data are $\mathbf{X}_u$) are estimated, provided that no interference occurs, by a direct least squares procedure [15,16,20,21]:

$$\boldsymbol{y}_u = \boldsymbol{S}_{cal}^{+}\, \mathrm{vec}(\boldsymbol{X}_u) \quad (5)$$

where $\mathbf{y}_u$ is the $1 \times N_c$ estimated concentration vector of the N_c analytes in $\mathbf{X}_u$, and $\mathbf{S}_{cal}$ is a calibration $JK \times N_c$ matrix given by:

$$\boldsymbol{S}_{cal} = [g_1(c_1 \otimes b_1)|g_2(c_2 \otimes b_2)|\ldots|g_{N_c}(c_{N_c} \otimes b_{N_c})] \quad (6)$$

where $\otimes$ indicates the Kronecker product.

When the calibrated analytes produce signals which are overlapped with those for interferences present in $\mathbf{X}_u$, a separate residual bilinearization (RBL) process is employed to find the interference profiles which are incorporated into an expanded version of $\mathbf{S}_{cal}$:

$$\boldsymbol{S}_{int} = [\boldsymbol{S}_{cal}|\boldsymbol{g}_{int}(\boldsymbol{c}_{int} \otimes \boldsymbol{b}_{int})] \quad (7)$$

where $\mathbf{g}_{int}$, $\mathbf{b}_{int}$ and $\mathbf{c}_{int}$ are obtained by SVD of a residual matrix ($\mathbf{E}_u$) computed while fitting the data to the sum of the various component contributions:

$$\boldsymbol{E}_u = \boldsymbol{X}_u - \sum_{n=1}^{N_c} \boldsymbol{g}_n \boldsymbol{b}_n (\boldsymbol{c}_n^{T}) \boldsymbol{y}_{u,n} \quad (8)$$

$$(\boldsymbol{b}_{int}, \boldsymbol{g}_{int}, \boldsymbol{c}_{int}) = \mathrm{SVD}_1(\boldsymbol{E}_u) \quad (9)$$

The RBL process can be performed by an iterative method [15,20,22] or by a Gauss–Newton minimization procedure [18,20]. It is important to note that in the BLLS model no initialization or constraining procedures are required, and that the second-order advantage is acquired by the RBL analysis of the residual matrix $\mathbf{X}_u$. The number of interferences present can be estimated by comparison of the residuals left out by the model in a prediction sample with the residuals in the calibration samples or with the instrumental noise level (obtained by suitable blank replication).

1.2. Figures of merit

The estimation of figures of merit is an active area of research in chemometrics, and these parameters are regularly employed for method comparison. For multivariate calibration these estimates are based on the concept of net analyte signal (NAS), first developed by Lorber [23]. For second or higher-order multivariate calibrations, two independent approaches to NAS computation were developed by Messick et al. [24] and by Ho et al. [25]. Recently, a general expression was derived to estimate the sensitivity of second-order bilinear calibration models, such as PARAFAC and BLLS, taking into account whether the second-order advantage is required or not [26]. Following this last approach, the sensitivity can be obtained as [26]:

$$\mathrm{SEN}_n = z_n \{[(B_{exp}^{T} P_{b,unx} B_{exp})(C_{exp}^{T} P_{c,unx} C_{exp})]^{-1}\}^{-1/2} \quad (10)$$

where B_{exp} and C_{exp} are the chromatographic and spectral profiles, respectively, for the calibrated analytes (provided by the PARAFAC and BLLS models); $P_{b,unx}$ and $P_{c,unx}$ are projection matrices, orthogonal to the space spanned by all unexpected components in each mode [26]:

$$P_{b,unx} = I - B_{unx} B_{unx}^{+} \quad (11)$$

$$P_{c,unx} = I - C_{unx}C_{unx}^{+} \quad (12)$$

and z_n is the g_n value obtained in Eq. (4). The SEN values depend on the presence of interferences and are sample-specific. Therefore, SEN cannot be defined for the whole multivariate method. In such cases, an average value for a set of samples can be estimated and reported.

The limit of detection (LOD) is an important figure of merit that has recently been discussed for several first and second-order multivariate techniques [27–29]. An approximation to the LOD can be obtained by the expression [17,29]:

$$\mathrm{LOD}_n = 3.3\frac{s_r}{\mathrm{SEN}_n} \quad (13)$$

where s_r is an estimative of the instrumental noise. Since the SEN is given as an average value, LOD is also reported as an average figure.

Another important figure of merit to be estimated is the standard error in the estimated concentrations, an active area of research in the second-order scenario. Mathematical expressions for sample-specific prediction uncertainty show consistent results in simulated data, and they are available for BLLS [15] models when they are not exploiting the second-order advantage. Hence, they are not applicable when a real sample such as wine is analyzed. An useful alternative for method comparison is to estimate a mean prediction error for a set of test samples. This can be achieved by the well-known parameter root mean square error of prediction (RMSEP):

$$\mathrm{RMSEP} = \sqrt{\sum_{n=1}^{I}\frac{(y_{ref,i} - y_{u,i})^2}{I}} \quad (14)$$

where y_{ref} is the reference concentration value for each of the I test samples. From RMSEP, a relative error of prediction (REP) can be obtained as [30]:

$$\mathrm{REP} = \sqrt{\sum_{n=1}^{I}\frac{(y_{ref,i} - y_{u,i})^2}{Iy_{ref,i}^2}} \times 100 \quad (15)$$

1.3. Genetic algorithm (GA)

The basic operations of GA involve five steps: codification of the variables, creation of the initial population, evaluation of every chromosome, crossing and mutation. The implementation of GA in the selection of variables is different from the applications normally carried out, in that it refers to codification of the problem and the response function, since the other stages remain unchanged. In the codification of the problem and selection of variables, it is considered that the chromosome has "p" genes, where each gene represents one of the variables of the analytical signal (i.e. spectra). Then, the chromosome will have the same number of variables as contained in this signal. In the selection of variables the binary code (0, 1) is used to codify the problem. Each gene can assume the value 1 or 0. If this gene is "0" the variable is not selected. Otherwise, if its value is "1", the variable is selected. Fig. 1 shows the codification of a chromosome

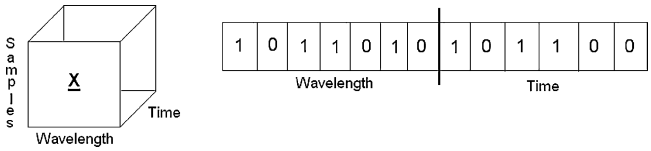


Fig. 1. Codification of the chromosome for variable selection in second-order calibration.

for the selection of variables in a second-order calibration with a HPLC–DAD system.

The GA program for selection of variables was built following the stages explained in detail below:

(a) Firstly, an initial population was created by a random generation of the matrix **R** with values 1 and 0, where each line is a chromosome and each column represents a variable. In second-order methods, there are two species of variable, which represent each dimension of the data matrix (i.e. spectral and time profiles), in this way the number of columns of **R** will be the sum of all variables in the time and spectral dimensions (Fig. 1). In this work an initial population of 100 chromosomes was used, and the initial number of selected variables was imposed to be around 10%.

(b) Each chromosome is constituted of two parts, corresponding to the variables of the first and second dimension, respectively. Then, the variables of the two dimensions that had been joined in the random generation are now separated. In all stages of the GA (generation, crossing and mutation) the variable in both dimensions will be joined and then separated again in the evaluation stage, as illustrated in Fig. 2. After the separation of the chromosome in $\mathrm{chromosome}_{way1}$ and $\mathrm{chromosome}_{way2}$, the data tensor has its dimensions reduced by elimination of variables in accordance with the value of their respective genes (0 or 1). Then the BLLS model is built for each chromosome and evaluated and the performance of the population is organized by the increase of the RMSEP for posterior crossings.

(c) The crossing is the most important stage of GA. Here, two chromosomes previously evaluated are combined to give origin to two new chromosomes. The crossing is carried out following a fixed order, and s based on the sequence that was established in the evaluation stage, random crossing is not used to avoid a possible precocious convergence of the population. The crossing between two chromosomes was

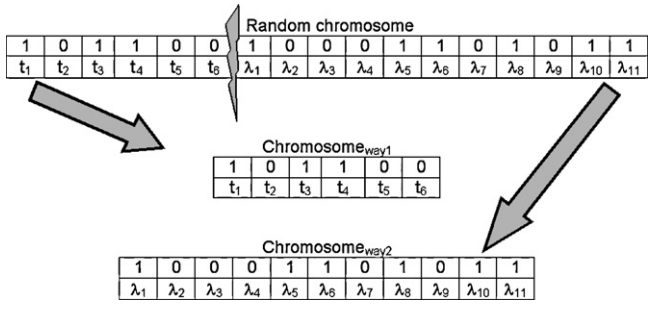


Fig. 2. Illustration of the division of a chromosome for the variable selection process in second-order calibration.

carried out by changing approximately 50% of its genes, by combining the two chromosomes involved with a vector randomly generated called "mask vector" that defines which genes will be changed (i.e. those genes that have 1 in mask). To generate the new population, the chromosomes were crossed in the following way, for a population with 100 chromosome: 1st with 36th, 2nd with 37th, ..., 35th with 70th. Each crossing originates two new chromosomes. The stage of crossing occurs with 90% probability, and when it does not occur, the old chromosomes are preserved in the new population. This process results in 70 chromosomes for the new population, the others 30 chromosomes are obtained by crossing the 15 better chromosomes with another 15 randomly chosen chromosomes. In this way, 100 chromosomes are generated in the new population.

(d) The mutation occurs after the crossings, and only a small percentage of chromosomes (around of 1%) suffer mutation in some of their genes (1% of the genes on average). This operation eliminates the possibility of all chromosomes having a gene with the same value (1 or 0), which will result in a gene where no possible crossing could modify it, leading to a less heterogeneous system.

(e) In this last stage, BLLS models are build for the new population (originating by crossing/mutation), as explained in 'b' and the algorithm repeats this process until a pre-established number of iterations or generations is reached. In all optimizations 100 generations were used.

In order to obtain robustness and easy of result interpretation, instead of using each gene separately, they were joined in groups of five genes. Using this artifice the isolated variables (genes) are eliminated, since they neither contribute nor disturb the model. Moreover, the number of possibilities of results is reduced, leading to assist in the convergence and the precision of GA.

In all situations two replicates of different samples were used for GA optimizations. This procedure was necessary since the GA should take into account interference in the samples to select the variables that will provide the lowest error of prediction in these replicates. For the situations where a BLLS model is built to predict two analytes simultaneously, the GA was optimized for each analyte individually.

2. Experimental

2.1. Apparatus

The HPLC system consisted of a Shimadzu VP Series Liquid Chromatograph equipped with a SIL-10AXL autosampler, a model LC-10ATVP solvent pump and an SPD-10AVP DAD. The data were acquired and exported with ClassVP software, Version 6.1. A Novapack C18 (4 μm) column (150 mm × 4.6 mm i.d.) from Waters and a similar guard column were used for the separations.

In the isocratic method (IM), the separations were carried out with 50:50 (v/v) acetonitrile:water as mobile phase, the water was acidified to pH 3.0 with phosphoric acid before mixing and a flow rate of 0.60 mL min^{-1} was used.

Table 1
Concentration ranges, in μg mL^{-1}, used for the analytes of interest

Analyte	Range
DMT	1.00–7.50
PTA	0.10–1.40
TIO	0.50–5.37
SIM	0.10–1.24
CBL	1.00–6.00

2.2. Reagents and standards

The solvents for preparation and chromatographic analysis were acetonitrile (HPLC-grade, Tedia), water (Milli-Q, Millipore), phosphoric acid (Merck), ethyl acetate (Tedia), methanol (HPLC-grade, Tedia) and isopropanol (Merck). They were filtered using a 0.45 μm poly(vinylidene) fluoride (PVDF) membrane (Millipore). Pesticides standards were simazine (SIM) (98.3%) obtained from Novartis, carbaryl (CBL) (99.8%) from Supelco, methyl thiophanate (TIO) (98.5%) and dimethoate (DMT) from Riedel-de-Häen. The metabolite was phthalimide (PTA) (99.9%) from Riedel-de-Häen.

Stock solutions of each analyte were prepared with acetonitrile in the following concentrations: 1046 μg mL^{-1} PTA, 1077 μg mL^{-1} CBL, 1028 μg L^{-1} TIO, 1011 μg mL^{-1} DMT and 402.8 μg mL^{-1} for SIM. Intermediate solutions of each analyte were obtained by appropriate dilutions with a solution 50:50 (v/v) acetonitrile:water of the stock solutions yielding 20.15 μg mL^{-1} for PTA, 105.7 μg mL^{-1}CBL, 39.63 μg mL^{-1} TIO, 99.21 μg mL^{-1} DMT and 19.40 μg mL^{-1} SIM. For analytes PTA, TIO and SIM two dilutions were performed. These solutions were stored at 4 °C in the dark.

For model development, six calibrations standards consisted of a mixture of all interest compounds were prepared daily, covering the analytical range presented in Table 1 and with concentrations distributed equally. The concentration ranges were established by preliminary runs with each analyte, obtaining the area of the chromatographic peak using the isocratic method (IM) at 220 nm, or following the recommendations of the Codex Alimentarious for maximum limits of residuals [31].

2.3. Wine samples

Wine samples were Juan Carrau red wine from Santana do Livramento, Rio Grande do Sul, Brazil. This wine was obtained from grapes that had not been treated with synthetic pesticides. Each wine sample was submitted to a solid phase extraction procedure (SPE) for cleanup. The SPE method employed 1.00 mL Oasis HLB cartridges (purchased from Waters) that were first conditioned with 2.50 mL of methanol and 2.50 mL of water. Then 2.50 mL of wine were added and allowed to percolate slowly. The cartridge was then washed with 1.50 mL of a 2% (v/v) isopropanol solution and dried for 20 min. The pesticides were directly eluted with 3.00 mL of ethyl acetate to a

laboratory-made Florisil cartridge, allowed to percolate through the cartridges under positive pressure, and collected in an assay tube. The solvent was evaporated to dryness at room temperature under a nitrogen stream. The dry sample was redissolved with 1.00 mL of acetonitrile, obtaining a concentration factor of 2.5. Finally, the solution was transferred to a vial for analysis.

Six extracts were spiked with the compounds of interest and analyzed by the isocratic procedure. For a better evaluation of the prediction errors of the second-order models, the analytes were spiked into the extracts after the SPE phase, therefore no loss in the SPE method was considered in these six samples. A volume of 100 μL was used to spike the analytes into the extract; therefore the interferences in these samples were diluted by 10%. All samples and standards were analyzed in duplicate.

2.4. Software

All calculations were performed using Matlab 6.5 [32]. The BLLS, time shift, baseline correction and GA routines were developed in our laboratory.

3. Results and discussion

Fig. 3 presents the chromatogram detected at 220 nm between 2 and 5 min. In the six calibration standards DMT and PTA are highly overlapped presenting just one peak, at approximately 2.7 min. For TIO and SIM a sight overlap is also observed, while CBL is resolved. It is also observed that when a wine sample is analyzed, interferences overlap with DMT, PTA TIO, SIM and CBL, providing three distinct situations where the application of BLLS is necessary. It is also observed in Fig. 3 that the wine sample presents a time shift in comparison with the calibration standards. This kind of deviation must be corrected before model development, since this produces a deviation of the bilinear structure of the data matrices. The time shifts were corrected in all situations based on the procedure proposed by Prazen et al. [33], where a data matrix N (taken as a reference) is moved in relation to another data matrix M until the minimum

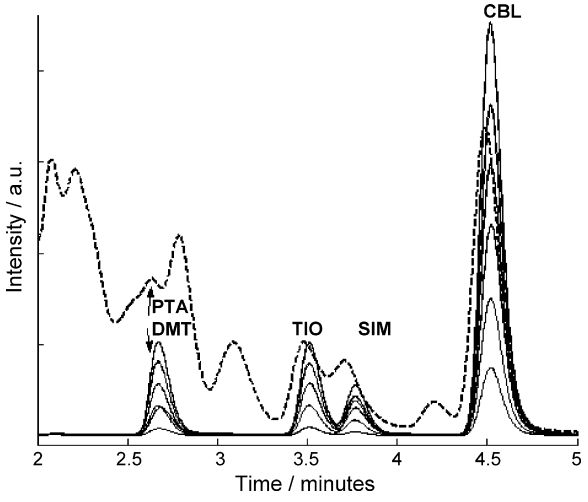


Fig. 3. Chromatogram at 220 nm in the region of the compounds of interest, (solid) calibration standards, (dashed) a wine sample.

residue of the singular value decomposition of a joint matrix $N|M$. For DMT and PTA an additional procedure for background correction was necessary due the high interference present. This correction was accomplished by subtracting a blank sample, weighted to avoid negative signals, from the wine sample.

Fig. 4 presents the plots of a wine sample in contour line for each analyte, where the variables selected by Gaare shown. For these five optimizations the GA provided a variable number reduction of approximately 80%, 90%, 85%, 95% and 80% for CBL, SIM, TIO, PTA and DMT, respectively.

For CBL, which represents the simplest situation, it is observed that 65% of the selected variables are in the time dimension, distributed in three main regions, in the beginning, middle and end of this dimension. On the other hand, the selected variables in the spectral dimension were distributed over all wavelengths. These results reflect an important aspect for variable selection in second-order calibration methods employing the second-order advantage that is the need to select some representative variables for each significant component to estimate the analyte and interference profiles and make possible determination when interferences are not present in the calibration. Due to this fact, in the GA optimization it is necessary to include some samples with interferences present.

For the other four analytes, the interpretation of the selected variables is more difficult, since they are two pairs of compounds of interest, one partially and the other highly overlapped (Fig. 4). For these four analytes it was observed that GA selected a larger number of variables in the spectral dimension than in the time dimension, which can be understood from the higher importance of this dimension for the determination of these analytes. For the optimization for SIM it is interesting to note that GA did not select the variables in the time dimension around the maximum of absorbance for this analyte but did select the maximum for TIO. By the way, the selected variables in the spectral dimension agree with the main signals for both analytes. For DMT and PTA, it was observed that DMT required a larger number of variables in the time dimension than PTA, which can be explained by the fact that this analyte presents a less characteristic UV spectrum. The previous observations about the variables selected by GA suggest that when GA is optimized to provide a minimum RMSEP the main factors that probably affect the selected variables are: the profile estimations and the relation between them.

Fig. 5 presents the variation of the sensitivity values in each situation for the six wine samples in duplicate. It is observed that for all analytes there is a decrease in the sensitivities, which represents an approximate decrease of four times for CBL, SIM, TIO and PTA, respectively, and 1.5 for DMT, in relation to the model developed with all variables. The limits of detection observed for the analytes were: 0.04, 0.05, 0.13, 0.04 and 0.17 $\mu g\,mL^{-1}$ with the selected variables and 0.01, 0.01, 0.04, 0.01 and 0.13 $\mu g\,mL^{-1}$ for models with all variables (for CBL, SIM, TIO, PTA and DMT, respectively), showing higher limits of detection, corresponding to sensitivity decrease. These results are also observed in first-order calibration, when optimization methods, such as GA, are used. Its main cause is probably the optimization criterion used in the GA algorithm, since it selects

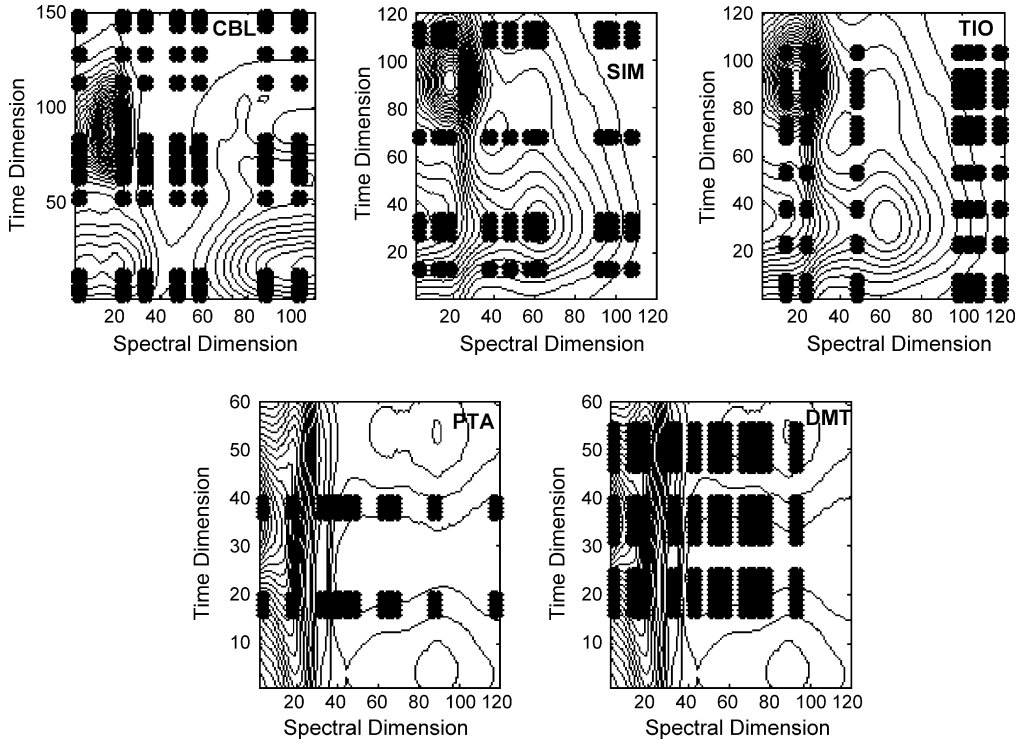


Fig. 4. Contour lines of a wine sample showing the variable selected by GA (black points) for the five compounds analyzed.

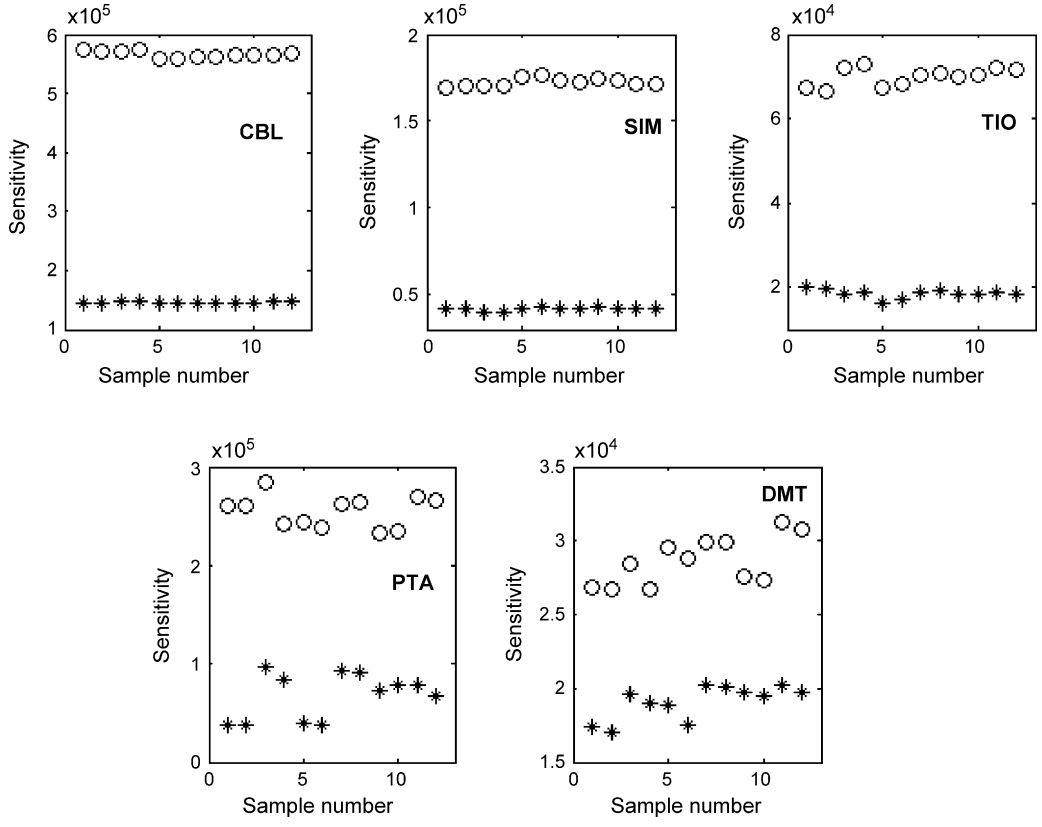


Fig. 5. Variation of the sensitivity between the samples when all variables are used in the model (◯) and for the variable selected by GA (*).

Table 2
Observed RMSEP and mean recovery values for each analyte of interest in the wine samples, when the BLLS method is applied with all variables and with the variable selected by GA

Pesticides	BLLS–RBL		BLLS–RBL–GA	
	Recoveries[a]	RMSEP[b]	Recoveries[a]	RMSEP[b]
CBL	98 (2)	0.09 (2.7)	101 (1)	0.09 (2.2)
SIM	92 (2)	0.08 (7.9)	99 (2)	0.02 (2.5)
TIO	98 (2)	0.13 (3.0)	100 (2)	0.07 (1.5)
PTA	105 (6)	0.08 (7.0)	102 (3)	0.03 (3.7)
DMT	101 (5)	0.27 (5.1)	102 (3)	0.19 (3.2)

[a] Recovery values in percent, with estimates of standard deviations in parenthesis.
[b] RMSEP in $\mu g\ mL^{-1}$ and percent relative errors of prediction (REP) in parenthesis.

the best variables that minimize the RMSEP, which may not include the most sensitive variables. The variation between the samples is due to the interferences in the sample, and it is more evident for PTA and DMT, where a more intense interference is observed.

In Table 2 the RMSEP, REP and recoveries values are presented for the variables selected by GA and for all variables used for model development. For SIM, TIO, PTA and DMT it was observed that the samples with lower concentrations present a significantly larger error in relation to the other samples. A comparison of the chromatograms at a fixed wavelength of this sample with the others shows that the time shift for this sample was not corrected due the low concentrations and the presence of interferences. This sample was then considered an outlier for these analytes. Therefore, the values presented in Table 2 were calculated based on six samples measured in duplicate just for CBL, and five samples for the other analytes where these data matrices were not used in the GA optimization. It can be observed from the RMSEP and REP values that there is an improvement in the results when the variables selected by GA were used for model development. The only exception is found for CBL, where no apparent improvement was observed. This can be explained by the error in the reference values in the wine samples that can be estimated by simple error propagation, considering all steps involved in sample preparation, which provide 0.08, 0.02, 0.08, 0.02 and 0.12 $\mu g\ mL^{-1}$ for CBL, SIM, TIO, PTA and DMT, respectively. Comparing these values with the RMSEP shown in Table 2, it can be observed that, for CBL using all variables or the selected variables the RMSEP already reach the limiting value of error in the reference values. For the other compounds this limit is not reached only for DMT. This result suggests that GA is a powerful tool for the minimization of the errors in the calibration model. In Table 2 the recoveries obtained for all analytes are also given, which are in a good agreement with the expected value of 100%.

In Fig. 6 the joint confidence regions based on bivariate least squares for the slope and intercept of the regression of predicted concentration versus reference values are presented [34]. It can be observed that for all BLLS models built with the selected values by GA the confidence regions contain the ideal point of unit and zero for slope and intercept, indicating that there is not significant bias for the prediction when GA was applied. By the way, using all variables only PTA and DMT contain this ideal point. However, the elliptic sizes indicate that for CBL, SIM and TIO the GA provide results lightly less precise for these

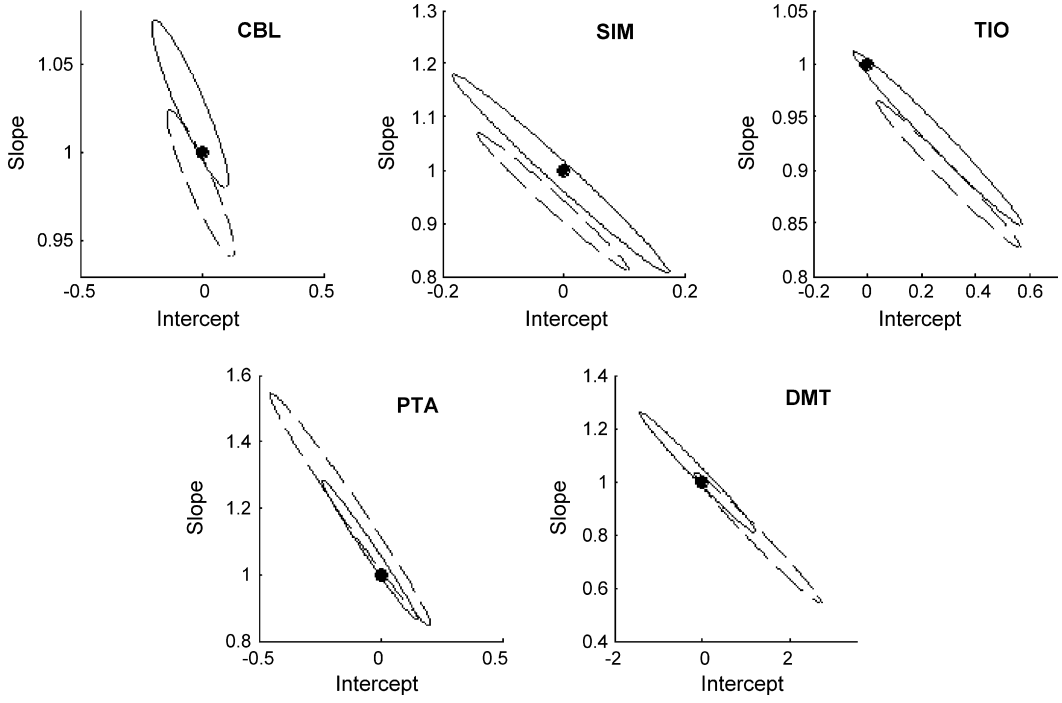


Fig. 6. Elliptical joint confidence regions for the slope and intercept of the regression of predicted concentrations vs. reference values in the spiked wine samples using bivariate least squares, (dashed) BLLS, (solid) BLLS–GA, (●) point of intercept equal to zero and slope equal to one.

compounds and more precise for PTA and DMT, compared with the elliptic sizes obtained with all variables.

4. Conclusions

The presented results suggest that GA can be a very useful tool for variable selection in second-order models, such as BLLS. Minimization of the error can be obtained even if the second-order advantage is active. However this requires the use of atleast some samples where interference is present for the optimization of GA. Decreases in the errors were observed for most of the analytes, where the observed RMSEP of the models built with GA reach the level of uncertainty estimated in the reference values. However, it is necessary a wider study, with a large number of samples in the test set, to confirm the observed results and conclusions about the performance of the GA in second-order calibration models.

The results for sensitivity showed that, as observed for first-order calibration, the sensitivity values obtained in BLLS models with selected variables can be lower than that obtained with all variables, leading to larger limits of detection. However, as the calibration models were used in a fixed range (delimited by the calibration standards) and in this region the errors obtained by GA were lower, the decrease of the sensitivity obtained with the selected variables is a not significant factor in the model. For applications where sensitivity is critical, GA can be optimized taking into account both sensitivity and the prediction errors by the used of a multivariate optimization function.

Acknowledgements

The authors acknowledge financial support and fellowships from the UNICAMP Graduate Instructors Program and from FAPESP (proc. 05/53280-4).

References

[1] S. Forrest, Science 261 (1993) 872.
[2] D.E. Goldberg, Genetic Algorithm in Search, Optimization and Machine Learning, Addison-Wesley, New York, 1989, p. 125.
[3] A.S. Bangalore, R.E. Shaffer, G.W. Small, M.A. Arnold, Anal. Chem. 68 (1996) 4200.
[4] Q. Ding, G.W. Smal, M.A. Arnold, Anal. Chem. 70 (1998) 4472.
[5] P.A. Costa Filho, R.J. Poppi, Anal. Chim. Acta 446 (2001) 39.
[6] M.A.A. Lomillo, O. Renedo, M.J.A. Martinéz, Anal. Chim. Acta 449 (2001) 167.
[7] J. Ghasemi, A. Niazi, R. Leardi, Talanta 59 (2003) 311.
[8] A. Abdollahi, L. Bagheri, Anal. Chim. Acta 514 (2004) 211.
[9] H.A. Dantas Filho, E.S.O.N. Souza, V. Visani, S.R.R.C. Barros, T.C.B. Saldanha, M.C.U. Araújo, R.K.H. Galvão, J. Braz. Chem. Soc. 16 (2005) 58.
[10] J. Ghasemi, D.M. Ebrahimi, L. Hejazi, R. Leardi, A. Niazi, J. Anal. Chem. 61 (2006) 92.
[11] V.G. Franco, J.C. Perin, V.E. Mantovani, H.C. Goicoechea, Talanta 68 (2006) 1005.
[12] W. Wu, Q. Guo, D.L. Massart, C. Boucon, S. de Jong, Chemometr. Intell. Lab. Syst. 65 (2003) 83.
[13] J.A. Lopes, J.C. Menezes, Chemometr. Intell. Lab. Syst. 68 (2003) 75.
[14] S. Gourvénec, X. Capron, D.L. Massart, Anal. Chim. Acta 519 (2004) 11.
[15] M. Linder, R. Sundberg, Chemometr. Intell. Lab. Syst. 42 (1998) 159.
[16] M. Linder, R. Sundberg, J. Chemometr. 16 (2002) 12.
[17] P.C. Damiani, A.J. Nepote, M. Bearzotti, A.C. Olivieri, Anal. Chem. 76 (2004) 2798.
[18] N.R. Marsili, A. Lista, B.S.F. Band, H.C. Goicoechea, A.C. Olivieri, Analyst 130 (2005) 1291.
[19] H.C. Goicoechea, A.C. Olivieri, Appl. Spectrosc. 59 (2005) 926.
[20] A. Haimovich, R. Orselli, G.M. Escandar, A.C. Olivieri, Chemometr. Intell. Lab. Syst. 80 (2006) 99.
[21] N.M. Faber, J. Ferré, R. Boqué, J.H. Kalivas, Chemometr. Intell. Lab. Syst. 63 (2002) 107.
[22] J. Öhman, P. Geladi, S. Wold, J. Chemometr. 4 (1990) 79.
[23] A. Lorber, Anal. Chem. 58 (1986) 1167.
[24] N.J. Messick, J.H. Kalivas, P.M. Lang, Anal. Chem. 68 (1996) 1572.
[25] C.N. Ho, G.D. Christian, E.R. Davidson, Anal. Chem. 52 (1980) 1071.
[26] A.C. Olivieri, N.M. Faber, J. Chemometr. 19 (2005) 583.
[27] R. Boqué, M.S. Larrechi, F.X. Rius, Chemometr. Intell. Lab. Syst. 45 (1999) 397.
[28] R. Boqué, J. Ferre, N.M. Faber, F.X. Rius, Anal. Chim. Acta 451 (2002) 313.
[29] A.C. Olivieri, N.M. Faber, Chemometr. Intell. Lab. Syst. 70 (2004) 75.
[30] N.J. Miller-Ihli, T.C. O'Haver, Spectrochim. Acta 39B (1984) 1603.
[31] http://www.codexalimentarius.net/mrls/pestdes/jsp/pest_q-e.jsp.
[32] MATLAB 6.5, The Mathworks Inc., Natick, MA, 2000.
[33] B.J. Prazen, R.E. Synovec, B.R. Kowalski, Anal. Chem. 70 (1998) 218.
[34] J. Riu, F.X. Rius, Anal. Chem. 68 (1996) 1851.