

REINALDO FRANCISCO TEÓFILO

Doutorado

**Métodos Quimiométricos em Estudos
Eletroquímicos de Fenóis sobre Filmes
de Diamante Dopado com Boro**

**Orientadora: Márcia Miguel Castro Ferreira
Co-Orientador: Lauro Tatsuo Kubota**

**Universidade Estadual de Campinas
Instituto de Química**

**Campinas - SP
Outubro de 2007**

**FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO INSTITUTO
DE QUÍMICA DA UNICAMP**

T264m	<p>Teófilo, Reinaldo Francisco. Métodos quimiométricos em estudos eletroquímicos de fenóis sobre filmes de diamante dopado com boro / Reinaldo Francisco Teófilo. -- Campinas, SP: [s.n], 2007.</p> <p>Orientadora: Márcia Miguel Castro Ferreira.</p> <p>Co-orientador: Lauro Tatsuo Kubota</p> <p>Tese - Universidade Estadual de Campinas, Instituto de Química.</p> <p>1. Quimiometria. 2. Eletroquímica. 3. Multimodos. 4. Fenóis. I. Ferreira, Márcia Miguel Castro. II. Kubota, Lauro Tatsuo. III. Universidade Estadual de Campinas. Instituto de Química. IV. Título.</p>
-------	--

Título em inglês: Chemometric methods in the electrochemical studies of phenols on boron-doped diamond films

Palavras-chaves em inglês: Chemometrics, Electrochemistry, Multiway, Phenols

Área de concentração: Química Analítica

Titulação: Doutor em Ciências

Banca examinadora: Márcia Miguel Castro Ferreira (orientadora), Claudia Longo (IQ-UNICAMP), Jarbas José Rodrigues Rohwedder (IQ-UNICAMP), Romeu Cardozo Rocha Filho (DQ-UFSCar), Roberto Kawakami Harrop Galvão (IEE-ITA)

Data de defesa: 02/10/2007

BANCA EXAMINADORA

Márcia Miguel Castro Ferreira – Orientadora / IQ-UNICAMP

Cláudia Longo / IQ-UNICAMP

Jarbas José Rodrigues Rohwedder / IQ-UNICAMP

Romeu Cardozo Rocha Filho / DQ-UFSCar

Roberto Kawakami Harrop Galvão / IEE-ITA

*“Nós somos o que repetidas vezes fazemos.
Excelência então, não é um ato, mas um hábito”
Aristóteles*

*Aos meus pais Paulo Teófilo e
Maria Elisa Migliorini Teófilo*

À minha namorada, Gilmare.

Agradecimentos

A Deus, por ter me presenteado com saúde e perseverança para realizar este trabalho, entre outras tantas coisas.

Aos meus pais, Maria Elisa Migliorini Teófilo e Paulo Teófilo, pelas orações, conselhos e apoio.

À Gilmare, que participou efetivamente de todos os acontecimentos desta escalada, sendo fundamental, pela sua afabilidade, dedicação, compreensão, companheirismo e, sobretudo, pelo júbilo de sua presença.

Aos meus irmãos, irmãs, sobrinhos, sobrinhas, cunhados e cunhadas que estiveram sempre presentes nos bons e maus momentos.

À Profa. Márcia M. C. Ferreira, pelos ensinamentos, dedicação, apoio, confiança, oportunidades e valiosas sugestões.

Ao Prof. Lauro T. Kubota, pelas oportunidades, ensinamentos, valiosas sugestões, apoio e atenção.

Ao Prof. Vitor Baranauskas, pela confiança, oportunidade, apoio, sugestões e pela frutífera cooperação.

Aos Pesquisadores Helder Ceragioli e Alfredo Peterlevitz pelo empenho em tornar possível a realização dos experimentos, além de toda amizade, confiança e atenção.

Ao Pesquisador Rudolf Kiralj pelas frutíferas discussões, cooperação e dedicação.

Ao companheiro João P. A. Martins, pelo auxílio, trocas de informações e frutíferas discussões.

Aos Pesquisadores Leonardo M. da Silva e Flávio Damos pelos auxílios e sugestões.

Aos membros da Banca, pela disponibilidade e dedicação.

Ao companheiro Juliano A. Andrade pela disposição e dedicação.

Aos Pesquisadores Rasmus Bro, Scott Ramos, Christopher D. Brown, Nicholas M. Faber e Ricard Boqué, pelos auxílios, informações, algoritmos e sugestões.

Aos membros do Laboratório de Quimiometria Teórica e Aplicada, pelo desvelo, auxílio e companheirismo.

Aos membros do Laboratório de Eletroquímica, Eletroanalítica e Desenvolvimento de Sensores, pelo companheirismo, auxílio, sugestões e colaborações.

À Universidade Estadual de Campinas e ao Instituto de Química, pelo apoio incondicional e pela oportunidade oferecida para a realização deste curso e deste trabalho.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo aporte financeiro, permitindo a dedicação à pesquisa e a oportunidade da contribuição científica.

Enfim, àqueles que embora não citados nominalmente, que nesta caminhada, em todos os momentos, me ajudaram, se não com ações, com exemplos e ensinamentos.

CURRICULUM VITAE

Reinaldo Francisco Teófilo

E-mail: rteofilo@gmail.com

1. FORMAÇÃO ACADÊMICA

- 08/2003 – 10/2007:** Doutor em Química – Área de concentração: Química Analítica. Universidade Estadual de Campinas, UNICAMP, São Paulo, Brasil. Bolsista do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).
- 08/2001 – 08/2003:** Mestre em Agroquímica – Área de concentração: Química Analítica. Universidade Federal de Viçosa, UFV, Minas Gerais, Brasil. Bolsista da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).
- 03/1996 – 08/2001:** Bacharel e Licenciado em Química – Bacharelado e Licenciatura. Universidade Federal de Viçosa, UFV, Minas Gerais, Brasil.

2. ATIVIDADES CIENTÍFICAS E ACADÊMICAS

- 10/2003 – 09/2004:** Representante dos estudantes na Coordenadoria de Pós-graduação do Instituto de Química da Unicamp. Carga horária quinzenal: 2h.
- 10/2002 – 06/2003:** Colaboração técnica e teórica para o desenvolvimento de uma metodologia para a determinação e especiação voltamétrica de Arsênio (III) e (V). Carga horária semanal: 20h. UFOP.
- 06/2002 – 07/2003:** Representante dos estudantes no Conselho de Pós-graduação do Curso de Mestrado em Agroquímica. Carga horária quinzenal: 2h. UFVViçosa.
- 10/2000 - 7/2001:** Tutor do programa de tutoria da UFVViçosa, Carga horária semanal: 12h.
- 9/1999 - 9/2000:** Aluno Bolsista de Iniciação Científica pelo CNPq/PIBIC. Carga horária semanal: 21h.
- 10/1998 - 12/1999:** Estagiário do Conselho Regional de Química de Minas Gerais - CRQ-MG. Carga horária mensal: 30h.

3. PRODUÇÃO CIENTÍFICA E TECNOLÓGICA

3.1. Resumos e trabalhos em congressos científicos

- 11 resumos publicados em anais de eventos internacionais;
- 22 resumos publicados em anais de eventos nacionais;
- 3 trabalhos completos publicados em anais de eventos internacionais;
- 3 trabalhos completos publicados em anais de eventos nacionais.

3.2. Artigos publicados em revistas científicas

1. C.M. Henrique, R.F. Teófilo, L. Sabino, M.M.C. Ferreira. Classification of cassava starch films by physicochemical properties and water vapor permeability quantification by FTIR and PLS. *J. Food Sci.* **2007**, 72(4), E184-E189.
2. K. Pasqualoto, R.F. Teófilo, Guterres M., F.S. Pereira. M.M.C. Ferreira. A study of physicochemical and biopharmaceutical properties of Amoxicillin tablets using full factorial design and PCA biplot. *Anal. Chim. Acta*, **2007**, 595, 216-220.
3. R.F. Teófilo, H.J. Ceragioli, A.C. Peterlevitz, L.M. da Silva, F.S. Damos, M.M.C. Ferreira, V. Baranauskas, L.T. Kubota, Improvement of the electrochemical properties of “as-grown” boron-doped polycrystalline diamond electrodes deposited on tungsten wires using ethanol. *J. Solid State Electrochem.* **2007**, 11, 1449-1457.

4. R.F. Teófilo, M.M.C. Ferreira. Quimiometria II: Planilhas Eletrônicas para Cálculos de Planejamentos Experimentais, um tutorial. *Quim. Nova* **2006**, 29, (2), 338-350.
5. R.F. Teófilo; E.L. Reis, C. Reis, G.A. da Silva, L.T. Kubota, Experimental Design Employed to Square Wave Voltammetry Response Optimization for the Glyphosate Determination. *J. Braz. Chem. Soc.* **2004**, 15, (6), 865-871.
6. F.B. Egreja Filho, R.F. Teófilo, E.L. Reis, C. Reis, M.P.F. Fontes. Determinação do ponto de carga zero por efeito salino (PCZES) em amostras de solos, por titulação contínua. *Tecno-Lógica*. **2004**, 8 (2), 51-59.
7. R.F. Teófilo, P.C. Braathen, M.M.M. Rubinger, Reação relógio iodo/iodeto com material alternativo de baixo custo e fácil aquisição. *Quím. Nova na Escola* **2002**, 16, 41-44.
8. E.R. Pinheiro, I.M.D.A. Silva, E.R. Amante, R.F. Teófilo, M.M.C. Ferreira, R.D.M.C. Amboni, Optimization of extraction of high-ester pectin from passion fruit peel (*Passiflora edulis flavicarpa*) with citric acid by using response surface methodology. *Bioresource* **2007**. *Aceito*.
9. R.F. Teófilo, R. Kiralj, H.J. Ceragioli, A.C. Peterlevitz, V. Baranauskas, L.T. Kubota, M.M.C. Ferreira. QSPR study of passivation by phenolic compounds at platinum and boron-doped diamond electrodes. **2008**. *Submetido*.
10. R.F. Teófilo, E.L. Reis, C. Reis, G.A. da Silva, L.T. Kubota, Glyphosate determination in soil, water and vegetables by DPV optimized by experimental design. **2007**. *Submetido*.
11. E. Kliemann, E.R. Amante, R.F. Teófilo, M.M.C. Ferreira, R.D.M.C. Amboni Optimization of pectin acid extraction from passion fruit peel (*Passiflora edulis flavicarpa*) using response surface methodology. **2007**. *Submetido*.
12. J.S. Ribeiro, R.F. Teófilo, F. Augusto, M.M.C. Ferreira. Exploratory study and chemometric discrimination of processed Brazilian commercial coffees using SPME-GC-FID for quality control. **2008**. *Submetido*.
13. R.F. Teófilo, J.P.A. Martins, M.M.C. Ferreira. Sorting variables using informative vectors as a strategy for feature selection in multivariate regression. **2008**. *Submetido*.

3.3. Co-autor em livro didático:

M.M.M. Rubinger, J.F. de Freitas, R.F. Teófilo, E.A. Carlos. *Tutoria em química analítica aplicada - Exercícios*. Editora UFV: Viçosa, 2005; p 93.

3.4. Patente depositada:

V. Baranauskas, A. Peterlevitz, H.J. Ceragioli, L.T. Kubota, R.F. Teófilo, Processo de fabricação de eletrodos de diamante dopado para eletroquímica usando etanol como fonte de carbono. INPI, *PI0600897-6*, 03/16/2006.

3.5. Registro de software requerido:

R.F. Teófilo, J.P.A. Martins, M.M.C. Ferreira, OPS Toolbox 1.0, Software para seleção de variáveis em calibração multivariada e em análise discriminante. INPI, 018070036875, 12/06/2007.

3.6. Software sem registro:

R.F. Teófilo, M.M.C. Ferreira, Planilhas eletrônicas para cálculo de planejamentos experimentais, 2002.

Disponível em <http://lqta.iqm.unicamp.br> desde março de 2006.

RESUMO – MÉTODOS QUIMIOMÉTRICOS EM ESTUDOS ELETROQUÍMICOS DE FENÓIS SOBRE FILMES DE DIAMANTE DOPADO COM BORO: O objeto deste trabalho é a aplicação de métodos quimiométricos na análise de dados obtidos a partir da eletroanálise ou eletrólise de misturas de compostos fenólicos em eletrodos de diamante dopado com boro (BDD). Filmes de BDD foram crescidos a partir do etanol e caracterizados usando diferentes tipos de técnicas. Este material indicou boa qualidade para fins eletroquímicos. Uma comparação entre os eletrodos de BDD e o de platina usando a técnica cronoamperométrica e a relação quantitativa estrutura propriedade (QSPR) foi realizada usando como propriedade de interesse o bloqueio eletroquímico devido à oxidação fenólica. Os resultados mostraram que o BDD é mais resistente ao bloqueio devido à fraca interação com os radicais, monômeros e polímeros formados durante a oxidação. Um eletrodo de BDD foi usado, então, para realizar eletroanálise de compostos fenólicos na presença de interferentes. Para realizar a quantificação foi empregado a voltametria de onda quadrada juntamente com a regressão por quadrados mínimos parciais (PLS). Uma nova estratégia para seleção de variáveis denominada Seleção das Variáveis Ordenadas (OPS) foi comparada com o Algoritmo Genético (GA). Os compostos guaiacol e cloroguaiacol foram determinados simultaneamente com sucesso diretamente em água contaminada por esgoto doméstico. Em outra determinação, a dopamina (DA), em baixas concentrações, foi determinada na presença de grandes quantidades de ácido úrico (AU) e ácido ascórbico (AA). Além disso, o AU e o AA também puderam ser determinados na mesma mistura usando o mesmo voltamograma. Os melhores resultados foram obtidos após a correção de linha de base e aplicando o método OPS. Parâmetros analíticos para ambas as análises foram calculados. Em uma última aplicação, a eletrólise de misturas de compostos fenólicos no anodo de BDD foi executada. Monitoramentos globais e individuais foram realizados via carbono orgânico total (TOC) e espectroscopia de fluorescência excitação-emissão (EEM), respectivamente. Métodos quimiométricos de ordem superior foram aplicados sobre os EEM para observar a degradação individual. Os resultados mostraram uma boa eficiência do BDD também para diminuir a toxicidade provocada por estes compostos. Todos os métodos quimiométricos usados foram fundamentais para se atingir os objetivos.

ABSTRACT – CHEMOMETRIC METHODS IN THE ELECTROCHEMICAL STUDIES OF PHENOLS ON BORON-DOPED DIAMOND FILMS:

The subject of this work is the application of chemometric methods in the analysis of the data obtained from electrochemical measurements of phenol compounds on boron-doped diamond electrodes (BDD). BDD films were grown from ethanol and characterized using different techniques. These materials showed good electrochemical qualities. A comparison between BDD and platinum electrodes, using the chronoamperometric technique and the Quantitative Structure Properties Relationship (QSPR) method was performed employing the electrochemical fouling property due to phenol oxidation. The results showed that BDD is much more resistant to fouling due to the weak interaction to the formed radicals, monomers and polymers on its surface during the oxidation. A BDD electrode was then used to carry out the electroanalysis of phenol compounds in presence of interference *i.e.* in low selectivity situation. Square Wave Voltammetry (SWV) was employed applying Partial Least Square regression (PLS) for data treatment. Ordered Predictors Selection (OPS), a new strategy for variable selection is shown and compared with Genetic Algorithm (GA). Guaiacol and chloroguaiacol were successfully determined directly in sewage, simultaneously. In other study, dopamine (DA), at low concentration levels, was determined in presence of excess of uric acid (AU) and ascorbic acid (AA). Besides, the AU and AA were also determined in the same analysis using the same voltammogram. The best results were obtained after baseline correction and applying the OPS method. Analytical parameters for both analyses were determined. In the last study, the electrolysis of phenol mixtures on BDD anode was performed. Global and individual monitoring were carried out using Total Organic Carbon (TOC) analysis and Excitation-Emission Fluorescence Spectroscopy (EEM), respectively. Higher-way chemometric approaches were applied on EEM to observe the individual degradations. The results showed a good efficiency of BDD to performing the electrolysis of these compounds. All used chemometric methods were fundamental to reach the thesis purposes.

Lista de abreviaturas

Abreviaturas	Termos em português	Termos em inglês
CNLS	Ajuste não Linear por Quadrados Mínimos Complexos	Complex Nonlinear Least Squares Fitting
GA	Algoritmo Genético	Genetic Algorithm
PLSBdg	Algoritmo PLS Bidiagonal	Bidiagonal PLS Algorithm
HCA	Análise de Agrupamentos Hierárquicos	Hierarchical Cluster Analysis
PCA	Análise de Componentes Principais	Principal Component Analysis
ANOVA	Análise de Variância	Variance Analysis
PARAFAC	Análise dos Fatores Paralelos	Parallel Factor Analysis
CSD	Base de Dados de Estruturas de Cambridge	Cambridge Structural Database
TOC	Carbono Orgânico Total	Total Organic Carbon
r	Coefficiente de Correlação	Correlation Coefficient
R^2	Coefficiente de Correlação Quadrático de Calibração	Quadratic Correlation Coefficient of Calibration
Q^2	Coefficiente de Correlação Quadrático de Validação Cruzada	Quadratic Correlation Coefficient of Cross-validation
CONAMA	Conselho Nacional de Meio Ambiente	Environment National Council
QSPC	Curvas de Polarização Quasi-Estabilizada	Quasi-Steady Polarisation Curves
SVD	Decomposição dos Valores Singulares	Singular Value Decomposition
CVD	Deposição de Vapor Químico	Chemical Vapor Deposition
HFCVD	Deposição de Vapor Químico com Filamento Quente	Hot Filament Chemical Vapor Deposition
BDD	Diamante Dopado com Boro	Boron-Doped Diamond
CPE	Elemento de Fase Constante	Constant Phase Element
SCE	Eletrodo de Calomelano Saturado	Saturated Calomel Electrode
RMSEP	Erro Quadrático Médio de Previsão	Root Mean Square Error of Prediction
RMSECV	Erro Quadrático Médio de Validação Cruzada	Root Mean Square Error of Cross Validation
EEM	Espectros de Fluorescência Excitação - Emissão	Excitation-Emission Fluorescence Spectra
EIS	Espectroscopia de Impedância Eletroquímica	Electrochemistry Impedance Spectroscopy
FV	Fonte de Variação	Variation Source
IMP	Gráfico da Identidade Emparelhada	Identity Match Plot
RIP	Gráfico de Influência de Reamostragem	Resample Influence Plot
LOD	Limite de Detecção	Detection Limit
MQ	Média Quadrática	Mean Square
RSM	Metodologia de Superfície de Resposta	Response Surface Metodology
SEM	Microscopia Eletrônica de Varredura	Scanning Electronic Microscopy
α	Nível de significância estatístico.	Significance Statistical Level (Alpha Level)
nGL	Número de Graus de Liberdade	Degree of Freedom Number
CORCONDIA	Porcentagem de Consistência do Núcleo	Core Consistency Diagnostic

CCD	Planejamento Composto Central	Central Composite Design
ALS	Quadrados Mínimos Alternantes	Alternating Least Squares
NIPALS	Quadrados Mínimos Iterativos não linear	Nonlinear Iterative Partial Least Squares
PLS	Quadrados Mínimos Parciais	Partial Least Squares
<i>M</i> -PLS	Quadrados Mínimos Parciais Matriciado	Matricizing (Unfolding) Partial Least Squares
<i>N</i> -PLS	Quadrados Mínimos Parciais Multilinear	Multilinear Partial Least Squares
MLR	Regressão Linear Múltipla	Multiple Linear Regression
PCR	Regressão por Componentes Principais	Principal Component Regression
QSAR	Relação Quantitativa Estrutura Atividade	Quantitative Structure-Activity Relationship
QSPR	Relação Quantitativa Estrutura Propriedade	Quantitative Structure-Property Relationship
OPS	Seleção das Variáveis Ordenadas	Ordered Predictors Selection
nOPS	Seleção das Variáveis Ordenadas Multilinear	Multilinear Ordered Predictors Selection
SEL	Seletividade	Selectivity
SEN	Sensibilidade	Sensitivity
NAS	Sinal Analítico Líquido	Net Analyte Signal
SQ	Soma Quadrática	Sums of Squares
DFT	Teoria de Densidade Funcional	Density Functional Theory
<i>p</i>	Valor <i>p</i> estatístico.	Statistical Significance (<i>p</i> Level)
CV	Voltametria Cíclica	Cyclic Voltammetry
SWV	Voltametria de Onda Quadrada	Square Wave Voltammetry
DPV	Voltametria de Pulso Diferencial	Differential Pulse Voltammetry

Índice Geral

<i>Índice de Figuras</i>	<i>xxi</i>
<i>Índice de Tabelas</i>	<i>xxix</i>
<i>Como Ler esta Tese</i>	<i>xxxiii</i>
Introdução	1
Referências	5
Parte Conceitual	9
Capítulo 1	11
Métodos Quimiométricos: Uma Visão Geral	11
Introdução.....	11
Notações, Convenções e alguns Pré-Tratamentos.....	12
Planejamento e Análise de Experimentos	19
Análise de Componentes Principais.....	46
Calibração Multivariada.....	49
Quadrados Mínimos Parciais (PLS).....	53
Seleção de Variáveis em Calibração Multivariada	63
Sinal Analítico Líquido e Parâmetros Analíticos.....	70
Métodos Quimiométricos de Ordem Superior	83
Referências	99
Capítulo 2	111
Diamante Dopado com Boro: Aspectos Eletroquímicos	111
Introdução.....	111
Deposição de Filmes Finos de Diamante	114
Referências	124
Parte Experimental	129
Capítulo 3	131
Propriedades dos Diamantes Dopados com Boro Crescidos do Etanol	131
Introduction	132
Experimental	134
Results and Discussion.....	137
Conclusions	147
References	148

Capítulo 4.....	153
Estudo QSPR do Bloqueio Eletroquímico Provocado por Fenóis em Eletrodos de Platina e Diamante Dopado com Boro.....	153
Introduction.....	154
Experimental.....	157
Results and Discussion.....	162
Conclusions.....	178
References.....	178
Capítulo 5.....	183
Determinação Simultânea de DA, AU e AA com Diamante Dopado com Boro usando Voltametria e Métodos Quimiométricos.....	183
Introdução.....	183
Experimental.....	186
Resultados e Discussão.....	190
Conclusões.....	211
Referências.....	212
Capítulo 6.....	219
Determinação Simultânea de Guaiacol e Cloroguaiacol com Diamante Dopado com Boro usando Voltametria e Métodos Quimiométricos.....	219
Introdução.....	219
Experimental.....	222
Resultados e Discussão.....	225
Conclusões.....	242
Referências.....	243
Capítulo 7.....	249
Monitoramento da Degradação Eletroquímica de Fenóis no Eletrodo de Diamante Dopado com Boro empregando Fluorescência e Métodos Multimodos.....	249
Introdução.....	249
Experimental.....	252
Resultados e Discussão.....	257
Conclusões.....	274
Referências.....	275
Conclusão Geral e Perspectivas.....	279
Apêndice.....	283
Tópicos de Álgebra Linear.....	283
Referências.....	292

Índice de Figuras

Parte Conceitual.....	9
Capítulo 1.....	11
Figura 1.1. Ilustração do escalar, vetor, matriz e tensor. O círculo fechado (●) representa as posições do elemento no vetor (1), na matriz (1,1) e no tensor (1,1,1). As setas indicam que há escalares em todas as outras posições.....	12
Figura 1.2. Ilustração do arranjo tensorial (três modos) em fatias (dois modos).....	16
Figura 1.3. Definição de linhas, colunas, e tubos em um arranjo de três dimensões.....	16
Figura 1.4. Matriciação de um arranjo em três dimensões \mathbf{X} para um de duas dimensões $\mathbf{X}_{(I \times JK)}$	17
Figura 1.5. Experimentos “selecionados” para um planejamento fatorial fracionário 2^{3-1}	24
Figura 1.6. Planejamento 2^2 com ponto central.....	28
Figura 1.7. Representação gráfica do valor p para um teste unilateral.....	35
Figura 1.8. Planejamentos compostos centrais: (A) $J = 2$ e $\alpha=1.414$, (B) $J = 3$ e $\alpha=1.682$. Os diferentes símbolos significam: (●) planejamento fatorial, (■) pontos axiais e (O) ponto central.....	37
Figura 1.9. Planejamento Doehlert com duas variáveis (—) com adição de novos experimentos (⋯).	40
Figura 1.10. Representação matricial da análise de componentes principais. ...	47
Figura 1.11. Representação geométrica de duas componentes principais em um espaço com três variáveis (3D).....	48
Figura 1.12. Representação de vários tipos de arranjos de dados para construção de modelos de calibração.....	51
Figura 1.13. Esquema usado para construir e validar o modelo.....	60
Figuras 1.14. Esquema do método ‘leave-one-out’ para validação cruzada.....	61
Figura 1.15. Exemplo de validação cruzada ‘leave-n-out’ para mais de uma amostra excluída.....	62
Figura 1.16. Etapas da seleção de variáveis usando o OPS.....	69
Figura 1.17. Ilustração da orientação e comprimento do vetor do sinal analítico líquido em relação ao plano dos interferentes e sinal analítico do componente puro	72

Figura 1.18. Representação esquemática da obtenção de uma matriz de dados com informações analíticas para uma amostra, apta a ser tratada com métodos multimodo.	83
Figura 1.19. Extensão da decomposição SVD para a decomposição PARAFAC. Na decomposição SVD (ver apêndice), as matrizes \mathbf{A}_{svd} e \mathbf{P} são obrigatoriamente ortogonais. O modelo PARAFAC não é seqüencial e a solução é única (não há liberdade rotacional). O PARAFAC é um caso especial do modelo Tucker3, onde o número de componentes em cada modo é igual e a matriz núcleo \mathbf{G} que é superdiagonal e unitária.	85
Figura 1.20. Partição de I amostras na análise <i>split-half</i> . As amostras são divididas em dois grupos A e B que são separados e divididas posteriormente em dois outros grupos, C e D, que também são separados.	92
Figura 1.21. Primeira etapa do algoritmo nOPS: Ordenando as variáveis.	97
Figura 1.22. Segunda etapa do algoritmo nOPS: Obtendo o vetor prognóstico.	98
Figura 1.23. Terceira etapa do algoritmo nOPS: Investigação das variáveis ordenadas no tensor.	99
Capítulo 2	111
Figura 2.1. Esquema de um típico reator empregado para a deposição de diamante. Neste esquema, especificamente, o etanol é a fonte de carbono e trióxido de boro a dopante.	115
Figura 2.2. Domínio do diamante no diagrama ternário C-H-O de acordo com Bachmann (1991) [20] e Ford (1995-96) [22]	118
Figura 2.3. Mecanismo sugerido para o crescimento do diamante a partir de metano altamente diluído em hidrogênio, usando o processo CVD.	119
Parte Experimental	129
Capítulo 3	131
Figure. 3.1 Typical Raman spectra of “as-grown” boron-doped diamond: solid line, type A; dashed line, type B	138
Figure. 3.2 SEM images of polycrystalline BDD films: a type A; b type B ...	138
Figure. 3.3 Cyclic voltammograms at 100 mV s ⁻¹ for 1×10 ⁻³ mol L ⁻¹ ferrocyanide in 0.1 mol L ⁻¹ KCl, solid line, “as-grown”; cathodic treatments: dashed line, 1, dotted line, 2, dash-dot line, 3, respectively, in electrodes type A (a) and type B (b)	139
Figure. 3.4 Cyclic voltammograms as functions of the electrolyte composition. a 0.1 mol L ⁻¹ H ₂ SO ₄ ; b 1 mol L ⁻¹ HNO ₃ for electrodes type A (solid line) and type B (dashed line). Scan rate: 100 mV s ⁻¹	140

Figure. 3.5 Cyclic voltammograms for three analytes: ferrocyanide (a), hexamineruthenium (b) and dopamine (c) at BDD type A (solid line) and type B (dashed line) electrodes. All redox analyte concentrations were $10^{-3} \text{ mol L}^{-1}$. The supporting electrolyte was $1 \text{ mol L}^{-1} \text{ KCl}$. Scan rate: 100 mV s^{-1} 143

Figure. 3.6 Equivalent circuit used for representing the impedance response in the presence of diffusion control 144

Figure. 3.7 Nyquist plots obtained for different electron-transfer reactions on BDD electrodes: a ferrocyanide, $10^{-3} \text{ mol L}^{-1}$; b hexamineruthenium, $10^{-3} \text{ mol L}^{-1}$; c dopamine, $10^{-3} \text{ mol L}^{-1}$. Supporting electrolyte was $1.0 \text{ mol L}^{-1} \text{ KCl}$. $T=25 \text{ }^\circ\text{C}$. Legend: Type A, experimental (circles) and curve fitting (solid line); type B, experimental (squares) and curve fitting (dashed line)..... 145

Capítulo 4.....153

Figure 4.1. Mechanism of electrooxidation of phenol [11-15]. Adsorption of phenol on an electrode (A-B), electrooxidation (B-C) and resonance structures of the formed phenoxy radical (C-F). 155

Figure 4.2. Passivation measured using chronoamperometric curves for Pt electrode..... 159

Figure 4.3. Structures of twenty phenolic compounds with measured passivation rates (training set), together with their names and used abbreviations, and reactivity markers (sites for nucleophilic aromatic substitution)..... 160

Figure 4.4. Structures of phenolic compounds without measured passivation rates (prediction set), together with their names and used abbreviations, and reactivity markers (sites for nucleophilic aromatic substitution)..... 162

Figure 4.5. Graphic representation of the phenolic passivation for different substituents on the phenol ring obtained from chronoamperometric measurements. Gray: Pt electrode; Dark gray: BDD electrode. Bar error is the standard deviation. 164

Figure 4.6. Scores plots (top figures) and HCA dendrograms (bottom figures) for molecular descriptors which were used to build the Pt model (left figures) and the BDD model (right figures). Compounds in the PCA plots are distinguished by the number of the most probable sites for nucleophilic aromatic substitution (zero and non-zero numbers are separated by the dashed line at $\text{PC1} = 0$ for the Pt data set) and are arranged in the order of appearance along PC1 for the Pt data set. HCA sub-clusters are based on passivation and structural features of the compounds. Phenols which form external validation sets are marked with solid squares in the dendrograms..... 170

Figure 4.7. Prediction capacity using PLS models (A–C) and relative error (B–D) for Pt model and BDD model, respectively. Key for graphics A and C: Predictions for calibration compounds (■); predictions for external compounds (●). Key for graphics B and D: light gray - relative error for calibration compounds; dark gray - relative error for external compounds..... 175

Figure 4.8. Validation of the Pt model (left figures) and the BDD model (right figures) by means of leave- N -out crossvalidation (upper figures) and Y-randomization (lower figures). Mean Q^2 values are shown by dashed lines for leave- N -out crossvalidation. Models resulting from Y-randomizations are shown by solid squares to be distinct from the validated PLS models (white squares). 177

Capítulo 5.....183

Figura 5.1. Níveis e concentrações estudadas para os três analitos. (CA) concentrações altas de AA e AU em relação à DA, (CB) concentrações próximas de uma mesma faixa de concentração. 188

Figura 5.2. Forma de aplicação do potencial na SWV e esquema de medida. (A) varredura do potencial em rampa, (B) varredura em pulsos constantes, (C) voltametria de onda quadrada. 191

Figura 5.3. Gráfico de probabilidade normal para a DA (A) e AU (B). Fatores fora da linha tracejada são significativos. 193

Figura 5.4. Valores medidos vs. previstos pelo modelo da DA (A) e do AU (C), respectivamente e valores medidos vs. resíduos para os modelos da DA (B) e AU (D), respectivamente. 195

Figura 5.5. Superfícies de resposta para os modelos quadráticos construídos para a DA e para o AU..... 196

Figura 5.6. Conjunto CA: Voltamogramas originais (A) e com correção de linha base (C) para DA(⋯), AU(– –), AA(—) e mistura(–·–); Voltamogramas originais das misturas (B) e com correção de linha de base com janela móvel (D). A largura da janela na correção de linha de base foi igual a 0,015 V. 197

Figura 5.7 Conjunto CB: Voltamogramas originais (A) e com correção de linha base (C) para DA(⋯), AU(– –), AA(—) e mistura(–·–); Voltamogramas originais das misturas (B) e com correção de linha de base (D). A largura da janela na correção de linha de base foi igual a 0,02 V. 197

Figura 5.8. Dados originais e vetores prognósticos usados para selecionar variáveis para a dopamina usando o método OPS. No eixo das variáveis são indicados os seus números e não os potenciais..... 200

Figura 5.9. Gráficos característicos obtidos usando o método OPS. As linhas horizontais indicam o valor de RMSECV com todas as variáveis e as linhas verticais indicam o conjunto escolhido. A, B, C são os resultados para DA, AU e AA para o conjunto CA, e D, E, F para o conjunto CB, respectivamente. Os símbolos ■ e ● indicam os voltamogramas sem e com correção de linha de base, respectivamente.201

Figura 5.10. Conjunto CA: Variáveis selecionadas pelo GA para os compostos DA (A), AU(B) e AA(C) e variáveis selecionadas pelo método OPS para os compostos DA(D), AU(E), AA(F). 205

Figura 5.11. Conjunto CB : Variáveis selecionadas pelo GA para os compostos DA (A), AU(B) e AA(C) e variáveis selecionadas pelo OPS para os compostos DA(D), AU(E), AA(F).....	206
Figura 5.12. Conjunto CA : Concentrações adicionadas vs. previstas para o modelo (■) e para o conjunto externo de previsão (●) para DA (A), AA (B) e AU(C). As barras verticais são os erros relativos para as amostras externas para DA (D), AA (E) e AU(F)	207
Figura 5.13. Conjunto CB : Concentrações adicionadas vs. previstas para o modelo (■) e para o conjunto externo de previsão (●) para DA (A), AA (B) e AU(C). As barras verticais são os erros relativos para as amostras externas para DA (D), AA (E) e AU(F)	208
Figura 5.14. Gráficos pseudo-univariados (referência vs. <i>NAS</i>) para o conjunto CA : DA(A), AU(B) e AA(C) e pra o conjunto CB : DA(D), AU(E), AA(F).	211
Capítulo 6.....	219
Figura 6.1. Representação geométrica do planejamento composto central usado para o estudo de três variáveis do sistema voltamétrico para a determinação de GUA e CGUA.	225
Figura 6.2. Gráfico de probabilidade normal para a GUA (A) e CGUA (B). Fatores fora da linha tracejada são significativos.	226
Figura 6.3. Superfícies de resposta extrapoladas para o GUA(A-C) e CGUA(B-D) obtidas através do planejamento CCD com a <i>Freq.</i> fixada em 35 Hz.	228
Figura 6.4. Pontos experimentais da matriz Doehlert representados geometricamente conforme as variáveis invetigadas.	229
Figura 6.5. Superfícies de resposta obtidas através do modelo Doehlert para duas variáveis. Superfícies em 2 dimensões (A) e em 3 dimensões (B). As linhas pretas internas à superfície A indicam a região ótima usada.....	231
Figura 6.6. Voltamogramas de onda quadrada obtidos nas condições otimizadas. Voltamogramas puros e para a mistura antes (A) e após correção da linha de base com janela móvel (B) e voltamogramas das misturas para a construção do modelo antes (C) e após correção de linha de base (D). Os voltamogramas da água do rio no tampão sem os guaiacóis são mostrados abaixo das curvas. A largura da janela na correção de linha de base foi igual a 0,005 V.....	232
Figura 6.7. Módulo dos vetores informativos para GUA (linha cinza) e CGUA (linha preta) usados pelo OPS na seleção de variáveis. (A) vetor de regressão, (B) vetor de correlações X-y e (C) combinações entre estes vetores. Estes vetores estão ou normalizados e/ou ponderados e as unidades de suas intensidades são arbitrárias.	233

Figura 6.8. Gráficos característicos usados para a seleção de variáveis com o método OPS. As linhas horizontais indicam o valor de RMSECV com todas as variáveis e as linhas verticais indicam o conjunto escolhido para seleção com correção. (A) seleções realizadas para o GUA e (B) seleções realizadas para o CGUA. As 45 amostras foram usadas nestas seleções e o número de componentes foram fixados em 6 para A e 5 para B. Os símbolos ● e ■ indicam os voltamogramas sem e com correção de linha de base, respectivamente.....	235
Figura 6.9. Variáveis selecionadas pelo GA para os compostos GUA (A), CGUA(B) e variáveis selecionadas pelo OPS para os compostos GUA(C) e CGUA (D). As linhas verticais indicam as variáveis selecionadas.	237
Figura 6.10. Concentrações referência vs. previstas para o modelo (■) e para o conjunto externo de previsão (●) para GUA (A) e CGUA (B). As barras verticais são os erros relativos para as amostras externas para GUA (C), CGUA (D).	239
Figura 6.11. Gráficos pseudo-univariados (<i>NAS</i> vs. referência) para o GUA(A), CGUA(B).....	240
Capítulo 7.....	249
Figura 7.1. Esquema do reator, na escala utilizada. (A) reator com dois eletrodos, (B) reator com quatro eletrodos.	253
Figura 7.2. Imagem de microscopia eletrônica de varredura (A) e espectro Raman (B) para um dos eletrodos usados.....	257
Figura 7.3. Remoções do sinal de emissão coletado a 320 nm (●) e do valor de TOC (■) obtidos durante a degradação para os conjuntos <i>MS</i> no nível A e B e para o conjunto <i>MT</i> no nível C. As medidas em cada ponto foram realizadas na mesma solução.	258
Figura 7.4. Comportamento típico dos espectros EEM durante as correções. Espectros originais (A), espectro com remoção dos espalhamentos Raman e Rayleigh (B) e espectro com interpolação sobre os dados faltantes (C).....	261
Figura 7.5. Resultados da análise com três diferentes inicializações (SVD, DTLTD, aleatória) e oito componentes.....	262
Figura 7.6. Gráficos RIP para os modos excitação e emissão. Os pontos nos gráficos representam os comportamentos das amostras e cada modo.....	263
Figura 7.7. Espectros puros (—) de excitação (acima) e emissão (abaixo) e respectivos <i>loadings</i> (····) obtidos pelo modelo PARAFAC.....	263
Figura 7.8. Gráfico IMP para HYDR (A), GUAI (B), PCRE (C), MCRE (D), CATE (E), PHEN (F).....	264
Figura 7.9. Análise diagnóstica para três diferentes inicializações (SVD, DTLTD, aleatória) e 5 componentes: (A) gráfico dos resíduos, (B) percentagem da análise Corcondia e (C) número de iterações	265
Figura 7.10. Gráficos RIP para o modo da excitação (A) e emissão (B).....	266

Figura 7.11. Espectros puros (—) para HYDR (A,D), PHEN (B,E) e CATE (C,F) e respectivas médias das estimativas dos <i>loadings</i> (····) para 38 segmentos <i>Jack-knife</i>	267
Figura 7.12. Análise <i>Split-half</i> para três componentes: <i>Loadings</i> dos modos excitação (A) e emissão (B) e espectros puros experimentais da excitação (C) e emissão (D). Os comprimentos de onda máximos de excitação e emissão são respectivamente em nm de: PHEN (270/300); CATE(275/320); HYDR(287/331).	267
Figura 7.13. Gráficos IMP para HYDR (A), PHEN (B) e CATE (C).	268
Figura 7.14. Análise diagnóstica do conjunto de degradação para três diferentes inicializações e seis componentes: (A) gráfico dos resíduos, (B) percentagem da análise Corcondia e (C) número de iterações	269
Figura 7.15. Variáveis selecionadas pelo nOPS, as linhas indicam as variáveis selecionadas em cada modo. HYDR(A,D); PHEN(B,E) e CATE (C,F), as imagens D-E mostram os espectros EEM puros de cada analito e as respectivas variáveis selecionadas.	271
Figura 7.16. Valores adicionados vs. previstos para os modelos de calibração N-PLS/nOPS (■) e amostras previstas (●) e erros relativos das amostras previstas para HYDR (A,D), PHEN (B,E) e CATE (C,F).....	272
Figura 7.17. Estimativas das degradações da HYDR(A), PHEN(B) e CATE (C) para os modelos PARAFAC, M-PLS e N-PLS/nOPS e ambas as estimativas para o modelo N-PLS/nOPS (D).....	273
Apêndice.....	283

Índice de Tabelas

Parte Conceitual.....	9
Capítulo 1.....	11
Tabela 1.1. Representação esquemática de uma matriz de dados.....	13
Tabela 1.2. Matriz de coeficientes de contraste para um planejamento fatorial 2^3	21
Tabela 1.3. Planejamento fatorial fracionário 2^{3-1} a partir de um planejamento fatorial completo 2^2	24
Tabela 1.4. Planejamento fatorial completo 2^3	24
Tabela 1.5. Coeficientes de contraste para um planejamento fatorial fracionário 2^{5-1}	26
Tabela 1.6. Alguns planejamentos CCD e suas características.	38
Tabela 1.7. Matrizes dos planejamentos compostos centrais com duas e três variáveis.	38
Tabela 1.8. Matrizes Doehlert para duas, três e quatro variáveis.	41
Tabela 1.9. Análise de variância para regressão múltipla.....	45
Capítulo 2.....	111
Parte Experimental.....	129
Capítulo 3.....	131
Table 3.1 CV and kinetic parameters obtained at BDD type A and type B electrodes for three redox systems in 1 mol L ⁻¹ KCl at 25 °C	143
Table 3.2 The double-layer capacity C_{dl} and the heterogeneous electron-transfer rate constant k^0 for BDD type A and type B electrodes obtained from the EIS data for different redox systems	146
Capítulo 4.....	153
Table 4.1. Experimental results obtained using cyclic voltammetry and chronoamperometry for the Pt and BDD electrodes.*	163
Table 4.2. Molecular descriptors and passivations used in QSPR.....	166
Table 4.3. Correlation matrix for dependent and independent variables used for QSPR.....	167
Table 4.4. Comparison of the Pt and BDD PLS models with basic statistics .	174
Table 4.5. Results for the QSPR models prediction capacity.	176
Capítulo 5.....	183
Tabela 5.1. Variáveis e níveis investigados do algoritmo genético.....	189

Tabela 5.2. Fatores, níveis investigados codificados e decodificados e respostas obtidas para a DA e AU.....	192
Tabela 5.3. Efeitos principais do planejamento fracionário 2^{4-1} para cada composto e seus respectivos valor p calculados a partir do erro puro.....	193
Tabela 5.4 Planejamento Doehlert com os níveis codificados e experimentais e as respectivas respostas observadas para as variáveis amplitude e frequência.	193
Tabela 5.5. Análises de variância para os modelos DA e AU	195
Tabela 5.6. Concentrações de cada componente nas misturas usadas nos conjuntos de calibração CA e CB , além das concentrações médias, mínimas e máximas, para cada componente em cada conjunto.	198
Tabela 5.7. Parâmetros usados para executar o algoritmo genético.	199
Tabela 5.8. Diferentes modelos PLS construídos para a quantificação simultânea de DA, AU e AA para o conjunto CA	202
Tabela 5.9. Diferentes modelos PLS construídos para a quantificação simultânea de DA, AU e AA para o conjunto CB	203
Tabela 5.10. Concentrações previstas pelos modelos do conjunto CA e seus respectivos erros relativos.....	207
Tabela 5.11. Concentrações previstas pelos modelos do conjunto CB e seus respectivos erros relativos.....	208
Tabela 5.12. Parâmetros analíticos para os modelos finais de ambos os conjuntos	209
Tabela 5.13. Faixa de parâmetros analíticos encontrada na literatura para determinação simultânea DA, AU e AA usando univariados em diferentes tipos de eletrodos modificados usando técnicas de pulso (SWV ou DPV).....	210
Capítulo 6.....	219
Tabela 6.1. Fatores, níveis investigados codificados e decodificados e respostas obtidas para a GUA e CGUA.	226
Tabela 6.2. Efeitos principais do planejamento fracionário CCD para cada composto, e seus respectivos valores p calculados a partir do erro puro.	227
Tabela 6.3. ANOVA para os modelos quadráticos obtidos com o planejamento CCD para GUA e CGUA.....	227
Tabela 6.4. Planejamento Doehlert para duas variáveis, em ordem aleatória, para guaiacol.	230
Tabela 6.5. Análise estatística dos coeficientes do modelo quadrático usando a matriz Doehlert codificada.....	230
Tabela 6.6. Análises de variância para o modelo do GUAI.....	230
Tabela 6.7. Parâmetros selecionados para executar o algoritmo genético.....	234
Tabela 6.8. Parâmetros estatísticos de diferentes modelos PLS construídos para a quantificação simultânea de GUA e CGUA.	236

Tabela 6.9. Concentrações previstas para GUA e CGUA e seus respectivos erros relativos.....	238
Tabela 6.10. Parâmetros analíticos para os modelos finais de ambos os conjuntos.....	239
Tabela 6.11. Limites máximos definidos pelo CONAMA para alguns compostos fenólicos e fenóis totais.....	241
Capítulo 7.....	249
Tabela 7.1. Níveis usados para a elaboração do planejamento fatorial parcial para as duas misturas <i>MS</i> e <i>MT</i>	255
Tabela 7.2. Custo energético para remoção de TOC e aromaticidade.....	259
Tabela 7.3. Diferentes modelos multimodos para realizar a previsão das concentrações durante a degradação individual.....	269
Tabela 7.4. Valores adicionados (Adic.) vs. previstos (Prev.) e erros relativos (Erros relat.) para as amostras externas usando o modelo <i>N</i> -PLS com seleção <i>n</i> OPS.....	272
Apêndice.....	283

Como Ler esta Tese

Esta tese está dividida em duas partes: uma conceitual e uma experimental. Na parte conceitual são descritos aspectos teóricos e práticos utilizados na parte experimental. As partes são independentes, porém, muitos conceitos citados na parte experimental são prontamente encontrados na parte conceitual. De outro modo, o leitor interessado apenas na parte conceitual é somente convidado a ler a parte experimental, uma vez que nesta parte são encontradas aplicações de métodos e processos discutidos na parte conceitual.

A organização da tese segue a seguinte ordem:

Introdução

Descreve de uma maneira ampla a proposta da tese e seu objetivo geral.

Parte conceitual

Capítulo 1

Neste capítulo são descritos brevemente os conceitos básicos referentes às metodologias quimiométricas empregadas. Para o leitor que não está familiarizado com estas metodologias, a leitura deste capítulo é fortemente recomendada. Uma extensa lista bibliográfica é também proporcionada para aprofundamento dos conceitos apresentados.

Capítulo 2

Este capítulo é dedicado ao diamante dopado e descreve de uma maneira introdutória fundamentos históricos, de produção e de aplicações eletroquímicas. O leitor interessado nesta área encontrará uma seleção de referências bibliográficas que abragem mais profundamente os aspectos deste material, principalmente para fins eletroquímicos.

Parte experimental

Capítulo 3

Nesta tese, diamantes dopados com boro foram produzidos para serem aplicados em eletroanálise e eletrólise. A síntese do diamante e sua caracterização foram realizadas antes de quaisquer aplicações e são apresentadas neste capítulo.

Capítulo 4

Uma desejável característica dos materiais eletródicos é a resistência ao bloqueio eletroquímico provocada principalmente por compostos orgânicos durante a sua oxidação na superfície do eletrodo. Na literatura é afirmado que o diamante dopado é altamente resistente a este bloqueio. Como neste trabalho os compostos estudados foram os fenólicos, que ao se oxidarem formam filmes passivantes sobre eletrodos, uma avaliação sobre a resistência do diamante é necessária. Este capítulo apresenta um estudo que relaciona quantitativamente a propriedade de bloqueio eletroquímico dos compostos fenólicos às suas estruturas químicas. Este é o primeiro capítulo onde diversas metodologias quimiométricas foram aplicadas e é recomendado para o leitor interessado em aplicações de metodologias quimiométricas em estudos fundamentais de sistemas químicos.

Capítulos 5 e 6

Os estudos realizados nos capítulos 3 e 4 dão condições experimentais para aplicações eletroanalíticas dos eletrodos de diamante dopado com boro. Nestes capítulos são mostradas aplicações analíticas deste eletrodo para a determinação simultânea de misturas de compostos usando voltametria e métodos quimiométricos, além disso, no capítulo 6, são mostradas medidas realizadas diretamente em água contaminada com esgoto doméstico.

Capítulo 7

Neste último capítulo o diamante dopado com boro foi utilizado para realizar a eletrodegradação de compostos fenólicos. O monitoramento da degradação eletroquímica no ânodo de diamante foi realizado empregando medidas de carbono orgânico total, espectroscopia de fluorescência e metodologias quimiométricas de ordem superior.

Recomenda-se este capítulo para os interessados no uso do diamante dopado com boro para fins de eletrodegradação de compostos persistentes e àqueles interessados na utilização de espectroscopia de fluorescência aliada aos métodos quimiométricos de ordem superior para o monitoramento da degradação de compostos fenólicos.

Conclusão geral e perspectivas

Apresenta uma conclusão geral da tese e perspectivas para continuidade do trabalho.

Apêndice

No apêndice desta tese alguns tópicos de álgebra linear são apresentados. Estes tópicos foram inseridos neste trabalho devido ao grande uso de termos, equações e conceitos que são usados em alguns capítulos, principalmente no capítulo 1. Este apêndice é fortemente recomendado ao leitor interessado no capítulo 1 e não está familiarizado com alguns conceitos fundamentais de álgebra linear.

Introdução

“Os dados conduzem para a informação, que orientam para o conhecimento”

Bruce Kowalski

A análise de dados multivariados tem se tornado, de modo crescente, uma importante área da química e de outras ciências [1,2]. Isto sugere que a tendência atual do pensamento científico envolve um raciocínio multivariado [1,3], principalmente devido ao desenvolvimento tecnológico atingido neste último século [1].

Historicamente, a análise de dados multivariados em química tem utilizado ferramentas desenvolvidas na área de ciências sociais como economia (econometria) e psicologia (psicometria). Apesar de serem interpretadas para objetivos específicos em cada área, a maioria destas ferramentas se baseia em sofisticados métodos matemáticos, estatísticos, computacionais e de lógica formal [2,4]. Em química, as análises multivariadas são úteis para (1) planejar ou selecionar procedimentos experimentais ótimos; (2) extrair o máximo da informação química relevante pela análise dos dados; e (3) obter conhecimentos sobre os sistemas químicos estudados [5].

Com a aplicação destas ferramentas sobre dados fundamentalmente obtidos nas áreas da química, nasceu uma nova disciplina designada por Svante Wold em 1972, de quimiometria [6]. De uma maneira filosófica, se partirmos do ponto de vista de que a quimiometria é, teoricamente, a aplicação de quaisquer métodos matemáticos e estatísticos para o tratamento de dados químicos, a quimiometria se inicia em 1772 com Lavoisier, que deve ser considerado o primeiro quimiometrista, por ter sido o primeiro a usar métodos quantitativos em química [6]. Porém, há um consenso geral de que a quimiometria se iniciou por volta de 1970 com diversos trabalhos científicos, livros e um simpósio, ocorrido em São Francisco, em 2 de

setembro de 1976 [7], todos direcionados para a análise multivariada de dados químicos. O grande divulgador e entusiasta internacional desta nova disciplina foi, sem dúvida, Bruce Kowalski [6,8]. Foram dele os primeiros trabalhos, a primeira comunicação por escrito dirigida aos químicos anunciando esta nova área e o primeiro livro editado na área [9].

A criação desta nova disciplina ocorreu, principalmente, devido à automação dos laboratórios com o aumento significativo no número de instrumentos analíticos que fornecem dados multivariados, ou seja, métodos que geram mais de uma medida para uma dada amostra [10]. Além disso, outras importantes contribuições foram: a acessibilidade a microcomputadores e a disponibilidade de metodologias e softwares dedicados à realização de cálculos com dados multivariados [4].

Atualmente a quimiometria está suficientemente estabelecida e seu uso disseminado [11]. Cada vez mais, metodologias quimiométricas estão sendo aperfeiçoadas e/ou desenvolvidas de maneira direcionada para os problemas de origem química [12,13].

Dentre as áreas tradicionais da química, a química analítica é, atualmente, a mais influenciada pela quimiometria [4]. Dentre as diferentes técnicas analíticas, as ferramentas quimiométricas se tornaram mais populares na área de espectroscopia [14]. Por outro lado, as aplicações das metodologias quimiométricas em eletroanalítica são relativamente escassas [10,14,15]. Sendo assim, um dos objetivos deste trabalho é divulgar algumas aplicações das metodologias quimiométricas em eletroanalítica e também, de um modo geral, em eletroquímica.

Algumas razões agindo de maneira simultânea, podem justificar a lenta difusão das ferramentas quimiométricas em eletroanalítica: (1) a baixa repetibilidade do sensor ou eletrodo para análises multipotenciais de uma grande quantidade de amostras; (2) a complexidade da interface eletrodo-solução para matrizes naturais devido à presença de muitas substâncias iônicas, de substâncias orgânicas que podem bloquear a superfície do eletrodo ou sensor, da sensível dependência da resposta às variações do pH e da atividade iônica do meio; (3) a complexidade do sistema devido a competição de analitos eletroativos pelos sítios ativos do eletrodo ou sensor, durante a eletroanálise; (4) a baixa seletividade do sinal, que na maioria das vezes é formado por apenas um único pico, que traz a

informação de múltiplos analitos. Estes fatores agindo simultaneamente ou individualmente tornam as análises multipotenciais pouco precisas e/ou limitadas.

Resta aos eletroquímicos focar os seus objetivos na utilização e/ou desenvolvimento de metodologias quimiométricas e torná-las uma aliada na solução de problemas complexos, que constantemente surgem em uma célula eletroquímica. Portanto, o eletroanalítico interessado em aplicar ferramentas quimiométricas precisa assegurar, tanto quanto possível, que a maior parte da informação do sistema eletroquímico seja extraída com precisão, principalmente garantindo uma alta robustez do sensor ou eletrodo para muitas análises multivariadas em matrizes complexas.

Porém, as metodologias quimiométricas para o eletroquímico não se resumem somente às aplicações em eletroanálises [14,15]. O eletroquímico pode aplicar muitas das metodologias disponíveis, ou desenvolver novas, para estudos fundamentais de substâncias eletroativas na interface eletrodo-solução [16,17]. Outro tipo de aplicação seria sobre outras técnicas, como por exemplo, as espectroscópicas [18], empregadas para entender um determinado sistema eletroquímico sob investigação.

Há, portanto, uma gama de metodologias quimiométricas disponíveis e uma série de técnicas eletroquímicas com alta potencialidade de fornecimento de dados multivariados. Porém, a dificuldade do eletroquímico atual está em buscar uma superfície eletródica capaz de ser robusta e resistente ao bloqueio eletroquímico, de maneira a realizar análises mais precisas e estáveis por longo período de tempo e, sobretudo, diretamente em matrizes naturais complexas.

Atualmente, um dos materiais que tem se destacado entre os eletroquímicos devido às suas características únicas e extremamente úteis é o Diamante Dopado com Boro (BDD).

O BDD possui várias propriedades eletroquimicamente importantes, incluindo inércia química, ampla janela de potencial em meio aquoso e não aquoso, baixa capacitância, insensibilidade ao oxigênio dissolvido, robustez e alta resistência ao bloqueio eletroquímico [19-23].

O interesse mundial no uso dos filmes de BDD em eletroanálise, eletrosíntese, proteção à corrosão, dispositivos para armazenamento de energia e para a limpeza

de águas contaminadas com rejeitos tóxicos tem rapidamente emergido, guiado pelas excelentes características deste material [20,24-28].

Neste trabalho, filmes de BDD foram crescidos, caracterizados e aplicados empregando técnicas eletroquímicas. Os dados obtidos foram tratados com ferramentas quimiométricas. Os compostos eletroativos escolhidos para os estudos a serem realizados foram os compostos fenólicos. A escolha destes compostos é justificada devido ao seu amplo uso em muitas áreas, tais como farmacêutica, alimentícia, análise clínica, agricultura, além de serem comumente encontrados em diversos produtos industriais [29-33]. Além disso, compostos fenólicos são substâncias endógenas em vegetais e animais [34-37] e são, muitas vezes, os responsáveis pela cor e/ou sabor de alimentos, água potável, bebidas, flores e frutos [38].

Um dos grandes desafios da eletroquímica é realizar estudos e análises de compostos fenólicos com precisão, ou realizar a degradação destes compostos com eficiência, empregando a eletrólise. Isto ocorre, pois a oxidação destes compostos sobre eletrodos sólidos convencionais produz radicais fenóxidos que se acoplam para formar filmes poliméricos que bloqueiam eletroquimicamente a superfície do eletrodo. Este filme polimérico é um problema em eletroanálise e eletrólise de fenóis (despoluição de águas residuárias), visto que rapidamente desativa o eletrodo e oxidações posteriores tornam-se menos ativas [39-43]. Porém, na literatura, estudos mostram que o eletrodo de BDD é resistente ao bloqueio eletroquímico dos compostos fenólicos [44,45], o que contribui para minimizar este indesejável fenômeno, que dificulta o trabalho dos eletroquímicos para algumas áreas.

Por outro lado, outro grande desafio está na análise de misturas de compostos fenólicos, visto que, muitos destes compostos se oxidam ou reduzem próximo a um mesmo potencial, o que dificulta uma análise seletiva empregando técnicas eletroanalíticas [46].

Tendo em vista tantos desafios, a sugestão de combinar as excelentes características do eletrodo de BDD com as potencialidades da quimiometria pode trazer grandes vantagens, o que possivelmente tornará promissora tal união.

Esta tese defenderá a aplicação de ferramentas quimiométricas sobre dados fundamentalmente obtidos empregando técnicas eletroquímicas com um eletrodo

de diamante dopado com boro, sejam para estudos fundamentais dos compostos fenólicos ou para fins eletroanalíticos de misturas destes compostos. Além disso, esta tese visa mostrar o potencial das ferramentas quimiométricas sobre dados obtidos por outras técnicas não eletroquímicas, que são aliadas para o entendimento de sistemas eletroquímicos, tais como a espectroscopia de fluorescência que pode ser empregada para o monitoramento de fenóis durante a eletrólise.

Referências

1. Rencher, A. C. *Methods of multivariate analysis*. 2nd ed.; John Wiley & Sons Inc: New York, 2002.
2. Phatak, A. Evaluation of some multivariate methods and their applications in chemical engineering. University of Waterloo, Ontario, 1993.
3. Workman, J. J.; Mobley, P. R.; Kowalski, B. R.; Bro, R. Review of chemometrics applied to spectroscopy: 1985-95 .1. *Appl. Spectrosc. Rev.* **1996**, 31, (1-2), 73-124.
4. Brereton, R. G. *Chemometrics: Data Analysis for the laboratory and chemical plant*. John Wiley & Sons Inc: Chinchester, 2003; p 407.
5. Massart, D. L.; Vandeginste, B. G. M.; Buydens, L. M. C.; de Jong, S.; Lewi, P. J.; Verbeke-Smeyers, J. *Handbook of Chemometrics and Qualimetrics*. Elsevier: Amsterdam, 1998; Vol. Part A, .
6. Esbensen, K.; Geladi, P. The start and early history of chemometrics - selected interviews .2. *J. Chemometr.* **1990**, 4, (6), 389-412.
7. Sheperd, P. Retrospective. *J. Chemometr.* **1987**, 1, 3-6.
8. Geladi, P.; Esbensen, K. The start and early history of chemometrics .1. selected interviews. *J. Chemometr.* **1990**, 4, (5), 337-354.
9. Kowalski, B. R. *Chemometrics: Theory and Application*. ACS Symp. Ser.: New York, 1977; .
10. Richards, E.; Bessant, C.; Saini, S. Multivariate data analysis in electroanalytical chemistry. *Electroanalysis* **2002**, 14, (22), 1533-1542.
11. Lavine, B. K.; Workman, J. Chemometrics: Past, present, and future. **2005**, 894, 1-13.
12. Hopke, P. K. The evolution of chemometrics. *Anal. Chim. Acta* **2003**, 500, (1-2), 365-377.
13. Hasegawa, T.; Ozaki, Y. New development in chemometrics. *Bunseki Kagaku* **2005**, 54, (1), 1-26.
14. Esteban, M.; Arino, C.; Diaz-Cruz, J. M. Chemometrics in electroanalytical chemistry. *Crit. Rev. Anal. Chem.* **2006**, 36, (3-4), 295-313.

15. Esteban, M.; Arino, C.; Diaz-Cruz, J. M. Chemometrics for the analysis of voltammetric data. *Trac-Trends Anal. Chem.* **2006**, 25, (1), 86-92.
16. Esteban, M.; Arino, C.; Diaz-Cruz, J. M.; Diaz-Cruz, M. S.; Tauler, R. Multivariate curve resolution with alternating least squares optimisation: a soft-modelling approach to metal complexation studies by voltammetric techniques. *Trac-Trends Anal. Chem.* **2000**, 19, (1), 49-61.
17. Cruz, B. H.; Diaz-Cruz, J. M.; Arino, C.; Esteban, M. Complexation of heavy metals by phytochelatins: Voltammetric study of the binding of Cd²⁺ and Zn²⁺ ions by the phytochelatin (gamma-Glu-Cys)(3)Gly assisted by multivariate curve resolution. *Environ. Sci. Technol.* **2005**, 39, (3), 778-786.
18. Herrero, A.; Zamponi, S.; Marassi, R.; Conti, P.; Ortiz, M. C.; Sarabia, L. A. Determination of the capability of detection of a hyphenated method: application to spectroelectrochemistry. *Chemometrics Intell. Lab. Syst.* **2002**, 61, (1-2), 63-74.
19. Teófilo, R. F.; Ceragioli, H. J.; Peterlevitz, A. C.; Da Silva, L. M.; Damos, F. S.; Ferreira, M. M. C.; Baranauskas, V.; Kubota, L. T. Improvement of the electrochemical properties of "as-grown" boron-doped polycrystalline diamond electrodes deposited on tungsten wires using ethanol. *J. Solid State Electrochem.* **2007**, 11, 1449-1457.
20. Compton, R. G.; Foord, J. S.; Marken, F. Electroanalysis at diamond-like and doped-diamond electrodes. *Electroanalysis* **2003**, 15, (17), 1349-1363.
21. Rao, T. N.; Fujishima, A. Recent advances in electrochemistry of diamond. *Diam. Relat. Mat.* **2000**, 9, (3-6), 384-389.
22. Granger, M. C.; Witek, M.; Xu, J. S.; Wang, J.; Hupert, M.; Hanks, A.; Koppang, M. D.; Butler, J. E.; Lucazeau, G.; Mermoux, M.; Strojek, J. W.; Swain, G. M. Standard electrochemical behavior of high-quality, boron-doped polycrystalline diamond thin-film electrodes. *Anal. Chem.* **2000**, 72, (16), 3793-3804.
23. Xu, J. S.; Granger, M. C.; Chen, Q. Y.; Strojek, J. W.; Lister, T. E.; Swain, G. M. Boron-doped diamond thin-film electrodes. *Anal. Chem.* **1997**, 69, (19), A591-A597.
24. Zhi, J. F.; Tian, R. H. Electrochemistry of diamond thin film. *Prog. Chem.* **2005**, 17, (1), 55-63.
25. Panizza, M.; Cerisola, G. Application of diamond electrodes to electrochemical processes. *Electrochim. Acta* **2005**, 51, (2), 191-199.
26. Gerger, I.; Haubner, R.; Kronberger, H.; Fafilek, G. Investigation of diamond coatings on titanium substrates for electrochemical applications. *Diam. Relat. Mat.* **2004**, 13, (4-8), 1062-1069.
27. Haenni, W.; Rychen, P.; Fryda, M.; Comninellis, C. Industrial applications of diamond electrodes. **2004**, 77, 149-196.

28. Robinson, V. S.; Show, Y.; Swain, G. M.; Reifenberger, R. G.; Fisher, T. S. Thermionic emission from surface-terminated nanocrystalline diamond. *Diam. Relat. Mat.* **2006**, 15, (10), 1601-1608.
29. Panizza, M.; Michaud, P. A.; Cerisola, G.; Comninellis, C. Anodic oxidation of 2-naphthol at boron-doped diamond electrodes. *J. Electroanal. Chem.* **2001**, 507, (1-2 SI Sp. Iss. SI), 206-214.
30. Puig, D.; Barcelo, D. Determination of phenolic compounds in water and waste water. *Trac-Trends Anal. Chem.* **1996**, 15, (8), 362-375.
31. Canizares, P.; Lobato, J.; Paz, R.; Rodrigo, M. A.; Saez, C. Electrochemical oxidation of phenolic wastes with boron-doped diamond anodes. *Water Res.* **2005**, 39, (12), 2687-2703.
32. Freire, R. S.; Duran, N.; Kubota, L. T. Development of a laccase-based flow injection electrochemical biosensor for the determination of phenolic compounds and its application for monitoring remediation of Kraft E1 paper mill effluent. *Anal. Chim. Acta* **2002**, 463, (2), 229-238.
33. Rosatto, S. S.; Freire, R. S.; Duran, N.; Kubota, L. T. Amperometric biosensors for phenolic compounds determination in the environmental interest samples. *Quim. Nova* **2001**, 24, (1), 77-86.
34. Kuch, H. M.; Ballschmiter, K. Determination of endocrine-disrupting phenolic compounds and estrogens in surface and drinking water by HRGC-(NCI)-MS in the picogram per liter range. *Environ. Sci. Technol.* **2001**, 35, (15), 3201-3206.
35. Robbins, R. J. Phenolic acids in foods: An overview of analytical methodology. *J. Agric. Food Chem.* **2003**, 51, (10), 2866-2887.
36. Lee, K. W.; Lee, H. J. The roles of polyphenols in cancer chemoprevention. *Biofactors* **2006**, 26, (2), 105-121.
37. Garg, R.; Kurup, A.; Hansch, C. Comparative QSAR: On the toxicology of the phenolic OH moiety. *Crit. Rev. Toxicol.* **2001**, 31, (2), 223-245.
38. Campanella, L.; Beone, T.; Sammartino, M. P.; Tomassetti, M. Determination of phenol in wastes and water using an enzyme sensor. *Analyst* **1993**, 118, (8), 979-986.
39. Gattrell, M.; Kirk, D. W. A fourier-transform infrared-spectroscopy study of the passive film produced during aqueous acidic phenol electrooxidation. *J. Electrochem. Soc.* **1992**, 139, (10), 2736-2744.
40. Gattrell, M.; Kirk, D. W. A study of electrode passivation during aqueous phenol electrolysis. *J. Electrochem. Soc.* **1993**, 140, (4), 903-911.
41. Gattrell, M.; Kirk, D. W. A study of the oxidation of phenol at platinum and preoxidized platinum surfaces. *J. Electrochem. Soc.* **1993**, 140, (6), 1534-1540.

42. Wang, J.; Jiang, M.; Lu, F. Electrochemical quartz crystal microbalance investigation of surface fouling due to phenol oxidation. *J. Electroanal. Chem.* **1998**, 444, (1), 127-132.
43. Ezerskis, Z.; Jusys, Z. Electropolymerization of chlorinated phenols on a Pt electrode in alkaline solution Part - I: A cyclic voltammetry study. *J. Appl. Electrochem.* **2001**, 31, (10), 1117-1124.
44. Muna, G. W.; Quaiserova-Mocko, V.; Swain, G. M. The analysis of chlorinated phenol solutions by capillary electrophoresis coupled with direct and indirect amperometric detection using a boron-doped diamond microelectrode. *Electroanalysis* **2005**, 17, (13), 1160-1170.
45. Sarada, B. V.; Rao, T. N.; Tryk, D. A.; Fujishima, A. Electroanalytical applications of conductive diamond electrodes. *New Diam. Front. Carbon Technol.* **1999**, 9, (5), 365-377.
46. De Carvalho, R. M.; Mello, C.; Kubota, L. T. Simultaneous determination of phenol isomers in binary mixtures by differential pulse voltammetry using carbon fibre electrode and neural network with pruning as a multivariate calibration tool. *Anal. Chim. Acta* **2000**, 420, (1), 109-121.

Parte Conceitual

“O mais difícil não é fazer o que é certo, mas saber o que é certo”

Lyndon Johnson

Capítulo 1

Métodos Quimiométricos: Uma Visão Geral

“Tanto quanto as leis da matemática referem-se à realidade, elas não são exatas; e tanto quanto elas são suficientemente exatas, elas não se referem à realidade”

Albert Einstein

Introdução

Embora muitas metodologias quimiométricas tenham originalmente sido desenvolvidas na área de ciências sociais, elas agora estão sendo aplicadas quase que como um procedimento de rotina na análise de dados oriundos de sistemas químicos. Assim, fica bem definido que a Quimiometria não é uma disciplina nem da matemática, nem da estatística ou da computação, mas sim da química. Os problemas que ela se propõe a resolver são de interesse e originados na química, ainda que as ferramentas de trabalho provenham principalmente da matemática, estatística e computação [1].

Informações químicas tais como, voltamogramas, espectros, cromatogramas, curvas de titulação e outras fontes podem ser digitalizadas formando uma série de números que podem ser representadas como vetores e matrizes [2]. Em quimiometria, os processamentos destes dados exigem, na maioria das vezes, operações algébricas bem definidas. Portanto, torna-se necessário padronizar as notações e convenções destas operações para em seguida apresentar as metodologias quimiométricas. Neste capítulo serão discutidas as notações, convenções e alguns pré-processamentos usados nesta tese e as seguintes metodologias quimiométricas: (1) planejamento e análise de experimentos; (2) análise de componentes principais (PCA); (3) regressão multivariada empregando quadrados mínimos parciais (PLS); (4) metodologias de seleção de variáveis em calibração multivariada; (5) cálculo do sinal analítico líquido (NAS) para

calibração multivariada inversa e estimativas de parâmetros analíticos (figuras de mérito); (6) análise dos fatores paralelos (PARAFAC) e (7) regressão em multimodos *i.e.*, PLS multilinear (*N*-PLS).

Nos subtítulos (4) e (7) serão apresentadas novas metodologias para seleção de variáveis em calibração multivariada (primeira e segunda ordem), elaboradas durante o desenvolvimento desta tese.

Notações, Convenções e alguns Pré-Tratamentos

Em toda a tese, as notações e convenções listadas abaixo serão seguidas. Outras notações especiais serão introduzidas quando necessárias. Ainda neste subtítulo, alguns pré-tratamentos que serão empregados em quase toda tese, serão descritos.

Variáveis e amostras

Vetores podem ser definidos como uma seqüência de escalares (números), matrizes como uma seqüência de vetores e tensores como uma seqüência de matrizes. Estas estruturas são apresentadas na Figura 1.1..

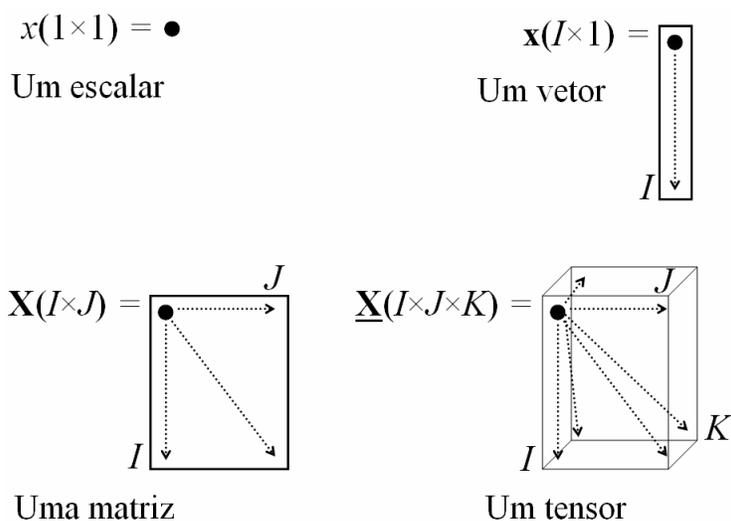


Figura 1.1. Ilustração do escalar, vetor, matriz e tensor. O círculo fechado (\bullet) representa as posições do elemento no vetor (1), na matriz (1,1) e no tensor (1,1,1). As setas indicam que há escalares em todas as outras posições.

A Tabela 1.1 representa uma matriz de dados com I linhas e J colunas. Por definição, nas colunas de uma matriz $\mathbf{X}(I \times J)$ estão dispostas as J variáveis e nas linhas estão as I diferentes amostras.

Usando voltamogramas como exemplo, a Tabela 1.1 pode ser construída da seguinte maneira: as variáveis $(1, 2, \dots, J)$ seriam os potenciais investigados; em cada posição das amostras $(1, 2, \dots, I)$ estariam as indicações dos diferentes voltamogramas e cada posição dentro do quadrado pontilhado (x_{ij}) estariam as correntes obtidas em cada potencial.

Somente os números presentes dentro do quadrado pontilhado na Tabela 1.1 fazem parte do único ou de um dos blocos considerados para a realização de cálculos quimiométricos.

Tabela 1.1. Representação esquemática de uma matriz de dados.

Amostras	Variáveis			
	1	2	...	J
1	x_{11}	x_{12}	x_{13}	x_{1J}
2	x_{21}	x_{22}	x_{23}	x_{2J}
\vdots	\vdots	\vdots	\vdots	\vdots
\vdots	\vdots	\vdots	\vdots	\vdots
I	x_{I1}	x_{I2}	x_{I3}	x_{IJ}

Escalares

Um escalar é um número que em química pode representar uma medida representativa (Figura 1.1). Em voltametria, por exemplo, ao obter uma corrente de pico, temos um número que é a corrente.

Os escalares serão indicados por letras minúsculas em itálico, podendo estar em letras gregas ou romanas.

a, b, c, ... , $\alpha, \beta, \gamma, \dots$

Vetores

Um vetor é uma seqüência de escalares que pode trazer alguma informação interpretável (Figura 1.1). Em voltametria, por exemplo, ao obter a corrente de pico de medidas diferentes (amostras), pode-se dispor cada escalar (corrente de pico individual) em uma seqüência de correntes. Ao observar esta seqüência de escalares no vetor pode-se inferir sobre as medidas e realizar interpretações.

Os vetores serão definidos como uma coluna de números e eles serão representados por letras minúsculas em negrito.

a, b, c, ... , α , β , γ , ...

Um vetor com I elementos terá uma dimensão ($I \times 1$) com seu i -ésimo elemento dado pela correspondente letra minúscula itálico com um subscrito apropriado. Por exemplo, o i -ésimo elemento do vetor \mathbf{x} é x_i . O vetor $\mathbf{1}$ define uma seqüência de números, todos eles iguais a 1.

Matrizes

Uma matriz é uma seqüência de vetores, todos eles com a mesma dimensão e organizados em uma tabela (Figura 1.1). Um voltamograma, por exemplo, é formado por uma seqüência de correntes obtidas com a varredura de potenciais. Esta seqüência de correntes constitui um vetor que traz toda a informação do voltamograma. Se mais de um voltamograma é obtido nas mesmas condições e estes voltamogramas forem dispostos seqüencialmente (amostras), forma-se uma tabela com tais voltamogramas (Tabela 1.1), onde cada ponto nesta tabela é uma corrente obtida no definido potencial. Esta tabela pode ser definida como uma matriz de dados contendo, neste caso, informações eletroquímicas.

Matrizes serão definidas por letras maiúsculas gregas ou romanas, em negrito.

A, B, C, ..., \mathbf{A} , \mathbf{B} , $\mathbf{\Phi}$, ...

Seus elementos serão representados pelas correspondentes letras minúsculas, em itálico com os subscritos definindo a linha e coluna, respectivamente, *e.g.*, os elementos da matriz \mathbf{X} são dados por x_{ij} .

Em quimiometria, os conjuntos de dados obtidos instrumentalmente são geralmente definidos pelo símbolo \mathbf{X} . No caso de dados de segunda ordem (*i.e.* uma matriz de dados) a letra \mathbf{X} maiúscula e em negrito será usada.

Algumas vezes, matrizes serão escritas como $\mathbf{X}(I \times J)$ para enfatizar que a matriz \mathbf{X} tem I linhas e J colunas. As matrizes identidades (Apêndice) de ordem J são escritas como \mathbf{I}_J , mas onde a ordem é óbvia a partir do contexto, o subscrito será omitido. Frequentemente será feita referência às colunas da matriz, e é, portanto conveniente definir as matrizes como uma série de vetores colunas. Assim, a matriz $\mathbf{X}(I \times J)$ também será definida como $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_J]$ em que \mathbf{x}_j , $j = 1, 2, \dots, J$ são as $(I \times 1)$ colunas de \mathbf{X} . Se for necessário considerar a matriz formada pelas primeiras S ($< J$) colunas de \mathbf{X} , a matriz será escrita como \mathbf{X}_S , em que $\mathbf{X}_S(I \times S) = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_S]$.

Tensores

Uma seqüência de matrizes de iguais dimensões organizadas de maneira a formar, por exemplo, um cubo de dados (arranjo de altas ordens), é chamado de tensor (Figura 1.1) [3]. Os tensores serão definidos por letras maiúsculas, em negrito e sublinhadas, podendo estar em letras gregas ou romanas, *i.e.*, $\underline{\mathbf{A}}, \underline{\mathbf{B}}, \underline{\mathbf{C}}, \dots, \underline{\mathbf{\Phi}}$.

Seus elementos serão representados como indicado para as matrizes, porém, os subscritos estarão definindo os modos do tensor *e.g.*, os elementos do tensor $\underline{\mathbf{X}}$ serão dados por x_{ijk} . Algumas vezes, tensores serão escritos como $\underline{\mathbf{X}}(I \times J \times K)$ para enfatizar que o tensor $\underline{\mathbf{X}}$ tem I matrizes no modo 1, J matrizes no modo 2 e K matrizes no modo 3 (Figura 1.1).

Um tensor é frequentemente considerado em termos de seu conjunto de matrizes. Estas matrizes formam as fatias horizontal, lateral e frontal de um tensor em três dimensões (Figura 1.2). Especificamente, a fatia horizontal i pertence às entidades $i = 1, \dots, I$ do modo 1, a fatia lateral j pertence às entidades $j = 1, \dots, J$ do modo 2, a fatia frontal k pertence às entidades $k = 1, \dots, K$ do modo 3.

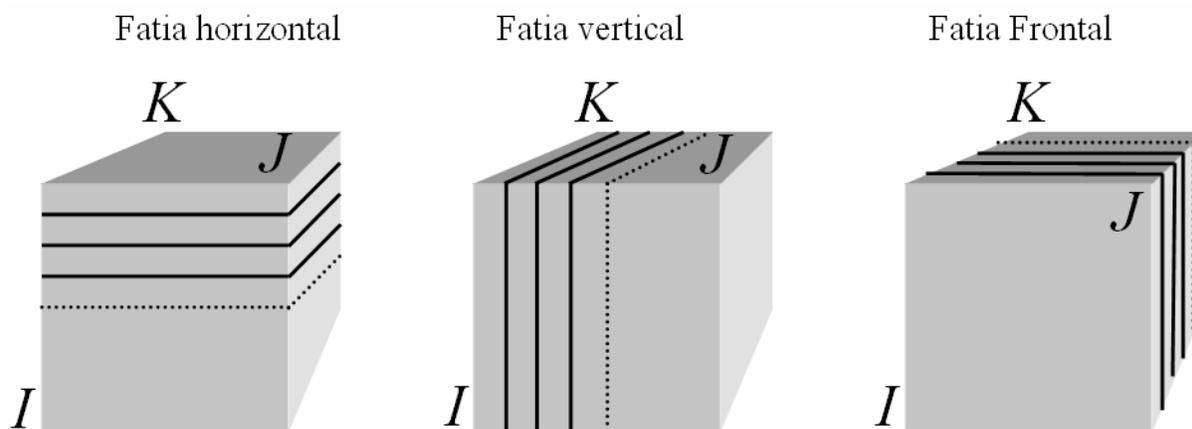


Figura 1.2. Ilustração do arranjo tensorial (três modos) em fatias (dois modos).

Os tensores, como as matrizes, são um bloco de dados. Assim, as definições para variáveis e amostras seguem a mesma idéia, porém mais de duas dimensões são consideradas para este tipo de dado e esta nova dimensão é definida como mais uma variável da dimensão adicional. Deste modo, as I linhas representam a dimensão 1, das amostras (modo 1), as J colunas representam as variáveis da dimensão 2 (modo 2) e os K tubos representam as variáveis da dimensão 3 (modo 3), conforme indicado na Figura 1.3.

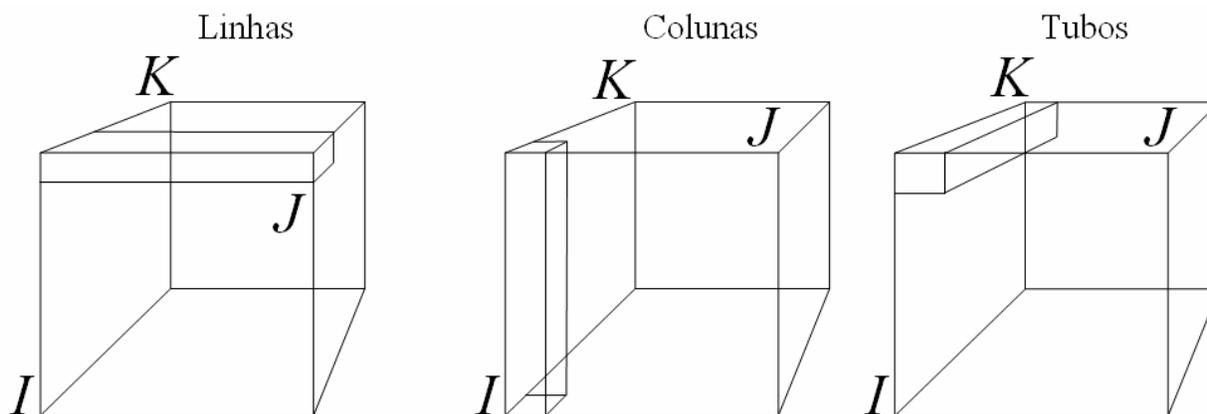


Figura 1.3. Definição de linhas, colunas, e tubos em um arranjo de três dimensões.

É, porém, útil coletar todos os modos em uma única matriz. O processo de rearranjo dos elementos de $\underline{\mathbf{X}}$ dentro de \mathbf{X} é frequentemente chamado em quimiometria de desdobramento, mas este termo é confuso, visto que em

psicometria o termo desdobramento é uma técnica particular para escalar dados multidimensionais [4].

Aqui este processo será denominado como matriciação de um tensor em uma matriz e o processo reverso será chamado de reforma de uma matriz em um tensor.

As matriciações podem ser realizadas em todos os modos (Figura 1.4). Assim, $J \times KI$, corresponde às entidades do modo 3 embutidas nas entidades do modo 1; $K \times IJ$, corresponde às entidades do modo 1 embutidas nas entidades do modo 2 e $I \times JK$, corresponde às entidades do modo 2 embutidas nas entidades do modo 3.

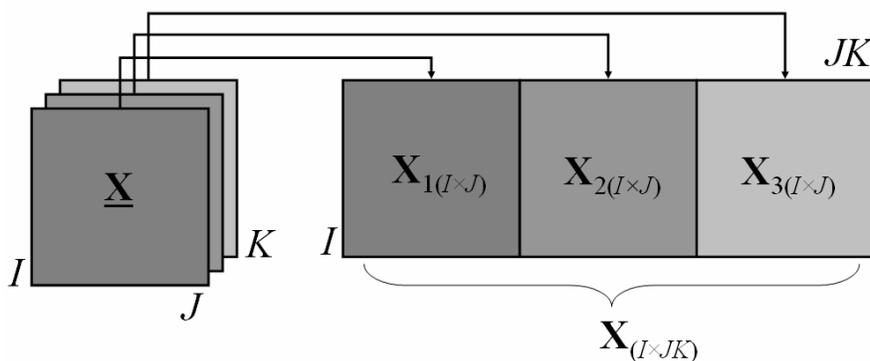


Figura 1.4. Matriciação de um arranjo em três dimensões $\underline{\mathbf{X}}$ para um de duas dimensões $\mathbf{X}_{(I \times JK)}$.

Símbolos

Os sobrescritos t , $+$ e -1 indicam transposição, pseudo-inversa e inversa, respectivamente. As barras $\| \cdot \|$ referem-se à norma euclidiana e \mathbf{I} simboliza uma matriz identidade dimensionada apropriadamente. O símbolo $\hat{\cdot}$ indica a matriz, vetor ou escalar estimado. Detalhes sobre transposição, pseudo-inversa, inversa e norma euclidiana estão descritos no apêndice.

Pré-Processamentos: Centrar na Média e Escalar pela Variância

Na maioria dos métodos quimiométricos que serão descritos nesta tese é assumido que as colunas das matrizes são centradas na média e, algumas vezes, que elas também foram escaladas pela variância da coluna. Quando estes dois pré-processamentos são realizados simultaneamente, o que também é comum de se

assumir; é dito que as colunas da matriz estão autoescaladas pela variância, ou somente, que a matriz está autoescalada.

Centrar os dados na média equivale geometricamente, a fazer uma translação do sistema de eixos ao longo do vetor das médias, para o centro do conjunto de dados.

Escalar os dados tem como objetivo colocar variáveis em uma mesma escala quando tais unidades apresentam diferentes unidades entre si ou quando a faixa de variação dos dados é grande.

Se x_{ij} representa o (ij) -ésimo elemento da matriz \mathbf{X} , então a centragem dos dados na média é realizada da seguinte maneira

$$x_{ij}^c = x_{ij} - \bar{x}_j \quad 1.1$$

em que $\bar{x}_j = \sum_{i=1}^I x_{ij} / I$ é a média aritmética dos elementos da j -ésima coluna, \mathbf{x}_j sendo I o número de elementos na coluna.

Quando as colunas de \mathbf{X} são escaladas pela variância temos:

$$x_{ij}^e = x_{ij} / s_j \quad 1.2$$

em que $s_j = \sqrt{\frac{\sum_{i=1}^I (x_{ij} - \bar{x}_j)^2}{I - 1}}$, que representa o desvio padrão da j -ésima coluna. O

autoescalamamento é obtido da seguinte maneira:

$$x_{ij}^a = (x_{ij} - \bar{x}_j) / s_j \quad 1.3$$

Durante a tese não será feita nenhuma distinção de notação entre a matriz \mathbf{X} original e a pré-processada por centragem ou por autoescalamamento.

Transformação: Correção de Linha de Base

Durante a obtenção dos dados, variações sistemáticas podem ocorrer e devem ser removidas ou corrigidas antes das análises [5,6].

Em análises eletroquímicas, muitas variáveis podem influenciar o sinal, o que pode comprometer a qualidade dos resultados. Podemos destacar como influências,

o deslocando da linha de base para maiores ou menores valores de corrente; deformações do sinal, saliências indesejáveis na parte íngreme do sinal, ou um sinal obtido sobre uma corrente de fundo inclinada.

Uma maneira de resolver parte destes problemas é através da subtração do branco da respectiva amostra. Entretanto, algumas saliências podem ainda permanecer e o deslocamento na linha de base não é completamente corrigido. Além disso, nem sempre é possível obter um sinal de branco para cada amostra.

Uma maneira prática e eficiente de resolver estes tipos de problemas é aplicando a correção de linha de base para cada amostra usando uma média móvel. Durante a aplicação deste método o número de pontos nos dados (número de variáveis J) é substituído temporariamente pelo cálculo da média dentro de uma janela de potenciais. A abertura da janela deve ser definida como tendo a menor largura do pico do voltamograma obtido. A linha de base é subsequentemente calculada comparando cada valor de corrente com o valor médio de seus dois vizinhos. Se o valor absoluto da média for menor, a corrente é substituída pelo valor atual da média. Esta operação é realizada repetidas vezes até que nenhum valor de corrente seja mais substituído. Depois que a correção da linha de base é realizada com sucesso, um sinal com um pico bem definido é obtido [7]. É importante ressaltar que ao final do processo o número de variáveis originais não é alterado.

Planejamento e Análise de Experimentos

Um experimento é apenas um teste de uma série de testes. Experimentos são executados em todas as disciplinas científicas e são uma importante parte do nosso aprendizado sobre o funcionamento dos sistemas e processos. A validade das conclusões que são inferidas de um experimento depende da escala em que o experimento foi conduzido. Portanto, o planejamento de experimentos representa um dos principais papéis na eventual solução do problema que inicialmente motivou o experimento [8].

Os experimentos podem ser executados em duas ou mais variáveis (ou fatores) que o experimentalista supõe serem importantes, como por exemplo, pH e tipo de enzima. Os planejamentos fatoriais é uma poderosa ferramenta para investigar este

tipo de problema. Geralmente, em um planejamento experimental, os ensaios são executados em todas as combinações dos níveis dos fatores, normalmente dois ou três.

Por outro lado, a metodologia de superfície de resposta (RSM) é uma coleção de técnicas matemáticas e estatísticas que são úteis para modelar e analisar sistemas onde a resposta de interesse é influenciada por várias variáveis, e cujo objetivo é otimizar a resposta. Neste caso, em relação aos planejamentos fatoriais, mais níveis das variáveis são investigados em uma maior região experimental.

Para analisar os resultados de ambas as metodologias, vários gráficos e tabelas são utilizadas. O método de regressão por quadrados mínimos é extremamente necessário e normalmente, o primeiro cálculo realizado. A análise de variância (ANOVA), em particular, é usada como uma das principais ferramentas para a análise estatística.

Nesta seção serão descritos os planejamentos fatoriais completos e fracionários, normalmente empregados na etapa de triagem. Para a etapa de otimização serão descritos os planejamentos compostos centrais e Doehlert, que se baseiam na metodologia de superfície de resposta.

Experimentos para Triagem

Muitas vezes em um sistema, diversos fatores ou variáveis (os termos fatores e variáveis serão usados nesta seção indistintamente) podem influenciar a resposta desejada. Um experimento para triagem é executado com o interesse em se determinar as variáveis experimentais e as interações entre variáveis que têm influência significativa sobre as diferentes respostas de interesse [8,9].

Após selecionar as variáveis que são possíveis de serem estudadas e que provavelmente interferem no sistema, é preciso avaliar a metodologia experimental (tempo, custo, etc.). As variáveis que não forem selecionadas devem ser fixadas durante todo o experimento.

Em uma etapa seguinte, deve-se escolher qual planejamento usar para estimar a influência (o efeito) das diferentes variáveis no resultado. No estudo de triagem, as interações entre as variáveis (interações principais) e de segunda ordem, obtidas

normalmente pelos planejamentos fatoriais completos ou fracionários, são de extrema importância para a compreensão do comportamento do sistema.

Planejamentos Fatoriais Completos

Em um planejamento fatorial são investigadas as influências de todas as variáveis experimentais de interesse e os efeitos de interação (sinergismo) entre elas na resposta ou respostas. Se a combinação de J fatores é investigada em dois níveis, um planejamento fatorial consistirá de 2^J experimentos. Normalmente, os níveis dos fatores quantitativos (*i.e.* concentrações de uma substância, valores de pH, etc.) são designados pelos sinais – (menos) para o nível mais baixo e + (mais) para o nível mais alto, porém o que importa é a relação inicial entre o sinal atribuído e o efeito obtido, não sendo um critério definido a nomeação dos sinais. Para fatores qualitativos (*i.e.* tipos de ácidos, tipos de catalisadores, etc.), como não existem valores altos ou baixos, fica a critério do experimentalista atribuir os seus níveis.

Os sinais para os efeitos de interação de 2ª ordem e de ordem superior entre todas as variáveis do planejamento, realizando todas as combinações possíveis, são obtidos pelo produto dos sinais originais das variáveis envolvidas. Desta maneira é possível construir as colunas de sinais para todas as interações e assim elaborar a matriz de coeficientes de contraste (Tabela 1.2).

Tabela 1.2. Matriz de coeficientes de contraste para um planejamento fatorial 2^3 .

Média	Variáveis			Interações*			
	x_1	x_2	x_3	x_{12}	x_{13}	x_{23}	x_{123}
+	-	-	-	+	+	+	-
+	+	-	-	-	-	+	+
+	-	+	-	-	+	-	+
+	+	+	-	+	-	-	-
+	-	-	+	+	-	-	+
+	+	-	+	-	+	-	-
+	-	+	+	-	-	+	-
+	+	+	+	+	+	+	+

* x_{12} , x_{13} e x_{23} , são os efeitos de interação de 2ª ordem; x_{123} é o efeito de interação de 3ª ordem.

Uma coluna de sinais + (mais) é adicionada à esquerda da matriz de coeficientes de contraste para o cálculo da média de todas as respostas observadas.

Suponha um planejamento fatorial com I ensaios e com y_i observações individuais (quando houver replicatas, considere a resposta média \bar{y}_i). Os efeitos para cada coluna da matriz de coeficientes de contraste (conforme Tabela 1.2) são dados pelas seguintes equações:

$$ef_{\text{médio}} = \frac{\sum_{i=1}^I y_i}{I} \quad 1.4$$

$$ef = \frac{\sum_{i=1}^{1/2} y_{i(+)} - \sum_{i=1}^{1/2} y_{i(-)}}{1/2} \quad 1.5$$

A equação 1.4 descreve o efeito para a média de todas as observações, enquanto a equação 1.5 descreve a diferença entre as médias das observações no nível mais ($y_{i(+)}$) e as médias das observações no nível menos ($y_{i(-)}$).

Outro método para o cálculo dos efeitos para um planejamento fatorial completo será descrito mais adiante no item: Modelos Empíricos em Estudo de Triagem.

Planejamentos Fatoriais Fracionários

O planejamento fatorial completo necessita de 2^J ensaios para sua execução, portanto sua principal desvantagem é o grande número de ensaios que devem ser realizados a cada fator adicionado ao estudo. Se considerarmos (e observarmos) que os efeitos de altas ordens, principalmente para planejamentos com $J > 4$, são quase sempre não significativos, a realização de ensaios para estimar tais efeitos de interação é desnecessária. Desta maneira, com um número menor de experimentos, é possível obter informações daqueles efeitos mais importantes e retirar, na maioria das vezes, as mesmas conclusões caso fosse realizado um planejamento fatorial completo. Os planejamentos que apresentam estas características são conhecidos como planejamentos fatoriais fracionários [9-13].

Há muitos e diferentes tipos de planejamentos fatoriais fracionários ($1/2$, $1/4$, $1/8$, $1/16 \dots 1/2^b$ de um planejamento 2^{J-b}) descritos na literatura [9-11], em que J é o número de variáveis investigadas e b é o tamanho da fração. O tamanho da fração

influenciará no possível número de efeitos a serem estimados, e conseqüentemente, no número de experimentos necessários [9].

Pode-se ainda dizer que há dois tipos de frações: aquelas cujo objetivo é obter somente os efeitos principais [12] e aquelas em que se adicionam experimentos para separar e estimar efeitos de interações, caso se assuma que tais interações apresentam influência sobre os resultados [9]. No primeiro caso é necessário executar, por exemplo, apenas 8 experimentos para investigar 7 variáveis, 12 experimentos para 11 variáveis, etc. Estes são conhecidos como planejamentos saturados [9-11] e planejamentos de Plackett-Burman [12].

Construção do planejamento fatorial fracionário

Normalmente, para a construção do planejamento fatorial fracionário desejado, utiliza-se um planejamento fatorial completo. Para exemplificar, consideremos um planejamento 2^{3-1} construído a partir de um planejamento fatorial completo com duas variáveis: V_1 e V_2 . A Tabela 1.3 apresenta o planejamento fatorial completo com duas variáveis. Serão necessários quatro experimentos para a execução deste planejamento 2^2 e a partir dos resultados podem-se obter os efeitos principais das duas variáveis e o efeito de interação (V_1V_2). A coluna da matriz de coeficientes de contraste responsável pela interação é obtida pela multiplicação dos elementos da coluna da variável V_1 com os respectivos elementos da variável V_2 . As colunas de V_1 , V_2 e V_1V_2 da matriz de coeficientes de contraste de um planejamento fatorial completo definem a configuração de um planejamento fatorial fracionário com três variáveis utilizando apenas os quatro ensaios destacados em negrito na Tabela 1.4, em que V_1 , V_2 e V_1V_2 serão substituídas pelas variáveis independentes x_1 , x_2 e x_3 .

Como o número de experimentos é a metade do completo, temos uma meia fração de um planejamento fatorial 2^3 ($\frac{1}{2}2^3 = 2^{-1}2^3 = 2^{3-1}$), conforme as Tabelas 1.3 e 1.4.

Tabela 1.3. Planejamento fatorial fracionário 2^{3-1} a partir de um planejamento fatorial completo 2^2 .

Nº exp. ^(*)	x_1 (V_1)	x_2 (V_2)	x_3 (V_1V_2)
5	-1	-1	1
2	1	-1	-1
3	-1	1	-1
8	1	1	1

(*) Ensaios correspondentes obtidos da Tabela 1.3.

Uma outra propriedade importante dos planejamentos fatoriais fracionários diz respeito aos experimentos selecionados que cobrem o volume máximo do domínio considerado em um número limitado de experimentos [9]. Note nas Tabelas 1.2 e 1.3 que os experimentos destacados são comuns aos dois planejamentos.

A Figura 1.5 destaca como os experimentos de meia fração “selecionados” se distribuem em um domínio experimental para três variáveis (Tabela 1.4).

Tabela 1.4. Planejamento fatorial completo 2^3 .

Nº exp.	x_1	x_2	x_3
1	-1	-1	-1
2	1	-1	-1
3	-1	1	-1
4	1	1	-1
5	-1	-1	1
6	1	-1	1
7	-1	1	1
8	1	1	1

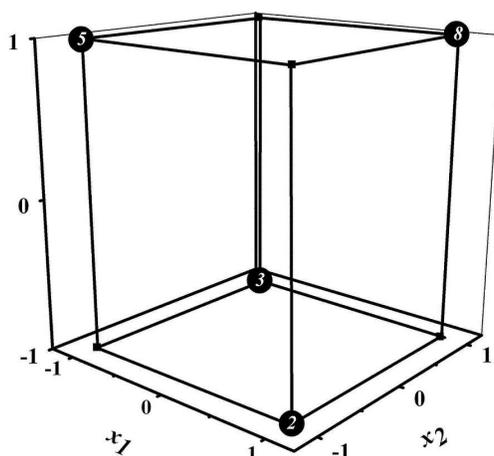


Figura 1.5. Experimentos “selecionados” para um planejamento fatorial fracionário 2^{3-1} .

Para a montagem de um planejamento fatorial fracionário saturado, considere o seguinte exemplo: sete variáveis podem ser estudadas, com um mínimo de experimentos, em um planejamento fatorial fracionário com fração 1/16, ou seja, 2^{7-4} ($\frac{1}{16}2^7 = 2^{-4}2^7 = 2^{7-4}$). Para este caso, o planejamento é definido pelo modelo de matriz para um planejamento fatorial 2^3 (Tabela 1.4).

Um planejamento fatorial completo com sete variáveis necessita de 128 experimentos. Sendo o planejamento fracionário 2^{7-4} , 1/16 do planejamento completo, ele necessitará de somente 8 experimentos.

Neste caso, para a elaboração da matriz de planejamento a partir da matriz mostrada na Tabela 1.4, as variáveis x_4 , x_5 , x_6 e x_7 , serão os produtos das colunas das variáveis x_1x_2 , x_1x_3 , x_2x_3 e $x_1x_2x_3$, respectivamente.

Efeitos confundidos e resolução

Certamente há perda de informações quando se realizar um planejamento fatorial fracionário. Os efeitos principais são misturados com os efeitos de interação e esta contaminação aumenta entre as interações quando se aumenta a fração do planejamento.

Para entender melhor a confusão causada por esta mistura, considere uma meia fração de um planejamento fatorial 2^{5-1} (Tabela 1.5). O objetivo é obter todos os efeitos principais e todos os possíveis efeitos de interação realizando apenas 16 ensaios. Uma matriz de planejamento é elaborada a partir de um planejamento 2^4 . A coluna da quinta variável (5) é obtida a partir da multiplicação de todas as outras, ou seja, **1234**. Assim **5 = 1234**, é chamado de *gerador* de um planejamento fracionário [11] [13]. Observa-se pela Tabela 1.5 que os coeficientes de contraste para o efeito de **123** têm os mesmos sinais de **45**. Isto mostra que os efeitos estimados por estas duas interações serão os mesmos, ou seja, serão confundidos, sendo assim chamados de pseudônimo um do outro.

A meia fração do planejamento mostrada na Tabela 1.5 foi obtida a partir do *gerador* (**5=1234**), no entanto, para conhecer a identidade de um pseudônimo é

realizada uma *definição de contrastes* a partir da relação apresentada pela equação 1.6.

$$C = 12345 \quad 1.6$$

O pseudônimo é obtido multiplicando o definido contraste C por cada um dos efeitos. Por exemplo: o pseudônimo de I é obtido multiplicando I por 12345 . Considere também, como regra, que se um mesmo termo aparecer mais de uma vez na multiplicação, este termo desaparece. Portanto, $112345 = 2345$. Para 12 o pseudônimo é $1212345 = 345$.

Tabela 1.5. Coeficientes de contraste para um planejamento fatorial fracionário 2^{5-1} .

Nº exp.	1	2	3	4	5	1234	123	45	2345
1	-	-	-	-	+	+	-	-	-
2	+	-	-	-	-	-	+	+	+
3	-	+	-	-	-	-	+	+	-
4	+	+	-	-	+	+	-	-	+
5	-	-	+	-	-	-	+	+	-
6	+	-	+	-	+	+	-	-	+
7	-	+	+	-	+	+	-	-	-
8	+	+	+	-	-	-	+	+	+
9	-	-	-	+	-	-	-	-	+
10	+	-	-	+	+	+	+	+	+
11	-	+	-	+	+	+	+	+	-
12	+	+	-	+	-	-	-	-	+
13	-	-	+	+	+	+	+	+	-
14	+	-	+	+	-	-	-	-	+
15	-	+	+	+	-	-	-	-	-
16	+	+	+	+	+	+	+	+	+

Se outro fator F é adicionado ao planejamento 2^{5-1} (Tabela 1.5) ele passa a ser um planejamento 2^{5-2} e, portanto, um quarto do fatorial completo. Para obter este novo fator, uma outra *definição de contrastes* é obtida de um gerador adicional. Assim $5 = 1234$ e $6 = 123$. Portanto, $C = 12345$ e $C = 1236$, respectivamente. Uma terceira *definição de contrastes* é então obtida multiplicando os dois anteriores conforme a equação 1.7.

$$C = 123451236 = 456 \quad 1.7$$

O pseudônimo para cada efeito pode agora ser obtido pela maneira usual. Para I temos: $I = 112345 = 11236 = 1456$ ou $I = 2345 = 236 = 1456$.

Assim, com um quarto do planejamento 2^5 , cada efeito apresenta três pseudônimos e a estimativa dos efeitos principais são individuais, no entanto eles se confundem com efeitos de interação de 2ª ordem.

Este tipo de confusão pode ocasionalmente causar dificuldades na interpretação dos resultados, mas isto é facilmente contornado por adição de poucos experimentos complementares a fim de separar efeitos confundidos.

Um importante conceito em planejamentos fatoriais fracionários é a resolução do planejamento, que define a ordem em que se negligenciam os efeitos e é definida por um número romano colocado depois do planejamento ou subscrito. Em geral, a resolução é o valor do número de fatores que compõem o termo de menor comprimento nas definições de contrastes C [11]. Para ilustrar:

Resolução III – Não confunde efeitos principais entre si, mas os confunde com efeitos de interação entre dois fatores.

Resolução IV – Não confunde efeitos principais entre si e nem com efeitos de interação entre dois fatores, mas confunde efeitos principais com efeitos de interação entre três variáveis e os efeitos entre duas variáveis se confundem com outros efeitos, inclusive entre eles.

Resolução V – Os efeitos principais são confundidos com efeitos de interação entre quatro variáveis e os efeitos de interação entre duas variáveis são confundidos com efeitos de interação entre três variáveis.

Planejamentos fatoriais fracionários com resolução maior que V são raramente usados em triagem [9].

Os cálculos para estimativa dos efeitos para um planejamento fatorial fracionário serão descritos mais adiante no item: Modelos Empíricos em Estudo de Triagem.

Planejamentos Fatoriais com Ponto Central

Em muitos casos, a realização de repetições autênticas pode ser algo inconveniente por diversas razões. Para contornar este infortúnio e obter uma boa estimativa dos erros, um experimento é normalmente incluído no centro do

planejamento (Figura 1.6), em que o valor médio dos níveis de todas as variáveis é empregado. São os conhecidos experimentos no ponto central (nível zero). Deste modo, é possível avaliar a significância dos efeitos ou coeficientes, tanto em planejamentos de triagem (completos ou fracionários) como em metodologias de superfície de resposta (discutidos mais adiante). Além desta vantagem, recomenda-se este tipo de experimento pelas seguintes razões [9]:

- o risco de perder a relação não linear entre os intervalos é minimizado;
- é possível estimar um modelo razoável e verificar se há falta de ajuste;

Logicamente não há como fugir das repetições, mas o número destas, na maioria dos casos, é significativamente reduzido.

No entanto, esta metodologia é somente possível de ser aplicada quando se utilizam variáveis quantitativas, visto que, para variáveis qualitativas não há como adicionar níveis no ponto central.

Os cálculos para estimativa de efeitos e coeficientes relacionados aos planejamentos com ponto central serão discutidos mais adiante no item: Modelos Empíricos em Estudo de Triagem.

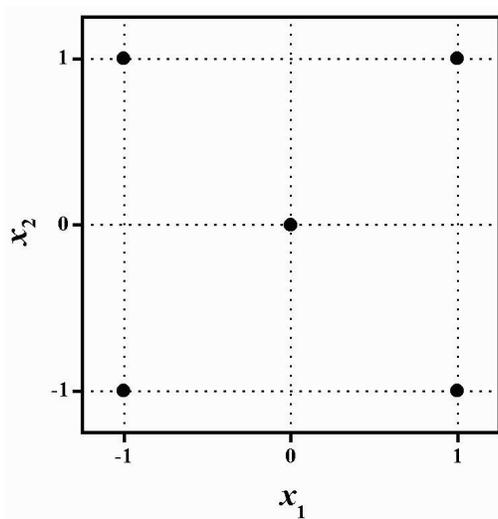


Figura 1.6. Planejamento 2^2 com ponto central.

Modelos Empíricos em Estudo de Triagem

Pode-se assumir, desde o início do experimento, que o sistema estudado (domínio experimental) é regido por uma expressão matemática que é função das

variáveis experimentais. Normalmente esta função pode ser aproximada por um polinômio, o qual pode fornecer uma boa descrição entre os fatores e a resposta obtida. A ordem deste polinômio é limitada pelo tipo de planejamento usado. Planejamentos fatoriais de dois níveis, completos ou fracionários, podem estimar apenas efeitos principais e interações. Planejamentos fatoriais de três níveis (ponto central) podem estimar, além disso, o grau de curvatura na resposta.

Para descrever tais modelos em um estudo de triagem, são utilizados os polinômios mais simples, ou seja, aqueles que contêm apenas termos lineares. Considerando um exemplo para três variáveis, x_1 , x_2 e x_3 dois polinômios seriam:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + e \quad 1.8$$

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3 + e \quad 1.9$$

Segundo a equação 1.8, o coeficiente b_0 é o valor populacional da média de todas as respostas obtidas, b_1 , b_2 e b_3 são os coeficientes relacionados com as variáveis x_1 , x_2 , e x_3 respectivamente, e é o erro aleatório associado ao modelo e, para o caso da equação 1.9, b_{12} , b_{13} , b_{23} são os coeficientes para as interações x_1x_2 , x_1x_3 , x_2x_3 e b_{123} é o coeficiente para a interação $x_1x_2x_3$.

A partir do planejamento montado, pode-se construir a matriz de coeficientes de contraste, de acordo com a Tabela 1.1. A matriz de coeficientes de contraste \mathbf{X} , juntamente com o vetor de respostas \mathbf{y} , obtido experimentalmente, será utilizada para o cálculo dos coeficientes do modelo, conhecidos também como *vetor de regressão*.

O modelo procurado, descrito pelas equações 5 e 6 pode ser representado matricialmente pela equação 1.10,

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{b} \quad 1.10$$

em que $\hat{\mathbf{y}}$ é o vetor das respostas estimadas pelo modelo e \mathbf{b} o vetor de regressão.

Uma maneira de determinar o vetor de regressão \mathbf{b} é através do *método dos quadrados mínimos* [14] [15], definido pela equação 1.13, em que \mathbf{X}^t indica a transposta de \mathbf{X} .

$$\mathbf{b} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y} \quad 1.11$$

em que $(\mathbf{X}^t\mathbf{X})^{-1}$ é a matriz inversa do produto da transposta da matriz \mathbf{X} com ela mesma.

Para um planejamento fatorial completo, a matriz $\frac{1}{\sqrt{I}}\mathbf{X}$ é quadrada e ortogonal onde I corresponde ao número de ensaios. Isto ocorre pois as colunas de \mathbf{X} não estão normalizadas, portanto $\mathbf{X}^{-1} = \frac{1}{I}\mathbf{X}^t$. Neste caso a equação 1.13 se resume à equação 1.12.

$$\mathbf{b} = \frac{1}{I}\mathbf{X}^t\mathbf{y} \quad 1.12$$

A equação 1.11 é geral e pode ser empregada para fazer a estimativa de efeitos e coeficientes para todos os planejamentos descritos neste artigo, no entanto a equação 1.12 é específica para o planejamento fatorial completo, não sendo aplicável aos outros planejamentos descritos.

Como as variáveis são estudadas em dois níveis codificados, cada efeito satisfaz à variação de duas unidades da variável correspondente. Se considerarmos os fatores individualmente, pode-se provar que o valor de cada coeficiente do modelo é a metade do valor do efeito correspondente, exceto para b_0 , cujo valor é o mesmo do calculado para seu efeito (equação 1.6). Desta maneira, o modelo empregado para descrever as respostas é elaborado em função dos efeitos por unidade de variação individual [10].

Estimativa dos Erros para os Efeitos através de Repetições

Normalmente, os resultados obtidos em experimentos de planejamento fatorial completo ou fracionário com repetições, consistem de uma pequena amostra de um hipotético conjunto maior, representado por uma população. Destes dados podemos obter a média amostral, a variância amostral e o desvio padrão amostral, como se segue:

$$\bar{y} = \frac{\sum_{r=1}^R y_r}{R} \quad 1.13$$

$$s^2 = \frac{\sum_{r=1}^R (y_r - \bar{y})^2}{(R-1)} \quad 1.14$$

$$s = \sqrt{\frac{\sum_{r=1}^R (y_r - \bar{y})^2}{(R-1)}} \quad 1.15$$

em que R é o número de replicatas, isto é, o número de ensaios realizados em um mesmo ponto experimental (nível), y_r são os valores de cada observação individual, \bar{y} é o valor médio, s^2 é a variância e s o desvio padrão.

A soma dos desvios da média amostral de R replicatas é necessariamente zero. Isto requer que $\sum (y_r - \bar{y}) = 0$ constitua uma restrição linear nos desvios usados no cálculo de s^2 . Está subentendido que com $R - 1$ replicatas é possível determinar a que falta. Os R resíduos $y - \bar{y}$ e consequentemente a soma dos seus quadrados juntamente com a variância amostral, são ditas ter, por esta razão, $R - 1$ graus de liberdade. A perda de um grau de liberdade está associada à necessidade de substituir a média populacional pela média amostral derivada dos dados. Desta maneira, quando repetições genuínas são realizadas em uma série de condições experimentais, a variação entre suas observações pode ser usada para estimar o desvio padrão de uma simples observação e consequentemente o desvio dos efeitos [11]. Todos os ensaios, inclusive repetições, devem ser realizados aleatoriamente, refazendo todas as etapas do experimento. As repetições realizadas desta maneira são consideradas genuínas.

Em um experimento em que cada ensaio foi realizado r vezes, se o valor de R for pequeno, por exemplo, $2 \leq R \leq 10$, haverá poucos graus de liberdade para o cálculo da variância. Uma maneira para se obter um maior número de graus de liberdade é realizando uma estimativa conjunta das variâncias conforme a equação 1.16.

$$s_c^2 = \frac{\sum_{i=1}^I [(R_i - 1)s_i^2]}{\sum_{i=1}^I (R_i - 1)} \quad 1.16$$

$$v = \sum_{i=1}^I (R_i - 1) \quad 1.17$$

O resultado da equação 1.16 reflete a variância conjunta (s_c^2) de cada observação individual y_i sendo o erro padrão igual à raiz quadrada do mesmo. A equação 1.17, que é exatamente o denominador da equação 1.16 é, portanto, o número de graus de liberdade da estimativa conjunta, designada como v .

Observa-se através das equações 1 e 2 e considerando a realização de repetições autênticas, que cada efeito é uma combinação linear dos valores de y_i dos I ensaios, com coeficientes c iguais a $+2/I$ e $-2/I$. Levando em consideração a aleatoriedade dos ensaios, tais valores são estatisticamente independentes e apresentam a mesma variância populacional σ^2 . Neste caso, por definição, as correlações entre todas as variáveis são nulas, e a variância da combinação linear das variáveis aleatórias pode ser dada por [10]:

$$\sigma^2 = \sum_i c_i^2 \sigma_i^2 \quad 1.18$$

Transportando a equação 1.18 para o nosso mundo amostral, pode-se provar que a variância de cada efeito é dada por [10]:

$$V(ef) = \frac{4s_c^2}{\sum_{i=1}^I R_i} \quad 1.19$$

em que s_c^2 é dada pela equação 1.16.

Finalmente, para calcular o valor do erro padrão de cada efeito, basta extrair a raiz quadrada de $V(ef)$.

Estimativa dos Erros para os Efeitos, sem Repetições

Uma maneira de estimar erros de efeitos sem a realização de repetições é supor que interações de altas ordens para $J \geq 3$ não são significativos e, portanto, são erros experimentais nos valores dos efeitos.

Aplicando a equação 1.16 sobre estes efeitos de interação e fazendo algumas considerações, obtém-se a variância dos efeitos conforme a equação 1.20.

$$V(ef) = \frac{\sum_{i=1}^l (efi_i)^2}{l} \quad 1.20$$

em que efi_i são os efeitos de interação considerados como erros experimentais e l é o número total de efeitos considerados.

É preciso estar atento ao utilizar este tipo de estimativa do erro. Nem sempre os efeitos de altas ordens são irrelevantes e, se tais efeitos forem incluídos no cálculo, os erros se tornam altos e desta maneira não é possível distinguir com confiabilidade aqueles que são realmente importantes.

O número de graus de liberdade utilizado para a avaliação dos efeitos agregados a estes erros é o mesmo número total de efeitos considerados como erros, isto é, o valor l da equação 1.20.

Estimativa dos Erros para os Efeitos e Coeficientes a partir das Repetições no Ponto Central

Conforme mencionado anteriormente, uma das grandes vantagens da inclusão de experimentos no centro do planejamento é devido a estimativa do erro com poucas repetições, normalmente entre 3 e 5.

A partir das repetições realizadas é possível obter a média e a variância das replicatas de acordo com as equações 1.13 e 1.14, respectivamente.

Por outro lado, sendo os coeficientes obtidos através da equação 1.11, nota-se que a matriz $(\mathbf{X}^t\mathbf{X})^{-1}$ apresenta grande influência na variância dos parâmetros de regressão. O produto desta matriz com o valor da variância obtida através das repetições no ponto central (equação 1.14) fornece a matriz $\mathbf{V}(b)$ conhecida como matriz de *variância-covariância* (equação 1.21). A matriz $\mathbf{V}(b)$ é simétrica e os elementos de sua diagonal são as variâncias dos parâmetros de regressão na mesma ordem em que elas aparecem na equação de regressão. Os elementos fora da diagonal são as covariâncias entre os parâmetros de regressão. A raiz quadrada dos elementos da diagonal principal da matriz $\mathbf{V}(b)$ determina os valores correspondentes dos erros padrão (s_{bi}) dos coeficientes calculados (equação 1.22).

$$\mathbf{V}(b) = (\mathbf{X}^t\mathbf{X})^{-1}s^2 \quad 1.21$$

$$s_{bi} = \pm\sqrt{v(b)_{ii}}, i=1, 2, \dots, I \quad 1.22$$

Identificação dos Efeitos e Coeficientes Significativos

Na literatura [9-11,13] há diferentes métodos para se avaliar efeitos e coeficientes significativos; entre os mais usados destaca-se a análise de variância (ANOVA), o gráfico de probabilidade (distribuição normal), e a comparação de efeitos com uma medida independente da variabilidade. Neste trabalho, as avaliações de significância para a decisão estatística, tanto para efeitos como para coeficientes de modelos, foram realizadas empregando o teste *t* (*distribuição de Student*), através do valor *p* [8,16].

Para realizar a interpretação correta utilizando o valor *p* é necessário compreender os testes de hipótese e significância, conforme descrito abaixo.

Em muitos casos, formula-se uma hipótese estatística com o propósito de rejeitá-la ou invalidá-la. Por exemplo: se o desejo é decidir se um sistema é diferente do que outro, formula-se a hipótese de que não há diferença entre os sistemas. Essa hipótese é denominada nula e representada por H_0 . Qualquer hipótese diferente da pré-fixada é denominada hipótese alternativa e é normalmente representada por H_1 [8,16].

Se uma hipótese for rejeitada quando deveria ser aceita, diz-se que foi cometido um erro tipo I. Se, por outro lado, uma hipótese for aceita quando deveria ter sido rejeitada, diz-se que foi cometido um erro do tipo II. Em ambos os casos foi tomada uma decisão errada ou foi cometido um erro de julgamento [8].

Para que qualquer teste de hipótese ou regra de decisão seja adequada, eles devem ser planejados de modo que os erros de decisão sejam reduzidos ao mínimo.

Ao testar uma hipótese estabelecida, a probabilidade máxima, representada freqüentemente por α , com a qual se sujeitaria a correr o risco de um erro do tipo I é denominada nível de significância do teste.

Se, por exemplo, é escolhido um nível de significância $\alpha = 0,05$ ou 5%, há então cerca de 5 chances em 100 da hipótese nula ser rejeitada, quando deveria ser aceita, isto é, há uma confiança de 95% de que se tomou uma decisão correta.

Nesses casos, diz-se que a hipótese é rejeitada ao nível de significância 0,05, o que significa que a probabilidade de erro seria de 0,05 [16].

Considerando a hipótese nula de que o valor do efeito se confunde com seu erro, pode-se formular a seguinte regra de decisão:

a) Rejeição da hipótese nula com 5% de significância, quando o valor de t_{calc} situar-se fora do intervalo entre $\pm t_{\alpha}$ ($|t_{calc}| \geq t_{\alpha}$). Isso equivale a dizer que o valor estatístico amostral observado é significativo no nível definido.

b) Aceitação da hipótese, caso contrário.

Uma maneira alternativa de concluir o teste de hipótese é comparar o valor p do teste estatístico amostral com o nível de significância α . O valor p do teste estatístico amostral é o menor nível de significância para que se rejeite H_0 [8]. Neste sentido, compara-se o valor p com α e, se o valor $p \leq \alpha$ rejeita-se H_0 , caso contrário H_0 é aceito. A vantagem de se conhecer o valor p está na possibilidade de se avaliar todos os níveis de significância para que o resultado observado possa ser estatisticamente rejeitado [8,16]. A representação gráfica do valor p é descrita na Figura 1.7.

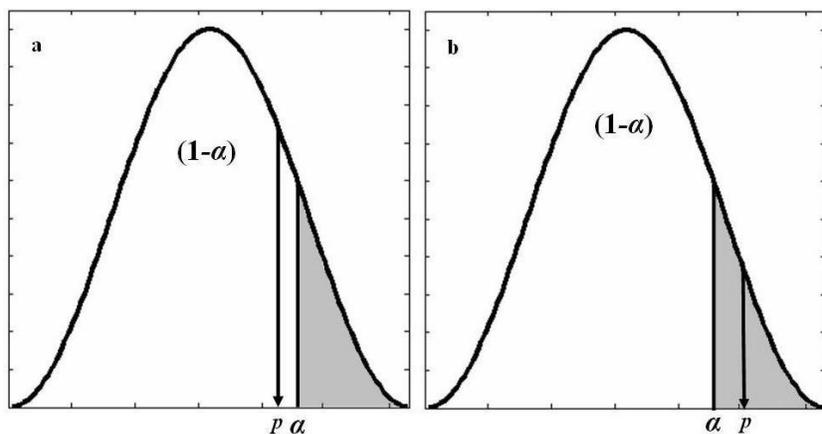


Figura 1.7. Representação gráfica do valor p para um teste unilateral.

Conforme as Figuras 1.7(a e b), o nível de significância α é a área hachurada no gráfico de distribuição. Na Figura 1.7a a área correspondente ao valor p é maior que o nível de significância, portanto, o valor calculado do teste estatístico está fora da região crítica, o que implica em aceitar H_0 . Para a Figura 1.7b a área do valor p é menor que o nível de significância e assim, o valor calculado do teste estatístico

está dentro da região crítica, o que implica e rejeitar H_0 . É importante ressaltar que esta discussão é válida tanto para testes unilaterais quanto para bilaterais.

Especificamente, o valor p representa a probabilidade de validade do erro envolvido no resultado observado, isto é, como representativo da população. Por exemplo: levando em consideração o valor de um efeito, se o valor do teste estatístico calculado (razão entre o efeito e o seu erro) apresentar um grande desvio da *distribuição de Student*, ele provavelmente descreve algo mais que o resíduo experimental. Desta maneira, será significativo dentro de um intervalo de confiança e assim, $|t_{calc}| \geq t_\alpha$ ou $p \leq \alpha$, para o número de graus de liberdade em questão.

Metodologia de Superfície de Resposta: Otimização

Após a realização dos experimentos de triagem, os fatores significativos são selecionados e uma metodologia de análise de superfícies de respostas pode ser executada para a otimização do experimento. Neste sentido, otimizar significa encontrar os valores das variáveis que irão produzir a melhor resposta desejada, isto é, encontrar a região ótima na superfície definida pelos fatores.

A metodologia de superfície de resposta baseia-se na construção de modelos matemáticos empíricos que geralmente empregam funções polinomiais lineares ou quadráticas, para descrever o sistema estudado e, conseqüentemente dão condições de explorar (modelar e deslocar) o sistema até sua otimização.

Um planejamento experimental construído para estimar coeficientes, segundo algum modelo aproximado, deve reunir certos critérios desejáveis, sendo os principais [14,15]: Proporcionar boas estimativas para todos os coeficientes, exigindo poucos experimentos; Fornecer condições de avaliação dos coeficientes e do modelo, ou seja, da regressão e da falta de ajuste.

Neste trabalho serão descritos os Planejamentos: Composto Central (*CCD - Central Composite Design*) [17] e Doehlert [18], visto que apresentam todos os critérios descritos acima, dentre outros e possivelmente são as classes de planejamentos mais utilizadas para ajustar modelos quadráticos (equação 1.23),

$$y = b_0 + \sum_{j=1}^J b_j x_j + \sum_{j=1}^J b_{jj} x_j^2 + \sum_{f < j} \sum_j b_{ff} x_f x_j + e \quad 1.23$$

na equação 1.23, J , é o número de variáveis.

Planejamento Composto Central

Os planejamentos compostos centrais (CCD) foram apresentados por Box e Wilson [17] em 1951 como uma evolução dos planejamentos 3^3 , que necessitavam de muitos experimentos para um pequeno número de fatores, mesmo para planejamentos fracionários. Outras vantagens, como rotabilidade e blocagem ortogonal, além do pouco número de ensaios, foram obtidas devido à presença das seguintes partes no planejamento [14]:

Um planejamento fatorial completo de dois níveis, podendo ser usado ainda, um planejamento fatorial fracionário de dois níveis;

Experimentos no ponto central, isto é, $x_i = 0$ para todo i ;

Experimentos nos pontos axiais em que $x_i = \pm \alpha$, com $x_j = 0$ para $j \neq i$, e $\alpha = \sqrt[4]{2^J}$. Estes pontos são situados nos eixos do sistema de coordenadas com distância $\pm \alpha$ da origem e formam a parte estrela do planejamento.

A Figura 1.8 ilustra os pontos experimentais no sistema de coordenadas, definidos pelos eixos x_i .

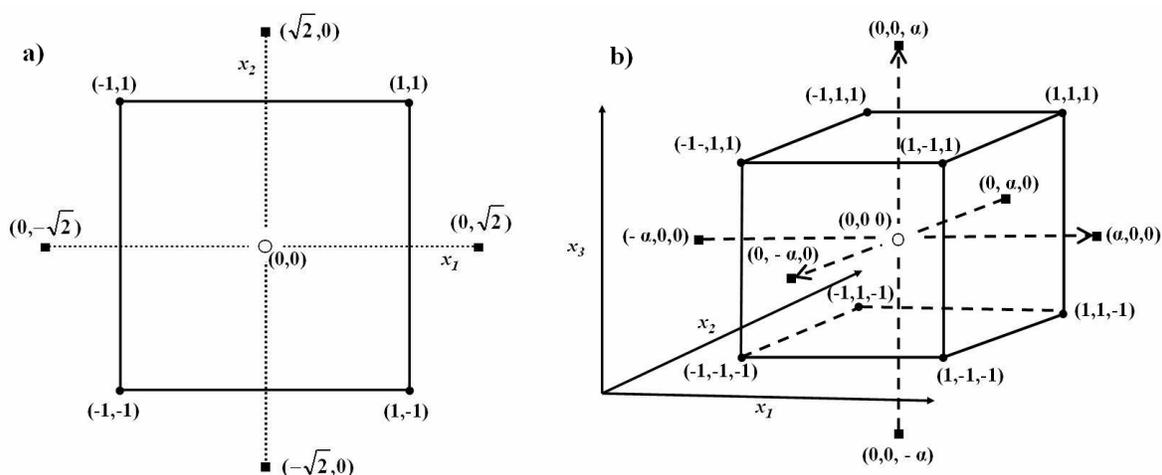


Figura 1.8. Planejamentos compostos centrais: (A) $J = 2$ e $\alpha=1.414$, (B) $J = 3$ e $\alpha=1.682$. Os diferentes símbolos significam: (●) planejamento fatorial, (■) pontos axiais e (○) ponto central.

Para construção de um planejamento CCD é necessário definir o número de variáveis a serem estudadas (J), qual planejamento fatorial será empregado (completo 2^J ou fracionário 2^{J-b}) e quantos experimentos serão realizados no ponto central (normalmente, $2J+1$). O número de experimentos a ser realizado é dado por: 2^J+2J+1 . A Tabela 1.6 descreve como o valor de α varia com o número de variáveis [14,15] e a Tabela 1.7 apresenta as matrizes de planejamento para 2 e 3 variáveis.

Tabela 1.6. Alguns planejamentos CCD e suas características.

Nº variáveis(J)	2	3	4	5	5	6	6
Valor de b para 2^{J-b}	0	0	0	0	1	0	1
Nº de exp. no planej. fatorial	4	8	16	32	16	64	32
Nº de exp. no ponto axial	4	6	8	10	10	12	12
Valor de α	1,414	1,682	2,000	2,378	2,000	2,828	2,378

Tabela 1.7. Matrizes dos planejamentos compostos centrais com duas e três variáveis.

Planejamento CCD – 2^2			Planejamento CCD – 2^3				
x_1	x_2		x_1	x_2	x_3		
-1	-1	Planejamento fatorial	-1	-1	-1	Planejamento fatorial	
1	-1		1	-1	-1		
-1	1		-1	1	-1		
1	1		1	1	-1		
0	0	Ponto central	-1	-1	1		
			1	-1	1		
$-\alpha$	0	Pontos axiais	1	1	1		
α	0						
0	$-\alpha$		0	0	0		Ponto central
0	α						
			$-\alpha$	0	Pontos axiais		
			α	0			
			0	$-\alpha$			
			0	α			
			0	0	$-\alpha$		
			0	0	α		

Observe que neste tipo de planejamento, os níveis α (o mesmo vale para qualquer valor codificado x_i) necessitam ser decodificados para os valores

experimentais dos níveis das variáveis a serem estudadas e para isso utiliza-se a equação 1.24:

$$x_i = \frac{z_i - \bar{z}}{\frac{\Delta z}{2}} \quad 1.24$$

em que x_i é o valor codificado do planejamento CCD, z_i o valor experimental do nível, \bar{z} o valor médio entre os níveis mais (+) e menos (-), que é exatamente o valor do nível zero (0) e Δz é a diferença entre os níveis mais (+) e menos (-).

Vale ressaltar também, que é aceitável o ajuste de α a um valor experimentalmente viável para o nível, desde que não ocorram grandes distorções no valor original.

Utilizando a Tabela 1.6 podem-se construir diferentes tipos de planejamentos CCD, conforme aqueles apresentados na Tabela 1.7. A partir do planejamento montado, pode-se obter a matriz de coeficientes de contraste, de acordo com a Tabela 1.2. O modelo apresentado na equação 1.23 pode ser representado matricialmente segundo a equação 1.10 e seus coeficientes são estimados através da equação 1.11.

A estimativa dos erros para os coeficientes a partir das replicatas no ponto central é realizada conforme as equações 1.21 e 1.22 e a avaliação dos coeficientes é realizada conforme o item: Identificação dos Efeitos e Coeficientes Significativos, descrito anteriormente.

Matriz Doehlert

O planejamento Doehlert ou Matriz Doehlert foi apresentado por David H. Doehlert em 1970 [18], sendo uma alternativa bastante útil e atrativa aos planejamentos experimentais de segunda ordem. Os pontos da matriz Doehlert correspondem aos vértices de um hexágono gerado de um simplex regular e em geral, o número total de pontos experimentais no planejamento é igual a $J^2 + J + pc$, em que J é o número de fatores e pc é o número experimentos no ponto central. Uma importante propriedade do planejamento Doehlert diz respeito ao número de níveis que cada variável possui. Com quatro variáveis, por exemplo, o número de níveis são 5, 7, 7 e 3, o que permite avaliar as variáveis consideradas mais

importantes, ou seja, que apresentam efeitos mais pronunciados em um número maior de pontos do espaço estudado [19]. Além disso, este tipo de planejamento necessita de um menor número de experimentos em relação ao planejamento Composto Central, sendo, portanto, mais eficiente. Esse menor número de experimentos para se chegar à região ótima vem do fato de que o domínio da vizinhança é facilmente explorado pelo ajuste de poucos experimentos, já que o próximo hexágono utiliza pontos experimentais já explorados pelo hexágono anterior, conforme Figura 1.9 [9].

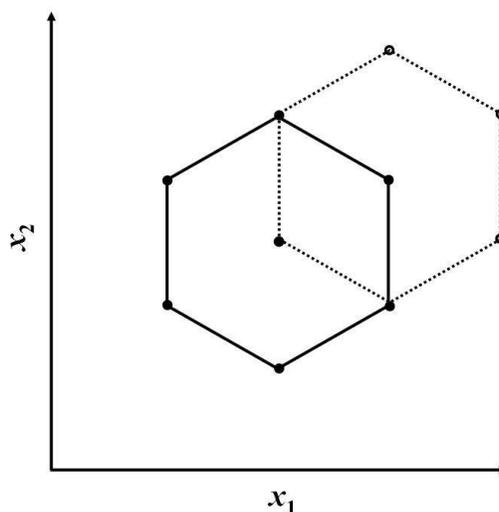


Figura 1.9. Planejamento Doehlert com duas variáveis (—) com adição de novos experimentos (.....).

O planejamento Doehlert descreve um domínio circular para duas variáveis, esférico para três e hipersférico para mais de três variáveis, o que acentua a uniformidade no espaço envolvido. Embora suas matrizes não sejam ortogonais nem rotacionais, elas não apresentam divergências significativas que comprometam a qualidade necessária para seu uso efetivo [20]. Para duas variáveis, a matriz Doehlert consiste de um ponto central e mais seis pontos adicionais formando um hexágono regular e, por esse motivo, situada sob um círculo (Figura 1.9). As matrizes do planejamento Doehlert para duas, três e quatro variáveis podem ser visualizadas na Tabela 1.8. As matrizes Doehlert apresentadas na Tabela 1.8 são pré-estabelecidas e suas construções não são triviais como os outros planejamentos

apresentados neste trabalho. Mais detalhes sobre a construção deste tipo de matriz podem ser obtidos nas referências 22 e 25.

Cada planejamento é definido considerando o número de variáveis e os valores codificados (x_i) da matriz experimental. A relação entre os valores experimentais e os valores codificados é dada pela Equação 1.25.

$$x_i = \left(\frac{z_i - z_i^0}{\Delta z_i} \right) \beta_d \quad 1.25$$

O termo x_i é o valor codificado para o nível do fator i ; z_i é o seu experimental; Δz_i é a distância entre o valor experimental no ponto central e o experimental no nível superior ou inferior, β_d é o maior valor limite codificado na matriz para cada fator e z_i^0 é o valor experimental no ponto central.

Tabela 1.8. Matrizes Doehlert para duas, três e quatro variáveis.

	Variáveis Experimentais – Matriz Doehlert								
	2 Variáveis		3 Variáveis			4 Variáveis			
	x_{1a}	x_{2a}	x_{1b}	x_{2b}	x_{3b}	x_{1c}	x_{2c}	x_{3c}	x_{4c}
1	0	0	0	0	0	0	0	0	0
2	1	0	1	0	0	1	0	0	0
3	0,5	0,866	0,5	0,866	0	0,5	0,866	0	0
4	-1	0	0,5	0,289	0,817	0,5	0,289	0,817	0
5	-0,5	-0,866	-1	0	0	0,5	0,289	0,204	0,791
6	0,5	-0,866	-0,5	-	0	-1	0	0	0
7	-0,5	0,866	-0,5	-	-	-0,5	-0,866	0	0
8			0,5	-	0	-0,5	-0,289	-	0
9			0,5	-	-	-0,5	-0,289	-	-
10			-0,5	0,866	0	0,5	-0,866	0	0
11			0	0,577	-	0,5	-0,289	-	0
12			-0,5	0,289	0,817	0,5	-0,289	-	-
13			0	-	0,817	-0,5	0,866	0	0
14						0	0,577	-	0
15						0	0,577	-	-
16						-0,5	0,289	0,817	0
17						0	-0,577	0,817	0
18						0	0	0,613	-
19						-0,5	0,289	0,204	0,791
20						0	-0,577	0,204	0,791
21						0	0	-	0,791

A maioria dos pacotes computacionais disponíveis atualmente, comerciais ou gratuitos, ainda não disponibilizou algoritmos para cálculos de planejamentos Doehlert sendo, portanto, um indicativo dentre outros, de que este planejamento está relativamente pouco difundido entre os pesquisadores.

Avaliação do Modelo

O modelo obtido pode não ser exatamente aquele que descreve a região estudada do sistema e, neste caso, não pode ser usado para fazer estimativas para deslocamento e muito menos para extrair conclusões sobre a região ótima. A maneira mais confiável de se avaliar a qualidade do ajuste do modelo é empregando a *análise de variância* (ANOVA) [8,16].

Na ANOVA a variação total da resposta é definida como a soma quadrática de dois componentes: a soma quadrática da regressão (SQ_{regr}) e a soma quadrática dos resíduos (SQ_{res}). A soma quadrática da variação total, corrigida para a média (SQ_{total}), pode assim ser escrita como a equação 1.26.

$$SQ_{\text{total}} = SQ_{\text{regr}} + SQ_{\text{res}} \quad 1.26$$

em que SQ_{regr} e SQ_{res} são apresentadas com mais detalhes nas equações 1.27 e 1.28, respectivamente.

$$SQ_{\text{regr}} = \sum_{m=1}^M \sum_{r=1}^{R_i} (\hat{y}_m - \bar{y})^2 \quad 1.27$$

$$SQ_{\text{res}} = \sum_{m=1}^M \sum_{r=1}^{R_i} (y_{mr} - \hat{y}_i)^2 \quad 1.28$$

Da equação 1.27, M é o número total de níveis do planejamento, isto é, pontos experimentais do planejamento; \hat{y}_m é o valor estimado pelo modelo para o nível m e \bar{y} é o valor médio das replicatas (R). No entanto, como há somente replicatas no ponto central, a média das replicatas nos níveis (+) mais e (-) menos é o próprio valor observado do ensaio naquele nível. Note que o segundo somatório indica que se deve fazer o quadrado das diferenças inclusive com cada repetição.

As replicatas realizadas no ponto central deixarão um resíduo para cada observação y_m . A soma quadrática destes resíduos fornece a soma quadrática residual somente no nível zero.

A equação 1.28 indica que o quadrado da diferença entre o valor de cada observação (y_m) e o valor estimado (\hat{y}_m) e inclusive das replicatas (y_r) em cada nível (m), fornece a soma quadrática residual de todos os níveis.

Quando algum modelo é ajustado aos dados, a soma quadrática do erro puro é sempre uma parte da soma quadrática dos resíduos. Então cada resíduo pode ser desmembrado em duas partes, isto é [15]:

$$(y_{mr} - \hat{y}_m) = (y_{mr} - \bar{y}_m) - (\hat{y}_m - \bar{y}_m) \quad 1.29$$

Pela equação 1.29, o primeiro termo da direita diz respeito à diferença entre o valor de cada observação individual no nível e a média de todas as observações naquele nível. Já o segundo termo corresponde à diferença entre o valor estimado no nível e a média de todas as observações naquele nível. A subtração entre estes dois termos fornece como resposta o resíduo de cada observação individual.

Para evitar os termos negativos na equação 1.29, tomam-se as suas diferenças quadráticas e obtém-se equação 1.30.

$$\sum_{m=1}^M \sum_{r=1}^{R_i} (y_{mr} - \hat{y}_m)^2 = \sum_{m=1}^M \sum_{r=1}^{R_i} (y_{mr} - \bar{y}_m)^2 + \sum_{m=1}^M (\hat{y}_m - \bar{y}_m)^2 \quad 1.30$$

O primeiro termo da direita é chamado de soma quadrática do erro puro e está relacionado exclusivamente com os erros aleatórios das replicatas. Já o segundo termo da direita é chamado de soma quadrática da falta de ajuste, pois ele é uma medida da discrepância entre a resposta do modelo de previsão (\hat{y}_m) e a média das replicatas (\bar{y}_m) realizadas no conjunto de condições experimentais.¹⁶

A equação 1.30 pode ser resumida da seguinte maneira:

$$SQ_{res} = SQ_{ep} + SQ_{faj} \quad 1.31$$

Para cada fonte de variação (regressão, resíduos, falta de ajuste, erro puro e total), é necessário obter o número de graus de liberdade. Não introduzindo detalhes, pode-se provar que o número de graus de liberdade para as equações 1.27, 1.28 e 1.30 são, $P - 1$, $I - P$ e $(I - M) + (M - P)$, respectivamente [11,15,20], em

que P é o número de parâmetros (coeficientes) do modelo, I é o número total de observações (ensaios) e M é o número de níveis do planejamento. O número de graus de liberdade para as outras fontes de variação pode ser obtido por cálculos algébricos simples.

A divisão da soma quadrática de cada fonte de variação pelo seu respectivo número de grau de liberdade fornece a média quadrática (MQ). A razão entre a média quadrática da regressão (MQ_{reg}) pela média quadrática dos resíduos (MQ_{res}), que nada mais é do que a razão entre duas variâncias, pode ser usada para comparar tais fontes de variação através do teste F (*distribuição de Fisher*), levando em consideração seus respectivos números de graus de liberdade. O mesmo pode ser feito para a razão entre a média quadrática da falta de ajuste (MQ_{faj}) pela média quadrática do erro puro (MQ_{ep}).

Assim, como foi realizada a avaliação dos efeitos e coeficientes empregando o teste t , através do valor p , o mesmo será feito para comparar as duas fontes de variação entre si, empregando-se neste caso o teste F (equação 1.32).

$$\frac{MQ_1}{MQ_2} \approx F_{v_1, v_2} \quad 1.32$$

em que MQ₁ e MQ₂ são as médias quadráticas das fontes de variação 1 e 2 respectivamente e, v_1 , v_2 são seus respectivos números de graus de liberdade.

Exemplificando, a razão entre as médias quadráticas da falta de ajuste e do erro puro é o valor calculado do teste estatístico (F_{calc}), que é usado para comparar qual é mais significativo. Pode-se formular uma hipótese nula (H_0) considerando que não há diferença entre as fontes de variação comparadas. Esta hipótese se reflete na seguinte regra de decisão:

- a) Rejeição da hipótese nula com 5% de significância, quando o valor de F_{calc} situar-se fora do intervalo definido por F_α ($|F_{calc}| \geq F_\alpha$), ou seja, $p \leq \alpha$. Isso equivale a dizer que o valor estatístico amostral observado é significativo no nível definido e as fontes de variação comparadas são diferentes.
- b) Aceitação da hipótese nula, caso contrário.

Se H_0 for rejeitada para MQ_{faj}/MQ_{ep} , isto é, $p \leq \alpha$, então há uma falta de ajuste significativa ao nível de probabilidade e número de graus de liberdade definidos (normalmente, $\alpha = 0,05$) e o modelo não é adequado.

Em termos práticos, um bom modelo necessita ter uma regressão significativa e uma falta de ajuste não significativa. Isto equivale a dizer que a maior parte da variação total das observações em torno da média deve ser descrito pelo modelo de regressão e o restante certamente, ficará com os resíduos. Da variação que fica com os resíduos é esperado que a maior parte fique com o erro puro, que é o erro experimental, e não com a falta de ajuste, que está diretamente relacionada com o modelo.

Outros parâmetros para observar se toda variação em torno da média foi explicada pela regressão são o valor do coeficiente de variação R^2 (equação 1.33) e o gráfico dos resíduos [10].

O valor de R^2 representa a fração da variação em torno da média que é explicada pela regressão (equação 1.33). Quanto mais próximo de 1 o valor do coeficiente estiver, melhor estará o ajuste do modelo às respostas observadas.

$$R^2_{expl.} = \frac{SQ_{regr}}{SQ_{total}} \quad 1.33$$

Note que o erro puro não explica nada do modelo, então o valor máximo possível de R^2 é:

$$R^2_{max} = \frac{SQ_{total} - SQ_{ep}}{SQ_{total}} \quad 1.34$$

A análise de variância (ANOVA), de acordo com as equações descritas para a avaliação do modelo, está resumida na Tabela 1.9.

Tabela 1.9. Análise de variância para regressão múltipla.

Fonte de variação	SQ	Nº de graus de liberdade	MQ	F
Regressão	SQ_{regr}	$P - 1$	MQ_{regr}	MQ_{regr}/MQ_{res}
Resíduos	SQ_{res}	$I - P$	MQ_{res}	
Falta de ajuste	SQ_{faj}	$M - P$	MQ_{faj}	MQ_{faj}/MQ_{ep}
Erro Puro	SQ_{ep}	$I - M$	MQ_{ep}	
Total	SQ_{total}	$I - 1$		

Uma boa prática é examinar a distribuição dos resíduos, pois ajuda a verificar se não há nada de errado com o modelo. Neste caso, pode-se fazer um gráfico dos valores estimados pelo modelo (\hat{y}_i) com os valores da diferença entre os valores observados experimentalmente (y_i) e os seus respectivos valores estimados, isto é, (\hat{y}_i) versus ($y_i - \hat{y}_i$). Se os resíduos não estiverem aleatoriamente distribuídos, pode-se desconfiar do modelo e investir em outros recursos para a sua melhoria.

Análise de Componentes Principais

A Análise de Componentes Principais (PCA) é provavelmente o método quimiométrico mais amplamente difundido, e devido à importância das medidas multivariadas em química, ela é considerada por muitos como o método que mais significativamente muda o ponto de vista dos químicos para a análise multivariada dos dados [6].

A aplicação mais freqüente do método PCA ocorre nos casos em que as colunas em \mathbf{X} são altamente colineares, *i.e.*, as variáveis são correlacionadas e apresentam redundâncias. Tais colinearidades indicam que a matriz \mathbf{X} apresenta algum tipo de variabilidade dominante que carrega a maioria da informação disponível. Estas redundâncias e pequenas variabilidades devem então ser removidas. A proposta da PCA é expressar as informações mais significativas contidas nas variáveis originais em um pequeno número de novas variáveis, as então chamadas componentes principais de \mathbf{X} [21].

Naturalmente, pode-se estender esta descrição para outros métodos multivariados, pois muitos diferem apenas no critério utilizado para determinar como as componentes são construídas [22]. Assim, a PCA constitui, em muitas maneiras, a base para a análise multivariada dos dados [1].

Uma variedade de algoritmos pode ser usada para calcular as componentes principais de \mathbf{X} . A decomposição por valores singulares (SVD) é um algoritmo rotineiramente empregado [23] e está descrito no apêndice. Usando este algoritmo, uma matriz de dimensão arbitrária pode ser escrita como produto de três outras matrizes, assim:

$$\mathbf{X} = \mathbf{ADP}^t \quad 1.35$$

Na literatura, as matrizes **A** ou **AD** são denominadas escores e contém as coordenadas das amostras nos eixos das componentes. Nesta tese, o produto **AD** será definido como escore e representado como a matriz **T**. A matriz **P** contém a informação sobre como as variáveis originais estão relacionadas às componentes e é denominada de matriz de *loadings*. A matriz **D** é diagonal (Apêndice) e contém informações sobre a quantidade de variância em cada componente.

O princípio matemático básico da PCA está em produzir uma aproximação da matriz de dados **X**, em termos do produto de duas matrizes **T** e **P^t** conforme Equação 1.36.

$$\hat{\mathbf{X}} = \mathbf{T}_h \mathbf{P}_h^t = \mathbf{A}_h \mathbf{D}_h \mathbf{P}_h^t \quad 1.36$$

em que o subscrito *h* representa o número do truncamento das matrizes *i.e.*, quantas componentes foram usadas para reconstruir a matriz **X**, definida por $\hat{\mathbf{X}}$.

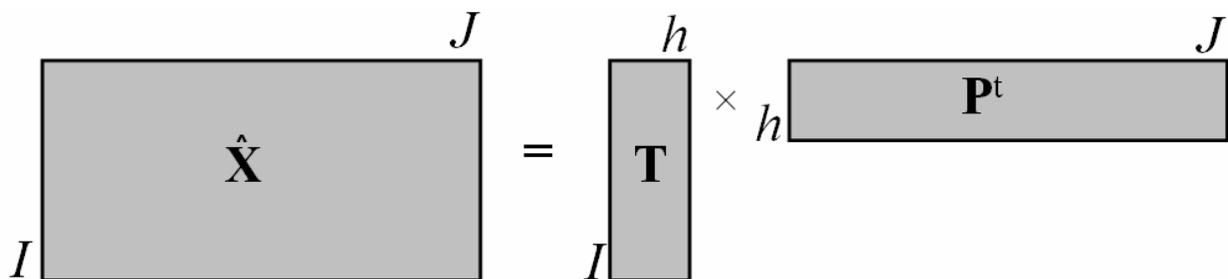


Figura 1.10. Representação matricial da análise de componentes principais.

Como mostrado na Figura 1.10, na matriz **T**, de escores, o número de linhas é igual ao número de linhas da matriz original, usualmente o número de amostras. O número de colunas é igual ao número *h* de componentes significativos nos dados, e pode ser qualquer número entre 1 e $\min\{I, J\}$. O número de componentes significativos é comumente definido como a dimensionalidade intrínseca dos dados ou posto químico da matriz original.

Na matriz **P^t**, de *loadings*, o número de colunas é igual ao número de colunas na matriz original, usualmente o número de variáveis. O número de linhas corresponde ao número de componentes principais *h*.

Geometricamente, as matrizes **A** e **P** geram um novo sistema de coordenadas formado pelas *h* componentes principais (colunas de **A** e **P**), que são os novos eixos (Figura 1.11). Assim, exatamente como as amostras têm coordenadas no espaço

original das linhas (definida pelas variáveis originais), elas também têm coordenadas em relação aos novos eixos *i.e.*, os escores. Estas coordenadas são as projeções ortogonais de cada amostra sobre os eixos das componentes, como representado pelas setas retas na Figura 1.7 e descrevem a representação das amostras padrões dominantes das linhas em **X**.

Cada componente é construída a partir da combinação das variáveis originais. Em termos matemáticos, a contribuição de cada eixo original para uma componente principal é o co-seno do ângulo entre o eixo da variável com o eixo da componente (ângulos ϕ 's na Figura 1.11). Se uma componente principal aponta exatamente na mesma direção de uma variável individual, o ângulo entre elas é zero e o co-seno é 1. Isto indica que a componente descreve toda a variação no eixo daquela variável. Similarmente, se uma componente é perpendicular (90°) ao eixo de uma determinada variável, o co-seno é 0 indicando que esta variável não fornece nenhuma contribuição à componente em questão. Os valores destes co-senos são os *loadings* e variam na faixa de -1 a 1 [24,25].

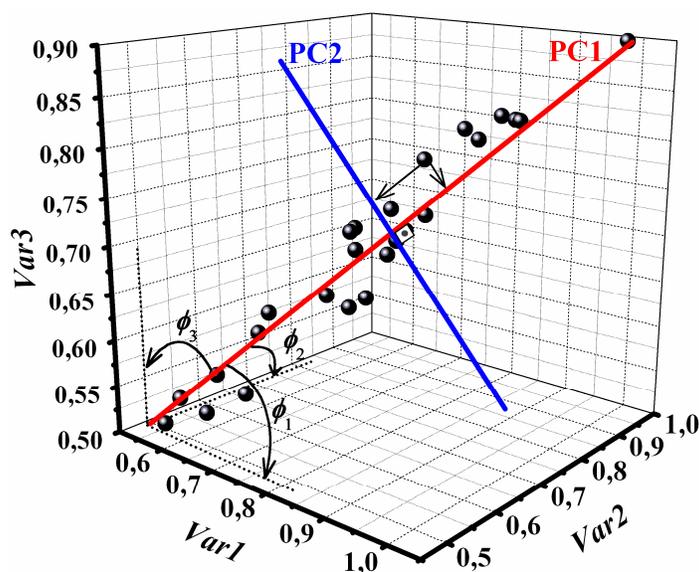


Figura 1.11. Representação geométrica de duas componentes principais em um espaço com três variáveis (3D).

A direção da primeira componente principal descreve o espalhamento máximo das amostras (linhas) *i.e.*, a quantidade máxima de variação possível dos dados e pode ser definida como o melhor ajuste da reta no espaço multidimensional. A

próxima componente, perpendicular à primeira (Figura 1.11), captura a variância comum em sua direção e em menor quantidade que a primeira e assim sucessivamente. As componentes principais são, portanto, calculadas em ordem decrescente de significância [25,26] e a percentagem da variação total nos dados descrita por qualquer componente pode ser calculada com precisão usando, por exemplo, informações da diagonal da matriz **D**. Porém, como as componentes são perpendiculares entre si, elas contêm informações diferentes e complementares em ordem decrescente de quantidade de variância.

Devido à reorientação das coordenadas do sistema de maneira a obter a máxima variância comum, a maioria da informação dos dados pode ser capturada por um pequeno número de componentes, que compacta as informações, eliminando as redundâncias e correlações nos dados originais [1,25,26]. Como resultado, as informações mais relevantes do ponto de vista da variância se tornam mais óbvias, exatamente porque foram compactadas.

Um importante conceito que se pode introduzir aqui é o conceito da unicidade da análise (ou do modelo). As componentes principais são únicas, mas não são únicas em fornecer uma base para o plano do melhor ajuste. Este plano pode ser também definido usando outra base, *e.g.*, **P** pode ser rodado por **Z** (em que **Z** é uma matriz $h \times h$ não singular (ver apêndice). Desse modo, usando $(\mathbf{Z}^t)^{-1}$ para contra rodar **T**, a solução \mathbf{TP}^t não muda, assim: $\mathbf{TP}^t = \mathbf{T}(\mathbf{Z}^t)^{-1}\mathbf{Z}^t\mathbf{P}^t = \tilde{\mathbf{T}}(\mathbf{PZ})^t = \tilde{\mathbf{T}}\tilde{\mathbf{P}}^t$, onde $\tilde{\mathbf{P}}$ é a nova base e $\tilde{\mathbf{T}}$ são os escores em relação à nova base. Esta propriedade é conhecida como *liberdade rotacional* [27].

Calibração Multivariada

O termo calibração, do ponto de vista da quimiometria é diretamente relacionado à construção de um modelo que relaciona respostas obtidas a partir de determinadas amostras às concentrações ou propriedades conhecidas destas amostras. Normalmente, o principal objetivo do modelo construído é realizar previsões de concentrações ou propriedades desconhecidas a partir de respostas obtidas a partir de novas amostras. O modelo obviamente é calibrado com um conjunto definido de amostras referência para as concentrações ou propriedades de interesse. As novas amostras, de onde serão obtidas as respostas previstas das

concentrações ou propriedades de interesse, normalmente apresentam as mesmas características das amostras de referência.

No contexto descrito acima (modelo inverso), as respostas obtidas são normalmente definidas como variáveis independentes e as concentrações ou propriedades são definidas como variáveis dependentes.

Na maioria das vezes os modelos construídos são baseados em um limitado conhecimento do investigador em relação ao sistema, mas onde uma grande quantidade de dados pode ser obtida. Este tipo de modelo é definido como *modelos flexíveis* em alusão aos *modelos rígidos* que são aqueles derivados a partir de teorias fundamentais do sistema investigado [22].

Diversos são os *modelos flexíveis* que podem ser empregados para relacionar as respostas com as concentrações/propriedades. Atualmente, com o advento da quimiometria, muitos modelos exigem certo tipo de arranjo das respostas para serem construídos. Estes arranjos dependem do tipo de resposta coletada e podem ser definidos da seguinte maneira: (1) quando uma resposta é obtida para cada amostra, a resposta é um escalar e pode-se definir como arranjo de ordem zero; (2) quando muitos escalares são obtidos para cada amostra *i.e.*, um vetor, define-se como arranjo de primeira ordem; (3) quando muitos vetores de mesmo comprimento são obtidos para cada amostra *i.e.*, uma matriz, define-se como arranjo de segunda ordem e assim, sucessivamente [28]. A Figura 1.12 apresenta a resposta para cada amostra e os arranjos de dados que podem ser obtidos para construção de modelos de calibração.

Para o arranjo de ordem zero o modelo é normalmente obtido através de quadrados mínimos, projetando a variável dependente sobre a variável independente. Um coeficiente (escalar) faz o ajuste entre estas duas variáveis. Este procedimento é conhecido como calibração univariada e é ainda, a mais usada pelos químicos analíticos apesar deste método necessitar de total seletividade da resposta frente ao analito ou propriedade de interesse.

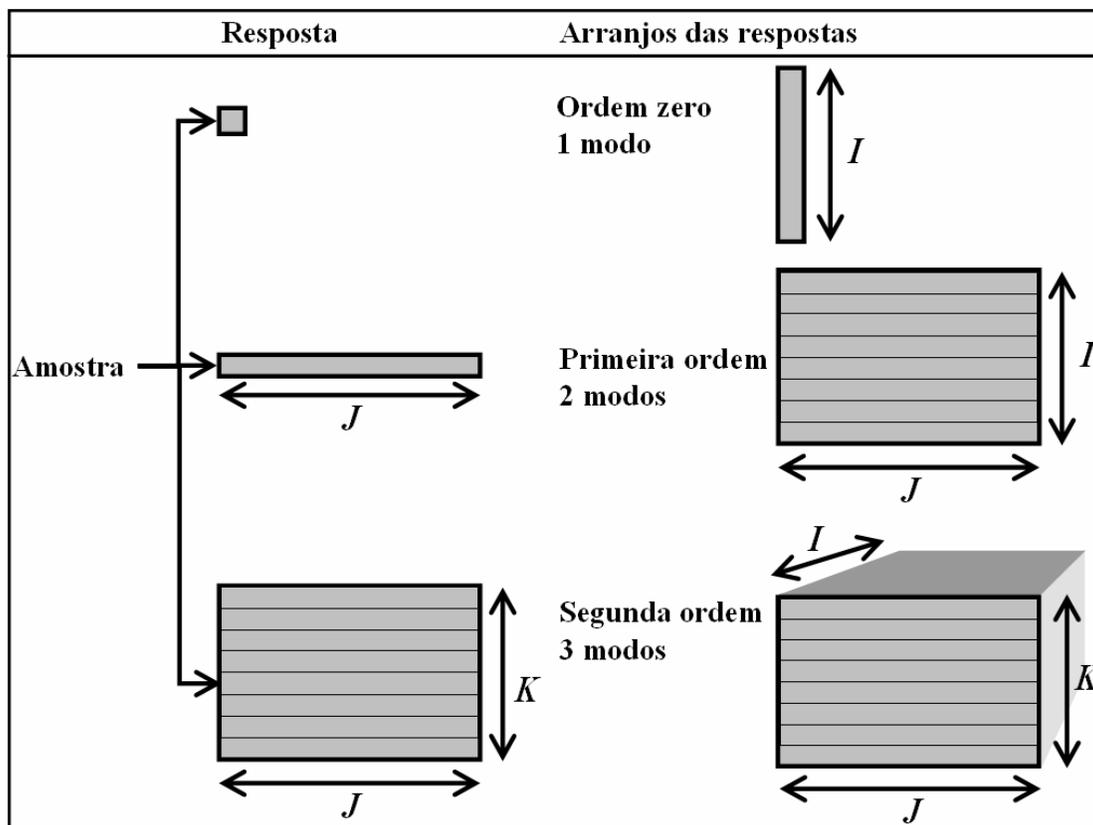


Figura 1.12. Representação de vários tipos de arranjos de dados para construção de modelos de calibração.

Arranjos de primeira, segunda e ordens superiores pertencem ao domínio multivariado e são processados por adequados procedimentos quimiométricos. A idéia central de todas estas metodologias é compensar a falta de seletividade do sinal analisado, usando eficientes e sofisticados algoritmos matemáticos para extrair a porção da massa de dados que pode ser empregada para fazer a previsão [21,26,28].

Há várias vantagens em se empregar o domínio multivariado em detrimento do univariado, como por exemplo [29]: (1) redução do ruído; (2) calibração na presença de interferentes; (3) manipulação de interferentes; (4) seleção de respostas para aumentar a seletividade; (5) controle de amostras anômalas; (6) aspectos exploratório dos dados e, com algumas metodologias quimiométricas baseadas em dados de segunda ordem ou ordens superiores é possível: (7) realizar separação cromatográfica de sinais e (8) possibilidade de realizar calibração sem a necessidade de modelar os possíveis interferentes.

O arsenal disponível para análise de dados de primeira ordem pode ser dividido entre dois tipos de regressão multivariada: o clássico (ou direto) e o inverso. A regressão multivariada clássica assume que as variáveis independentes são função da variável dependente e a inversa o contrário. Destes dois tipos, a calibração multivariada inversa é certamente a mais utilizada atualmente [1,6].

Entre as metodologias de calibração multivariada inversa a mais simples é a regressão linear múltipla (MLR), entretanto, algumas desvantagens comprometem a simplicidade deste método e seu emprego é limitado. Na MLR, a matriz de variáveis independentes \mathbf{X} é idealmente relacionada à variável dependente \mathbf{y} pela equação 1.10. O modelo é obtido calculando o vetor de regressão \mathbf{b} através de quadrados mínimos, conforme equação 1.11.

Um problema com esta aproximação é que a matriz $\mathbf{X}^t\mathbf{X}$, quando há mais colunas (variáveis) que linhas (amostras) em \mathbf{X} ($I < J$), não tem inversa. Na prática, devido a fatores como ruídos e interferentes desconhecidos, uma inversa pode ser calculada, mas ela não é significativa, pois a solução estará sendo influenciada por tais fatores. O uso da inversa de $\mathbf{X}^t\mathbf{X}$ é apenas praticável se: (1) o número de linhas *i.e.*, amostras, e o número de colunas, *i.e.* variáveis, são no mínimo igual ao número de analitos na mistura; (2) o número de experimentos deve ser no mínimo igual e de preferência maior que o número de variáveis ($I > J$) e (3) as colunas da matriz \mathbf{X} devem fornecer informações únicas *i.e.* não pode haver colinearidade (colunas linearmente dependentes).

Uma das maneiras de resolver o problema de colinearidade é através do emprego de métodos de compressão de dados. Estes métodos consistem em obter poucas combinações lineares (componentes principais) das variáveis independentes originais. Entre os métodos de compressão que constroem novas variáveis estão a regressão por componentes principais (PCR) e a regressão PLS, sendo esta última a mais usada atualmente [1,21,30].

Os métodos de compressão de dados tais como PCR e PLS são consideravelmente mais eficientes para lidar com ruídos experimentais, colinearidade e no caso do PLS, com leves não linearidades [21,31]. Nestes métodos, todas as variáveis originais relevantes podem ser incluídas nos modelos, o que implica que a calibração pode ser realizada eficientemente mesmo na presença

de interferentes se estes forem modelados [21]. Além disso, não há necessidade do conhecimento do número e natureza das variáveis originais. Estes métodos são muitas vezes robustos, *i.e.*, seus parâmetros praticamente não se alteram com a inclusão de novas amostras no conjunto de calibração [1,21,31]. A base fundamental dos métodos de compressão é a análise de componentes principais e será descrito com mais detalhes, para o método PLS, na seção seguinte [1].

Finalmente, se para uma amostra for possível produzir dados de segunda ordem (matrizes) ou de ordens superiores (conjunto de matrizes), então a estratégia multivariada é conhecida como técnicas multimodos [3,32-34]. A descrição das técnicas multimodos empregadas nesta tese está descrita na última seção deste capítulo.

Quadrados Mínimos Parciais (PLS)

De todos os métodos multivariados que serão discutidos nesta tese, o PLS é talvez o menos entendido [22]. Entretanto, ele é amplamente usado como método de regressão multivariada e tem guiado muitos trabalhos para o esclarecimento de suas propriedades matemática e estatística [35-41].

Herman Wold [42] foi o grande responsável pelo desenvolvimento do PLS. Isto ocorreu por volta do ano de 1975 quando ele trabalhava com dados complicados na área de econometria, os '*path models*' [43].

A origem do termo '*Partial Least Squares*', que define PLS, é um pouco obscura. De acordo com Wold [42], o procedimento é chamado '*Partial*' pois, embora ele consista de muitas otimizações locais, nenhum critério é completamente maximizado ou minimizado. De outro modo, Lohmöller [44] escreveu que o termo '*Partial*' refere-se ao particionamento dos parâmetros em subconjuntos estimáveis.

Alguns dos primeiros trabalhos empregando PLS em química usaram conceitos dos '*path models*' como em econometria [45,46]. Porém, o desenvolvimento do PLS nesta área é associado em particular, ao trabalho de Harald Martens (1985) [47], cuja tese contém muitos exemplos do uso da regressão PLS. Os trabalhos de Frank et al. (1983) [48] e de Svante Wold (1983) [49], filho de Herman Wold, foram também de grande contribuição para esta área. Hoje as aplicações do método PLS são vastas como, por exemplo, na indústria de

alimentos, de produtos farmacêuticos, petroquímica etc; em análise clínica; em estudos teóricos que usam Relações Quantitativas Estrutura-Atividade/Propriedade (QSAR/QSPR), dentre outras. Em todas estas áreas há dois pontos em comum em relação ao conjunto de dados usados para executar o PLS: eles consistem de dois blocos de variáveis (independentes e dependentes) e dentro de cada bloco, as variáveis são frequentemente altamente correlacionadas. Na maioria dos casos em calibração multivariada, o bloco das variáveis dependentes é composto de apenas uma coluna.

Atualmente há vários algoritmos para executar a regressão PLS e todos eles fornecem uma mesma previsão para uma única variável dependente [50]. São eles: (1) o algoritmo NIPALS (*non-linear iterative partial least squares*), proposto por Wold et al. [51]; (2) o algoritmo baseado na decomposição bidiagonal (PLSBdg) apresentado por Manne [40]; (3) o algoritmo SIMPLS, proposto por De Jong [52] e (4) o algoritmo Kernel, proposto por Lindgren et al. [53]. Outros algoritmos estão disponíveis na literatura, mas a maioria são adaptações ou modificações destes últimos citados.

Teófilo et al. [50] realizaram uma série de testes com estes algoritmos em relação aos seus desempenhos computacionais e concluíram que existe uma pequena diferença entre os métodos PLSBdg e SIMPLS, mas ambos foram extremamente mais eficientes que os outros dois, principalmente para grandes conjuntos de dados. O algoritmo Kernel mostrou ser extremamente dispendioso computacionalmente.

Apesar de não ter um alto desempenho computacional, o algoritmo NIPALS é hoje o mais conhecido e talvez o mais amplamente usado, porém, os algoritmos SIMPLS e principalmente o PLSBdg tendem a se tornar os mais empregados no futuro.

Nesta tese, a discussão será realizada sobre os algoritmos NIPALS e PLSBdg. Na discussão, será assumido que as matrizes \mathbf{X} ($I \times J$) e \mathbf{y} ($I \times 1$) foram centradas na média e que elas são as variáveis independentes e dependentes, respectivamente. Além disso, para evitar confusão, os termos variáveis latentes e fatores serão substituídos por componentes principais. Deste modo, $\min(I - 1, J)$ componentes

podem ser extraídas. Entretanto, em aplicações reais, poucas destas componentes são extraídas.

Embora no algoritmo NIPALS os passos individuais sejam muito simples, não é imediatamente óbvio justificar o que está sendo otimizado ou o que exatamente está ocorrendo. Muitos esforços têm sido feitos para explicar como o algoritmo funciona; em particular os trabalhos de Naes e Martens (1985) [54], Manne (1987) [40], Lorber et al. (1987) [55], Helland (1988) [56], Phatak (1993) [22] e mais recentemente Ergon [57] e Pell et al. [58] tentam esclarecer a confusão e colocar o PLS dentro de uma base matemática e estatística firme e consolidada.

Os detalhes do algoritmo NIPALS (versão ortogonal) podem ser descritos como segue [22]:

1. $\hat{\mathbf{w}}_h = \mathbf{X}^t \mathbf{y} / \sqrt{\mathbf{y}^t \mathbf{X} \mathbf{X}^t \mathbf{y}}$, cálculo do vetor de pesos normalizado para $h=1$;
2. $\hat{\mathbf{t}}_h = \mathbf{X} \hat{\mathbf{w}}_h$, cálculo do vetor de escores ortogonais usando os pesos;
3. $\hat{\mathbf{p}}_h = \mathbf{X}^t \hat{\mathbf{t}}_h / \hat{\mathbf{t}}_h^t \hat{\mathbf{t}}_h$, cálculo do vetor de *loading* normalizado;
4. $\mathbf{E}_X = \mathbf{X} - \hat{\mathbf{t}}_h \hat{\mathbf{p}}_h^t$, cálculo dos resíduos da matriz \mathbf{X} ;
5. Incremente h , substitua \mathbf{X} por \mathbf{E}_X na etapa 1 e continue até atingir o número desejado de componentes.
6. $\hat{\mathbf{b}}_h = \hat{\mathbf{W}}_h (\hat{\mathbf{W}}_h^t \mathbf{X}^t \mathbf{X} \hat{\mathbf{W}}_h)^{-1} \hat{\mathbf{W}}_h^t \mathbf{X}^t \mathbf{y}$, cálculo do vetor de regressão depois de calcular o número de componentes desejados.

No passo 1, os pesos são calculados usando o vetor de variáveis dependentes como inicialização do algoritmo, o que difere do algoritmo NIPALS para o cálculo das componentes principais na PCA, onde a inicialização é realizada, por exemplo, com a coluna de maior soma quadrática em \mathbf{X} .

A justificativa para a equação do vetor de regressão no passo 6 pode ser descrita da seguinte maneira:

Da mesma maneira como no PCA, em PLS a matriz \mathbf{X} pode ser decomposta em duas matrizes, porém, na maioria das vezes, as colunas destas matrizes trazem informações diferentes daquelas descritas pela análise PCA.

No PLS, os vetores \mathbf{t}_i formam uma base que é gerada pelas colunas de \mathbf{X} e, visto que \mathbf{t}_i são ortogonais usando o algoritmo discutido acima, a base é também

ortogonal. Conseqüentemente ela deve ser expressa como uma combinação linear de \mathbf{X} . É verificado que o espaço gerado pelas primeiras h componentes \mathbf{t}_i é também gerado pelas colunas da matriz \mathbf{XW}_h , em que as colunas de \mathbf{W}_h são as primeiras h componentes \mathbf{w}_i . Conseqüentemente, pode-se mostrar que as primeiras colunas \mathbf{t}_i podem ser escritas como [22]:

$$[\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_h] = \mathbf{T}_h = \mathbf{XW}_h(\mathbf{P}_h^t \mathbf{W}_h)^{-1} \quad 1.37$$

Se considerarmos que $\mathbf{C}_h = \mathbf{W}_h(\mathbf{P}_h^t \mathbf{W}_h)^{-1}$, então sua i -ésima coluna, \mathbf{c}_i , fornece um vetor ponderado que nos permite calcular \mathbf{t}_i de \mathbf{X} . No algoritmo SIMPLS de De Jong (1993) [52], os vetores \mathbf{c}_i são calculados diretamente de uma maneira muito elegante.

Após entrar no núcleo iterativo do algoritmo e observá-lo como um problema de autovetores e então esboçar algumas das propriedades dos pesos e dimensões associadas com a matriz \mathbf{X} , é possível ver com maior clareza o que está sendo otimizado no PLS. Para a primeira componente, observando a combinação linear \mathbf{Xc} , conclui-se que de todas as possíveis combinações lineares construídas desta maneira, maximiza a soma quadrática da covariância com \mathbf{y} , ou com as colunas de \mathbf{Y} , para o PLS2. A solução dada pelo primeiro vetor de pesos, \mathbf{w}_1 , e a soma quadrática das covariâncias de $\mathbf{Xw}_1 (= \mathbf{t}_1)$ com \mathbf{y}_i é, portanto, a maior possível.

Está claro que o algoritmo resolve um problema de maximização para cada componente extraída, embora não seja claro se h componentes tomadas como um grupo satisfaz algum critério ótimo global. Além disso, diferente de outros métodos, não é possível derivar os vetores pesos como sucessivos autovetores da mesma matriz.

O valor predito de \mathbf{y} , após h componente terem sido extraídas, é dado pela equação 1.38, visto que ele é expresso em termos de \mathbf{t}_i . Assim, ele pode ser escrito como uma combinação linear de \mathbf{X} , da seguinte maneira:

$$\hat{\mathbf{y}} = \mathbf{T}_h(\mathbf{T}_h^t \mathbf{T}_h)^{-1} \mathbf{T}_h^t \mathbf{y} \quad 1.38$$

que mostra que os valores ajustados é simplesmente a projeção de \mathbf{y} dentro do espaço gerado por \mathbf{t}_i . Além disso, foi mostrado que o espaço gerado pelas colunas de \mathbf{T} é também gerado pelas colunas de \mathbf{XW} , conseqüentemente, $\hat{\mathbf{y}}$ pode ser escrito como

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{W}_h(\mathbf{W}_h^t\mathbf{X}^t\mathbf{X}\mathbf{W}_h)^{-1}\mathbf{W}_h^t\mathbf{X}^t\mathbf{y} \quad 1.39$$

o que esclarece o cálculo do vetor de regressão apresentado na etapa 6.

O algoritmo SIMPLS é diferente do algoritmo NIPALS em dois pontos: os sucessivos \mathbf{t}_i são calculados explicitamente como combinação linear de \mathbf{X} e a matriz \mathbf{X} não é subtraída da matriz de resíduos em nenhum passo do algoritmo. Estes dois pontos tornam o algoritmo SIMPLS muito mais eficiente que o algoritmo NIPALS para o cálculo do PLS1.

De outro modo, Manne (1987) [40] demonstrou que o PLS1 é equivalente a um algoritmo desenvolvido por Golub e Kahan (1965) [59] para a *bidiagonalização* de matrizes. A *bidiagonalização* de matrizes é uma decomposição útil que frequentemente é utilizada como inicialização rápida em algoritmos para o cálculo da decomposição de valores singulares de matrizes [23].

Este método considera que qualquer matriz \mathbf{X} ($I \times J$) pode ser escrita como:

$$\mathbf{X} = \mathbf{U}\mathbf{R}\mathbf{V}^t \quad 1.40$$

em que $\mathbf{U}(I \times J)$, $\mathbf{V}(J \times J)$ são matrizes com colunas ortonormais (Apêndice), *i.e.*, elas satisfazem $\mathbf{U}^t\mathbf{U} = \mathbf{V}^t\mathbf{V} = \mathbf{I}$, e $\mathbf{R}(J \times J)$ é uma matriz bidiagonal.

Vários trabalhos na literatura descrevem a interessante relação do PLS1 ao método de decomposição bidiagonal [40], Lorber (1988) [60], Phatak (1993) [22,39,58]. A decomposição bidiagonal direta que equivale ao método PLS1 é realizada pelo algoritmo PLSBdg, e como no algoritmo NIPALS e SIMPLS, durante a decomposição bidiagonal, a informação em \mathbf{y} é considerada. O algoritmo PLSBdg pode ser seguido da seguinte formulação resumida [39,60].

1. inicialize o algoritmo para a primeira componente,

$$v_1 = \mathbf{X}^t\mathbf{y} / \|\mathbf{X}^t\mathbf{y}\| ; \alpha_1 u_1 = \mathbf{X}v_1$$
2. para $i = 2, \dots, h$ componentes
 - 2.1. $\gamma_{i-1} v_i = \mathbf{X}^t u_{i-1} - \alpha_{i-1} v_{i-1}$
 - 2.2. $\alpha_i u_i = \mathbf{X}v_i - \gamma_{i-1} u_{i-1}$

$$\text{Com } \mathbf{V}_h = (v_1, \dots, v_h), \mathbf{U}_h = (u_1, \dots, u_h) \text{ e } \mathbf{R}_h = \begin{pmatrix} \alpha_1 & \gamma_1 & & & \\ & \alpha_2 & \gamma_2 & 0 & \\ & & \ddots & \ddots & \\ & 0 & & \alpha_{k-1} & \gamma_{k-1} \\ & & & & \alpha_{k-1} \end{pmatrix}$$

prova-se ainda que $\mathbf{XV}_h = \mathbf{U}_h\mathbf{R}_h$ e, portanto, $\mathbf{R}_h = \mathbf{U}_h^t\mathbf{XV}_h$.

Assim, com as matrizes \mathbf{U} , \mathbf{V} e \mathbf{R} calculadas para h componentes, pode se estimar a pseudoinversa Moore-Penrose de \mathbf{X} e resolver o problema de quadrados mínimos, como mostrado a seguir:

$$\mathbf{y} = \mathbf{Xb} \quad \rightarrow \quad \mathbf{y} = \mathbf{U}_h\mathbf{R}_h\mathbf{V}_h^t\mathbf{b} \quad \rightarrow \quad \hat{\mathbf{b}} = \mathbf{V}_h\mathbf{R}_h^{-1}\mathbf{U}_h^t\mathbf{y} \quad 1.41$$

As matrizes da bidiagonalização são análogas àquelas derivadas pelo método SVD (Apêndice), conforme equação 1.35 ($\mathbf{X} = \mathbf{ADP}^t$). As partes em comum são descritas pela equação 1.42.

$$\mathbf{A} \equiv \mathbf{U}, \quad \mathbf{D} \equiv \mathbf{R}, \quad \mathbf{P} \equiv \mathbf{V} \quad 1.42$$

Observe que as matrizes não são idênticas, o que significa que os escores (\mathbf{UR}) e *loadings* (\mathbf{V}) do PLSbdg são diferentes dos escores (\mathbf{AD}) e *loadings* (\mathbf{P}) da PCA e portanto, do método PCR. Entretanto, eles são frequentemente muito similares.

Um detalhe conceitual importante é que os escores calculados pelo algoritmo de bidiagonalização diferem daqueles produzidos pelos algoritmos NIPALS ou SIMPLS por um fator de normalização [58]. Os vetores em \mathbf{V} são normalizados e ortogonais, e pode se mostrar que ambos os algoritmos fornecem as mesmas colunas $\{\mathbf{v}\}$. Porém, Ergon(2002) [57] e Pell (2007) [58] mostraram que há uma diferença entre a matriz reconstruída pelo método PLSbdg com aquela reconstruída pelos métodos NIPALS e SIMPLS devido a não realização de uma truncção implícita (número do posto químico escolhidos) na matrix \mathbf{X} durante a implementação dos algoritmos NIPALS e SIMPLS. Isto indica uma maior consistência algébrica do algoritmo PLSbdg, visto que sua base matemática é melhor definida [23,59].

Phatak (1993) [22] sugeriu, sem detalhes, uma relação entre as quantidades calculadas pelo método NIPALS e a decomposição bidiagonal, assim:

$$\mathbf{U} = \mathbf{T} |(\mathbf{T}^t \mathbf{T})|^{-1/2}; \quad \mathbf{R} = |(\mathbf{T}^t \mathbf{T})|^{1/2} \mathbf{P}^t \mathbf{W}; \quad \mathbf{V} = \mathbf{W} \quad 1.43$$

em que o subscrito $1/2$ indica que deve ser extraído a raiz quadrada dos elementos da matriz $\mathbf{T}^t \mathbf{T}$.

Embora existam relações, os algoritmos NIPALS e SIMPLS usam um espaço para calcular o vetor de regressão, mas outro espaço para representar os dados reconstruídos. A magnitude desta diferença depende do grau de truncção do espaço que é geralmente maior para maiores truncções (poucas componentes) [58].

Apesar de o vetor de regressão obtidos pelo algoritmo NIPALS, SIMPLS ou PLSBgd serem os mesmos [39,58], a aproximação da matriz \mathbf{X} pelo método bidiagonal é mais confiável para cálculos de reconstrução da matriz \mathbf{X} com h componentes para, por exemplo, detecção de amostras anômalas quando se usa a matriz reconstruída [57,58,61].

Os escores do PLSbdg ($\underline{\mathbf{T}} = \mathbf{UR}$) podem ser usados para o cálculo do vetor de pseudo-valores singulares, que pode ser calculado conforme equação 1.44:

$$\underline{\lambda} = \text{diag}(\underline{\mathbf{T}}^t \underline{\mathbf{T}}) \quad 1.44$$

Os pseudo-valores singulares do PLSbdg são similares aos valores singulares determinados pela PCA. Por isso o uso do sublinhado na equação 1.44.

Comparações com o método PCR foram realizadas e em todas elas o PLS1 apresenta duas grandes vantagens [35,55]: (1) o PLS1 ajusta melhor que PCR e (2) a regressão PLS1 usando os algoritmos SIMPLS ou PLSBgd é muito mais eficiente computacionalmente que o algoritmo que realiza a decomposição PCR. Assim, fica esclarecido o grande uso deste método em várias áreas científicas. Entre os algoritmos PLS, tem sido provado que o algoritmo PLSBgd apresenta eficiência computacional equivalente ou levemente maior que o algoritmo SIMPLS, sendo este último mais eficiente que o NIPALS [50].

Construindo e validando o modelo

Do ponto de vista prático, o modelo para calibração multivariada deve ser validado durante a sua construção. Este tipo de validação, realizada durante a construção do modelo é conhecida como validação interna. Durante a validação interna é possível escolher com maior precisão o número de componentes do modelo, detectar amostras anômalas e verificar preliminarmente sua robustez e capacidade de previsão. Em uma segunda etapa, o modelo construído é validado para amostras externas, ou seja, amostras que não foram usadas durante a sua construção. Esta etapa é conhecida como validação externa e tem como único objetivo verificar a capacidade de previsão do modelo e garantir previsões confiáveis para concentrações/propriedades de amostras desconhecidas.

Neste sentido, dois conjuntos devem ser separados para realizar as duas etapas. Um deles para a construção, conhecido também como conjunto de calibração ou treinamento e outro conhecido como conjunto de previsão ou teste (Figura 1.13).

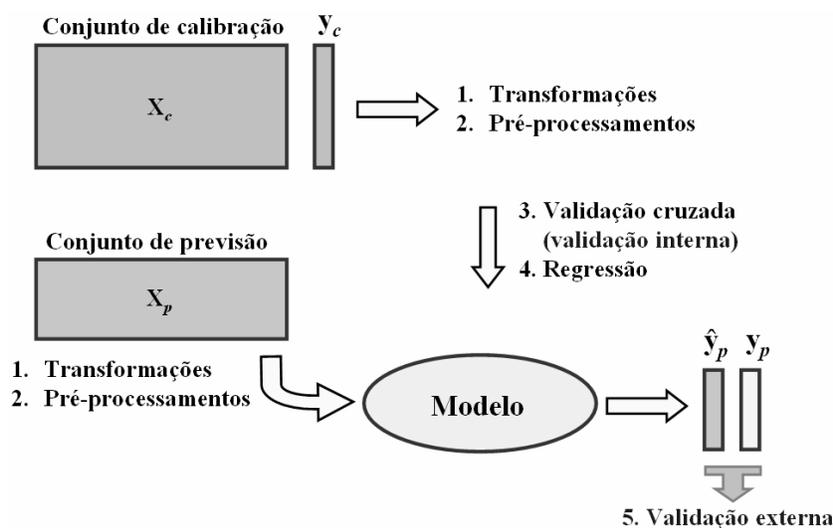
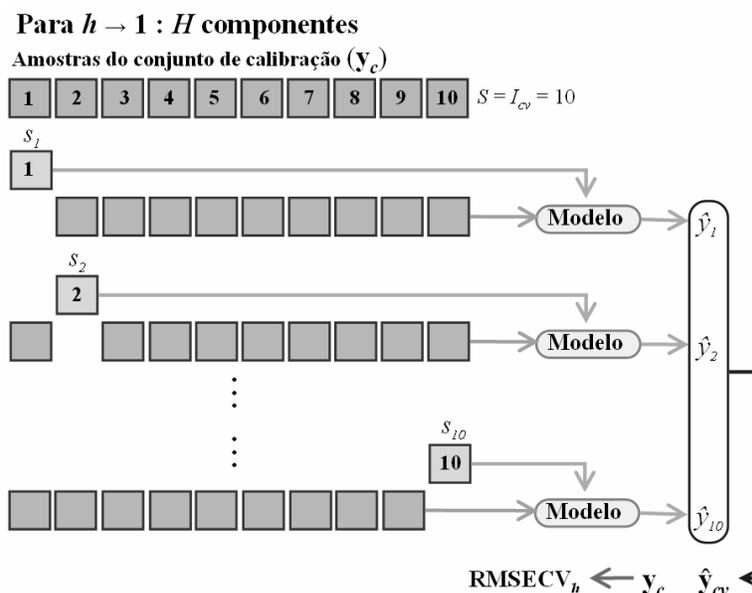


Figura 1.13. Esquema usado para construir e validar o modelo.

Apesar de ser óbvio, vale ressaltar que todas as transformações e pré-processamentos são realizados antes da execução da validação interna e externa.

Diversos modelos de regressão são construídos na validação cruzada, porém, somente um, *i.e.*, o ótimo, é construído e usado para previsões de conjunto externo.

Na etapa de validação interna o procedimento conhecido como validação cruzada é normalmente empregado [62,63]. Basicamente, a validação cruzada é executada dividindo os dados com I_{cv} amostras em s_i ($i = 1, 2, 3...S$) subconjuntos de aproximadamente igual tamanho. Cada subconjunto s_i pode ter, por exemplo, uma dimensão de 1 a $I_{cv}/2$ amostras removidas aleatoriamente ou não. Para cada componente, modelos são construídos S vezes, sendo que em cada vez, deixa-se um subconjunto s_i separado para previsão no respectivo modelo construído sem ele (Figuras 1.14 e 1.15). Isto ocorre até que todos os subconjuntos em s_i sejam testados. Se S é igual a I_{cv} , o método é conhecido como *leave-one-out*, conforme ilustrado na Figura 1.14. Embora o *leave-one-out* seja um método bastante usado, ele não é recomendado quando se tem muitas amostras, visto que modelos com sobreajuste podem ser obtidos com a escolha errada do número de componentes [43,63].



Figuras 1.14. Esquema do método ‘*leave-one-out*’ para validação cruzada.

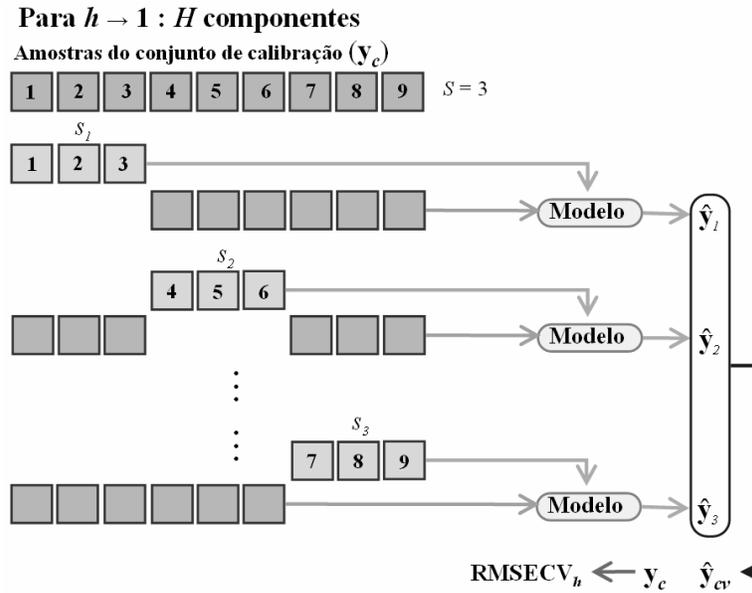


Figura 1.15. Exemplo de validação cruzada ‘leave-n-out’ para mais de uma amostra excluída.

As amostras previstas para os modelos durante a validação cruzada são comparadas com seus valores conhecidos. Parâmetros estatísticos são calculados para avaliar em que situação o modelo apresentou maior capacidade de previsão.

Os parâmetros usados em toda tese foram a raiz do erro quadrático médio de validação cruzada (RMSECV) apresentada na equação 1.45 e o coeficiente de correlação de validação cruzada (r_{cv}) apresentada na equação 1.46. Ambos obtidos a partir dos valores estimados por validação cruzada e os valores conhecidos.

$$\text{RMSECV}_h = \sqrt{\frac{\sum_{i=1}^{I_{cv}} (y_i - \hat{y}_i)_h^2}{I_{cv}}} \quad 1.45$$

$$r_{cv}^h = \frac{\sum_{i=1}^{I_{cv}} (\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{I_{cv}} (\hat{y}_i - \bar{\hat{y}})^2 (y_i - \bar{y})^2}} \quad 1.46$$

em que \hat{y}_i é o valor estimado para a amostra i , não incluída na construção do modelo, h é o número de componentes no modelo, I_{cv} é o número de amostras da validação interna, \bar{y} é a média dos valores em \mathbf{y} .

Na etapa de validação externa, um conjunto de variáveis independentes é fornecido ao modelo com h componentes para verificar o seu poder de previsão. A habilidade de previsão do modelo foi calculada em toda a tese usando erro quadrático médio de previsão (RMSEP) conforme é apresentado na equação 1.47 e o coeficiente de correlação de previsão (r_p) apresentada na equação 1.48

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{I_p} (y_i - \hat{y}_i)^2}{I_p}} \quad 1.47$$

$$r_p = \frac{\sum_{i=1}^{I_p} (\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{I_p} (\hat{y}_i - \bar{\hat{y}})^2 (y_i - \bar{y})^2}} \quad 1.48$$

em que \hat{y}_i é o valor estimado pelo modelo final para a amostra i , I_p é o número de amostra no conjunto de previsão, \bar{y} é a média dos valores em \mathbf{y} . Uma vez o modelo ter sido construído e validado, ele pode ser usado para previsões de novas amostras.

Seleção de Variáveis em Calibração Multivariada

Apesar de métodos para calibração multivariada tais como PLS e PCR serem aptos a trabalhar com uma grande quantidade de variáveis altamente correlacionadas, independente do número de amostras, está bem definido e aceito que as melhores previsões são obtidas quando variáveis apropriadas são selecionadas [64-67]. A seleção de variáveis é uma técnica que ajuda a identificar um subconjunto de variáveis que são, para um dado problema, as mais úteis para um modelo de regressão mais preciso e exato. Além disso, o subconjunto selecionado pode ajudar a tornar o modelo quimicamente mais facilmente interpretável o que é algo muito importante na área de análise sensorial, nos estudos de relação quantitativa estrutura-atividade dentre outras [66,68-70]. Isto

ocorre, pois, em um conjunto de centenas ou milhares de variáveis, algumas delas podem ser ruídos e/ou conter informações irrelevantes. Assim, o objetivo da seleção de variáveis é reduzir significativamente o número de variáveis para obter modelos mais simples, interpretáveis e robustos [71].

Um grande número de procedimentos para seleção de variáveis está disponível na literatura [70,72], a maioria focada em seleção de comprimentos de onda em espectroscopia. Estes procedimentos podem ser distinguidos um do outro pelo método de busca que localiza um subconjunto ótimo e pelo método objetivo para encontrar as variáveis ótimas [68].

Aqui será apresentada uma nova estratégia que será empregada para selecionar variáveis em dados voltamétricos e espectroscópicos. Em eletroanálise, este método será apresentado como uma alternativa ao algoritmo genético (GA), que tem sido usado nesta área, como único método capaz de selecionar variáveis com sucesso [71,73,74].

O GA é um método de seleção de variáveis parcimonioso baseado na pesquisa sistemática probabilística inspirada no mecanismo de seleção natural [75]. Em várias áreas da química o GA tem sido amplamente usado como seleção de variáveis [76]. Aqui ele será usado apenas para fins de comparação com o método proposto.

Algoritmo Genético para Seleção de Variáveis

Embora haja uma variedade de métodos de seleção de variáveis, o GA fornece um método direto baseado na *sobrevivência das variáveis mais aptas* como uma aproximação para modelar os dados.

A técnica de *seleção natural* para selecionar variáveis em calibração multivariada pode ser descrita da seguinte maneira [77]: dado uma matriz \mathbf{X} de variáveis independentes e um vetor \mathbf{y} de variável dependente, o algoritmo escolhe um subconjunto de variáveis em \mathbf{X} e, usando algum método de regressão multivariada (aqui, o método PLS) e alguma estratégia de validação cruzada, é determinado o valor de RMSECV para o subconjunto de variáveis. O algoritmo genético usa este conceito iterativamente para localizar o subconjunto ou subconjuntos que fornece(m) o menor valor de RMSECV.

A primeira etapa do GA é gerar um grande número (*e.g.*, 32, 64, 128) de seleções aleatórias das variáveis independentes (*termos iniciais*) e calcular o RMSECV para cada subconjunto. O RMSECV usado para cada modelo PLS é o menor obtido entre todos os calculados, limitado pelo número de componentes previamente definido. Cada subconjunto de variáveis é chamado de *indivíduos* ou *cromossomos* e o código sim(1)/não(0), que indica quais variáveis serão usadas é o gene para aquele indivíduo. Normalmente, 1 indica a maior aptidão da variável e 0, o contrário. A combinação de todos os indivíduos testados é a *população*. Os valores de RMSECV, descritos como aptidão de um indivíduo, indica a capacidade de previsão das variáveis selecionadas naquele indivíduo. Devido à variação dos níveis de ruído e variáveis interferentes, a aptidão dos diferentes indivíduos na população se estenderá por uma faixa de valores.

A segunda etapa do GA é a seleção. Nesta etapa, descartam-se os indivíduos com valores de RMSECV maiores que o valor de RMSECV mediano (obtido a partir de todos os modelos). Neste ponto, a população é reduzida para a metade de seu tamanho original.

Para substituir os indivíduos descartados, o GA *cruza* os indivíduos restantes. Isto é feito por um dos dois métodos: ‘*cross-over*’ simples ou duplo. No ‘*cross-over*’ simples, os genes de dois indivíduos aleatoriamente selecionados são divididos em algum ponto aleatório do gene. A primeira parte do gene do indivíduo A é trocada com a primeira parte do indivíduo B, e os dois genes híbridos formam dois novos indivíduos (C e D) para a população. No ‘*cross-over*’ duplo, o cruzamento é muito similar ao simples exceto que, no duplo, dois pontos aleatórios no gene são selecionados e a porção do meio dos dois genes é trocada. A grande diferença na prática é que o ‘*cross-over*’ duplo gera tipicamente novos subconjuntos de variáveis com maior número de variáveis iguais aos dos cromossomos de origem.

Após a adição dos novos indivíduos à população, todos os genes individuais têm a chance de sofrer mutação aleatória (*taxa de mutação*). Isto permite, dentro de uma chance finita de adição ou remoção, considerar variáveis que possam estar super- ou sub- representadas na população. As mutações são necessárias para superar alguns problemas que podem ocorrer durante a seleção. O problema mais

essencial a ser resolvido é que se uma variável não for selecionada em qualquer cromossomo original, ela nunca será selecionada nas gerações futuras se a mutação não existir. Uma mutação é simplesmente uma inversão do gene em um indivíduo.

Finalmente, depois de todos os indivíduos terem sido emparelhados e cruzados, a população retorna para o tamanho original e o processo pode novamente continuar na etapa de avaliação de aptidão (RMSECV) *i.e.*, na segunda etapa.

A finalização do GA pode ocorrer depois de um número de iterações finito (*número máximo de gerações*) ou depois de alguma percentagem de indivíduos na população estarem usando subconjunto de variáveis idênticas (*percentagem para convergência*).

Há três grandes problemas na implementação do GA, que são a inicialização, o sobre ajuste e o tempo computacional [76]. Na inicialização, uma alta correlação entre variáveis codificadas em pontos adjacentes pode impedir o GA de remover variáveis indesejadas, especialmente quando o *cross-over* simples é empregado. O sobreajuste é um risco bem conhecido do GA. A previsão casuística acoplada com a tendência do GA em transmitir trechos dos cromossomos através de recombinações, faz dele predisposto a obter excelentes valores de RMSECV, mas as variáveis selecionadas mostram um comportamento pobre para amostras externas. O último deles torna o GA praticamente inviável para grandes conjuntos de dados. Neste ponto, seleções prévias podem ajudar a resolver o problema, mas mesmo para pequenos conjuntos de dados, ele consome, relativamente, muito tempo para sua execução.

Seleção dos Preditores Ordenados – OPS[®]

É prática comum entre os quimiometristas visualizar gráficos de *vetores informativos ou prognósticos* para extrair informações desejadas em dados multivariados. Os *vetores informativos* são aqueles obtidos a partir de algum tratamento matemático dos dados, que tornam mais visíveis comportamentos padrões das variáveis, sejam das independentes, das dependentes ou da relação entre as duas.

Em calibração multivariada, as regiões onde há maior intensidade do sinal de algum vetor considerado informativo são intuitivamente relacionadas com possíveis regiões de variáveis dos dados originais que podem melhorar as previsões. De fato, se o vetor visualizado trazer a informação procurada, isto realmente é verdade e a seleção de variáveis está realizada, mesmo sem uma escolha criteriosa da região desejada.

Uma característica desejável em um método de seleção de variáveis, no que se refere à aplicabilidade e eficiência sobre diferentes tipos de dados, são suas características universais. Em química, por exemplo, diferentes técnicas analíticas fornecem diferentes tipos de variáveis, cada uma com características bem definidas e peculiares, como altas ou baixas correlações e com diferentes comportamentos (banda única, muitas bandas e picos, muitos picos, mal comportadas ou aleatórias). O uso de vetores prognósticos, por exemplo, é universal, visto que a informação extraída por eles é intrínseca das variáveis, livre de sua natureza.

Muitos autores [56,72,78] e o *software* Pirouette™ [61] apresentam o vetor de regressão com potencial ferramenta informativa para auxiliar na seleção de variáveis em calibração multivariada. Variáveis com pequenos coeficientes de regressão não contribuem significativamente para a previsão e, portanto, podem ser eliminadas. O vetor de regressão pode ser considerado ainda, como uma soma ponderada dos *loadings* incluídos no modelo [61]. Portanto, o formato do vetor muda dramaticamente com o número de componentes quando informações importantes são incluídas e poucas mudanças ocorrem quando ruídos passam a ser modelados e toda informação já está nas componentes.

Outros vetores informativos em calibração multivariada para seleção de variáveis são aqueles que relacionam, de maneira univariada, cada variável independente com a variável dependente e retira parâmetros desta relação, como por exemplo, a correlação [61,70,79-81]. Normalmente, variáveis com pobres correlações não trazem informações importantes e podem ser excluídas. Entretanto, uma atenção maior deve ser considerada, visto que este vetor traz informações univariadas e não multivariadas.

Outros vetores apresentados na literatura [72,82,83] são guiados pela mesma finalidade e se diferenciam somente na maneira de busca objetiva das informações.

Ambos são focados em realçar quais são as variáveis originais mais importantes para se obter boas previsões da variável dependente em questão.

A estratégia intuitiva de usar um vetor informativo é mais objetiva e conduz mais diretamente para a interpretação das variáveis selecionadas e também, para boas previsões de novas variáveis externas *i.e.*, o risco de sobre ajuste é menor que no GA, por exemplo.

Embora o uso de vetores prognósticos apresente grande potencial para a seleção de variáveis, pouco tem sido feito em automatização de algoritmos, definições de critérios de seleção e combinações entre vetores. Estes fatores certamente tornarão a seleção ainda mais seletiva e objetiva.

A metodologia apresentada aqui supre esta necessidade, introduzindo uma estratégia simples e intuitiva para seleção de variáveis, baseada nos vetores informativos e nas combinações entre eles.

De um modo geral, o núcleo do método está na obtenção de um vetor que contenha a informação da localização das melhores variáveis independentes para previsão. Normalmente, este vetor pode ser obtido diretamente a partir de cálculos realizados com ambas as variáveis, *i.e.* independentes e dependente ou a partir de combinações de distintos vetores, obtidos com o mesmo objetivo. Logicamente, a dimensão do vetor deve ser igual ao número de variáveis independentes e cada posição do vetor deve estar alinhada com a sua variável independente correspondente. Em uma segunda etapa, a partir do vetor prognóstico gerado, as variáveis independentes originais (colunas da matriz \mathbf{X}) são decrescentemente ordenadas de acordo com a ordem dos seus valores correspondentes, contidos no vetor. Os maiores valores no vetor informativo indicam a posição das variáveis independentes originais mais importantes (Figura 1.16). Como consequência, o algoritmo foi nomeado como *seleção dos preditores ordenados* (*OPS*[®], de ‘*Ordered Predictors Selection*’).

Na terceira etapa, as variáveis ordenadas são avaliadas usando alguma estratégia de validação cruzada com algum método de regressão multivariada. Os modelos são construídos e avaliados sobre uma janela de variáveis e posteriormente sobre a janela acrescida de um incremento fixo de variáveis e assim sucessivamente, até que todas as variáveis ou uma percentagem do total, sejam

adicionadas à janela. Parâmetros de qualidade dos modelos são obtidos a cada avaliação, para posterior comparação.

Na quarta e última etapa, os conjuntos de variáveis avaliados (janela inicial e janela mais incrementos) são comparados utilizando os parâmetros de qualidade calculados. O conjunto de variáveis que apresentarem os melhores parâmetros de qualidade contém as variáveis que apresentam a melhor capacidade de previsão para o modelo construído e, portanto, são estas as variáveis selecionadas.

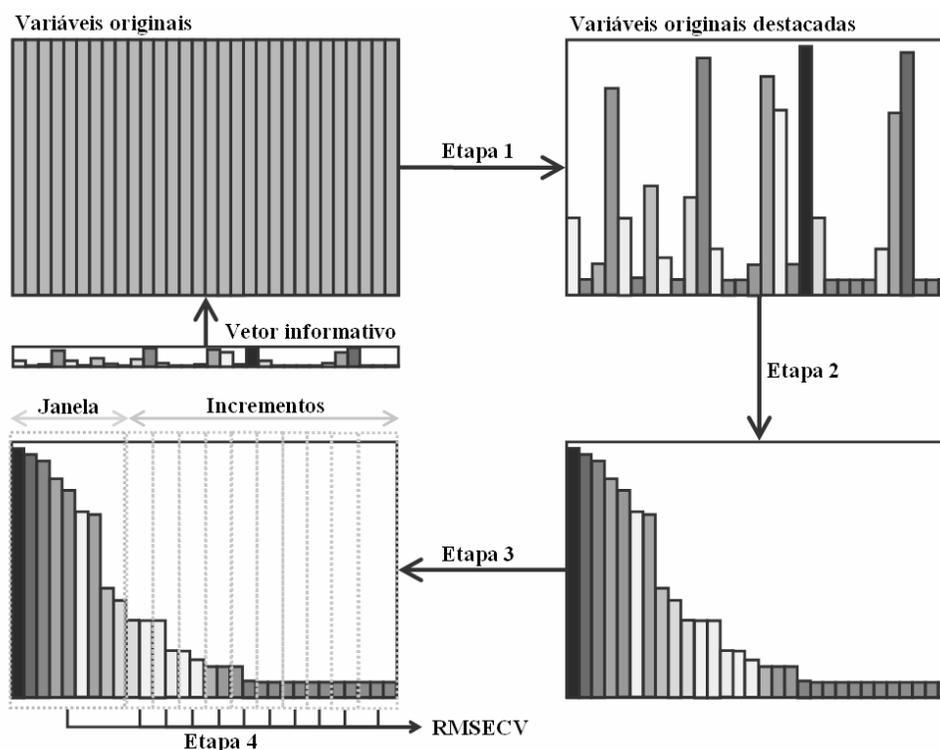


Figura 1.16. Etapas da seleção de variáveis usando o OPS

Especificamente, o algoritmo usado nesta tese apresenta as seguintes características:

1. Obtenção dos vetores informativos:

i. vetor de regressão para o modelo PLS (equação 1.41);

ii. vetor de correlação de cada coluna em \mathbf{X} (autoescaladas) com \mathbf{y} (autoescalado) (equação 1.49);

$$r_{\mathbf{x}_i, \mathbf{y}} = \frac{\mathbf{X}_{I \times J}^t \mathbf{y}_{I \times 1}}{I - 1} \quad 1.49$$

iii. o inverso das somas quadráticas das colunas da matriz de resíduos \mathbf{X} com h componentes definida como \mathbf{q} conforme equação 1.50.

$$\mathbf{E}_h = \mathbf{X} - \hat{\mathbf{X}}_h \Rightarrow q_j = \frac{1}{e_j e_j^t} \quad 1.50$$

em que e_j é a j -ésima coluna de \mathbf{E}_h . O inverso é usado pois espera-se que variáveis informativas apresentem baixos valores na matriz de resíduos;

iv. combinações entre estes vetores;

2. Método de regressão: modelo PLS usando o algoritmo bidiagonal;
3. Validação cruzada: método 'leave-N-out';
4. Parâmetros de qualidade avaliados: RMSECV e r_{cv} , obtidos para o número de componentes com menor valor da razão entre RMSECV e r_{cv} .

As vantagens deste método são: (1) aplicabilidade para qualquer tipo de dado (voltamétrico, espectroscópico, cromatográficos, variáveis de processo; QSAR/QSPR, espectrometria de massa, ressonância nuclear magnética, etc); (2) objetividade na seleção de variáveis que apresentam informações químicas relevantes, visto que os vetores informativos são aqueles selecionados; (3) necessidade de poucos parâmetros de entrada para executar a seleção, ao contrário do GA; (4) método computacionalmente eficiente e com aptidão para se tornar completamente automático; (5) capacidade de ser adaptado para seleção de variáveis em dados de ordem superior (Capítulo 7), em seleção de variáveis por intervalos e em análise discriminante, como o PLS-DA, por exemplo.

A aplicação deste método será realizada na *Parte Experimental* desta tese e seus resultados serão comparados com os resultados do GA.

Sinal Analítico Líquido e Parâmetros Analíticos

Os modelos de calibração multivariada podem usar sinais (*e.g.* voltamogramas) que não são seletivos para uma propriedade de interesse (*e.g.* concentração de certo analito). [84]. Porém, pesquisas têm sido realizadas para descrever a parte do sinal modelado que está quantitativamente relacionada à propriedade a ser predita. Esta teoria foi definida como Sinal Analítico Líquido (NAS) e tem levantado muito interesse em calibração multivariada.

O primeiro trabalho que apresentou um conceito similar à do NAS foi realizado por Morgan [85] em 1977 para calibração multivariada direta (clássica). Este trabalho contém alguns erros e não foi muito bem divulgado. Em 2003, um trabalho de Ferré e Faber [84], reconheceu o trabalho de Morgan como o pioneiro, mas não como o primeiro divulgador na introdução do conceito do NAS. Isto se deve, pois, em 1986, o trabalho de Lorber [86] apresentou uma elegante interpretação desta função para calibração multivariada por quadrados mínimos clássicos e em 1997, Lorber et al. [87] apresentaram, pela primeira vez, o cálculo do NAS para calibração multivariada inversa. Assim, os trabalhos de Lorber são reconhecidos pela comunidade científica especializada como os germinais para a teoria do NAS.

Atualmente muitas aplicações deste vetor em calibração multivariada têm sido descritas, tais como: seleção de variáveis [88]; detecção de amostras anômalas [87] [88]; desenvolvimento de novos métodos de calibração [89]; desenvolvimento de pré-tratamentos de dados [90]; avaliação do desempenho de modelos de calibração [91] e cálculo de parâmetros analíticos (figuras de mérito) [87,92-94].

Nesta tese, cálculos de alguns parâmetros analíticos em calibração multivariada inversa serão realizados com o auxílio do vetor NAS e, portanto, uma breve descrição dos métodos para obter este vetor será realizada.

Geometricamente, o vetor NAS representa a porção do vetor da mistura que reside no espaço ortogonal aos sinais dos componentes puros de todas as espécies interferentes em um sistema aditivo linear [95], conforme apresentado na Figura 1.17. Outra definição para o vetor NAS que segue o mesmo conceito anterior é: o vetor NAS para um componente é a parte do seu sinal na mistura que é ortogonal ao sinal de outros componentes (considerado para a calibração) [84].

Estas duas definições são corretas quando se considera a calibração direta *i.e.* quadrados mínimos clássico. Para calibração inversa (PLS, PCR, MLR) estas definições não são prontamente aptas e a mais indicada é: O vetor multivariado NAS é a parte do sinal da mistura que é útil para previsão [84].

O sinal analítico líquido é uma propriedade única do problema de calibração que é independente do modelo estimado e por alguma extensão, independente também do método de calibração. Por essa razão, pode ser impossível comparar

diferentes modelos de calibração competitivos pelos seus vetores NAS, visto que eles serão os mesmos, independente do método. Neste contexto, muitas diferentes aproximações podem ser razoáveis para estimar o vetor NAS, mas elas são todas interpretadas como tentativa para estimar o mesmo e verdadeiro sinal analítico líquido.

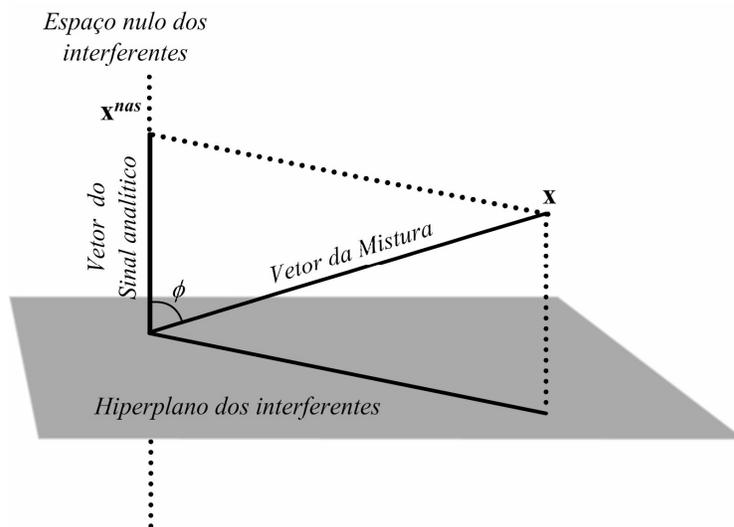


Figura 1.17. Ilustração da orientação e comprimento do vetor do sinal analítico líquido em relação ao plano dos interferentes e sinal analítico do componente puro

Aqui serão apresentados todos os cálculos disponíveis na literatura para se calcular o vetor NAS para calibração multivariada inversa. Dois tipos de métodos estão disponíveis: os baseados na anulação de posto e os baseados na projeção ortogonal do sinal sobre o vetor de regressão. Posteriormente, o uso do NAS será mostrado para o cálculo dos parâmetros analíticos, que serão aplicados nesta tese.

Métodos baseados na anulação de posto

Cálculo original: o método de Lorber

O cálculo do vetor NAS em calibração multivariada inversa apresentado por Lorber et al. (1997) [87] pode ser descrito como segue.

O objetivo inicial é eliminar a contribuição do k -ésimo constituinte de interesse presente nos sinais *i.e.*, em \mathbf{X} . Para realizar este cálculo os dados \mathbf{X} e \mathbf{y} são suficientes. Assim, é possível calcular o espaço do sinal gerado por todos os

constituintes exceto do k -ésimo analito. Esta eliminação pode ser realizada resolvendo o problema de anulação de posto.

Primeiro a matriz \mathbf{X} é reconstruída usando h componentes significativas via PLS (algoritmo bidiagonal) ou PCR, o que fornece uma matriz $\hat{\mathbf{X}}$ estimada.

A anulação de posto em um espaço de dimensão h é dada por

$$\hat{\mathbf{X}}_{-k} = \hat{\mathbf{X}} - \alpha \hat{\mathbf{y}}_k \hat{\mathbf{x}}^t \quad 1.51$$

em que $\hat{\mathbf{y}}_k$ é o vetor de concentrações estimadas ($\hat{\mathbf{y}}_k = \hat{\mathbf{X}}\hat{\mathbf{X}}^+ \mathbf{y}_k$) para o k -ésimo analito e $\hat{\mathbf{x}}$ é uma combinação linear das linhas de $\hat{\mathbf{X}}$, que deve incluir a contribuição do sinal do k -ésimo constituinte. Na configuração regular da anulação de posto, o vetor $\hat{\mathbf{x}}$ é o sinal puro do analito k . Porém, alguns autores [92,96] comentam que qualquer sinal \mathbf{x} , com grande contribuição do analito k obtido de \mathbf{X} , pode ser usado para esta proposta. Porém, é fácil provar que a escolha de diferentes combinações lineares deve afetar apenas o valor de α . A anulação de posto fará $\hat{\mathbf{X}}_{-k}$ livre da contribuição do k -ésimo componente. O posto de $\hat{\mathbf{X}}_{-k}$ é $h-1$. O escalar α é calculado da seguinte maneira:

$$\alpha = \frac{1}{\hat{\mathbf{x}}^t \hat{\mathbf{X}}^+ \hat{\mathbf{y}}_k} \quad 1.52$$

O vetor $\hat{\mathbf{X}}^+ \hat{\mathbf{y}}_k$ é o vetor de regressão estimado em calibração multivariada inversa.

O sinal analítico líquido para uma amostra é dado por:

$$\hat{\mathbf{x}}_{k,i}^{nas} = (\mathbf{I} - \hat{\mathbf{X}}_{-k}^t (\hat{\mathbf{X}}_{-k}^t)^+) \mathbf{x}_i \quad 1.53$$

Visto que $\hat{\mathbf{x}}_{k,i}^{nas}$ é livre de interferentes, é possível representá-lo geometricamente sem perda de informação, *e.g.* usando a sua norma euclidiana.

$$N\hat{A}S_{k,i} = \|\hat{\mathbf{x}}_{k,i}^{nas}\| \quad 1.54$$

Como pode ser observado, o valor calculado $N\hat{A}S_{k,i}$ é um escalar que representa a contribuição quantitativa do analito k para a amostra i . Assim, o valor do escalar NAS é calculado para cada amostra e, portanto, uma representação pseudo-univariada, rigorosamente provada como exata [97], pode ser realizada.

Portanto, para o modelo inverso, a concentração do analito é expressa como função do escalar *NAS*. Neste caso, a inclinação da reta do gráfico de calibração é identificada como a sensibilidade inversa e, conseqüentemente, um baixo valor da sensibilidade inversa é desejado, visto que orienta para uma pequena quantidade de propagação de erro [97]. Entretanto, como apontado por Faber [98], esta definição não é válida para modelos não centrados em situação onde a resposta é prevista com valores abaixo de zero.

O método de Xu e Schechter

Alguns meses após Lorber et al. (1997) terem apresentado seu método [87], Xu e Schechter (1997) [89], contrerrâneos de Lorber, propuseram uma nova metodologia para se obter o vetor *NAS* em calibração multivariada inversa.

Nesta proposta o vetor de concentrações \mathbf{y}_k é usado para definir \mathbf{X}_{-k} . A matriz de calibração \mathbf{X} é escalada dividindo cada linha da matriz \mathbf{X} pelo correspondente valor de \mathbf{y}_k , de tal maneira que cada vetor linha contenha a mesma contribuição do analito.

$$\mathbf{x}_i^{esc} = \frac{\mathbf{X}_i}{y_i} \quad 1.55$$

No próximo passo a matriz escalada é centrada na média removendo a contribuição constante do analito.

$$\mathbf{x}_{-k,i} = \mathbf{x}_i^{esc} - \mathbf{1}^t \bar{\mathbf{x}}^{esc} \quad 1.56$$

em que $\bar{\mathbf{x}}^{esc}$ é um escalar correspondente à média dos elementos do vetor \mathbf{x}^{esc} .

A partir daqui, retorna-se à equação 1.53 até o cálculo do escalar *NAS*.

Na equação 1.56 os autores não recomendam escolher um número ótimo de componentes para reconstruir a matriz \mathbf{X} , mas utilizar a matriz original. Desta forma, o trabalho propõe um método de calibração multivariada livre da escolha de um número de fatores ótimos, tendo como objetivo uma análise automatizada.

O método de Goicoechea e Olivieri

Goicoechea e Olivieri (1999) [92] propuseram uma aproximação similar à descrita por Xu e Schechter (1997) [89].

O método consiste em obter uma média dos sinais de calibração (equação 1.57) e a partir desta média uma aproximação da contribuição do analito de interesse é estimada e subtraída da matriz de dados \mathbf{X} , conforme equação 1.58.

$$\mathbf{x}_m = \frac{1}{I} \sum_{i=1}^I \mathbf{x}_i \quad 1.57$$

$$\mathbf{X}_{-k} = \mathbf{X} - \frac{\mathbf{y}\mathbf{x}_m^t}{\bar{y}} \quad 1.58$$

em que \bar{y} é um escalar correspondente à média dos elementos do vetor \mathbf{y} .

Para o cálculo do vetor NAS, O resultado da equação 1.58 é utilizado na equação 1.53.

Métodos baseados no vetor de regressão

O método de Faber

Klaas Faber (1998) [99] apresentou uma maneira altamente eficiente para realizar o cálculo do vetor NAS, que não realiza a anulação de posto. Este método baseia-se no uso do vetor de regressão uma vez que este vetor é ortogonal ao espectro dos interferentes. (Se os espectros dos interferentes variam devido à restrição de ortogonalidade o vetor de regressão para o analito *contra varia*) [99]. Isto é consistente com a idéia que os interferentes são implicitamente modelados em calibração inversa.

O produto interno na equação $y_i = \mathbf{b}^t \mathbf{x}_i + \varepsilon$ (ε nesta equação é o resíduo) dará apenas a contribuição de uma parte do espectro da mistura \mathbf{x} , que é ortogonal ao espectro dos interferentes; esta parte é o vetor NAS. Consequentemente o sinal da mistura pode ser dividido em duas contribuições mutuamente ortogonais como:

$$\mathbf{x}_i = \mathbf{x}_i^{nas} + \mathbf{x}_i^\perp \quad 1.59$$

em que \mathbf{x}_i^\perp é o vetor ortogonal ao vetor NAS.

Se $y_i = \mathbf{b}^t (\mathbf{x}_i^{nas} + \mathbf{x}_i^\perp) + \varepsilon \Rightarrow y_i = \mathbf{b}^t \mathbf{x}_i^{nas} + \varepsilon$, o vetor NAS pode ser calculado usando a seguinte equação:

$$\mathbf{x}_{k,i}^{nas} = y_i (\mathbf{b}^t)^+ = (y_i / \mathbf{b}^t \mathbf{b}) \mathbf{b} \quad 1.60$$

Originalmente, y_i foi sugerido como o valor esperado, mas na prática foi substituído pelo estimado [84,96]. Assim, a equação que deve ser aplicada é:

$$\mathbf{x}_{k,i}^{nas} = \hat{y}_i (\mathbf{b}^t)^+ = (\hat{y}_i / \mathbf{b}^t \mathbf{b}) \mathbf{b} \quad 1.61$$

Bro e Andersen: provando o método de Faber

Recentemente, Bro e Andersen (2003) [96] refinaram o cálculo proposto por Faber (1998) [99]. Na realidade, o método é conceitualmente o mesmo, entretanto estes autores fizeram uma descrição mais detalhada, provando a veracidade deste cálculo.

Neste método, as matrizes \mathbf{X}_k e \mathbf{X}_{-k} podem ser definidas, dentro dos fundamentos de um dado modelo de calibração (posto reduzido), da seguinte maneira: O espaço nulo do modelo pode ser indiferente. Para PCR, por exemplo, o espaço nulo é dado pela matriz de projeção $\mathbf{I} - \mathbf{P}\mathbf{P}^+$, em que \mathbf{P} ($J \times h$) é a matriz de loadings com h dimensões.

Ao invés de dividir os dados em duas diferentes partes como é usualmente feito, na derivação do NAS, o modelo é dividido em três contribuições, todas ortogonais entre si, i.e:

$$\mathbf{X} = \mathbf{X}_k + \mathbf{X}_{-k} + \mathbf{X}_e \quad 1.62$$

A matriz \mathbf{X}_k é a parte do sinal do analito que é ortogonal ao sinal do interferente e ao sinal do ruído (\mathbf{X}_e). Enquanto \mathbf{X}_{-k} é a parte de \mathbf{X} que está no subespaço gerado pelos interferentes (incluindo a parte do sinal analítico dentro deste subespaço).

Dentro do espaço do modelo dado por $\mathbf{P}\mathbf{P}^+$, a parte dada pelo vetor de regressão \mathbf{b} é assumida estar na mesma direção do vetor NAS. Por outro lado, as predições são influenciadas por outro vetor, além daquele relacionado com variação do analito. Conseqüentemente o espaço nulo do modelo deve ser equivalente ao espaço gerado por \mathbf{X}_{-k} . Portanto, a matriz \mathbf{X}_{-k} pode ser encontrada de:

$$\mathbf{X}_{-k} = \mathbf{X}\mathbf{P}\mathbf{P}^+ (\mathbf{I} - \mathbf{b}\mathbf{b}^+) \quad 1.63$$

Como \mathbf{b} é definido como sendo o espaço gerado pelas colunas de \mathbf{P} , assim $\mathbf{P}\mathbf{P}^+ \mathbf{b} = \mathbf{b}$ e conseqüentemente,

$$\mathbf{X}_k = \mathbf{XPP}^+\mathbf{bb}^+ = \mathbf{Xbb}^+ \quad 1.64$$

Isto significa que o vetor NAS é simplesmente encontrado pela projeção ortogonal dos sinais dentro do espaço gerado pelo vetor de regressão. Desta forma, o vetor NAS do sinal de certa amostra, \mathbf{x}_i é dado por:

$$\mathbf{x}_{k,i}^{nas} = \mathbf{bb}^+\mathbf{x}_i = \mathbf{b}(\mathbf{b}^t\mathbf{b})^{-1}\mathbf{b}^t\mathbf{x}_i = \mathbf{b}(\mathbf{b}^t\mathbf{b})^{-1}\hat{y}_i \quad 1.65$$

Note que a equação 1.65 e a equação 1.61 são idênticas.

Bro e Andersen (2003) em seu trabalho mostraram que os métodos baseados na anulação de posto apresentam algumas desvantagens em relação ao método baseado no vetor de regressão. A primeira delas é que o método baseado no vetor de regressão é mais geral e teoricamente mais consistente. A segunda é que os outros métodos não podem ser usados para acessar diferenças nos parâmetros analíticos entre, por exemplo, regressões PLS e PCR, visto que eles proporcionariam resultados muito parecidos para ambos os modelos, mesmo embora isto seja obviamente não razoável.

Tendo em vista estas considerações, nesta tese será empregado o método proposto por Faber (1998) [99] e refinado por Bro e Andersen (2003) [96].

Parâmetros analíticos em calibração multivariada inversa

Figuras de mérito tem se tornado muito importantes a fim de caracterizar, comparar e desenvolver novos métodos multivariados. Muitas delas estão estritamente relacionadas ao conceito do sinal analítico líquido, tais como a sensibilidade e a seletividade e outros como os limites de detecção (LOD) e quantificação (LOQ) são calculados seguindo os últimos avanços para a avaliação de incertezas em calibração multivariada.

O conceito de NAS aparece um tanto naturalmente em calibração multivariada considerando que uma amostra prevista apresente uma contribuição variada de outros componentes. Portanto, quando os vetores NAS das amostras diferem em forma, eles podem estar trazendo partes diferentes do espaço gerado. Conseqüentemente, qualquer parâmetro analítico calculado a partir destes vetores para representar a qualidade do modelo pode de fato ser enganosa e afetada pela variação que não é capturada pelo modelo de regressão [96]. Assim, para evitar

imprecisões nos cálculos dos parâmetros analíticos que usam o vetor NAS, deve-se encontrar e remover todas as amostras anômalas.

Os parâmetros analíticos sensibilidade, seletividade e limite de detecção, para calibração multivariada, serão descritos a seguir.

Sensibilidade

A sensibilidade refere-se à medida de quão fortemente o analito de interesse responde para uma específica medida. Em outras palavras, a sensibilidade diz em que extensão a resposta, devido a um particular analito, varia como função de sua concentração. Assim, a unidade da sensibilidade é sinal/concentração. Para o modelo inverso, a sensibilidade é o inverso da inclinação do gráfico pseudo-univariado [93].

Em termos do vetor de regressão a sensibilidade é dada pela equação 1.66 [93]

$$\text{SEN} = \frac{1}{\|\mathbf{b}\|} \quad 1.66$$

Em termos do NAS, a sensibilidade é dada pela equação 1.67 [87]

$$\mathbf{sen}_i = \frac{\mathbf{x}_i^{nas}}{y_i} \Rightarrow \text{SEN}_i = \|\mathbf{sen}_i\| \quad 1.67$$

Observe que na equação 1.66 a sensibilidade é calculada diretamente, enquanto que na equação 1.67 o cálculo é realizado para cada amostra. Ambos os cálculos atingem resultados quase idênticos, mas nesta tese o cálculo será realizado empregando a equação 1.67. A média de todas as sensibilidades será usada.

Seletividade

A IUPAC define a seletividade como uma medida da confiabilidade de um método usado para determinar um analito de interesse em uma mistura ou matriz sem a interferência de outras espécies de comportamento similar. Em uma linguagem mais precisa, um método é completamente seletivo contra um particular interferente se o resultado fornecido por ele é independente da concentração do interferente [100].

A seletividade varia entre 0 e 1, em que 0 significa que a análise é impossível visto que o sinal do analito é igual a uma combinação linear do espectro do interferente. Um valor igual a 1 indica que não há interferentes [101].

É importante notar que a seletividade definida aqui é dependente da amostra e será determinada pela quantidade de analito de interesse relativa aos interferentes.

Em termos de vetor de regressão a seletividade é dada pela equação 1.68 [93]:

$$SEL_i = \frac{1}{\|\mathbf{b}\| \cdot \|\mathbf{x}_i\|} \quad 1.68$$

Em termos do NAS, a seletividade é dada pela equação 1.69 [87]:

$$SEL_i = \frac{NAS_i}{\|\mathbf{x}_i\|} \quad 1.69$$

Novamente este parâmetro é calculado para cada amostra e, para divulgar a seletividade como um único número, a média será calculada. Espera-se que o uso da média descreva bem o valor da seletividade do modelo se as amostras apresentarem um comportamento homogêneo [92,102,103].

Limite de detecção (LOD)

A norma ISO 11843 define o limite de detecção (LOD) como a verdadeira concentração líquida do analito em um material a ser analisado que conduz, com probabilidade maior que zero, para a correta conclusão que a concentração no material analisado é diferente daquela em um material livre do analito analisado.

De todos os parâmetros analíticos calculados para calibração multivariada o cálculo do LOD é o mais questionado e ainda não bem definido até o momento [93] [104]. Booksh e Kowalski (1994) [93], apesar de apresentarem uma expressão para o cálculo do LOD, afirmam que este cálculo não faz sentido, visto que a quantidade de analito que pode ser detectado é uma função da concentração de cada interferente.

Recentemente, vários trabalhos têm avançado propondo cálculos mais precisos deste parâmetro em calibração multivariada [104-107], inclusive para calibração de dados de segunda ordem [104,108].

Nesta tese quatro diferentes estimativas serão realizadas para o LOD de maneira a observar os méritos e falhas de cada uma.

LOD usando o coeficiente de regressão e o ruído

O cálculo é uma extensão dos métodos usados para o cálculo do LOD em calibração univariadas [93]:

$$\text{LOD} = 3 \cdot \|\delta r\| \cdot \|\mathbf{b}\| \quad 1.70$$

em que $\|\delta r\|$ é a norma do vetor de desvios padrões das colunas da matriz de ruídos e $\|\mathbf{b}\|$ é a norma do vetor de regressão.

LOD usando as hipóteses de falso positivo e falso negativo

Método de Ortiz

Este método se baseia na extrapolação da equação de regressão obtida com os valores previstos e medidos da calibração multivariada para o valor zero de concentração. A base estatística para este método foi elaborada por Clayton et al. (1987) [109] e divulgada primeiramente para modelos de ordem zero [110] e recentemente Sanz et al. (2001) [107], mostraram a possibilidade de aplicar este métodos para modelos PLS com dados de voltametria de pulso diferencial (DPV). Ortiz et al. (2003) [104] provaram com uma série de diferentes tipos de dados, que esta maneira de estimar o LOD não se limita à calibração univariada (ordem zero), mas é também apta a ser aplicada a sinais não específicos de primeira, segunda e altas ordens [111,104] com dados voltamétricos ou espectroscópicos.

O método consiste em construir um modelo de calibração e aplicar o teste de Neyman-Pearson para a concentração y do analito.

Hipótese nula: $H_o : y = y_o$ i.e., concentração zero.

Hipótese alternativa: $H_A : y > y_o$

Probabilidade do erro tipo I: $\alpha = \text{pr}\{y > y_o \text{ quando } y = y_o\}$

Probabilidade do erro tipo II: $\beta = \text{pr}\{y = y_o \text{ quando } y > y_o\}$

O modelo univariado a ser construído é: $x = a + by + e$, em que x é a variável independente (sinal) e y a concentração.

Se a decisão for estabelecida na base da resposta x , a região crítica do teste pode ser escrita como $CR = \{x > x_c\}$, em que x_c é o valor crítico da resposta.

Levando esta decisão para a curva de calibração e estimando a concentração a partir do sinal, obtemos a expressão para o LOD [111,104]:

$$LOD = \frac{\Delta(\alpha, \beta) w_{y_o} \hat{\sigma}}{\hat{b}} \quad 1.71$$

em que $\hat{\sigma}$ é o desvio padrão residual da regressão (Equação 1.72), \hat{b} é o coeficiente da regressão univariada, w_{y_o} (Equação 1.73) é uma função da posição dos padrões, y_i , na curva de calibração e é inversamente proporcional ao número de replicatas da amostra, K e ao número de padrões no modelo de calibração, I . De uma maneira mais direta o produto $w_{y_o} \hat{\sigma}$ é uma estimativa do erro no intercepto. O fator $\Delta(\alpha, \beta)$ depende da probabilidade α e β (Equação 1.74) e do número de graus de liberdade $(I - 2)$. Este parâmetro pode ser obtido ou numericamente [112] ou através de tabelas estatísticas e representa o parâmetro de não-centralidade de uma distribuição t não central.

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^I (y_i - \hat{y}_i)^2}{I - 2} \quad 1.72$$

$$w_{y_o}^2 = \frac{1}{K} + \frac{1}{I} + \frac{(y_o - \bar{y})^2}{\sum_{i=1}^I (y_i - \bar{y})^2} \quad 1.73$$

$$\Delta(\alpha, \beta) = \frac{y - \hat{y}_o}{w_o \hat{\sigma}} \quad 1.74$$

O LOD será determinada com $y_o = 0$ e α e β serão as probabilidades de falso positivo e falso negativo, respectivamente.

Esta extensão é baseada na idéia de que o LOD não varia com a transformação linear na variável independente. Assim, o LOD poderá ser estimado de acordo com

a IUPAC e a norma ISO, por meio da regressão entre as concentrações estimadas pelos modelos multivariados (bilinear e trilinear PLS) vs. os valores padrões.

Método de Boqué e Rius

A seguinte equação resume a reformulação do LOD de Boqué e Rius (1997) [112] também divulgada por Lorber (1997) [87] para modelos de calibração multivariada inversa:

$$\text{LOD} = \Delta(\alpha, \beta) \cdot \delta r \quad 1.75$$

em que δr é uma estimativa do desvio padrão dos ruídos. Aqui será empregada a média dos desvios padrões de cada coluna da matriz de ruídos.

Método proposto

O método de Clayton et al. (1987) [109] infelizmente apresenta uma falha quando o modelo é construído longe da região do limite de detecção. A falha ocorre no cálculo de w_x que atinge altos valores para modelos longe do LOD. Aqui será proposta uma correção para esta falha que se baseia em um dos métodos proposto pela IUPAC para o cálculo do LOD que utiliza o modelo de regressão conforme a seguinte equação [113]:

$$\text{LOD} = \frac{\hat{a} + (3 \cdot \sigma_a)}{\hat{b}} \quad 1.76$$

em que \hat{a} é o intercepto na equação de regressão e \hat{b} é o coeficiente de regressão e σ_a é o erro relacionado ao intercepto, calculado a partir de repetições. A proposta apresentada aqui consiste em realizar a estimativa do erro no intercepto usando o cálculo da matriz de variância e covariância (equação 1.21) considerando os valores estimados como variável independente no modelo de regressão. O intercepto é estimado com uma coluna de uns, adicionada ao lado esquerda dos valores estimados. O valor do desvio padrão s é substituído na equação 1.21 pelo valor de $\Delta(\alpha, \beta)$. A variância do intercepto é calculada através da equação 1.22 com $i = 1$ e será definida aqui como e_a .

Assim, na equação 1.76, o termo $3 \sigma_a$ será substituído pelo termo e_a o que fornece a seguinte proposta para o cálculo do limite de detecção para modelos de calibração multivariada inversa:

$$\text{LOD} = \frac{\hat{a} + e_a}{\hat{b}} \quad 1.77$$

Este método será comparado aos outros descritos, na parte de aplicações desta tese.

Métodos Quimiométricos de Ordem Superior

Quando para cada amostra uma determinada variável (*variável 2* na Figura 1.18) é investigada em toda sua extensão de níveis desejados, dentro de cada nível de outra variável (*variável 1* na Figura 1.18), que também é investigada em toda sua extensão, obtêm-se como resposta uma matriz de dados. Os métodos de ordem superior ou multimodos são aqueles que trabalham este tipo de dado.

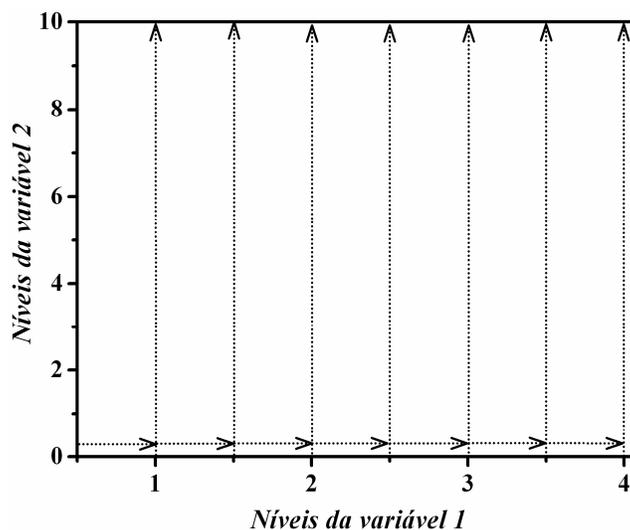


Figura 1.18. Representação esquemática da obtenção de uma matriz de dados com informações analíticas para uma amostra, apta a ser tratada com métodos multimodo.

Os então chamados *métodos hifenados*, como por exemplo, espectroeletróquímica, cromatografia gasosa com detecção por espectrometria de massas (CG-MS), cromatografia líquida de alta eficiência com detecção ultravioleta (UV) por arranjo de diodos (HPLC-DAD), etc, são fonte de dados para

os métodos multimodos. Outras técnicas aptas a fornecerem este tipo de dado são a espectroscopia de fluorescência, em que para cada comprimento de onda de excitação obtêm-se o espectro de emissão, as chamadas matrizes excitação/emissão (EEM) além das técnicas invasivas, como por exemplo, obter voltamogramas ou espectros em diferentes valores de pH.

Os métodos multimodos tiveram sua origem com o trabalho do psicometrista Cattell (1944) [114], que introduziu o conceito de '*perfis proporcionais paralelos*', seguido dos trabalhos de Tucker (1963 e 1966) [115,116] e dos trabalhos apresentados independentemente pelos psicometristas Harshman (1970) [117] e Carroll (1970) [118], que introduziram, respectivamente, os métodos: Análise dos Fatores Paralelos (PARAFAC) e Decomposição Canônica (CANDECOMP), ambos com o mesmo conceito.

Em química, esta área surgiu com os trabalhos apresentados por Ho et al., que datam de 1978 [119,120] e focavam o uso da anulação de posto. Outros métodos como o de Tucker e PARAFAC foram introduzidos na química entre meados dos anos 80 e início dos 90 [121-125].

Diversos métodos de ordem superior foram propostos e outros, modificados. Dentre estes métodos, o PARAFAC é atualmente o mais atraente [126] e tem sido amplamente aplicado em diversas áreas da química [127-130].

Em química analítica principalmente, um método que tem sido aplicado juntamente com o PARAFAC é a versão multimodo para PLS, conhecido também como PLS multilinear ou *N*-PLS [131-134]. Este método de regressão foi apresentado e bem divulgado por Bro (1996) [32], mas um método semelhante já havia sido proposto anteriormente por Ståhle (1989) [135]. O *N*-PLS utiliza a informação multimodo dos dados, ao contrário do PLS com os dados multimodos matriciados (desdobrados), o *M*-PLS. Teoricamente [32], o *N*-PLS apresenta desempenho superior ao *M*-PLS e seu vetor de regressão traz informações interpretativas de ambos os modos.

Nesta tese estes dois métodos serão aplicados sobre espectros EEM, obtidos durante a eletrólise de misturas de compostos fenólicos no eletrodo de diamante dopado com boro para realizar a previsão da degradação individual de cada composto na mistura, assim, os dois serão descritos com mais detalhes a seguir.

PARAFAC

O método PARAFAC é um método de decomposição, que pode ser comparado ao PCA, porém a decomposição do PARAFAC é realizada em tríades ou componentes trilineares. Ao invés de um vetor de escore e um vetor de *loadings* como em PCA, cada componente no PARAFAC consiste de um vetor de escores e dois vetores de *loadings*. É comum na prática da análise multimodos não fazer distinção entre escores e *loadings*, visto que são tratados igualmente em termos numéricos. Aqui o modo das amostras será tratado como escores.

A importante diferença entre o PCA e o PARAFAC é que no PARAFAC não há necessidade de exigir ortogonalidade para identificar o modelo.

A Figura 1.19 mostra a diferença estrutural dos métodos PCA e PARAFAC, considerando o PCA como a decomposição dos valores singulares (SVD).

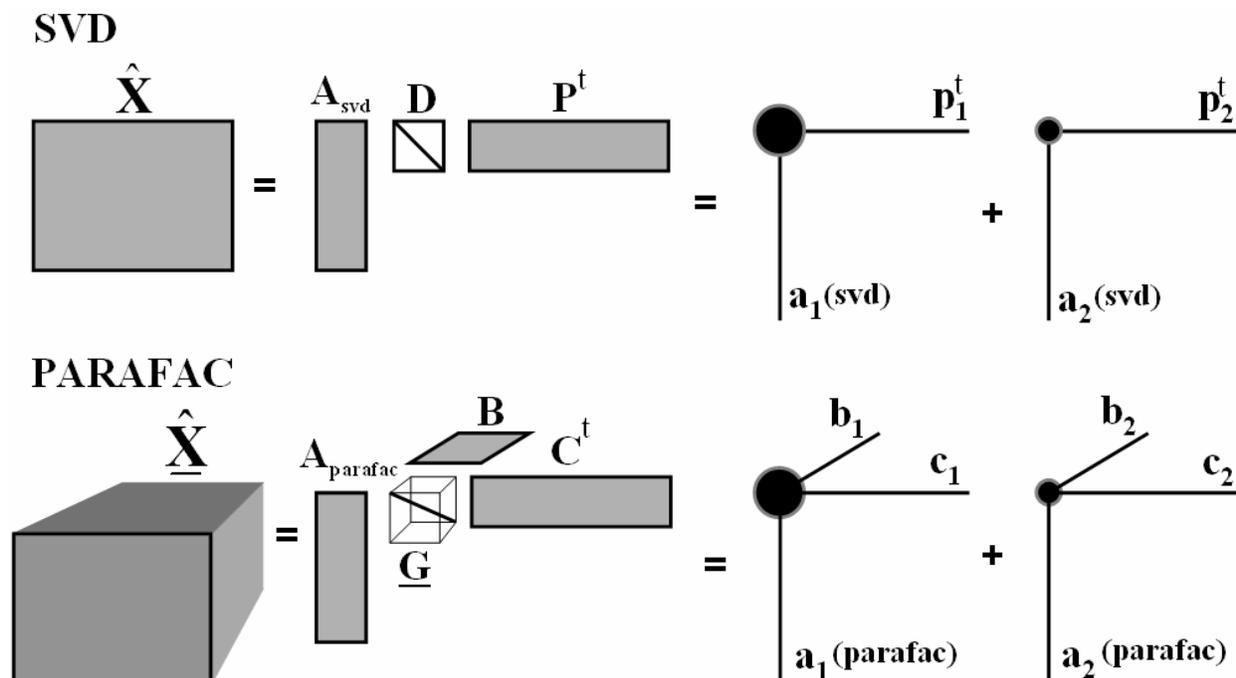


Figura 1.19. Extensão da decomposição SVD para a decomposição PARAFAC. Na decomposição SVD (ver apêndice), as matrizes \mathbf{A}_{svd} e \mathbf{P} são obrigatoriamente ortogonais. O modelo PARAFAC não é seqüencial e a solução é única (não há liberdade rotacional). O PARAFAC é um caso especial do modelo Tucker3, onde o número de componentes em cada modo é igual e a matriz núcleo \mathbf{G} que é superdiagonal e unitária.

A diferenciação entre a matriz \mathbf{A}_{svd} e a matriz $\mathbf{A}_{\text{parafac}}$ mostrada na Figura 1.19 usando um subscrito, foi realizada aqui para evitar confusões e isto será feito quando necessário, porém, em todo texto a distinção não será explicitamente realizada uma vez que o próprio modelo define a matriz \mathbf{A} .

Formalmente o modelo PARAFAC é dado pela seguinte equação:

$$x_{ijk} = \sum_{i=1}^h a_{ih} b_{jh} c_{kh} + e_{ijk} \quad 1.78$$

em que a_{ih} , é o elemento da matriz de escore do modo $\mathbf{A}(I \times h)$ e b_{jh} e c_{kh} são os elementos das matrizes de *loadings* dos modos $\mathbf{B}(J \times h)$ e $\mathbf{C}(K \times h)$, respectivamente e e_{ijk} é o elemento da matriz de resíduos. \mathbf{A} , \mathbf{B} e \mathbf{C} são escolhidas de maneira a minimizar a soma quadrática dos resíduos. Assim, o modelo PARAFAC em termos matriciais é apresentado da seguinte maneira:

$$\hat{\mathbf{X}}_{I \times JK} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})^t \quad 1.79$$

O símbolo \odot é o produto Khatri-Rao (Apêndice)

Entretanto, outra maneira de representar o modelo PARAFAC é através da representação do modelo Tucker3 restrito (não apresentados aqui).

$$\hat{\mathbf{X}}_{I \times JK} = \mathbf{A}\mathbf{G}(\mathbf{C} \otimes \mathbf{B})^t \quad 1.80$$

em que $\mathbf{G}(h \times hh)$ é o núcleo do arranjo (Figura 1.19) e o símbolo \otimes é o produto Kronecker (Apêndice). A matriz \mathbf{G} é idealmente para o modelo PARAFAC, um arranjo binário com zeros em todos os lugares exceto em sua superdiagonal, que contém apenas uns.

As características do modelo PARAFAC são: (1) o modelo é trilinear; (2) o número de componentes h são iguais em todos os modos; (3) o modelo não é seqüencial *i.e.* há a necessidade de reestimar completamente o modelo quando um diferente número de componentes é desejado; (4) não há interação permitida entre as componentes dos diferentes modos *i.e.* as informações nas componente são independentes; (5) a matriz núcleo \mathbf{G} , (Figura 1.19) é uma matriz identidade *i.e.* com uns em sua superdiagonal principal. Na equação 1.76 estes elementos são ‘armazenados’ em c ; (6) cada componente dos modos podem ser obtidos de tal maneira que eles não sejam necessariamente ortogonais; (7) a solução é dependente

da rotação. A realização de rotação destrói a soma quadrática mínima da solução ótima. Consequentemente a solução é única; (8) a solução por ser única pode ser interpretada.

Estas características mostram que, apesar de ser equivalente ao PCA para dados bilineares, o PARAFAC é diferente para dados trilineares e apresenta características superiores em relação ao PCA, principalmente devido a sua unicidade de solução.

No modelo PCA há liberdade rotacional, o que é considerado um problema para os quimiometristas. Os *loadings* no modelo PCA refletem os sinais puros dos analitos medidos, mas não é possível sem uma informação externa encontrar os sinais puros devido ao problema de rotação. Este problema não ocorre no modelo PARAFAC. Na maioria das circunstâncias o modelo tem identidade única da estrutura dos dados e consequentemente nenhum tratamento posterior é necessário quando modelo é o melhor em termos de problemas de quadrados mínimos. Se os dados são aproximadamente trilineares, o verdadeiro fenômeno essencial será encontrado se o número correto de componentes for usado e a razão sinal ruído for apropriada [136].

A unicidade do PARAFAC tem sido amplamente explorada em química analítica [93,137-139] e chamada *vantagem de segunda ordem*. Isto ocorre, pois, em uma decomposição única é possível obter os perfis puros dos sinais em ambos os modos. Um dos perfis em química analítica está correlacionado às amostras e, portanto, às concentrações dos analitos na amostra. As concentrações para cada componente na amostra podem ser extraídas diretamente, mesmo na presença de componentes não calibrados [93]. Logicamente, para cada interferente a mais na amostra, mais componentes serão necessários para ajustar o modelo.

Algoritmos para ajustar o modelo PARAFAC

Em termos computacionais o ajuste do modelo PARAFAC tem se restringido a duas classes de algoritmos: aqueles que empregam os quadrados mínimos alternantes (ALS) [117,118,136] que é iterativo, e aqueles que usam o procedimento direto baseado na resolução dos problemas de autovalores

generalizado [126] e.g. método de anulação de posto generalizado (GRAM) [140] ou decomposição trilinear direta (DTLD) [141].

Vários outros métodos têm sido testados, mas apenas alguns poucos conseguiram atingir o excelente desempenho do PARAFAC-ALS [142,143].

Porém, o desempenho de cada método é influenciado por vários aspectos que pertencem aos dados e não somente ao algoritmo por si só [126]. Conseqüentemente, fatores como desvio de posto, colinearidade entre componentes, pré-tratamentos adequados, quantidade e tipo de ruídos e presença e padrões de valores faltantes, desempenham um papel importante na escolha de qual algoritmo deve ser usado [143].

Porém, a escolha de um algoritmo adequado é muito complicada se for considerado que outro critério diferente do ALS possa ser mais apropriado para um específico tipo de dado. Além disso, se os dados necessitarem de restrições, outros métodos não são adequados, além do ALS.

Portanto, na maioria dos casos o algoritmo ALS resolve o problema, sendo, porém muito lento para dados com dificuldade de ajuste.

O ALS é um método antigo proposto por Yates (1933) [144], e conhecido no campo da matemática aplicada à computação como algoritmo não linear de Gauss-Seidel [126].

O problema não linear representado pelo ajuste do modelo PARAFAC é dividido em pequenos problemas lineares que são resolvidos iterativamente. Em cada etapa, um subconjunto dos parâmetros do modelo é estimado condicionalmente ao restante e uma iteração simples é completada quando todos os sub-problemas forem resolvidos. Conseqüentemente, cada completa iteração consiste de um número de sub-etapas que dependem do particionamento dos parâmetros do modelo.

O algoritmo PARAFAC-ALS consiste das seguintes etapas [136]

Defina o critério de convergência: ε

Inicialize \mathbf{B} e \mathbf{C}

1. $\mathbf{Z} = (\mathbf{C} \odot \mathbf{B})$

$$\mathbf{A} = \mathbf{X}_{I \times JK} \mathbf{Z}(\mathbf{Z}^t \mathbf{Z})^+$$

2. $\mathbf{Z} = (\mathbf{C} \odot \mathbf{A})$

$$\mathbf{B} = \mathbf{X}_{J \times IK} \mathbf{Z}(\mathbf{Z}^t \mathbf{Z})^+$$

3. $\mathbf{Z} = (\mathbf{B} \odot \mathbf{A})$

$$\mathbf{C} = \mathbf{X}_{K \times IJ} \mathbf{Z}(\mathbf{Z}^t \mathbf{Z})^+$$

4. Calcule o critério de convergência:

$$SQR = \left\| \mathbf{X}_{I \times JK} - \mathbf{A}(\mathbf{C} \odot \mathbf{B})^t \right\|^2 \text{ e } \sigma = \frac{SQR_{iter-1} - SQR_{iter}}{SQR_{iter-1}}$$

5. Se $\sigma \geq \varepsilon$, retorne à etapa 1

Uma boa inicialização do algoritmo pode ajudar no desempenho final e também ajuda a assegurar que o mínimo global do modelo seja encontrado [136]. Normalmente a inicialização é realizada com o auxílio de métodos diretos, como o DTLD por exemplo.

Algumas vantagens do ALS são: (1) simplicidade; (2) pode ser estendido para n modos; (3) pode manipular dados faltantes; (4) pode ser adaptado para aceitar restrições tais como não negatividade; unimodalidade; ortonormalidade; etc.

Validação de modelos PARAFAC

Como bem descrito por Smilde et al. (2004) [27], a análise dos dados é frequentemente baseada na avaliação de modelos ajustados. Estes modelos fornecem aproximações e interpolações do sistema real em estudo. Há frequentemente uma ampla escolha de modelos e a qualidade destes depende também de certo número de fatores, pois modelos matemáticos e estatísticos baseados em dados reais têm propriedades que dependem não apenas do sistema em estudo, mas também da amostragem, das propriedades do modelo, dos algoritmos etc.

Uma importante parte da validação é usar de diagnósticos para caracterizar a qualidade da solução. Nesta tese as ferramentas diagnósticas empregadas para validar o modelo foram aquelas baseadas em muitos ajustes usando apenas um conjunto ou múltiplos conjuntos de dados.

Escolha do número de componentes

Uma maneira de escolher o número correto de componentes é iniciar (repetir) o ajuste para o modelo PARAFAC de diferentes estratégias, como por exemplo, usando as decomposições DTLD, SVD e a inicialização aleatória. Para cada inicialização e para cada componente calculam-se parâmetros para o modelo que descreva a sua qualidade como, por exemplo, usando a soma quadrática residual, o número de iterações para atingir o ajuste e a percentagem de *consistência do núcleo* (CORCONDIA) [145]. Ao plotar os valores para cada um dos parâmetro obtidos de diferentes inicializações e para cada componente, é possível concluir com boa precisão quantas componentes são necessárias para ajustar o modelo PARAFAC observando o comportamento destes valores no gráfico simultaneamente.

Para analisar a soma quadrática residual, podemos considerar que um inesperado aumento no número de mínimos locais seja uma forte indicação que o ruído está sendo modelado e que as variações para um mesmo número de componentes possam ser observadas *i.e.* as repetições usando diferentes inicializações não são idênticas. Para este caso, um número ótimo de componentes é aquele que tem a soma quadrática residual pequena e que se repita para as diferentes inicializações.

Já, um aumento inesperado no número de iterações necessárias, sugere que muitas componentes estão sendo usadas, mas também pode indicar um difícil problema, por exemplo, falta de trilinearidade, grande quantidade de dados faltantes, baixa relação sinal/ruído, entre outros. Espera-se que o número ótimo aconteça antes de uma variação brusca no número de iterações para uma próxima componente.

A ferramenta diagnóstica CORCONDIA é um parâmetro diagnóstico que pode ser obtido com bastante rapidez para determinar o número correto de componentes, mas não pode ser aplicado sozinho [145].

Para entender o princípio do CORCONDIA, considere as equações 1.79 e 1.80. Se o núcleo do arranjo, representado por $\underline{\mathbf{G}}$, não estiver consistente com o modelo PARAFAC, o modelo não estará bem ajustado. Sabe-se que o núcleo do modelo Tucker3 definido por $\underline{\mathbf{T}}$, será idêntico ao núcleo $\underline{\mathbf{G}}$ (superdiagonal binário) se o modelo PARAFAC estiver perfeitamente ajustado. O modelo Tucker3 é por definição a representação ótima do arranjo dos dados em relação ao subespaço definido pelos componentes, enquanto que o modelo PARAFAC especificamente desconsidera qualquer variação associada com os elementos fora da superdiagonal, visto que são igualados a zero. Portanto, se os elementos da superdiagonal de $\underline{\mathbf{G}}$ estiverem próximos aos seus elementos correspondentes em $\underline{\mathbf{T}}$ e os elementos fora da superdiagonal são todos próximos de zero, o modelo é apropriado. Portanto, o CORCONDIA trata-se de um método de comparação entre os elementos de $\underline{\mathbf{T}}$ e de $\underline{\mathbf{G}}$. Uma maneira de quantificar estas similaridades é através da equação 1.81.

$$\text{CORCONDIA} = 100 \left(1 - \frac{\sum_{d=1}^h \sum_{e=1}^h \sum_{f=1}^h (t_{def} - g_{def})^2}{h} \right) \quad 1.81$$

em que t são os elementos de $\underline{\mathbf{T}}$ e g são os elementos de $\underline{\mathbf{G}}$.

Assim, se o modelo PARAFAC é válido então o CORCONDIA estará muito próximo de 100%. Se os dados não podem ser descritos por um modelo trilinear ou se muitas componentes estão sendo usadas, o CORCONDIA terá com valores próximos de zero ou mesmo negativos. Se a consistência estiver com valor próximo de 50% o modelo é instável.

Entretanto, Bro (1998) [146] fez a seguinte importante consideração: “o CORCONDIA é apenas uma ferramenta. Ele não prova nada. Ele é baseado em algumas suposições e pode como qualquer outra ferramenta ser enganosa. Então não é recomendável propor um modelo usando apenas a consistência do núcleo como uma alegação persuasiva. É recomendo o uso de várias ferramentas e visualizações para provar realmente que o modelo está ajustado com o número correto de componentes.”

Analizando o modelo ajustado

Duas outras ferramentas diagnósticas que podem e devem ser empregadas para analisar o modelo ajustado com h componentes são as técnicas ‘*split-half*’ e ‘*Jack-knife*’.

Alguns pesquisadores como R.A. Harshman [147] e R. Bro [136] defendem que o uso da análise *split-half* é uma das melhores maneiras de determinar se o posto químico do modelo PARAFAC está adequado e também, verificar se o modelo está corretamente ajustado.

Na análise *split-half* diferentes subconjuntos dos dados são analisados independentemente (Figura 1.20). Devido à unicidade do modelo PARAFAC, se o número correto de componentes foi escolhido, os mesmos resultados – mesmos *loadings* – deverão ser obtidos para cada subconjunto e entre os subconjuntos e o modelo completo. Se um número maior ou menor de componentes é escolhido em relação ao número ótimo, os parâmetros dos modelos construídos para cada subconjunto diferirão entre si e, portanto, os *loadings* serão diferentes.

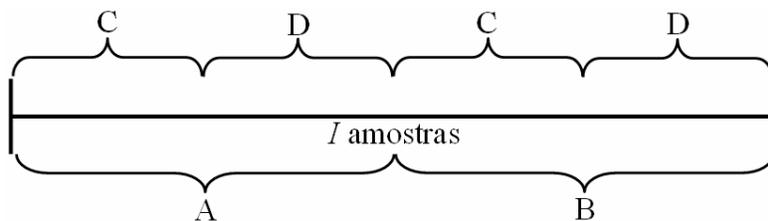


Figura 1.20. Partição de I amostras na análise *split-half*. As amostras são divididas em dois grupos A e B que são separados e divididas posteriormente em dois outros grupos, C e D, que também são separados.

Para julgar se os modelos são iguais, a indeterminação em modelos multilinear tem que ser respeitada. A ordem e escala dos *loadings* podem mudar se o algoritmo não for fixado. Se um modelo é estável na análise *split-half* é uma clara indicação que o modelo é *real*, *i.e.*, o modelo captura a variação essencial, que não pertence apenas a algumas amostras específicas. Se, por outro lado alguns componentes não são estáveis em uma análise *split-half*, é um indicativo que eles refletem o ruído ou que poucas componentes foram extraídas. Como consequência o modelo não é válido. Pode acontecer de alguns subconjuntos não serem estáveis. Isto pode indicar a presença de possíveis amostras anômalas nos subconjuntos. Portanto,

instabilidade em uma análise *split-half* não é sempre conclusiva, como acontece com a estabilidade.

Recentemente, Riu e Bro [148] mostraram, usando uma estratégia muito similar à análise *split-half* conhecida como *Jack-knife*, que também é possível estimar erros relacionados aos parâmetros e detectar amostras anômalas em modelos PARAFAC. Aplicando a técnica *jack-knife* para o modelo PARAFAC para um conjunto de I amostras, I submodelos PARAFAC são obtidos. Os submodelos são comparados entre si e com o modelo completo e assim pode-se verificar se há ou não amostras anômalas.

Cada um dos I submodelos (dado pelos parâmetros \mathbf{A}_{-m} , \mathbf{B}_{-m} e \mathbf{C}_{-m} em que o subscrito $-m$ significa que o modelo foi obtido deixando fora uma amostra m) tem a ele associado indeterminações de escala e de permutação que são removidas antes da estimativa *jack-knife* para ser diretamente comparado com os outros submodelos. Os detalhes da remoção destas indeterminações são bem descritas no trabalho de Riu e Bro (2003) [148].

Para detectar amostras anômalas e observar o comportamento dos modelos a partir dos resultados da análise *jack-knife*, dois gráficos são recomendados a serem observados: o gráfico de influência de reamostragem (RIP) e o gráfico da identidade emparelhada (IMP) [148].

No gráfico RIP, a soma dos quadrados dos resíduos da m -ésima amostra para o modelo calculado na m -ésima iteração *jack-knife* é mostrada em um dos eixos versus a soma dos quadrados da diferença entre os *loadings* (segundo e terceiro modo) obtido com cada segmento *jack-knife* e com o modelo PARAFAC com todas as amostras. Por exemplo, para o segundo modo (J) e o m -ésimo segmento *jack-knife*, a soma dos quadrados da diferença entre os *loadings* (b) para h componentes é dada por:

$$\sum_{i=1}^h \sum_{j=1}^J (b_{\text{completo},ji} - b_{-m,ji})^2 \quad 1.82$$

O gráfico RIP pode ser interpretado olhando as amostras que tem altos resíduos e/ou altas diferenças entre *loadings* calculados com ou sem a amostra. Este

gráfico ajuda a detectar amostras que diferem mais da população completa em relação ao modelo PARAFAC.

No gráfico IMP, em um dos eixos estão os escores obtidos com o modelo completo, $\mathbf{A}_{\text{completo}}$ e no outro eixo os escores previstos para a m -ésima amostra usando o modelo PARAFAC obtido na m -ésima iteração (*i.e.* o modelo obtido deixando fora a m -ésima amostra), $\mathbf{A}_{\text{previsto},m}$. Os escores previstos para as m -ésimas amostras são encontrados usando a seguinte expressão:

$$\mathbf{A}_{\text{previsto},m} = \mathbf{x}_m \cdot ((\mathbf{C}_{-m} \odot \mathbf{B}_{-m})^t)^+ \quad 1.83$$

em que \mathbf{x}_m corresponde ao dado experimental vetorizado para a m -ésima amostra. No gráfico IMP para cada analito, a amostra anômala será aquela que estiver longe da linha de identidade ideal.

Todos os métodos de validação descritos acima foram aplicados e os resultados são apresentados no Capítulo 7.

***N*-PLS, o PLS Multilinear**

Diferentes métodos podem ser empregados para realizar a calibração multivariada PLS ou PCR usando dados de ordem superior. Um deles é fazer a matriciação dos dados e construir o modelo PLS ou PCR normalmente, considerando os dados como de primeira ordem. Outra possibilidade é usar, por exemplo, o método PARAFAC para decompor o arranjo e então, subsequentemente, realizar a regressão PLS ou PCR da variável dependente sobre os escores, *i.e.* o modo das amostras. Entretanto, nenhum destes métodos utiliza o arranjo trilinear como um todo (senso trilinear) para extrair a informação necessária para a calibração. Porém, Ståhle (1989) [135] e posteriormente Bro (1996) [32], apresentaram uma metodologia que considera esta possível vantagem, estendendo diretamente para N -modos a regressão PLS aplicada em dois modos, o então conhecido N -PLS.

Apesar de o algoritmo de Ståhle (1989) [135] ser baseado na decomposição linear em três modos (LTD) aplicando a aproximação ALS, diferente do algoritmo de Bro (1996), ambos guiam para um modelo de calibração equivalente e fornecem as mesmas previsões [36]. O Algoritmo de Ståhle (1989), porém, fornece um

modelo complicado de se interpretar visto que o modelo torna-se um modelo como se os dados estivessem matriciados [27]. Apesar de atingirem os mesmos resultados, o algoritmo proposto por Bro (1996) apresenta algumas vantagens como: (1) o critério de otimização é explícito; (2) é numericamente mais confiável; (3) é usualmente mais rápido [36].

A principal característica do método de Bro (1996) [32] é que ele produz vetores de escores que em um senso trilinear apresenta a máxima covariância com a parte da variável dependente não explicada. A solução é fácil de explicar comparada com os métodos que realizam a matriciação. Isto é especialmente importante quando o número de variáveis é alto. O algoritmo é rápido devido ao número de parâmetros a serem estimados, que é menor e ao fato de que ele se resume a problemas de autovalores [32]. Há, porém uma perda no ajuste em relação aos modelos bilineares, devido às restrições mais severas [136] e também há perda da vantagem de segunda ordem, visto que as componentes são extraídas via decomposição SVD e, portanto, são sequenciais e o modelo não é único.

O algoritmo proposto por Bro(1996) [32] foi comentado por Smilde (1997) [149], que apresentou as equações para as etapas de previsão. Em 1998, de Jong [36] divulgou o algoritmo completo levemente modificado, em linguagem Matlab™. Entretanto, o algoritmo mostrado por de Jong apresenta vários erros e aqui é apresentado o algoritmo *N*-PLS corrigido por Martins [150].

1. $\mathbf{y}_e = \mathbf{y}$, inicialize a variável dependente;

Para cada componente:

2. $\mathbf{X}_{I \times JK} \leftarrow \mathbf{X}_{I \times J \times K}$, matriciação nos modos *JK* (variáveis independentes);

3. $\mathbf{c}_{1 \times JK} = \mathbf{y}_e^t \mathbf{X}$, vetor de covariância;

4. $\mathbf{Z}_{JK} \leftarrow \mathbf{c}_{1 \times JK}$, vetor de covariância convertido para matriz;

5. $[\mathbf{w}_j, \mathbf{s}, \mathbf{w}_k] \leftarrow \text{svd}(\mathbf{Z})$, decomposição SVD (Apêndice). Gera os pesos dos modos *J* e *K* (usa-se apenas a primeira componente);

6. $\mathbf{T} = \mathbf{X}(\mathbf{w}_k \otimes \mathbf{w}_j)$, cálculo dos escores;

7. $\mathbf{b} = (\mathbf{T}^t \mathbf{T})^{-1} \mathbf{T}^t \mathbf{y}$, relaciona os escores à variável dependente *y*.

8. $\mathbf{y}_e = \mathbf{y} - \mathbf{T}\mathbf{b}$, calcula o resíduo de *y*.

Para cada componente:

9. $\mathbf{b}_{\text{npls}} = (\mathbf{w}_k \otimes \mathbf{w}_j)\mathbf{b}$, coeficiente de regressão

Observe que o algoritmo não trabalha diretamente no tensor de segunda ordem, mas na sua forma matriciada e por isso usa a decomposição SVD.

Uma importante diferença do algoritmo N -PLS em relação ao PLS bilinear é que o cálculo da relação dos escores à variável dependente, \mathbf{b} , é realizado em cada etapa, não importando o número de componentes a serem calculados. Esta limitação é uma consequência da não ortogonalidade das componentes no algoritmo, ao contrário do PLS convencional [26].

Em relação à construção e validação do modelo de calibração usando o método N -PLS, basicamente os mesmos procedimentos já descritos nesta tese para PLS podem ser empregados, porém o algoritmo N -PLS apresenta a seguinte modificação após a etapa 8:

Para 1 até o número de componentes atual

$$9. \mathbf{b}_{NPLS} = \mathbf{b}_{NPLS} + (\mathbf{w}k \otimes \mathbf{w}j)\mathbf{b}$$

Finaliza o laço interno

$$10. \mathbf{B}_{nppls} = [\mathbf{B}_{nppls} \quad \mathbf{b}_{NPLS}]$$

Finaliza o laço externo

Com esta modificação no algoritmo N -PLS, o vetor de regressão é atualizado para cada uma das componentes desejadas e assim pode ser usado em um algoritmo de validação cruzada. O algoritmo anterior, diferentemente, calcula o vetor de regressão apenas para a última componente.

Seleção de variáveis em dados de segunda ordem

Dados multimodos, normalmente, são arranjos de um grande conjunto de dados, pois se trata de uma coleção de matrizes. Para processar tais informações, os algoritmos tendem a consumir muito tempo computacional e o custo da análise se torna alto e às vezes inviável computacionalmente. Além disso, dependendo da análise empregada, o modelo torna-se difícil de ser interpretado e muitas vezes, por processar variáveis não relevantes pode apresentar baixa capacidade de previsão.

Uma maneira de resolver estes problemas é selecionar as variáveis que realmente são importantes e assim: (1) diminuir o tamanho do conjunto de dados e

minimizar o tempo e custo computacional; (2) aumentar ou manter o poder de previsão e; (3) a partir das variáveis selecionadas, tornar o modelo interpretável.

Seleção de variáveis em dados multimodos é um campo de pesquisa muito recente dentro da quimiometria. Até o presente momento, há apenas quatro trabalhos que utilizaram seleção de variáveis em dados multimodos [151-154] e todos eles empregaram o algoritmo genético como selecionador de variáveis. O método proposto aqui é portanto o primeiro que usa uma estratégia diferente do GA para selecionar variáveis em dados multimodos.

O método de seleção das variáveis ordenadas (OPS[®]) proposto por Teófilo et al. (2006) [155,156] para dados em dois modos, será empregada aqui como uma extensão para dados em três modos e será denominado nOPS. A estratégia deste algoritmo é obter o vetor de regressão do modelo *N*-PLS, organizar este vetor em uma matriz (*matriz de regressão*) com *I* linhas e *J* colunas, em que cada linha é relacionada às matrizes de um dos modos e cada coluna está relacionada a cada variável do mesmo modo. Assim, um vetor é obtido para cada matriz de um dos modos. A transposta da *matriz de regressão* é relacionada com o outro modo das variáveis (Figuras 1.21).

Etapa 1. Vetores informativos individuais e execução do OPS[®]

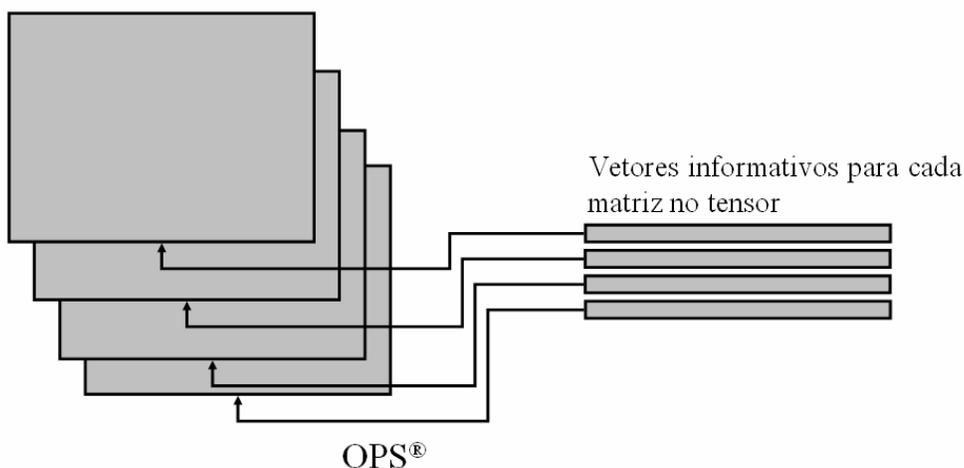


Figura 1.21. Primeira etapa do algoritmo nOPS: Ordenando as variáveis.

Considera-se aqui que cada linha da *matriz de regressão* e combinações destas linhas com outros vetores obtidos para cada matriz, traz informações das variáveis mais importantes nas suas respectivas matrizes no tensor de dados. Posteriormente

estes vetores são usados como entrada no algoritmo OPS[®], que é aplicado sobre cada matriz no tensor usando seu respectivo vetor (Figura 1.21).

Os conjuntos das variáveis selecionadas em cada matriz do tensor são armazenados e cada conjunto é repetido R vezes, de acordo com seus resultados estatísticos. Uma contagem das variáveis usadas em todos os conjuntos é realizada e o vetor com as variáveis usadas é decrescentemente ordenado de acordo com seu número de repetições (Figura 1.22).

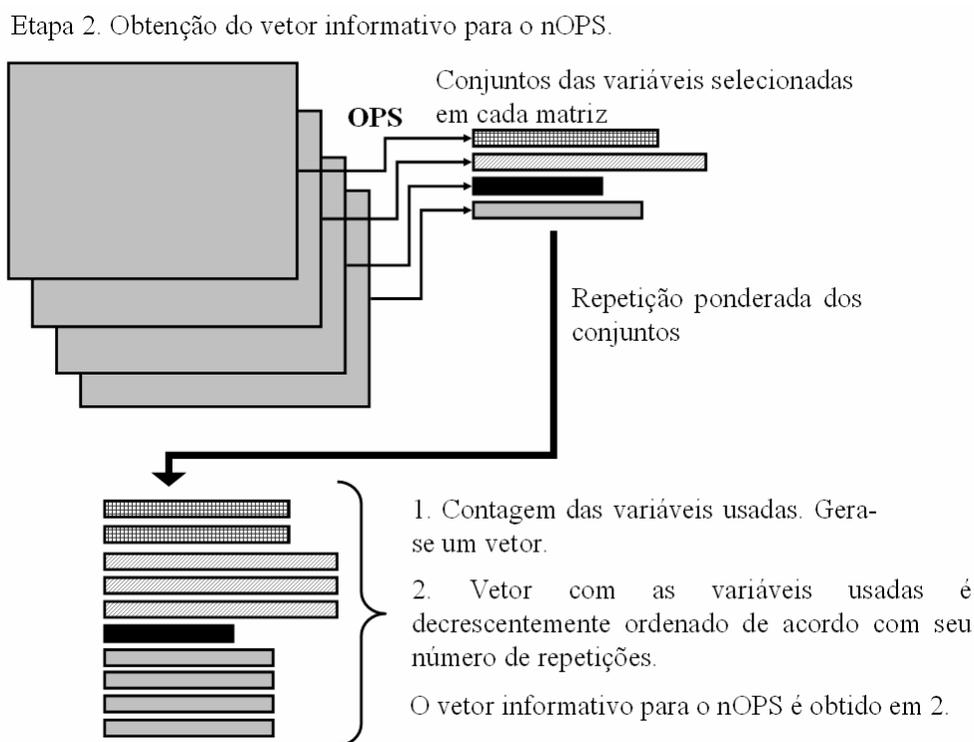


Figura 1.22. Segunda etapa do algoritmo nOPS: Obtendo o vetor prognóstico.

Este último vetor é usado como vetor informativo e empregado como no método OPS[®] para ordenar as variáveis. Aqui, o algoritmo empregado para analisar a janela e os incrementos das variáveis é o N -PLS (Figura 1.23).

Na janela e em cada janela com incrementos adicionados, obtêm-se parâmetros tais como RMSECV e r_{cv} , que são usados para decidir qual conjunto de variável apresenta maior capacidade de previsão.

Etapa 3. Execução do nOPS

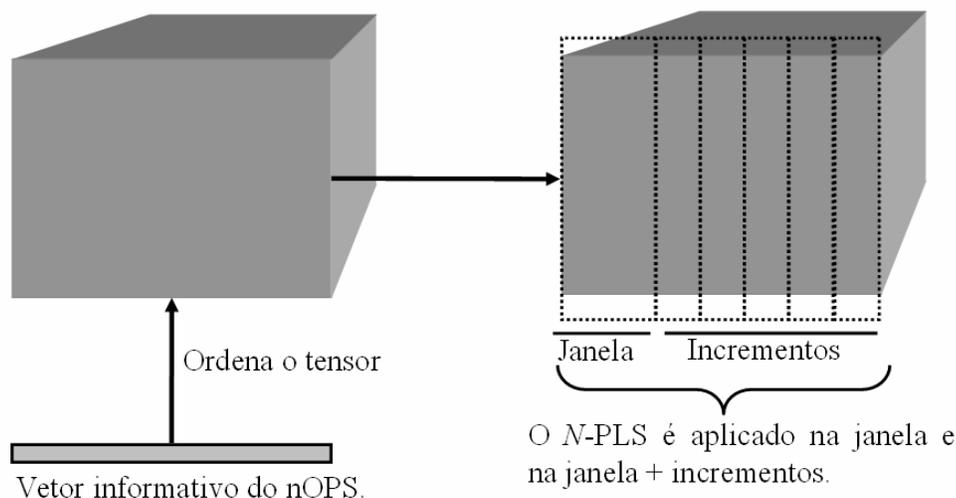


Figura 1.23. Terceira etapa do algoritmo nOPS: Investigação das variáveis ordenadas no tensor.

Após a seleção das variáveis em um dos modos, pode ser realizado a seleção das variáveis do outro modo sobre o conjunto já selecionado, ou sobre o conjunto completo. Portanto, as variáveis não são obrigatoriamente selecionadas em ambos os modos, o que pode ser uma vantagem em casos específicos. Esta propriedade do nOPS difere do método GA para selecionar variáveis em dados multimodos, visto que ao usar o GA, obrigatoriamente seleciona-se variáveis em ambos os modos, simultaneamente.

Referências

1. Ferreira, M. M. C.; Antunes, A. M.; Melgo, M. S.; Volpe, P. L. O. Chemometrics I: Multivariate calibration, a tutorial. *Quim. Nova* **1999**, 22, (5), 724-731.
2. Winefordner, J. D. *Chemometrics: From basics to wavelet transform*. John Wiley & Sons Inc: New Jersey, 2004; Vol. 164, p 292.
3. Bro, R. Review on multiway analysis in chemistry - 2000-2005. *Crit. Rev. Anal. Chem.* **2006**, 36, (3-4), 279-293.
4. Kiers, H. Towards a standardized notation and terminology in multiway analysis. *J. Chemometr.* **2000**, 14, (3), 105-122.
5. Chau, F. T.; Liang, Y. Z.; Gao, J.; Shao, X. G. Chemometrics: From Basics to Wavelet Transform. In *Chemical analysis*, 1 ed.; Winefordner, J. D., Ed. John Wiley & Sons Inc: New Jersey, 2004, Vol. 164.

6. Brereton, R. G. *Chemometrics: Data Analysis for the laboratory and chemical plant*. John Wiley & Sons Inc: Chinchester, 2003; p 407.
7. Eco Chemie, B. V. *General purpose electrochemical system (GPES) for windows*, 4.9.004; Eco Chemie: Utrecht, 2002.
8. Montgomery, D. C.; Runger, G. C. *Applied Statistics and Probability for Engineers*. 3rd ed.; John Wiley & Sons Inc: New York, 2003; p 703.
9. Lundstedt, T.; E., S.; Abramo, L.; Thelin, B.; Nyström, A.; Pettersen, J.; Bergman, R. Experimental design and optimization. *Chemometrics Intell. Lab. Syst.* **1998**, 42, 3-40.
10. Neto, B. B.; Scarminio, I. S.; Bruns, R. E. *Como fazer experimentos: pesquisa e desenvolvimento na ciência e na indústria*. Editora Unicamp: Campinas, 2001; p 401.
11. Box, G. E. P.; Hunter, W. G.; Hunter, J. S. *Statistic for Experimenters: An Introduction to Design, Data Analysis and Model Building*. 1 ed.; John Wiley & Sons Inc: New York, 1978.
12. Plackett, R. L.; Burman, J. P. The design of optimum multifactorial experiments. *Biometrika* **1946**, 33, (4), 305-325.
13. Morgan, E.; Burton, K. W.; Church, P. A. Pratical Exploratory Experimentatl Designs. *Chemometrics Intell. Lab. Syst.* **1989**, 5, 283-302.
14. Myers, R. H.; Montgomery, D. C. *Response Surface Methodology*. John Wiley & Sons Inc: New York, 2002.
15. Box, G. E. P.; Draper, N. R. *Empirical Moodel-Building and Response Surfaces*. 1 ed.; John Wiley & Sons Inc: New York, 1987.
16. Christensen, R. *Analysis of Variance, design and regression*. CRC Press Inc: New York, 2000.
17. Box, G. E. P.; Wilson, K. B. On the experimental attainment of optimum conditions. *J. Roy. Statist. Soc., Ser. B13* **1951**, 1-38.
18. Doehlert, D. H. Uniform shell designs. *Appl. Stat.* **1970**, 19, 231-239.
19. Ferreira, S.; dos Santos, W.; Quintella, C. M.; Neto, B. B.; Bosque-Sendra, J. A. Doehlert matrix: a chemometric tool for analytical chemistry - review. *Talanta* **2004**, 63, (4), 1061-1067.
20. Massart, D. L.; Vandeginste, B. G. M.; Buydens, L. M. C.; de Jong, S.; Lewi, P. J.; Verbeke-Smeyers, J. *Handbook of Chemometrics and Qualimetrics*. Elsevier: Amsterdam, 1998; Vol. Part A.
21. Martens, H.; Naes, T. *Multivariate Calibration*. John Wiley & Sons: New York, 1989; p 419.

22. Phatak, A. Evaluation of some multivariate methods and their applications in chemical engineering. University of Waterloo, Ontario, 1993.
23. Golub, G. H.; Van Loan, C. F. *Matrix computation*. 2nd ed.; Johns Hopkins University Press: Baltimore, 1989.
24. Beebe, K. R.; Pell, R. J.; Seasholtz, M. B. *Chemometrics: A practical guide*. 1 ed.; John Wiley & Sons Inc: New York, 1998; p 348.
25. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemometrics Intell. Lab. Syst.* **1987**, 2, (1-3), 37-52.
26. Brereton, R. G. Introduction to multivariate calibration in analytical chemistry. *Analyst* **2000**, 125, (11), 2125-2154.
27. Smilde, A.; Bro, R.; Geladi, P. *Multi-way analysis with applications in the chemical science*. 1 ed.; John Wiley & Sons Inc: Chichester, 2004; p 379.
28. Escandar, G. M.; Damiani, P. C.; Goicoechea, H. C.; Olivieri, A. C. A review of multivariate calibration methods applied to biomedical analysis. *Microchem J.* **2006**, 82, (1), 29-42.
29. Bro, R. Multivariate calibration - What is in chemometrics for the analytical chemist? *Anal. Chim. Acta* **2003**, 500, (1-2), 185-194.
30. Naes, T.; Isaksson, T.; Fearn, T.; Davies, T. *Multivariate calibration and classification*. Nir publications: Chichester, 2002; p 344.
31. Geladi, P.; Kowalski, B. R. Partial least-squares regression - a tutorial. *Anal. Chim. Acta* **1986**, 185, 1-17.
32. Bro, R. Multiway calibration. Multilinear PLS. *J. Chemometr.* **1996**, 10, (1), 47-61.
33. Andersen, C. M.; Bro, R. Practical aspects of PARAFAC modeling of fluorescence excitation-emission data. *J. Chemometr.* **2003**, 17, (4), 200-215.
34. Bro, R. PARAFAC. Tutorial and applications. *Chemometrics Intell. Lab. Syst.* **1997**, 38, (2), 149-171.
35. Phatak, A.; de Hoog, F. Exploiting the connection between PLS, Lanczos methods and conjugate gradients: alternative proofs of some properties of PLS. *J. Chemometr.* **2002**, 16, (7), 361-367.
36. De Jong, S. Regression coefficients in multilinear PLS. *J. Chemometr.* **1998**, 12, (1), 77-81.
37. Helland, I. S. Some theoretical aspects of partial least squares regression. *Chemometrics Intell. Lab. Syst.* **2001**, 58, (2), 97-107.
38. Phatak, A.; De Jong, S. The geometry of partial least squares. *J. Chemometr.* **1997**, 11, (4), 311-338.

39. Elden, L. Partial least-squares vs. Lanczos bidiagonalization - I: analysis of a projection method for multiple regression. *Comput. Stat. Data Anal.* **2004**, 46, (1), 11-31.
40. Manne, R. Analysis of 2 partial-least-squares algorithms for multivariate calibration. *Chemometrics Intell. Lab. Syst.* **1987**, 2, (1-3), 187-197.
41. Barker, M.; Rayens, W. Partial least squares for discrimination. *J. Chemometr.* **2003**, 17, (3), 166-173.
42. Wold, H. The basic design and some extensions. In *Systems under indirect observation*, G., J. K. and Wold, H., Eds. North-Holland: Amsterdam, 1982, Vol. II, pp 1-53.
43. Wold, S.; Sjostrom, M.; Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometrics Intell. Lab. Syst.* **2001**, 58, (2), 109-130.
44. Lohmöller, J.-B. *Latent variable path modeling with partial least squares*. 1 ed.; Physica-Verlag: Heidelberg, 1989.
45. Kowalski, B. R.; Gerlach, R. W.; Wold, H. Chemical systems under indirect observations. In *Systems under indirect observation*, G., J. K. and Wold, H., Eds. North-Holland: Amsterdam, 1982, Vol. 2, pp 191-207.
46. Geladi, P. Notes on the history and nature of partial least squares (PLS). *J. Chemometr.* **1988**, 2, 231-246.
47. Martens, H. Multivariate Calibration: quantitative interpretation of non-selective chemical data. PhD thesis, Trondheim, 1985.
48. Frank, I. E.; Kalivas, J. H.; Kowalski, B. R. Partial least-squares solutions for multicomponent analysis. *Anal. Chem.* **1983**, 55, (11), 1800-1804.
49. Wold, S.; Martens, H.; Wold, H. The multivariate calibration problem in chemistry solved by partial least squares. In *Lecture notes in mathematics: Matrix Pencils*, 1 ed.; A., R. and B., K., Eds. Springer-Verlag: Heidelberg, 1983, pp 286-283.
50. Teófilo, R. F.; Martins, J. P. A.; Ferreira, M. M. C. In *Study of the computational performance of PLS algorithms using experimental design*, 10th Scandinavian Symposium on Chemometrics, Lappeenranta, Finland, Lappeenranta, Finland, 2007, PO35.
51. Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. J. The collinearity problem in linear-regression - the partial least-squares (PLS) approach to generalized inverses. **1984**, 5, (3), 735-743.
52. De Jong, S. SIMPLS - an alternative approach to partial least-squares regression. *Chemometrics Intell. Lab. Syst.* **1993**, 18, (3), 251-263.
53. Lindgren, F.; Geladi, P.; Wold, S. The kernel algorithm for PLS. *J. Chemometr.* **1993**, 7, (1), 45-59.

54. Naes, T. Multivariate calibration when the error covariance-matrix is structured. *Technometrics* **1985**, 27, (3), 301-311.
55. Lorber, A.; Wangen, L. E.; Kowalski, B. R. A theoretical foundation for the PLS algorithm. *J. Chemometr.* **1987**, 1, 19-31.
56. Helland, I. S. On the structure of partial least squares regression. *Commun. Stat.-Simul. Comput.* **1988**, 17, 581-607.
57. Ergon, R. PLS score-loading correspondence and a bi-orthogonal factorization. *J. Chemometr.* **2002**, 16, (7), 368-373.
58. Pell, R. J.; Ramos, L.; Manne, R. The model space in partial least square regression. *J. Chemometr.* **2007**, 21, (3-4), 165-172.
59. Golub, G. H.; Kahan, W. Calculating the singular values and pseudo-inverse of a matrix. *SIAM J. Num. Anal. Ser. B.* **1965**, 2, 205-224.
60. Lorber, A.; Kowalski, B. R. A note on the use of the partial least-squares method for multivariate calibration. *Appl. Spectrosc.* **1988**, 42, (8), 1572-1574.
61. Pirouette, T. *Pirouette: Multivariate data analysis*, 3.11; Infometrix: Woodinville, 2003.
62. Efron, B. Estimating the error rate of a prediction rule: Improvement on cross-validation. *J. Am. Stat. Assoc.* **1983**, 78, 316-331.
63. Baumann, K. Cross-validation as the objective function for variable-selection techniques. *Trac-Trends Anal. Chem.* **2003**, 22, (6), 395-406.
64. Xu, L.; Schechter, I. Wavelength selection for simultaneous spectroscopic analysis. Experimental and theoretical study. *Anal. Chem.* **1996**, 68, (14), 2392-2400.
65. Nadler, B.; Coifman, R. R. The prediction error in CLS and PLS: the importance of feature selection prior to multivariate calibration. *J. Chemometr.* **2005**, 19, (2), 107-118.
66. Ferreira, M. M. C. Multivariate QSAR. *J. Braz. Chem. Soc.* **2002**, 13, (6), 742-753.
67. Jouanrimbaud, D.; Walczak, B.; Massart, D. L.; Last, I. R.; Prebble, K. A. Comparison of multivariate methods based on latent vectors and methods based on wavelength selection for the analysis of near-infrared spectroscopic data. *Anal. Chim. Acta* **1995**, 304, (3), 285-295.
68. Jiang, J. H.; Berry, R. J.; Siesler, H. W.; Ozaki, Y. Wavelength interval selection in multicomponent spectral analysis by moving window partial least-squares regression with applications to mid-infrared and near-infrared spectroscopic data. *Anal. Chem.* **2002**, 74, (14), 3555-3565.
69. Centner, V.; Massart, D. L.; de Noord, O. E.; De Jong, S.; Vandeginste, B. M.; Sterna, C. Elimination of uninformative variables for multivariate calibration. *Anal. Chem.* **1996**, 68, (21), 3851-3858.

70. Forina, M.; Lanteri, S.; Oliveros, M.; Millan, C. P. Selection of useful predictors in multivariate calibration. *Anal. Bioanal. Chem.* **2004**, 380, (3), 397-418.
71. Herrero, A.; Ortiz, M. C. Qualitative and quantitative aspects of the application of genetic algorithm-based variable selection in polarography and stripping voltammetry. *Anal. Chim. Acta* **1999**, 378, (1-3), 245-259.
72. Chong, I. G.; Jun, C. H. Performance of some variable selection methods when multicollinearity is present. *Chemometrics Intell. Lab. Syst.* **2005**, 78, (1-2), 103-112.
73. Arcos, M. J.; Alonso, C.; Ortiz, M. C. Genetic-algorithm-based potential selection in multivariate voltammetric determination of indomethacin and acetaminophen by partial least squares. *Electrochim. Acta* **1998**, 43, (5-6), 479-485.
74. Majidi, M. R.; Jouyban, A.; Asadpour-Zeynali, K. Genetic algorithm based potential selection in simultaneous voltammetric determination of isoniazid and hydrazine by using partial least squares (PLS) and artificial neural networks (ANNs). *Electroanalysis* **2005**, 17, (10), 915-918.
75. Leardi, R. Genetic algorithms in chemometrics and chemistry: a review. *J. Chemometr.* **2001**, 15, (7), 559-569.
76. Goicoechea, H. C.; Olivieri, A. C. A new family of genetic algorithms for wavelength interval selection in multivariate analytical spectroscopy. *J. Chemometr.* **2003**, 17, (6), 338-345.
77. Wise, B. M.; Gallagher, N. B.; Bro, R.; Shaver, J. M. *PLS_Toolbox 3.01*, Eigenvector Research, Inc.: Manson, 2003.
78. Williams, R. P.; Swinkels, A. J.; Maeder, M. Identification and application of a prognostic vector for use in multivariate calibration and prediction. *Chemometrics Intell. Lab. Syst.* **1992**, 15, 185-193.
79. Hoskuldsson, A. Variable and subset selection in PLS regression. *Chemometrics Intell. Lab. Syst.* **2001**, 55, (1-2), 23-38.
80. Brown, P. J. Wavelength selection in multicomponent near-infrared calibration. *J. Chemometr.* **1992**, 6, (3), 151-161.
81. Miller, C. E. The use of chemometric techniques in process analytical method development and operation. *Chemometrics Intell. Lab. Syst.* **1995**, 30, (1), 11-22.
82. Wold, S.; Johansson, E.; Cocchi, M. 3D QSAR. In *Drug Design: Theory, Methods and Applications*, 1 ed.; ESCOM, T., Ed. Holland: Leiden, 1993, pp 523-550.
83. Dodds, S. A.; Heath, W. P. Construction of an online reduced-spectrum NIR calibration model from full-spectrum data. *Chemometrics Intell. Lab. Syst.* **2005**, 76, (1), 37-43.

84. Ferre, J.; Faber, N. M. Net analyte signal calculation for multivariate calibration. *Chemometrics Intell. Lab. Syst.* **2003**, 69, (1-2), 123-136.
85. R., M. D. Spectral absorption pattern detection and estimation. I. Analytical techniques. *Appl. Spectrosc.* **1977**, 31, (5), 404-415.
86. Lorber, A. Error propagation and figures of merit for quantification by solving matrix equations. *Anal. Chem.* **1986**, 58, (6), 1167-1172.
87. Lorber, A.; Faber, K.; Kowalski, B. R. Net analyte signal calculation in multivariate calibration. *Anal. Chem.* **1997**, 69, (8), 1620-1626.
88. Ferre, J.; Rius, F. X. Detection and correction of biased results of individual analytes in multicomponent spectroscopic analysis. *Anal. Chem.* **1998**, 70, (9), 1999-2007.
89. Xu, L. A.; Schechter, I. A calibration method free of optimism factor number selection for automated multivariate analysis. Experimental and theoretical study. *Anal. Chem.* **1997**, 69, (18), 3722-3730.
90. Goicoechea, H. C.; Olivieri, A. C. A comparison of orthogonal signal correction and net analyte preprocessing methods. Theoretical and experimental study. *Chemometrics Intell. Lab. Syst.* **2001**, 56, (2), 73-81.
91. Wentzell, P. D.; Andrews, D. T.; Kowalski, B. R. Maximum likelihood multivariate calibration. *Anal. Chem.* **1997**, 69, (13), 2299-2311.
92. Goicoechea, H. C.; Olivieri, A. C. Enhanced synchronous spectrofluorometric determination of tetracycline in blood serum by chemometric analysis. Comparison of partial least-squares and hybrid linear analysis calibrations. *Anal. Chem.* **1999**, 71, (19), 4361-4368.
93. Booksh, K. S.; Kowalski, B. R. Theory of analytical chemistry. *Anal. Chem.* **1994**, 66, (15), 782A-791A.
94. Damiani, P. C.; Ribone, M. E.; Olivieri, A. C. Rapid-determination of paracetamol in blood-serum samples by first-derivative uv absorption-spectroscopy. *Anal. Lett.* **1995**, 28, (12), 2219-2226.
95. Brown, C. D. Discordance between net analyte signal theory and practical multivariate calibration. *Anal. Chem.* **2004**, 76, (15), 4364-4373.
96. Bro, R.; Andersen, C. M. Theory of net analyte signal vectors in inverse regression. *J. Chemometr.* **2003**, 17, (12), 646-652.
97. Faber, N. M. Exact presentation of multivariate calibration model as univariate calibration graph. *Chemometrics Intell. Lab. Syst.* **2000**, 50, (1), 107-114.
98. Faber, N. M. Mean centering and computation of scalar net analyte signal in multivariate calibration. *J. Chemometr.* **1998**, 12, (6), 405-409.
99. Faber, N. M. Efficient computation of net analyte signal vector in inverse multivariate calibration models. *Anal. Chem.* **1998**, 70, (23), 5108-5110.

100. Brown, C. D.; Ridder, T. D. Framework for multivariate selectivity analysis, part I: Theoretical and practical merits. *Appl. Spectrosc.* **2005**, 59, (6), 787-803.
101. Booksh, K. S.; Kowalski, B. R. Theory of analytical-chemistry. *Anal. Chem.* **1994**, 66, (15), A782-A791.
102. Mirmohseni, A.; Oladegaragoze, A. Application of the quartz crystal microbalance for determination of phenol in solution. *Sens. Actuator B-Chem.* **2004**, 98, (1), 28-36.
103. De la Pena, A. M.; Espinosa-Mansilla, A.; Valenzuela, M.; Goicoechea, H. C.; Olivieri, A. C. Comparative study of net analyte signal-based methods and partial least squares for the simultaneous determination of amoxicillin and clavulanic acid by stopped-flow kinetic analysis. *Anal. Chim. Acta* **2002**, 463, (1), 75-88.
104. Ortiz, M. C.; Sarabia, L. A.; Herrero, A.; Sanchez, M. S.; Sanz, M. B.; Rueda, M. E.; Gimenez, D.; Melendez, M. E. Capability of detection of an analytical method evaluating false positive and false negative (ISO 11843) with partial least squares. *Chemometrics Intell. Lab. Syst.* **2003**, 69, (1-2), 21-33.
105. Boque, R.; Larrechi, M. S.; Rius, F. X. Multivariate detection limits with fixed probabilities of error. *Chemometrics Intell. Lab. Syst.* **1999**, 45, (1-2), 397-408.
106. Felipe-Sotelo, M.; Cal-Prieto, M. J.; Ferre, J.; Boque, R.; Andrade, J. M.; Carlosena, A. Linear PLS regression to cope with interferences of major concomitants in the determination of antimony by ETAAS. *J. Anal. At. Spectrom.* **2006**, 21, (1), 61-68.
107. Sanz, M. B.; Sarabia, L. A.; Herrero, A.; Ortiz, M. C. Capability of discrimination: application to soft calibration methods. *Anal. Chim. Acta* **2001**, 446, (1-2), 297-311.
108. Faber, N. M.; Bro, R. Standard error of prediction for multiway PLS 1. Background and a simulation study. *Chemometrics Intell. Lab. Syst.* **2002**, 61, (1-2), 133-149.
109. Clayton, C. A.; Hines, J. W.; Elkins, P. D. Detection limits with specified assurance probabilities. *Anal. Chem.* **1987**, 59, 2506-2514.
110. Sarabia, L.; Ortiz, M. C. DETARCHI: a program for detection limits with specified assurance probabilities and characteristic curves of detection. *Trac-Trends Anal. Chem.* **1994**, 13, (1), 1-6.
111. Garcia, I.; Sarabia, L. A.; Ortiz, M. C. Detection capability of tetracyclines analysed by a fluorescence technique: comparison between bilinear and trilinear partial least squares models. *Anal. Chim. Acta* **2004**, 501, (2), 193-203.
112. Boque, R.; Rius, F. X. Computing detection limits in multicomponent spectroscopic analysis. *Trac-Trends Anal. Chem.* **1997**, 16, (8), 432-436.
113. Cuadrado, J. A. P.; Forn, M. P. *Validación de métodos analíticos*. 1 ed.; A.E.F.I.: Barcelona, 2001; Vol. 1, p 315.

114. Cattell, R. B. Parallel proportional profiles and others principles for determining the choice of factors by rotation. *Psychometrika* **1944**, 9, 267.
115. Tucker, L. R. Implications of factors analysis of three-way matrices for measurement of change. In *Problems of Measuring Change*, 1 ed.; Harris, C. W., Ed. University of Wisconsin Press: Madison, 1963, pp 122-137.
116. Tucker, L. R. Some mathematical notes on three-mode factor analysis. *Psychometrika* **1966**, 31, 279-311.
117. Harshman, R. A. Foundations of the PARAFAC procedure: Model and conditions for an 'explanatory' multi-mode factor analysis. *UCLA Working Paper in Phonetics* **1970**, 16, 1.
118. Carroll, J. D.; Chang, J. Analysis of individual differences in multidimensional scaling via an N-way generalization of and Eckart-Young decomposition. *Psychometrika* **1970**, 35, 283.
119. Ho, C. N.; Christian, G. D.; Davidson, E. R. Application of method of rank annihilation to quantitative-analyses of multicomponent fluorescence data from video fluorometer. *Anal. Chem.* **1978**, 50, (8), 1108-1113.
120. Ho, C. N.; Christian, G. D.; Davidson, E. R. Application of the method of rank annihilation to fluorescent multicomponent mixtures of polynuclear aromatic-hydrocarbons. *Anal. Chem.* **1980**, 52, (7), 1071-1079.
121. De Ligny, C.; Spanjer, M.; van Houwelingen, J.; Weise, H. Three-mode factor analysis of data on retention in normal-phase high-performance liquid chromatography. *J. Chromatogr. A* **1984**, 301, 311-324.
122. Wold, S.; Geladi, P.; Esbensen, K.; J., Ö. Multi-way principal components analysis and PLS analysis. *J. Chemometr.* **1987**, 1, 41-56.
123. Geladi, P. Analysis of Multi-Way (Multi-Mode) data. *Chemometrics Intell. Lab. Syst.* **1989**, 7, 11-30.
124. Smilde, A. K.; Doornbos, D. A. 3-way methods for the calibration of chromatographic systems - comparing PARAFAC and 3-way PLS. *J. Chemometr.* **1991**, 5, (4), 345-360.
125. Smilde, A. K. 3-WAY Analyses - problems and prospects. *Chemometrics Intell. Lab. Syst.* **1992**, 15, (2-3), 143-157.
126. Tomasi, G. Practical and computational aspects in chemometric data analysis. PhD thesis, The Royal Veterinary and Agricultural University, Frederiksberg, 2006.
127. Diez, R.; Sarabia, L.; Ortiz, M. C. Rapid determination of sulfonamides in milk samples using fluorescence spectroscopy and class modeling with n-way partial least squares. *Anal. Chim. Acta* **2007**, 585, (2), 350-360.

128. Hall, G. J.; Kenny, J. E. Estuarine water classification using EEM spectroscopy and PARAFAC-SIMCA. *Anal. Chim. Acta* **2007**, 581, (1), 118-124.
129. Holbrook, R. D.; Yen, J. H.; Grizzard, T. J. Characterizing natural organic material from the Occoquan Watershed (Northern Virginia, US) using fluorescence spectroscopy and PARAFAC. *Sci. Total Environ.* **2006**, 361, (1-3), 249-266.
130. Zhang, Q. Q.; Zhang, F.; Wang, W. G.; Wang, X. L. Synthesis, crystal structure and DNA binding studies of a binuclear copper(II) complex with phenanthroline. *J. Inorg. Biochem.* **2006**, 100, (8), 1344-1352.
131. Valverde, R. S.; Garcia, M.; Galera, M. M.; Goicoechea, H. C. Highly collinear three-way photoinduced spectrofluorimetric data arrays modelled with bilinear least-squares: Determination of tetracyclines in surface water samples. *Talanta* **2006**, 70, (4), 774-783.
132. Wang, T.; Zhang, L. D.; Lao, C. L.; Zhao, L. L.; Li, J. H.; Wang, Z. Y.; Huang, W. J.; Tao, A. W.; Lu-da, Z.; Cai-lian, L.; Long-lian, Z.; Jun-hui, L. i.; Zhong-yi, W.; Wen-jiang, H. Study on NPLS model with three-dimensional hyperspectrum for assessing per-mu-yield of winter wheat. *Spectrosc. Spectr. Anal.* **2006**, 26, (10), 1915-1917.
133. Hall, G. J.; Clow, K. E.; Kenny, J. E. Estuarial fingerprinting through multidimensional fluorescence and multivariate analysis. *Environ. Sci. Technol.* **2005**, 39, (19), 7560-7567.
134. Selli, E.; Zaccaria, C.; Sena, F.; Tomasi, G.; Bidoglio, G. Application of multi-way models to the time-resolved fluorescence of polycyclic aromatic hydrocarbons mixtures in water. *Water Res.* **2004**, 38, (9), 2269-2276.
135. Stahle, L. Aspects of the analysis of 3-way data. *Chemometrics Intell. Lab. Syst.* **1989**, 7, 95-100.
136. Bro, R. Multi-way analysis in the food industry. PhD thesis, Royal Veterinary and Agricultural University, Frederiksberg, 1998.
137. Arancibia, J. A.; Escandar, G. M. Two different strategies for the fluorimetric determination of piroxicam in serum. *Talanta* **2003**, 60, (6), 1113-1121.
138. Zhang, Y.; Wu, H. L.; Xia, A. L.; Zhu, S. H.; Han, Q. J.; Yu, R. Q. Fluorescence determination of metoprolol in human plasma by trilinear decomposition-based calibration techniques. *Anal. Bioanal. Chem.* **2006**, 386, (6), 1741-1748.
139. Xia, A. L.; Wu, H. L.; Fang, D. M.; Ding, Y. J.; Hu, L. Q.; Yu, R. Q. Determination of daunomycin in human plasma and urine by using an interference-free analysis of excitation-emission matrix fluorescence data with second-order calibration. *Anal. Sci.* **2006**, 22, (9), 1189-1195.
140. Sanchez, E.; Kowalski, B. R. Generalized rank annihilation factor-analysis. *Anal. Chem.* **1986**, 58, (2), 496-499.

141. Sanchez, E.; Kowalski, B. R. Tensorial resolution: a direct trilinear decomposition. *J. Chemometr.* **1990**, 4, 29-45.
142. Faber, N. M.; Bro, R.; Hopke, P. K. Recent developments in CANDECOMP/PARAFAC algorithms: a critical review. *Chemometrics Intell. Lab. Syst.* **2003**, 65, (1), 119-137.
143. Tomasi, G.; Bro, R. A comparison of algorithms for fitting the PARAFAC model. *Comput. Stat. Data Anal.* **2006**, 50, (7), 1700-1734.
144. Yates, F. The analysis of replicated experiments when the field results are incomplete. *The empire journal of experimental agriculture* **1933**, 1, 129-135.
145. Bro, R.; Kiers, H. A new efficient method for determining the number of components in PARAFAC models. *J. Chemometr.* **2003**, 17, (5), 274-286.
146. Bro, R. *N-Way tutorial: curso on-line*. Royal Veterinary and Agricultural University: Frederiksberg, pp 1-72.
147. Harshman, R. A. How can I know if it's 'real' ? In *Research methods for multimode data analysis*, 1 ed.; Law, H. G., Snyder, C. W., Hattie, J. A. and MacDonald, R. P., Eds. Praeger: New York, 1984, pp 566.
148. Riu, J.; Bro, R. Jack-knife technique for outlier detection and estimation of standard errors in PARAFAC models. *Chemometrics Intell. Lab. Syst.* **2003**, 65, (1), 35-49.
149. Smilde, A. K. Comments on multilinear PLS. *J. Chemometr.* **1997**, 11, (5), 367-377.
150. Martins, J. P. A. unpublished results. Campinas, 2007.
151. Wu, W.; Guo, Q.; Massart, D. L.; Boucon, C.; de Jong, S. Structure preserving feature selection in PARAFAC using a genetic algorithm and Procrustes analysis. *Chemometrics Intell. Lab. Syst.* **2003**, 65, (1), 83-95.
152. Lopes, J. A.; Menezes, J. C. Industrial fermentation end-product modelling with multilinear PLS. *Chemometrics Intell. Lab. Syst.* **2003**, 68, (1-2), 75-81.
153. Gourvenec, S.; Capron, X.; Massart, D. L. Genetic algorithms (GA) applied to the orthogonal projection approach (OPA) for variable selection. *Anal. Chim. Acta* **2004**, 519, (1), 11-21.
154. Carneiro, R. L.; Braga, J. W. B.; Bottoli, C. B. G.; Poppi, R. J. Application of genetic algorithm for selection of variables for the BLLS method applied to determination of pesticides and metabolites in wine. *Anal. Chim. Acta* **2007**, 595, 51-58.
155. Teófilo, R. F.; Martins, J. P. A.; Ferreira, M. M. C. *OPS Toolbox 1.0*, Registro de Software no INPI - 0000270703255138: Brasil, 2007.
156. Teófilo, R. F.; Martins, J. P. A.; Ferreira, M. M. C. In *Ordered predictors Selection: an intuitive method to find the most relevant variable in multivariate calibration*, 10th International Conference on Chemometrics in Analytical Chemistry, Águas de Lindóia, Brazil, Águas de Lindóia, Brazil, 2006, P066.

Capítulo 2

Diamante Dopado com Boro: Aspectos Eletroquímicos

*“As coisas são mais como são agora do
que jamais foram antes”
Dwight D. Eisenhower*

Introdução

O diamante é conhecido desde o ano 3.000 a.C. como um cristal misterioso, indestrutível e de brilho atraente, que esteve sempre associado à sorte, fidelidade, riqueza, ostentação, entre outras [1]. Em 1772, o químico Antoine Lavoisier foi o primeiro a desvendar a sua composição. Concluiu, acertadamente, que o diamante era composto unicamente de carbono.

Através dos séculos, nenhum material foi tão apreciado ou estimado como o diamante. No entanto, até relativamente pouco tempo atrás (1940), os diamantes eram vistos somente como pedras preciosas para fabricação de jóias, porém, as notáveis combinações de propriedades do diamante o tornaram um material desejado para aplicações mecânicas, óticas, eletrônicas e mais recentemente, em eletroquímica [1,2].

Na natureza o diamante é formado em condições extremas de temperatura e pressão. Entretanto, guiados por avidez bem como por desafios, os cientistas, por mais de um século, têm se esforçado para sintetizar diamantes. A fascinante história destes esforços é eloqüente e cronologicamente descrita no livro ‘*The Diamond Makers*’ de Robert M. Hazen [3].

Os esforços para produzir diamante iniciaram nos anos de 1820 e se estenderam até o início de 1950. Até então, nenhum sucesso havia sido alcançado.

Porém, durante a década de 1950, a empresa *General Electric* embarcou em um ambicioso programa para produzir diamantes. O time de pesquisadores entre 1954-55 realizou um excelente trabalho, que necessitou dos principais avanços em materiais de alta pressão e temperatura e dispositivos de alta tecnologia, bem como novos discernimentos dentro de um apropriado ambiente químico para a realização do crescimento do diamante. Finalmente, em 1957, a companhia iniciou a venda de diamantes feitos pelo homem. Em 2007, portanto, o diamante sintético comemora as suas bodas de ouro com a indústria.

Apesar deste grande avanço, o equipamento necessário para crescimento em alta pressão e temperatura era caro e muito pesado para produção em escala. Então, pesquisas científicas para produzir diamante à baixa pressão usando equipamentos pequenos e de menor custo se iniciaram durante a primeira metade da década de 1950. William Eversole, um cientista da *Union Carbide*, inventou, entre os anos de 1957-58, o primeiro equipamento reprodutível de baixa pressão para síntese de diamantes. Esforços paralelos foram realizados por dois principais grupos na década de 60: o de Derjaguin, na antiga União Soviética, e o de Angus na Universidade de Case Western Reserve, nos Estados Unidos (EUA). As pesquisas destes grupos mostraram que era possível realizar o crescimento do diamante a baixa pressão (0,05 a 0,1 atm) e temperaturas médias (500 a 1000 °C) empregando a técnica de deposição de vapor químico (CVD). Estes resultados deram origem à fundação da moderna tecnologia empregada nos dias de hoje para o crescimento do diamante a baixas pressões.

A alta mobilidade eletrônica e de vacâncias e estabilidade a altas temperaturas, são algumas propriedades que fazem o diamante atrativo para uso na indústria eletrônica de alta velocidade. Mas infelizmente, durante este período, dificuldades encontradas em obter diamantes depositados com baixos defeitos, a um custo efetivamente baixo, além de controle preciso sobre dopagens tipos *p*- e *n*-, impediram as aplicações de se tornarem uma realidade.

Entretanto, as pesquisas avançaram nas décadas de 1970, 1980 e 1990. Principalmente nesta última década, onde as pesquisas se concentraram menos nos EUA e mais na Europa e Japão, o custo de produção foi significativamente

reduzido devido a melhorias na nucleação e taxas de crescimento e, além disso, mais se aprendeu sobre o mecanismo e cinética da síntese do diamante.

Atualmente, o uso de diamantes tem apresentado grandes potencialidades em duas principais áreas: emissão de elétrons, com uso de diamantes adequadamente tratados e em eletroquímica, com o uso de diamantes condutores.

Na eletroquímica, objeto desta tese, o uso do diamante condutor foi usado pela primeira vez no ano de 1987 por Pleskov et al. [4].

O diamante é um material com amplo *bandgap* ($E_{\text{gap}} = 5,5 \text{ eV}$), sendo, normalmente, um magnífico isolante elétrico. Entretanto, quando o material é dopado com impurezas tais como o boro, a condutividade elétrica é introduzida de uma maneira controlada. A um nível de dopagem abaixo de 10^{19} B/cm^3 , o material comporta-se mais como um semicondutor e quando é dopado a níveis mais altos, comporta-se como um semimetal. Condutividades elétricas de $100\text{-}1000 \text{ S cm}^{-1}$ podem ser introduzidas dentro do material sem comprometer as suas outras propriedades únicas, fazendo ele suficientemente condutor para muitas aplicações eletroquímicas [2,5,6].

O diamante dopado possui um grande número de importantes propriedades tecnologicamente importantes para fins eletroquímicos, tais como [2,5-8]: (1) inércia química; (2) resistência à corrosão; (3) extrema dureza; (4) ampla janela de transparência ótica; (5) excelente condutividade térmica; (6) corrente de fundo voltamétrica e amperométrica baixa e estável; (7) baixa capacitância; (8) magnífica estabilidade microestrutural e morfológica mesmo sob condições eletroquímicas severas; (9) boas respostas para vários analitos redox em meio aquosos e não aquosos, sem qualquer pré-tratamento convencional; (10) alta resistência ao bloqueio eletroquímico por moléculas polares ou por potenciais contaminantes de superfície; (11) estabilidade por longo período de tempo e (12) insensibilidade ao oxigênio dissolvido.

Nenhum outro material mostra tanta versatilidade para fins eletroquímicos quanto o diamante dopado, mais precisamente o diamante dopado com boro (BDD). O material pode ser usado em eletroanálise com aumento de estabilidade, precisão e proporcionando baixos limites de detecção; em eletrólise, mesmo para altas densidades de corrente ($1\text{-}10 \text{ A cm}^{-2}$) e em ambiente com soluções agressivas,

ele pode ser empregado sem qualquer degradação morfológica ou microestrutural. Ele também pode ser empregado como eletrodo opticamente ativo (OTE) [9] para medidas espectroeletróquímica na região espectral do ultravioleta e infravermelho, explorando aqui a sua característica de ampla janela de transparência ótica do ultravioleta próximo ao infra-vermelho longínquo (0,225 – 100 μm) [2].

O BDD fornece aos cientistas e engenheiros um material que reúne as exigências para uma ampla faixa de aplicações. Este campo tem crescido significativamente nos últimos 10 anos. Se olharmos para o número de publicações em periódicos de impacto, segundo o ‘*Web of Science*’, o aumento foi de 53 em 1996 para próximo de 277 em 2006 e em junho de 2007 (data presente) já chega a 119 artigos! Outro indicativo está no número de fabricantes de BDD para fins comerciais, atualmente pelo menos 5 entidades comerciais existem em todo o mundo. As maiores são:

- Sumitomo no Japão (<http://www.sei.co.jp/>);
- Condias na Alemanha (<http://www.condias.de/>);
- Windsor na Inglaterra (<http://www.windsorscientific.co.uk/index.php>);
- CSEM na Suíça (<http://www.csem.ch/fs/water.htm>) e
- MagneTon na Holanda (<http://www.magnetochemie.com/uk/index.html>).

Infelizmente no Brasil, não há atualmente empresas dedicadas para a produção de filmes de diamante em escala comercial para fins eletroquímicos; porém, alguns poucos grupos de pesquisas produzem e estudam este material há pelo menos 15 anos, obtendo bons resultados com os diamantes produzidos aqui [10-13]. Entretanto, a transformação do conhecimento em riqueza é extremamente necessária tendo em vista a grande aplicabilidade deste material.

Deposição de Filmes Finos de Diamante

Filmes de diamante podem ser crescidos usando um de vários métodos de ativação, incluindo filamento quente, plasma de microondas ou combustão de chama assistida, ambos por CVD. O método mais comum é o plasma de microondas, visto que tais reatores são disponíveis comercialmente apesar do seu alto custo. Entretanto, o método que emprega filamento quente, o HFCVD é relativamente barato e fácil de usar [14,15], porém, ainda, não é disponível

comercialmente. Os principais objetivos dos métodos são: produzir hidrogênio atômico e ativar uma fonte carbonácea gasosa produzindo um precursor de crescimento de carbono sp^3 nas proximidades da superfície de um substrato [2,16].

Um reator CVD consiste de um compartimento onde ocorre o crescimento, associado com equipamentos que estão relacionados ao método de ativação (*e.g.* fonte de corrente para aquecer o filamento), bem como vários acessórios, tais como controladores de fluxo de massa para regular o fluxo da fonte de gás, válvula de exaustão, controladores para regular a pressão do sistema, termômetros, fonte de gás para a reação e compartimento das fontes de carbono (e dopante), caso seja líquida. A Figura 2.1 mostra um típico sistema HFCVD, para crescimento de filmes de diamante que, aliás, é o esquema do reator empregado para crescimento dos diamantes utilizados nesta tese.

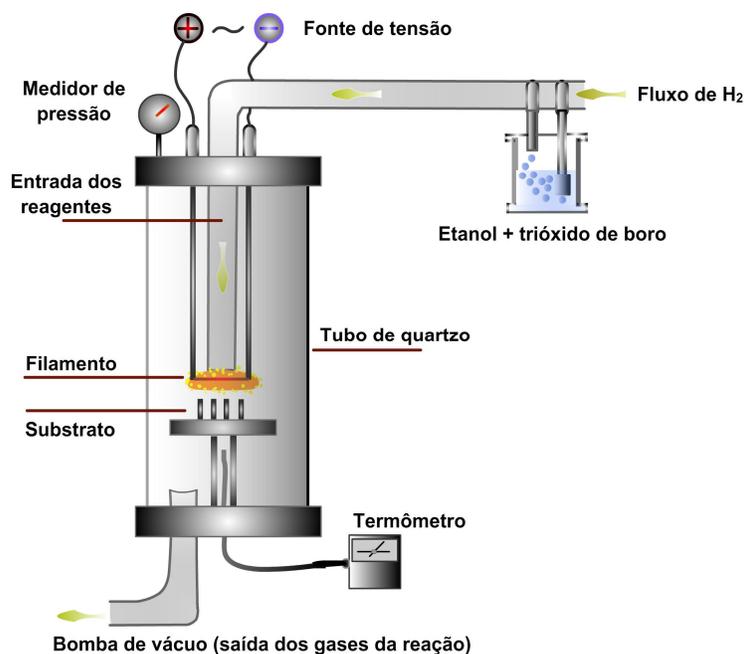


Figura 2.1. Esquema de um típico reator empregado para a deposição de diamante. Neste esquema, especificamente, o etanol é a fonte de carbono e trióxido de boro a dopante.

Os substratos usualmente empregados são Si, W, Mo, Pt e Ti. Os materiais que formam carburetos tendem a ser bons substratos para o crescimento do diamante. Além disso, eles devem possuir uma temperatura de fusão bem acima da temperatura de deposição e coeficiente de expansão térmica similar ao do

diamante. Geralmente, quando o diamante se deposita completamente sobre a superfície do substrato e a adesão é boa, as propriedades e respostas do eletrodo de diamante são independentes do material do substrato. Entretanto, o material do substrato poderá influenciar a resposta do filme (espessura $< 1\mu\text{m}$) no mínimo de três maneiras: (1) furos e pequenas fissuras entre os grãos; (2) impurezas do substrato ou das paredes do reator que podem se incorporar dentro do filme durante à deposição; (3) filmes de diamante depositados em substratos não-diamantes com significativo estresse interno.

A taxa de nucleação do diamante sobre um substrato liso e não-tratado é freqüentemente baixa. Portanto, o pré-tratamento do substrato é um pré-requisito para a deposição do diamante. Um método comum de pré-tratamento é semear a superfície do substrato ou por polimento mecânico com um pó de diamante de pequenos diâmetros ($0,01\text{-}1\mu\text{m}$) ou colocar o substrato em ultra-som na presença de pó de diamante. A aderência das partículas de diamante serve como centros de nucleação para crescimento do filme.

A fim de ter suficiente condutividade elétrica para muitas medidas eletroquímicas ($<0,1\text{ ohm-cm}$), filmes de diamante devem ser dopados com boro (dopante mais utilizado até o momento) em uma concentração de $1 \times 10^{19}\text{ cm}^{-3}$ ou maior. O boro é adicionado na fonte na forma de B_2O_3 , B_2H_6 ou $\text{B}(\text{CH}_3)_3$ a uma quantidade que pode variar de 10 a 10000 ppm. Assim, esta dopagem confere aos filmes uma condutividade elétrica e, devido à dopagem com o boro, um semicondutor tipo *p* é obtido [2,17,18].

Várias técnicas analíticas podem ser empregadas para caracterizar as propriedades morfológicas, ópticas, químicas e eletrônicas do diamante, tais como microscopia eletrônica de varredura, espectroscopia Raman, difração de raios X, espectroscopia fotoeletrônica de raios X, espectrometria de massas do íon secundário, voltametria cíclica, espectroscopia de impedância eletroquímica, entre outras [2]. A microscopia eletrônica de varredura (SEM), usada para observar a morfologia e a textura do filme, e a espectroscopia Raman, usada para a microestrutura e o estresse do filme, são as mais empregadas na caracterização *ex situ* do diamante. As técnicas eletroquímicas como voltametria cíclica (CV) e espectroscopia de impedância eletroquímica (EIS) devem ser empregadas como

caracterização *in situ* para observar a transferência eletrônica (CV e EIS), presença de carbono sp^2 (CV) e estimativa da concentração de boro no diamante (EIS) [2,6, 13,18,19].

O mecanismo do crescimento do diamante

Filmes de diamante podem ser crescidos a partir de uma fonte de carbono na presença de hidrogênio usando alguma tecnologia CVD. As fontes de carbono são tipicamente hidrocarbonetos pequenos, altamente diluídos em hidrogênio. Porém, como apresentado por Bachmann et al. (1991) [20], variações têm incluído monóxido de carbono, quantidades traço de oxigênio, gases combustíveis, halogênios e halocarbonos. Bachmann et al. (1991), após realizarem mais de 70 experimentos de deposição em diferentes reatores e usando diferentes processos gasosos, apresentaram o famoso diagrama ternário C-H-O (Figura 2.2). A partir deste diagrama foi mostrado que, independente do sistema de deposição ou mistura gasosa, o diamante deve crescer dentro de uma região saturada de hidrogênio e com concentrações diluídas de carbono na presença ou ausência de oxigênio. Isto significa que diamante pode ser crescido independente da natureza do precursor, desde que se respeite as proporções indicadas no diagrama. Ford (1995, 1991) [21,22], obteve teoricamente, a partir de modelos fenomenológicos, os resultados experimentais obtidos por Bachmann et al. (1991) [20].

Antes de difundir em direção à superfície do substrato, a mistura gasosa passa por uma região ativa, *e.g.* o filamento quente (~ 1000 °C), que fornece energia para as espécies presentes no gás se fragmentarem em radicais ativos que, por fim, atingem a superfície do substrato. Na superfície, mecanismos de adsorção, difusão, reações e dessorções de várias espécies ocorrem orientados para a nucleação de partículas de diamante, suprimindo a fase grafítica (carbono sp^2) e, enfim, o crescimento de um filme contínuo de diamante. Isto é um processo complexo com muitas variações e em que muitos mecanismos específicos ocorrem [16].

A maioria dos mecanismos apresentados na literatura é dedicada ao metano, visto que tal hidrocarboneto é a fonte de carbono mais empregada até o momento [2,16]. Além disso, a maioria desconsidera a presença de oxigênio no meio reacional. O mecanismo para outras fontes de carbono que inclui o oxigênio no

sistema ainda não foi descrito na literatura, mas a partir de estudos sobre o mecanismo do metano é possível inferir sobre o mecanismo que inclui, por exemplo, a presença do radical OH^\cdot no sistema.

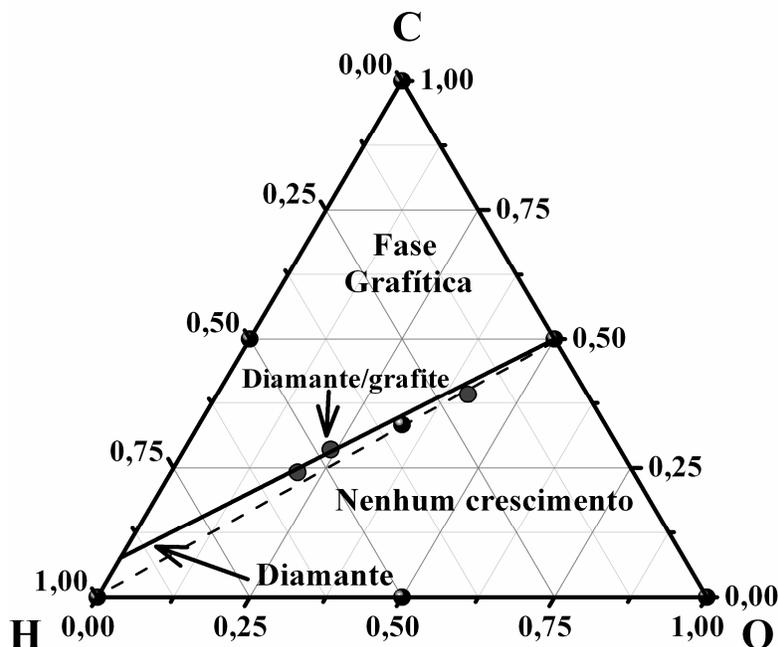


Figura 2.2. Domínio do diamante no diagrama ternário C-H-O de acordo com Bachmann (1991) [20] e Ford (1995-96) [22]

O mecanismo descrito por Butler e Woodin (1993) [16] é atualmente o mais aceito para descrever o crescimento do diamante a partir do metano diluído em hidrogênio e sobre uma superfície que contenha diamante semeado. O mecanismo é descrito da seguinte forma (Figura 2.3): (1) remoção de um hidrogênio do CH_4 pelo hidrogênio atômico para produzir CH_3^\cdot , que é o precursor primário de crescimento; (2) remoção de hidrogênios terminais da superfície do diamante semeado pelo hidrogênio atômico, gerando sítios ativos; (3) ocupação dos sítios ativos (adsorção) sobre a superfície pelos precursores primários; (4) remoção de um hidrogênio dos grupos metila adsorvidos, para produzir sítios ativos para futuras adições de precursores primários; (5) subseqüentes remoções de hidrogênio dos grupos metila acontecem e os grupos radiculares formados ou se ligam entre si, ou se ligam diretamente em sítios formados na superfície. Nesta etapa formam-se carbonos sp^3 que dão origem ao diamante. O crescimento do diamante pode ser considerado

como adições graduais de átomos de carbono, catalisado pela presença de excesso de hidrogênio atômico.

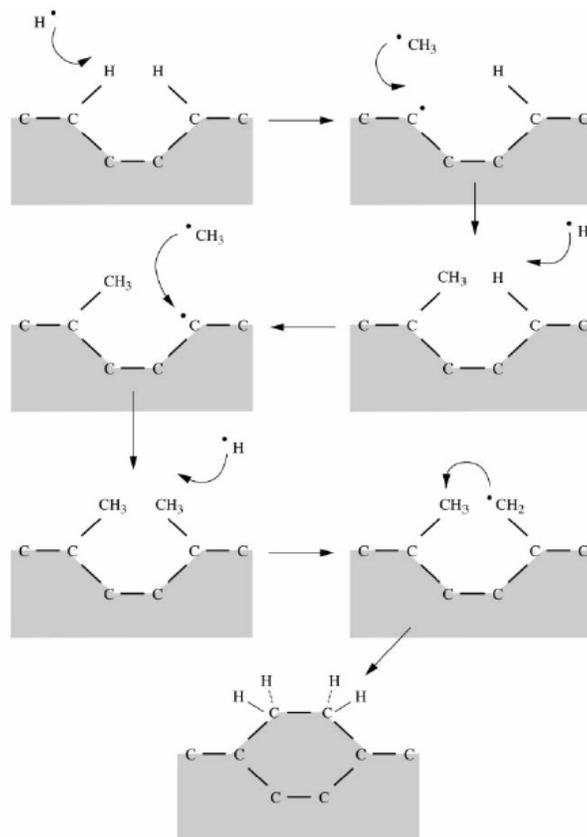
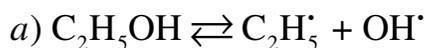


Figura 2.3. Mecanismo sugerido para o crescimento do diamante a partir de metano altamente diluído em hidrogênio, usando o processo CVD.

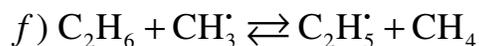
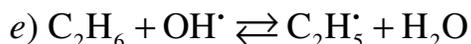
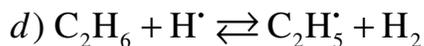
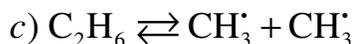
Nesta tese, o diamante foi crescido usando como fonte de carbono o etanol. O etanol tem apresentado algumas vantagens sobre o metano, como por exemplo, na obtenção de diamante com menor quantidade de fase gráfitica sobre a superfície [11,13,23-26], devido à presença do grupo OH na molécula do etanol.

Para descrever o mecanismo do crescimento do diamante a partir do etanol, primeiramente é necessário conhecer as reações desse composto na fase gasosa quando é ativado no filamento quente, por exemplo.

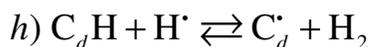
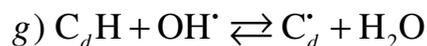
Possivelmente as primeiras reações que ocorrem são:



O trabalho de Frenklach e Wang (1991) [27] apresenta várias reações, de diversos reagentes que ocorrem neste sistema. Em relação às equações das reações anteriores, as seguintes reações são descritas:



Baseado nestas reações e no mecanismo já descrito para o metano pode-se deduzir as etapas do crescimento do diamante a partir do etanol. As reações que ocorrem na superfície do substrato são descritas a seguir:



A principal diferença entre o mecanismo do metano e a do etanol está na presença do radical OH^\cdot . Este radical, mesmo em concentração traço, possivelmente apresenta maior reatividade que o hidrogênio atômico nas mesmas condições.

Sabe-se que o hidrogênio atômico atua desativando ligações sp^2 pendentes, minimizando a reconstrução destas ligações e também, gaseificando rapidamente qualquer ligação sp^2 formada [16,27]. Um modelo formulado por Butler e Woodin (1993) [16], propõe que: (1) a fração de sítios radicais na superfície é uma função somente da temperatura e não depende do fluxo de hidrogênio atômico; (2) o crescimento de hidrocarbonetos adsorvidos para dentro da estrutura depende do fluxo de hidrogênio; (3) há uma limitação do crescimento do diamante pela geração de radicais a baixas temperaturas e (4) a qualidade é proporcional ao fluxo de hidrogênio e inversamente proporcional ao fluxo de hidrocarbonetos reativos.

Apesar de este modelo ser dedicado ao efeito do hidrogênio, possivelmente a presença do radical hidroxila no meio apresenta as mesmas características, porém com maior eficiência na remoção de ligações sp^2 . Isto justifica a menor quantidade de domínios grafíticos no diamante crescido do etanol em relação àqueles crescidos do etano e, portanto, filmes de diamante com melhor qualidade [27].

A remoção de carbonos sp^2 que se formam sobre a superfície do diamante é fundamental em eletroquímica, visto que domínios grafíticos na superfície do

diamante afetam as excelentes características do diamante, como microestrutura e morfologia estável para altas densidades de corrente em eletrólise. Além disso, em eletroanalítica, estes domínios afetam a sensibilidade, a reprodutibilidade e a estabilidade da resposta oferecida pelo carbono sp^3 [13,19,28,29].

Teófilo et al. [13] (Capítulo 2) mostraram que o uso de etanol é uma excelente alternativa para se obter diamantes dopados com boro para fins eletroquímicos diretamente após o crescimento, com quantidades extremamente baixas de grafite sobre a superfície do diamante, o que indica o papel do radical hidroxila no meio reacional.

Aplicações eletroquímicas do diamante dopado com boro

Filmes de diamante dopado com boro possuem um grande número de importantes e práticas propriedades eletroquímicas (já descritas). As áreas da eletroquímica que mais exploram as vantagens dos filmes de diamante são a eletroanálise e a eletrólise para limpeza de águas poluídas por compostos orgânicos.

Aplicações eletroanalíticas

Eletrodos de diamante oferecem vantagens sobre eletrodos convencionais usados na eletroanalítica em termos de faixa linear, limite de detecção, tempo de resposta, precisão e estabilidade da resposta [2,29], além de serem teoricamente apto a analisar metais e uma ampla variedade de compostos orgânicos e inorgânicos eletroativos. Uma grande vantagem deste material frente aos eletrodos sólidos convencionais em eletroanalítica está em poder analisar compostos que apresentam alto potencial de bloqueio eletroquímico (Capítulo 4).

Entre os diversos compostos determinados estão a do dinucleotídeo nicotinamida adenina (NADH) [30], determinações diretas e simultâneas de ácido úrico, ácido ascórbico e dopamina [31] (Capítulo 5) e de guaiacol e cloroguaiacol [32] (Capítulo 6), análise direta de ácido úrico na presença de ácido ascórbico [33] e de histamina e serotonina [34], e análise direta de citocromo c [35], e de L-cisteína em meio alcalino [36]. Além disso, tem sido usado em análises empregando cromatografia líquida de alta eficiência (HPLC) com detecção

eletroquímica (ED), tais como análise de sulfas (sulfadiazina, sulfamerazina e sulfametazina) [37], de clorofenóis em amostras ambientais [38] e de pesticidas carbamatos (carbaril, carbofuram, metil 2-benzimidazolcarbamato, bendiocarbo) [39].

Estes exemplos colocam o BDD como um material extremamente promissor em eletroanalítica e sugerem o seu uso em potencial para análise de rotina.

Aplicações em eletrólise

A outra importante tecnologia em que os eletrodos de BDD têm apresentado um significativo impacto é na área de descontaminação e purificação de água através da eletrólise. Os materiais eletródicos aliados às ferramentas eletroquímicas são extremamente atrativos para limpeza de compostos tóxicos em água, mas são frequentemente limitados pela estabilidade da microestrutura, morfologia e sensibilidade do eletrodo.

Operações bem sucedidas de qualquer tipo de estratégia de descontaminação baseada em métodos de oxidações eletroquímicas requerem: (1) um anodo dimensionalmente estável; (2) potencial de operação em ambientes agressivos; (3) uma sensibilidade constante; (4) resistência a bloqueios eletroquímicos por componentes da matriz (matéria orgânica); (5) amplo sobrepotencial para evolução do oxigênio e; (6) eficiência para suportar reações de transferência anódica de oxigênio. Os anodos de diamante dopado com boro reúnem todas estas exigências [2,40-42].

De um modo geral, a descontaminação e purificação de descartes tóxicos de compostos orgânicos e inorgânicos podem ser efetuadas por reações de eletrooxidação diretas ou indiretas, o que depende do material do eletrodo [2,41,43]. Contaminantes químicos podem ser diretamente oxidados no anodo ou indiretamente, através da eletrogeração de oxidantes (*e.g.* OH^\bullet), que subsequentemente oxidam quimicamente os contaminantes. Um mecanismo para oxidação de contaminantes químicos é através da reação com o radical OH^\bullet gerado durante o estágio inicial da evolução do oxigênio. Idealmente, compostos orgânicos e inorgânicos são completamente oxidados a CO_2 , ou no mínimo parcialmente oxidados para produtos que são menos tóxicos que os originais [2,41,43,44].

Dependendo da interação entre os radicais e a superfície do eletrodo, os anodos podem ser distinguidos como materiais *ativos* ou *não ativos* [42,43]. Nos eletrodos ativos, há uma forte interação dos radicais OH^\bullet com a superfície do eletrodo. O radical se adsorve e interage com o eletrodo com a possível transição do oxigênio do radical para o anodo formando o chamado *óxido superior*. Nos eletrodos não ativos há uma fraca interação entre os radicais OH^\bullet e a superfície do eletrodo. A oxidação dos compostos orgânicos é mediada pelos radicais, que pode resultar na completa oxidação *i.e.* a formação de CO_2 como produto final. Os anodos não ativos têm sido definidos como um eletrodo que não participa da reação anódica e não fornece sítios catalíticos ativos para adsorção de reagentes e/ou produtos em meio aquoso. Intermediários ativos, tais como os radicais OH^\bullet produzidos pela descarga da água no eletrodo não ativo são considerados os principais responsáveis pela oxidação de compostos orgânicos em meio aquoso [43].

O BDD apresenta uma característica inércia química e, devido a isto, é bem conhecida a sua excelente resistência à adsorção de compostos orgânicos em sua superfície. Isto sugere que este material é um eletrodo não ativo ideal, no qual a oxidação de compostos orgânicos e a evolução de oxigênio podem ser obtidas via a formação de radicais OH^\bullet . Além disso, anodos de BDD apresentam uma baixíssima atividade para evolução de O_2 [43].

Atualmente, diversos trabalhos têm apresentado o potencial deste material para oxidação de compostos orgânicos e inorgânicos. Alguns exemplos são: conversão direta de sulfeto para sulfato [45], tratamento de lixívias de aterro municipal [46], degradação de 17 beta-estradióis [47], oxidação do azul de metileno [48], do metil paration [49] e de diferentes tipos de compostos fenólicos (Capítulo 7) tais como 2-naftol mediado por cloreto ativo [50], misturas destes compostos, como fenol, ácido para-hidroxibenzóico, catecol e hidroquinona [51], e 4-clorofenol, 2,4-diclorofenol e 2,4,6-triclorofenol [44].

É divulgado na literatura que o BDD oferece significativas vantagens sobre outros materiais em termos de eficiência de corrente e estabilidade para uma variedade de processos eletroquímicos. Entretanto, apesar das numerosas vantagens, eles ainda não encontraram ampla aplicação industrial, principalmente

devido às dificuldades em encontrar um substrato apropriado para realizar o depósito dos filmes de diamante [41].

Referências

1. Baranauskas, V. O Futuro brilhante dos diamantes. *Ciência Hoje*, 40-46.
2. Swain, G. Electrically conducting diamond thin films: advanced electrode materials for electrochemical technologies. In *Electroanalytical chemistry: a series of advances*, Bard, A. J. and Rubinstein, I., Eds. Marcel Dekker Inc: New York, 2004, Vol. 22, pp 181-277.
3. Hazen, R. M. *The diamond makers*. Cambridge Univ Press: Cambridge, 1999; p 258.
4. Pleskov, Y. V.; Sakharova, A. Y.; Krotova, M. D.; Bouilov, L. L.; Spitsyn, B. V. Photoelectrochemical properties of semiconductor diamond. *J. Electroanal. Chem.* **1987**, 228, (1-2), 19-27.
5. Compton, R. G.; Foord, J. S.; Marken, F. Electroanalysis at diamond-like and doped-diamond electrodes. *Electroanalysis* **2003**, 15, (17), 1349-1363.
6. Xu, J. S.; Granger, M. C.; Chen, Q. Y.; Strojek, J. W.; Lister, T. E.; Swain, G. M. Boron-doped diamond thin-film electrodes. *Anal. Chem.* **1997**, 69, (19), A591-A597.
7. Pleskov, Y. V. Synthetic diamond in electrochemistry. *Uspekhi Khimii* **1999**, 68, (5), 416-429.
8. Rao, T. N.; Fujishima, A. Recent advances in electrochemistry of diamond. *Diam. Relat. Mat.* **2000**, 9, (3-6), 384-389.
9. Stotter, J.; Show, Y.; Wang, S. H.; Swain, G. Comparison of the electrical, optical, and electrochemical properties of diamond and indium tin oxide thin-film electrodes. *Chem. Mat.* **2005**, 17, (19), 4880-4888.
10. Baranauskas, V.; Fukui, M.; Rodrigues, C. R.; Parizotto, N.; Trava-Airoldi, V. J. Direct observation of chemical vapor-deposited diamond films by atomic force microscopy. *Appl. Phys. Lett.* **1992**, 60, (13), 1567-1589.
11. Li, B. B.; Baranauskas, V.; Peterlevitz, A.; Chang, D. C.; Doi, I.; Trava-Airoldi, V. J.; Corat, E. J. Annealing-induced enhancement in the activation energy of heavily boron-doped polycrystalline diamond. *Diam. Relat. Mat.* **1998**, 7, (9), 1259-1262.
12. Ferreira, N. G.; Silva, L.; Corat, E. J.; Trava-Airoldi, V. J.; Iha, K. Electrochemical characterization on semiconductors p-type CVD diamond electrodes. *Braz. J. Phys.* **1999**, 29, (4), 760-763.
13. Teófilo, R. F.; Ceragioli, H. J.; Peterlevitz, A. C.; Da Silva, L. M.; Damos, F. S.; Ferreira, M. M. C.; Baranauskas, V.; Kubota, L. T. Improvement of the

- electrochemical properties of "as-grown" boron-doped polycrystalline diamond electrodes deposited on tungsten wires using ethanol. *J. Solid State Electrochem.* **2007**, 11, 1449-1457.
14. Trava-Airoldi, V. J.; Corat, E. J.; Baranauskas, V. Diamond Chemical Vapor Deposition: Emerging technology for tooling applications. **1998**, 138-1, 195-244.
 15. Pastor-Moreno, G. Electrochemical applications of CVD Diamond. PhD thesis, Bristol University, Bristol, 2002.
 16. Butler, J. E.; Woodin, R. L. Thin-film diamond growth mechanisms. *Philos. Trans. R. Soc. Lond. Ser. A-Math. Phys. Eng. Sci.* **1993**, 342, (1664), 209-224.
 17. Angus, J. C.; Pleskov, Y. V.; Eaton, S. C. Electrochemistry of diamond. **2004**, 77, 97-119.
 18. Pleskov, Y. V. Electrochemistry of diamond: A review. *Russ. J. Electrochem.* **2002**, 38, (12), 1275-1291.
 19. Granger, M. C.; Witek, M.; Xu, J. S.; Wang, J.; Hupert, M.; Hanks, A.; Koppang, M. D.; Butler, J. E.; Lucazeau, G.; Mermoux, M.; Strojek, J. W.; Swain, G. M. Standard electrochemical behavior of high-quality, boron-doped polycrystalline diamond thin-film electrodes. *Anal. Chem.* **2000**, 72, (16), 3793-3804.
 20. Bachmann, P. K.; D., L.; Lydtin, H. Towards a general concept of diamond chemical vapour deposition. *Diam. Relat. Mat.* **1991**, 1, 12.
 21. Ford, I. J. Model of the competitive growth of amorphous-carbon and diamond films. *J. Appl. Phys.* **1995**, 78, (1), 510-513.
 22. Ford, I. J. Boundaries of the diamond domain in the C-H-O diagram of carbon film deposition. *J. Phys. D-Appl. Phys.* **1996**, 29, (9), 2229-2234.
 23. Farabaugh, E. N.; Robins, L.; Feldman, A.; Johnson, C. E. Growth and oxidation of boron-doped diamond films. *J. Mater. Res.* **1995**, 10, (6), 1448-1454.
 24. Baranauskas, V.; Ceragioli, H. J.; Peterlevitz, A. C.; Tosin, M. C.; Durrant, S. F. Effects of argon dilution of an ethanol/hydrogen gas feed on the growth of diamond by hot-filament chemical vapor deposition. *Thin Solid Films* **2000**, 377, 303-308.
 25. Baranauskas, V.; Ceragioli, H. J.; Peterlevitz, A. C. Growth of self-supported diamond tubes in the hot-filament chemical vapor deposition system. *Diam. Relat. Mat.* **2003**, 12, (3-7), 346-349.
 26. Baranauskas, V.; Peterlevitz, A. C.; Ceragioli, H. J.; Kubota, L. T.; Teófilo, R. F. Processo de fabricação de eletrodos de diamante dopado para eletroquímica. Patente depositada no INPI, PI0600897-6, 03/16/2006.
 27. Frenklach, M.; Wang, H. Detailed surface and gas-phase chemical-kinetics of diamond deposition. *Phys. Rev. B* **1991**, 43, (2), 1520-1545.

28. Mahe, E.; Devilliers, D.; Comninellis, C. Electrochemical reactivity at graphitic micro-domains on polycrystalline boron doped diamond thin-films electrodes. *Electrochim. Acta* **2005**, 50, (11), 2263-2277.
29. Granger, M. C.; Xu, J. S.; Strojek, J. W.; Swain, G. M. Polycrystalline diamond electrodes: basic properties and applications as amperometric detectors in flow injection analysis and liquid chromatography. *Anal. Chim. Acta* **1999**, 397, (1-3), 145-161.
30. Rao, T. N.; Yagi, I.; Miwa, T.; Tryk, D. A.; Fujishima, A. Electrochemical oxidation of NADH at highly boron-doped diamond electrodes. *Anal. Chem.* **1999**, 71, (13), 2506-2511.
31. Teófilo, R. F.; Ceragioli, H. J.; Peterlevitz, A. C.; Baranauskas, V.; Ferreira, M. M. C.; Kubota, L. T. In *Quantificação simultânea de dopamina, ácido úrico e ácido ascórbico com eletrodo de diamante dopado com boro usando voltametria e pls*, XVI Simpósio Brasileiro de Eletroquímica e Eletroanalítica, Águas de Lindóia, Brazil, Águas de Lindóia, Brazil, 2007, p ^pp 76.
32. Teófilo, R. F.; Ceragioli, H. J.; Peterlevitz, A. C.; Baranauskas, V.; Ferreira, M. M. C.; Kubota, L. T. In *Simultaneous determination of guaiacol and chloroguaiacol by SWV using boron-doped diamond electrode and PLS algorithms*, 10th International Conference on Chemometrics in Analytical Chemistry, Águas de Lindóia, Brazil, Águas de Lindóia, Brazil, 2006, p ^pp P009.
33. Popa, E.; Kubota, Y.; Tryk, D. A.; Fujishima, A. Selective voltammetric and amperometric detection of uric acid with oxidized diamond film electrodes. *Anal. Chem.* **2000**, 72, (7), 1724-1727.
34. Sarada, B. V.; Rao, T. N.; Tryk, D. A.; Fujishima, A. Electrochemical oxidation of histamine and serotonin at highly boron doped diamond electrodes. *Anal. Chem.* **2000**, 72, (7), 1632-1638.
35. Marken, F.; Paddon, C. A.; Asogan, D. Direct cytochrome c electrochemistry at boron-doped diamond electrodes. *Electrochem. Commun.* **2002**, 4, (1), 62-66.
36. Spataru, N.; Sarada, B. V.; Popa, E.; Tryk, D. A.; Fujishima, A. Voltammetric determination of L-cysteine at conductive diamond electrodes. *Anal. Chem.* **2001**, 73, (3), 514-519.
37. Rao, T. N.; Sarada, B. V.; Tryk, D. A.; Fujishima, A. Electroanalytical study of sulfa drugs at diamond electrodes and their determination by HPLC with amperometric detection. *J. Electroanal. Chem.* **2000**, 491, (1-2), 175-181.
38. Terashima, C.; Rao, T. N.; Sarada, B. V.; Tryk, D. A.; Fujishima, A. Electrochemical oxidation of chlorophenols at a boron-doped diamond electrode and their determination by high-performance liquid chromatography with amperometric detection. *Anal. Chem.* **2002**, 74, (4), 895-902.

39. Rao, T. N.; Loo, B. H.; Sarada, B. V.; Terashima, C.; Fujishima, A. Electrochemical detection of carbamate pesticides at conductive diamond electrodes. *Anal. Chem.* **2002**, 74, (7), 1578-1583.
40. Chen, G. H. Electrochemical technologies in wastewater treatment. *Sep. Purif. Technol.* **2004**, 38, (1), 11-41.
41. Panizza, M.; Cerisola, G. Application of diamond electrodes to electrochemical processes. *Electrochim. Acta* **2005**, 51, (2), 191-199.
42. Alfaro, M.; Ferro, S.; Martinez-Huitle, C. A.; Vong, Y. M. Boron doped diamond electrode for the wastewater treatment. *J. Braz. Chem. Soc.* **2006**, 17, (2), 227-236.
43. Marselli, B.; Garcia-Gomez, J.; Michaud, P. A.; Rodrigo, M. A.; Comninellis, C. Electrogeneration of hydroxyl radicals on boron-doped diamond electrodes. *J. Electrochem. Soc.* **2003**, 150, (3), D79-D83.
44. Canizares, P.; Garcia-Gomez, J.; Saez, C.; Rodrigo, M. A. Electrochemical oxidation of several chlorophenols on diamond electrodes - Part I. Reaction mechanism. *J. Appl. Electrochem.* **2003**, 33, (10), 917-927.
45. Waterston, K.; Bejan, D.; Bunce, N. J. Electrochemical oxidation of sulfide ion at a boron-doped diamond anode. *J. Appl. Electrochem.* **2007**, 37, (3), 367-373.
46. Cabeza, A.; Urtiaga, A. M.; Ortiz, I. Electrochemical treatment of landfill leachates using a boron-doped diamond anode. *Ind. Eng. Chem. Res.* **2007**, 46, (5), 1439-1446.
47. Murugananthan, M.; Yoshihara, S.; Rakuma, T.; Uehara, N.; Shirakashi, T. Electrochemical degradation of 17 beta-estradiol (E2) at boron-doped diamond (Si/BDD) thin film electrode. *Electrochim. Acta* **2007**, 52, (9), 3242-3249.
48. Panizza, M.; Barbucci, A.; Ricotti, R.; Cerisola, G. Electrochemical degradation of methylene blue. *Sep. Purif. Technol.* **2007**, 54, (3), 382-387.
49. Diagne, M.; Oturan, N.; Oturan, M. A. Removal of methyl parathion from water by electrochemically generated Fenton's reagent. *Chemosphere* **2007**, 66, (5), 841-848.
50. Panizza, M.; Cerisola, G. Electrochemical oxidation of 2-naphthol with in situ electrogenerated active chlorine. *Electrochim. Acta* **2003**, 48, (11), 1515-1519.
51. Polcaro, A. M.; Mascia, M.; Palmas, S.; Vacca, A. Electrochemical oxidation of phenolic and other organic compounds at boron doped diamond electrodes for wastewater treatment: Effect of mass transfer. *Ann. Chim.* **2003**, 93, (12), 967-976.

Parte Experimental

*“Palavras que não se traduzem em fatos
não têm importância”
Che Guevara*

Capítulo 3

Propriedades dos Diamantes Dopados com Boro Crescidos do Etanol

“Você é realmente quem você diz que é?”

Antes de realizar qualquer aplicação, filmes de diamante dopado com boro (BDD) foram crescidos e caracterizados quanto à autenticidade e quanto as suas propriedades eletroquímicas. Neste capítulo será apresentada a caracterização e as propriedades eletroquímicas de dois filmes policristalinos de BDD crescidos sobre fio de tungstênio usando etanol como um precursor. Os resultados obtidos mostram que o uso de etanol melhora as propriedades eletroquímicas do BDD que são usados logo após o crescimento para medidas eletroquímicas, *i.e.*, sem nenhum tratamento químico posterior antes do uso. Isto acontece, pois durante o processo de crescimento, o etanol minimiza a fase gráfitica sobre a superfície do BDD. Os eletrodos de BDD foram caracterizados por espectroscopia Raman, microscopia eletrônica de varredura (SEM), voltametria cíclica (CV) e espectroscopia de impedância eletroquímica (EIS). Uma estimativa do nível de dopagem com boro dos filmes indicou um valor de $\sim 10^{20}$ B/cm³. O comportamento eletroquímico foi avaliado usando os pares redox $\text{Fe}(\text{CN})_6^{3-/4-}$ e $\text{Ru}(\text{NH}_3)_6^{3+/2+}$ e também a dopamina. As constantes de transferência eletrônica aparente, k_{app}^0 foram determinadas para estes sistemas redox usando as técnicas CV e EIS. Valores de k_{app}^0 na faixa de 0,01-0,1 cm s⁻¹ foram observados para os pares redox $\text{Fe}(\text{CN})_6^{3-/4-}$ e $\text{Ru}(\text{NH}_3)_6^{3+/2+}$, enquanto no caso especial da dopamina, um baixo valor de k_{app}^0 de 10⁻⁵ cm s⁻¹ foi encontrado. Os resultados obtidos mostram que o uso de CH₃CH₂OH (etanol) como fonte de carbono constitui uma alternativa promissora para a fabricação de eletrodo de BDD para aplicações eletroquímicas.

O estudo completo será apresentado em idioma inglês conforme artigo 2º da informação CCPG/002/06 da Unicamp, que permite a inclusão de artigos científicos publicados (ver abaixo) ou submetidos para publicação como capítulo ou apêndice da tese.

J Solid State Electrochem (2007) 11:1449–1457
DOI 10.1007/s10008-007-0319-z

ORIGINAL PAPER

Improvement of the electrochemical properties of “as-grown” boron-doped polycrystalline diamond electrodes deposited on tungsten wires using ethanol

Reinaldo F. Teófilo • Helder J. Ceragioli •
Alfredo C. Peterlevitz • Leonardo M. Da Silva •
Flavio S. Damos • Márcia M. C. Ferreira •
Vitor Baranauskas • Lauro T. Kubota

Received: 5 March 2007 / Accepted: 26 March 2007 / Published online: 15 May 2007
© Springer-Verlag 2007

Introduction

The development of chemical and biological sensors of high sensitivity and stability has attracted great attention to electroanalytical applications of boron-doped polycrystalline diamond electrodes. Boron-doped diamond (BDD) has demonstrated to be an attractive material for electroanalytical applications, in view of its unique properties such as robustness, chemical inertness, resistance to electrode fouling, wide potential window, and low background currents [1,2,3].

However, several factors influence the electrical conductivity within BDD films and the charge transfer rates across the BDD/solution interface [4]. As a result, the electrochemical properties are influenced mainly by: (1) doping type and level, (2) morphology features (*e.g.*, grain boundaries and point defects), (3) nondiamond impurity content, (4) crystallographic orientation, and surface group functionalities (H vs O). The extent of each one of these factors affecting the

electrochemical response depends significantly on the reaction mechanism for the particular redox system [4-7].

The carbon source can be an important variable in the characteristics of BDD of electroanalytical interest, although methane is the most often employed carbon source [1-8].

Ford (1996). [9] showed, using a phenomenological model, how the growth of BDD films from a C–H–O mixture can lead to the formation of high-quality materials if the reagent concentration was optimized. According to this author [9], the presence of oxygen in the gas phase ties up a proportion of the available carbon in the form of CO, which reduces the flux of reactive carbonaceous groups to the surface. Experiments showed that using a microwave-plasma enhanced chemical vapor deposition (CVD) reactor and small amounts of oxygen in the methane and hydrogen mixture may increase the rate of the film growth but also the impurity content [10,11]. On the other hand, Baranauskas et al. [12,13] have obtained a high-quality diamond deposited on silicon or tungsten with ethanol and hydrogen using hot filament-assisted CVD (HFCVD), with a low quantity of reactive carbonaceous groups. Indeed, several different reports [14,15] have pointed out that ethanol as a carbon source to grow diamonds comprises a good alternative in obtaining high-quality diamond films. However, only a few papers have been devoted to obtain BDD films from alcohol (ethanol or methanol) [16-18].

The diamond growth using different carbon sources apart from methane is seldom exploited in electrochemical researches, probably because of the difficulty in obtaining a material that presents the required qualities, because not always a conducting diamond originated from the C–H–O mixture supplies all the characteristics required for the electrochemical applications. Ethanol was chosen because of its characteristic in decreasing the graphitic phase during growth. This characteristic is extremely important in electrochemistry because film growth from other sources require graphite removing by different drastic methods, chemical [4,8] or electrochemical [8], after growing.

The presence of nondiamond carbon, mostly constituted by graphitic domains, could play the role of electrocatalytic centers and bring functional groups on the electrode surface [8]. However, those Csp^2 impurities tend to be formed in the grain

boundaries and defects during the CVD process [7]. The magnitude of the background current, working potential window, and the voltammetric features are all sensitive to the presence of nondiamond (amorphous or graphitic) carbon impurities [1,3,4]. In a very general way, adsorbed contaminants can either block specific surface sites or increase the electron-tunneling distance for redox analytes [8].

The objective of this study is the investigation of the electrochemical properties of “as-grown” BDD films prepared using ethanol, applying different in situ (cyclic voltammetry [CV] and electrochemical impedance spectroscopy [EIS]) and ex situ (scanning electronic microscopy [SEM] and Raman) techniques, showing that the graphite should be responsible for the decrease in the electroactivity of the BDD electrode.

Experimental

Diamond deposition

Polycrystalline BDD films were produced in a previously described all-quartz cylindrical HFCVD reactor [19]. A boron-doping source was performed dissolving solid B_2O_3 in ethanol in the liquid reservoir before starting up the reactor. Tungsten wires with a diameter of 238 μm and length of 30 mm were used as substrates. The wires were mounted onto a polished silicon wafer placed below the hot filament and parallel to the central axis of the filament coil. Deposition temperatures were measured by a thermocouple underside the silicon wafer. Before deposition, the tungsten wires were dipped in a colloidal mixture of diamond dust (0.25 μm in diameter) dispersed by ultrasonic vibration in n-hexane. This ‘seeding’ is a used procedure to increase the nucleation density of diamond films. The diamond films were deposited employing a mixture of ethanol ($\text{CH}_3\text{CH}_2\text{OH}$) vapor diluted in hydrogen gas (99.5% vol) at a volumetric flow rate of 100 standard cm^3 per min, regulated by precision mass flow meters, and a total pressure kept at 20 Torr. We classified our samples in two typical types, A and B, according to their respective deposition temperatures: Type A were grown at deposition temperatures set at 1,050 K for 30 h, and type B were grown at 910 K for 35 h. The boron doping

concentrations, estimated using Mott–Schottky plots, were $\sim 2.9 \times 10^{20}$ and $\sim 4.0 \times 10^{18}$ B cm⁻³ (5,000 ppm or 0.5% B/C) for type A and B, respectively. Electric connections were placed on both types by wire wrapping a nickel–chromium wire and sealing with silver paint. The electrodes were insulated by a Teflon band and epoxy resin and then inserted into polypropylene tubes, exposing only the diamond area on the top of the electrodes for the electrochemical studies.

BDD electrode pretreatments

BDD films “as grown” received a mild cathodic treatment in 1.0 mol L⁻¹ HNO₃. The cathodic treatment, a modified treatment presented by Marken et al. [20], was carried out by cycling the potential in the range from 0 to -4 V (vs a saturated calomel electrode [SCE]) for at least three times.

We have observed that this pretreatment is also effective for recovery of diamond electrodes after extensive times of analysis.

BDD film characterizations

The BDD films were characterized using different in situ and ex situ techniques. For the ex situ characterization, the scanning electron microscopy (SEM) was used to examine the film morphology using a JEOL 6360–LV instrument. The Raman spectroscopy technique was employed to investigate the structural properties of the BDD films at ambient temperature (~ 25 °C) using a Jobin Yvon T64000 Raman spectrometer, employing an Argon laser for excitation (514.5 nm wavelength) at a laser power of about 6 mW.

Electrochemical studies: The samples were characterized using the CV and the EIS techniques.

CV study: Voltammograms were obtained using 5 ml of the supporting electrolyte containing different electroactive compounds (Fe(CN)₆^{3-/4-}, Ru(NH₃)₆^{3+/2+} and dopamine) with a concentration of 1.0×10^{-3} mol L⁻¹. CV was performed at different scan rates, between 0.02 and 0.1 V s⁻¹. In all cases, the potential range was cycled for three times, and the last cycle was recorded. Pure N₂ was bubbled through the sample solutions for 5 min before each measurement.

Quasi-steady polarization curves: In the special case of dopamine, the Quasi-steady polarisation curves (QSPC) were carried out in a range of 200 mV in relation to the open circuit potential (equilibrium potential). The experimental data were corrected for ohmic drop. The supporting electrolyte was 1 mol L⁻¹ KCl solution, and the dopamine concentration was 1×10⁻³ mol L⁻¹.

EIS study In all cases, the electrochemical investigation in the frequency domain was carried out by applying a d.c. potential at the open circuit potential. Impedance spectra were recorded covering the 10-mHz to 100-kHz frequency interval. To ensure the linearity of the electrode response, a small amplitude sine wave (5 mV [p/p]) was used throughout. As required by the linear system theory, validation of the EIS data was performed applying the Kramers–Kronig test using a software acquired from the AUTOLAB electrochemical system. The equivalent circuit, EC, proposed in the literature [2,7] for the BDD electrode in the presence of diffusion control, was applied to fit the impedance data using the EQUIVCRT program elaborated by Boukamp [21]. Pure N₂ was bubbled through the sample solutions for 5 min before each measurement.

The electrochemical studies were carried out using an all-glass electrochemical cell (10 cm³). A platinum wire was used as a counter electrode, and all potentials were recorded against a SCE. All measurements were carried out at 25 °C. An AUTOLABTM (Eco Chemie, The Netherlands) electrochemical system (GPES), PGSTAT-30 model, was used throughout.

Solutions

In both cases, the electrochemical behavior associated with the electron-transfer reaction at the BDD/solution interface was investigated using aqueous solutions containing one of the Fe(CN)₆^{3-/4-} (Aldrich) or Ru(NH₃)₆^{3+/2+} (Aldrich) redox couples or dopamine (Acros). The supporting electrolytes in each case were composed of 1 mol L⁻¹ KCl (Synth), 1 mol L⁻¹ HNO₃ (Merck), or 0.1 mol L⁻¹ H₂SO₄ (Merk) solutions.

All reagents were analytical grade without additional purification. All solutions were prepared with ultra pure water from a Milli-QTM purification system (>18 MΩ cm).

Results and Discussion

Raman spectroscopy of Diamond films synthesized by CVD with methane and other carbon sources, commonly, contain graphitic carbon [4,7]. Raman spectroscopy is one of the best tools to assess the qualities of BDD films because graphitic carbon has a distinctly different Raman peak that is more intense than the diamond peak, thus allowing a small amount of graphitic carbon to be recognized in BDD films. Figure 3.1 shows the typical Raman spectra of the “as-grown” samples. There is a single sharp peak at $1,330\text{ cm}^{-1}$ that corresponds to the C–C sp^3 bonds. The presence of a broader band around $1,550\text{ cm}^{-1}$ of very low intensity can be attributed to the presence of graphitic carbon impurities (sp^2 carbon) contamination. The ratio of graphite to the diamond scattering intensities ($I_{1,590}/I_{1,330}$) counts is very low (<0.05), and it is in agreement with the literature describing high-quality BDD films [22], thus revealing that only a very small amount of impurity is present after the BDD preparation via HFCVD using ethanol as source of carbon. This result shows that ethanol can significantly decrease the graphitic carbon impurities in situ, differently of the other sources.

SEM analysis

Figure 3.2 shows the typical SEM images of polycrystalline BDD films of types A (Figure 3.2a) and B (Figure 3.2b), respectively. Grains of an average size of $36\text{-}21\text{ }\mu\text{m}$ for type A (Figure 3.2a) and $8\text{-}4.5\text{ }\mu\text{m}$ for type B (Figure 3.2b) are clearly observed. The SEM images show that the tungsten-supported polycrystalline BDD films are mainly characterized by a uniform distribution of homogenous grains. Each grain presents a well-defined, randomly oriented appearance. In addition, the diamond films are without noticeable cracks, thus ensuring the necessary protection of the tungsten support for electrochemical applications. As discussed by Pleskov [2], BDD films not only withstand corrosion, but they protect the metal substrate against chemical attack.

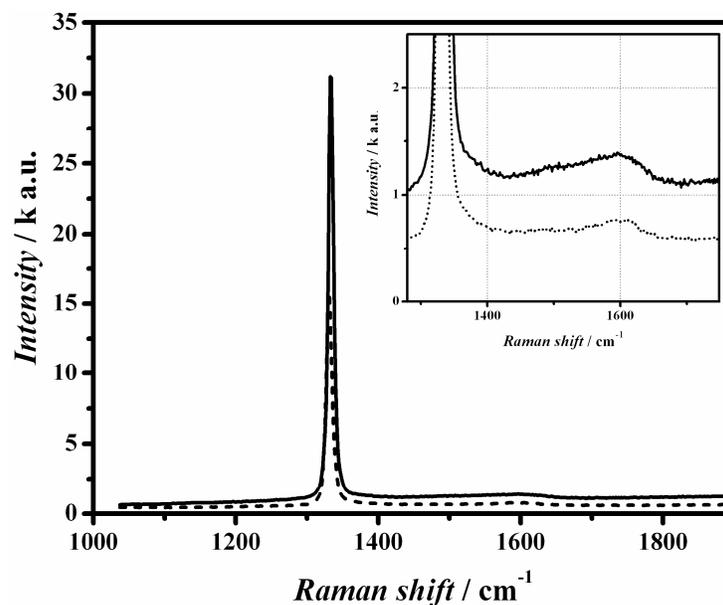


Figure. 3.1 Typical Raman spectra of “as-grown” boron-doped diamond: solid line, type A; dashed line, type B

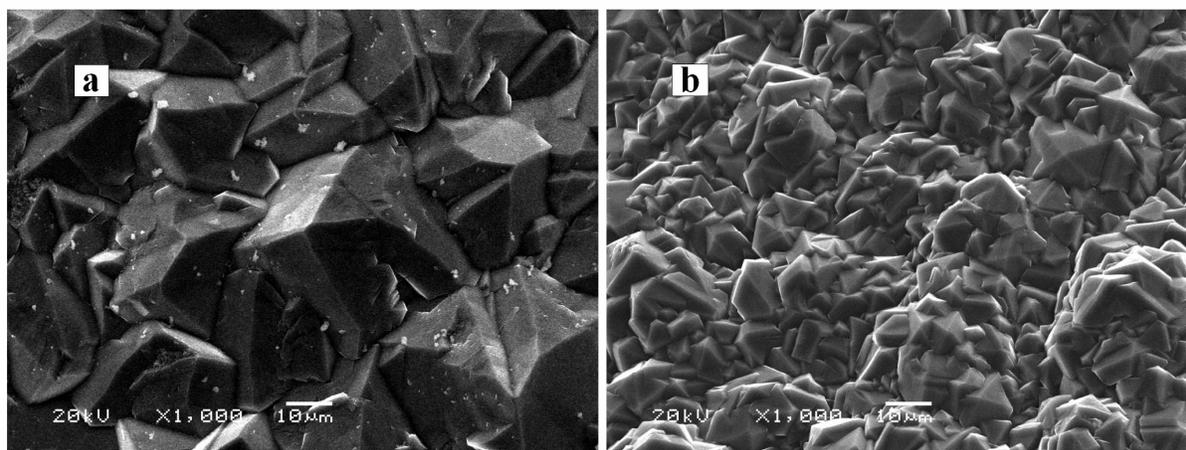


Figure. 3.2 SEM images of polycrystalline BDD films: **a** type A; **b** type B

Electrochemical characterization

Cathodic treatments were carried out to improve the charge transfer of the “as-grown” diamonds. The cyclic voltammograms are presented in Figure. 3.3. It can be observed that a significant increase in the electron transfer is obtained for two samples.

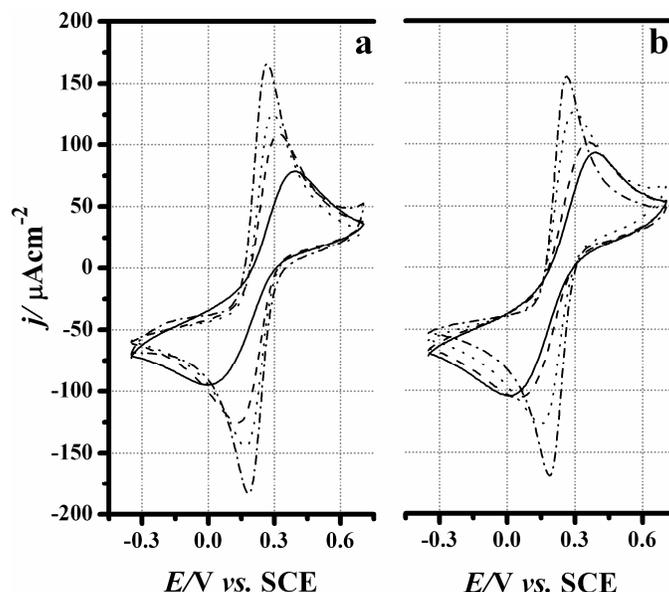


Figure. 3.3 Cyclic voltammograms at 100 mV s^{-1} for $1 \times 10^{-3} \text{ mol L}^{-1}$ ferrocyanide in 0.1 mol L^{-1} KCl, solid line, “as-grown”; cathodic treatments: dashed line, 1, dotted line, 2, dash-dot line, 3, respectively, in electrodes type A (a) and type B (b)

Recently, it was found that the carrier density near to the BDD electrode surface may be controlled by changing surface terminations [23]. The hydrogen-terminated diamond shows p-type semiconductivity with holes strongly limited to the surface region [24]. Hydrogenated samples exhibit a p-type conductivity, with an activation energy one order of magnitude lower than that of boron. Several models have been proposed for the source of conductivity. One suggestion is that hydrogen is directly responsible for the acceptor level [25]. A more likely model involves the transfer of electrons from the diamond to surface adsorbates [26], but the experimental evidence is far from clear. As a final point, it has been suggested that hydrogen is added by self-traps implantation, implying the simultaneous formation of an immobile hydrogen aggregate or a lattice defect [27]. We suppose that during the cathodic treatment, the generated hydrogen species are trapped on a boron acceptor on the surface of the diamond, in this manner there is the hole generation process, and thus the conductivity can be correlated by the presence of a hydrogenated surface. Different works have shown that BDD with a hydrogen terminated surface is, generally, the most active and gives the most reproducible response [1,4,7,8]. In this study, diamonds with surfaces with very low quantities of

nondiamond carbon phases and hydrogen terminated were electrochemically characterized.

Background cyclic voltammograms can be very informative about diamond quality and electrical conductivity [28]. The presence of nondiamond carbon impurity at the surface can be detected, with higher sensitivity than Raman, from background cycling in a strong acid media, especially with the effect on the width of the potential domain where the electrode is ideally polarizable. The background for types A and B electrodes are featureless, low, and ideally polarizable in the respective potential ranges and electrolytes: (1) -650 - $1,780$ and -950 - $2,025$ mV in $0.1 \text{ mol L}^{-1} \text{ H}_2\text{SO}_4$ solution and (2) -820 - $2,080$ and -880 - $2,110$ mV in $1 \text{ mol L}^{-1} \text{ HNO}_3$ solution (see Figure 3.4).

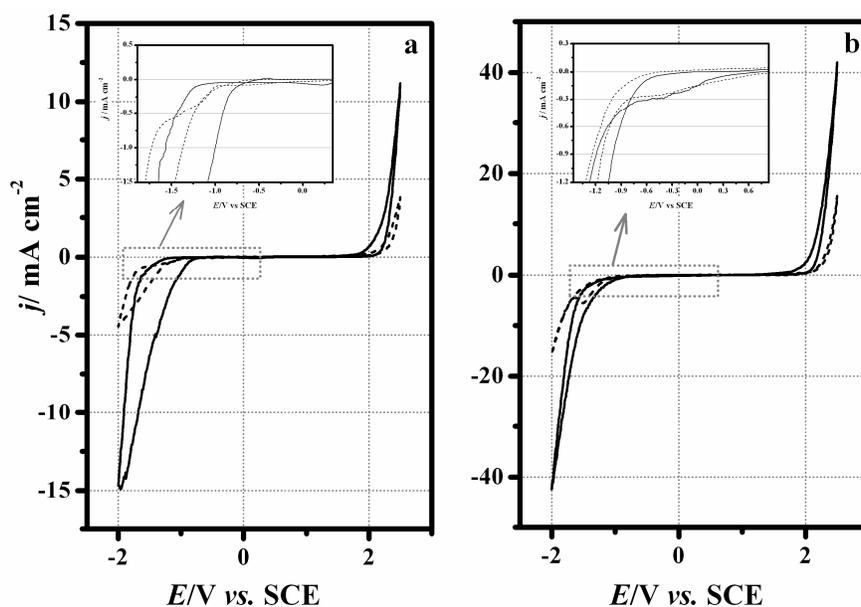


Figure. 3.4 Cyclic voltammograms as functions of the electrolyte composition. a $0.1 \text{ mol L}^{-1} \text{ H}_2\text{SO}_4$; b $1 \text{ mol L}^{-1} \text{ HNO}_3$ for electrodes type A (solid line) and type B (dashed line). Scan rate: 100 mV s^{-1}

The exposed geometric areas were 0.030 and 0.036 cm^{-2} for types A and B electrodes, respectively, and all currents were normalized to these areas. The working potential window obtained from the CV profiles are 3.28 and 3.5 V ($\pm 300 \mu\text{A cm}^{-2}$) in H_2SO_4 and 4.27 and 3.83 V ($\pm 9.0 \text{ A cm}^{-2}$) in HNO_3 for types A and B electrodes, respectively. The potential window can be related, in part, to the lower fraction of exposed grain boundaries [1,4]. The electrodes of type B showed

to have a larger potential window, presenting a lower fraction of exposed grain boundaries, which is in agreement with the literature [1,4]. The voltammograms also indicate that no significant level of nondiamond carbon impurities is present.

Cyclic voltammetry was also used to study the electrode response for different redox couples ($\text{Fe}(\text{CN})_6^{3-/4-}$, $\text{Ru}(\text{NH}_3)_6^{3+/2+}$) and for dopamine. These redox systems were chosen because several works with sp^2 carbon (*e.g.*, graphite, glassy carbon, etc.) and doped sp^3 carbon (*e.g.*, diamond) have used them [1,4,7,28,29,30]. In this study, the heterogeneous electron-transfer reactions were considered because the extent the physicochemical properties of sp^2 and sp^3 carbon electrodes influence the response depends on the mechanistic aspects for the particular redox analyte [1,4,8,31].

The ferri/ferrocyanide system proceeds through a more inner-sphere electron transfer pathway, and the electrode kinetics are highly sensitive to the surface termination of diamond [1,4,6,7,32-34]. Presumably, the reaction proceeds through a specific surface site that is blocked by oxygen. The surface chemical effect responsible for the low kinetics of the $\text{Fe}(\text{CN})_6^{3-/4-}$ redox reaction is directly related to the carbon–oxygen functionalities because of the presence of graphitic microdomains (carbon sp^2) or the activation of these microdomains during the anodic polarization [1,4,6,7]. Granger and Swain [6] showed that cyclic voltammetric curves are much more reversible after acid washing followed by surface rehydrogenation. After these treatment, the removal of the carbon sp^2 is expected [1,4,[6]. The reactivity of the BDD electrode to the ferri/ferrocyanide system is also re-established because such reactivity is strongly associated with p-type surface conductivity, in which the hole generation process is correlated by the presence of a hydrogenated surface [1,4,6,7].

Mahé et al. [7] showed that a drastic electrochemical polarization can clean the graphitic microdomains and accomplish the rehydrogenation of the surface, contributing to an increase in the kinetics of the redox reaction of $\text{Fe}(\text{CN})_6^{3-/4-}$. Thus, some authors have been reactivating the surface through cathodic treatments [7]. In this work, a significant improvement in the electron transfer for the ferri/ferrocyanide system after a mild cathodic treatment was also verified (see

Figure 3.3). In the literature [6], the values of the heterogeneous electron-transfer rate constant k^0 vary from 10^{-2} to 10^{-1} cm s⁻¹ for the case of a BDD electrode.

The reduction of $\text{Ru}(\text{NH}_3)_6^{3+/2+}$ is also commonly used to characterize the quality and properties of a diamond film and allows the comparison with data obtained in electrochemical experiments with other types of diamond and carbon electrode materials [35]. The redox system $\text{Ru}(\text{NH}_3)_6^{3+/2+}$ proceeds by outer-sphere electron transfer and the electrode kinetic is relatively insensitive to the surface microstructure, surface oxides, and adsorbed monolayers on sp^2 carbon electrodes [1,5,36]. The most important factor affecting the reaction rate is the electronic properties of the electrode, specifically the density of electronic states close to the formal potential of the redox system [4]. Heterogeneous electron-transfer rate constants between 0.01 and 0.2 cm s⁻¹ are commonly observed for conducting polycrystalline films with extensive pretreatment [1,4,8,28]. Dopamine belongs to inner-sphere system and exhibits much more electrochemical irreversibility, as evidenced by the relatively large ΔE_p . This more irreversible behavior is characteristic of all the catechols and catecholamines investigated so far on diamond electrodes [3,4,37,38]. The value of k^0 for dopamine, as well as for other catecholamines, the electron-transfer reaction on the BDD electrode surface, resides in the 10^{-4} to 10^{-6} cm s⁻¹ interval at hydrogen-terminated diamond [1,4]. It is postulated that the slow kinetics for this analyte result from weak adsorption on the diamond surface [1,4,30]. It is assumed that π - π interactions through the extended electronic conjugation are perhaps more influential than surface oxides in promoting molecular adsorption for some classes of analytes on these surfaces (*e.g.*, catechols) [1].

The kinetic parameters for the system $\text{Fe}(\text{CN})_6^{3-/4-}$ and $\text{Ru}(\text{NH}_3)_6^{3+/2+}$ were obtained according to Nicholson's method [39], assuming a value of the cathodic electron transfer coefficient $\alpha=0.5$ and diffusion coefficients for oxidized (D_{Ox}) and reduced species (D_{Red}) in 1 mol L⁻¹ KCl.

Figure 3.5 shows CV obtained for different electron transfer reactions taking place at the BDD electrode.

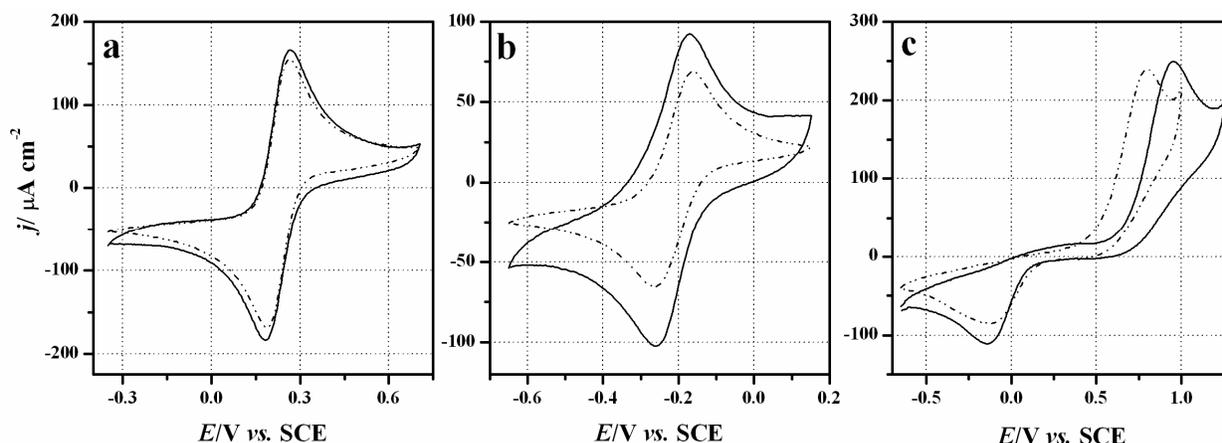


Figure. 3.5 Cyclic voltammograms for three analytes: ferrocyanide (a), hexamineruthenium (b) and dopamine (c) at BDD type A (solid line) and type B (dashed line) electrodes. All redox analyte concentrations were $10^{-3} \text{ mol L}^{-1}$. The supporting electrolyte was $1 \text{ mol L}^{-1} \text{ KCl}$. Scan rate: 100 mV s^{-1}

In all cases, the CV profiles furnished a linear response for the dependence of the peak current on the square root of the scan rate, thus revealing that the voltammetric current is controlled by the semi-infinite linear diffusion of the redox species at the BDD/solution interface. In the special case of dopamine, the kinetic parameter was obtained from the *Tafel* plot (E vs $\log i$) [40] because of the irreversible behavior presented by this compound. CV and kinetic parameters are presented in Table 3.1.

Table 3.1 CV and kinetic parameters obtained at BDD type A and type B electrodes for three redox systems in $1 \text{ mol L}^{-1} \text{ KCl}$ at 25°C

System	$\Delta E_p / \text{mV}$		$I_p^a / \mu\text{A cm}^{-2}$		$k^0 / \text{cm s}^{-1}$		α	
	A	B	A	B	A	B	A	B
$\text{Fe}(\text{CN})_6^{3-/4-}$	83.0	68.0	174.27	166.11	0.010	0.028	0.50	0.50
$\text{Ru}(\text{NH}_3)_6^{3+/2+}$	92.0	92.0	92.33	68.06	0.008	0.008	0.50	0.50
Dopamine	804.0	648.0	133.27	101.11	2.08×10^{-5}	6.60×10^{-5}	0.24	0.22

The results presented in Table 3.1 are coherent with those found in the literature [1,4,6,7,28]. In a comparative analysis among the electrodes, the ΔE_p values are smaller for type B electrodes, which supplied larger k^0 values. Although type A presented smaller values of k^0 in relation to type B, such electrodes are within the standard parameters for electroanalytical purposes. On the other hand,

note that the oxidation current was larger for type A electrodes for all analytes, indicating a higher sensitivity. Electrodes prepared from ethanol as the carbon source showed excellent electrochemical parameters and thus become an alternative to diamond electrodes produced with methane.

Electrochemical impedance spectroscopy As previously discussed by Mahé et al. [7], the electrochemical kinetic parameters determined using EIS are more reliable than those from CV because EIS constitutes a more efficient and more accurate experimental approach than CV. Therefore, in the present case, we have performed EIS analysis to compare the results with those obtained from CV.

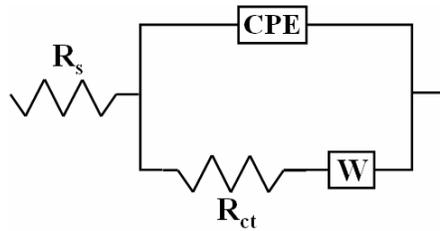


Figure. 3.6 Equivalent circuit used for representing the impedance response in the presence of diffusion control

EIS analysis of BDD electrodes can be carried out in the presence of diffusion control using the Randles–Ershler’s circuit presented in Figure 3.6 [2,7].

In the equivalent circuit, R_s is the uncompensated ohmic resistance ($=R_{\text{solution}}$); CPE is the constant phase element used for representing the capacitive behavior presented by the double layer capacitance; R_{ct} is the charge transfer resistance, and W is the Warburg impedance associated with diffusion control at the BDD/solution interface.

Using the treatment proposed by Brug et al. [41], the double layer capacitance C_{dl} was obtained from the CPE using the equation:

$$C_{dl}^{\phi} = \frac{Z_{CPE}^{-1} (j\omega)^{-\phi}}{[(R_s)^{-1} + (R_{ct})^{-1}]^{(1-\phi)}} \quad 3.1$$

where the parameters are obtained from impedance data using the complex nonlinear least square fitting (CNLS) [21]. The power ϕ is a dimensionless parameter related to the depression angle originated from the non-ideal electric behaviour presented by any electrode/electrolyte interface.

In the present case, the EIS study was carried out in solutions having a $10^{-3} \text{ mol L}^{-1}$ concentration of the different redox systems ($\text{Fe}(\text{CN})_6^{3-/4-}$, $\text{Ru}(\text{NH}_3)_6^{3+/2+}$ or dopamine) under natural convection, using 1 mol L^{-1} KCl as the supporting electrolyte. In all cases, the EIS spectrum was recorded at the open circuit potential (equilibrium potential), which for the type A and B BDD electrodes and the $\text{Fe}(\text{CN})_6^{3-/4-}$, $\text{Ru}(\text{NH}_3)_6^{3+/2+}$ and dopamine systems, were 223/226, -233/-153, and 809/636 mV(vs SCE), respectively.

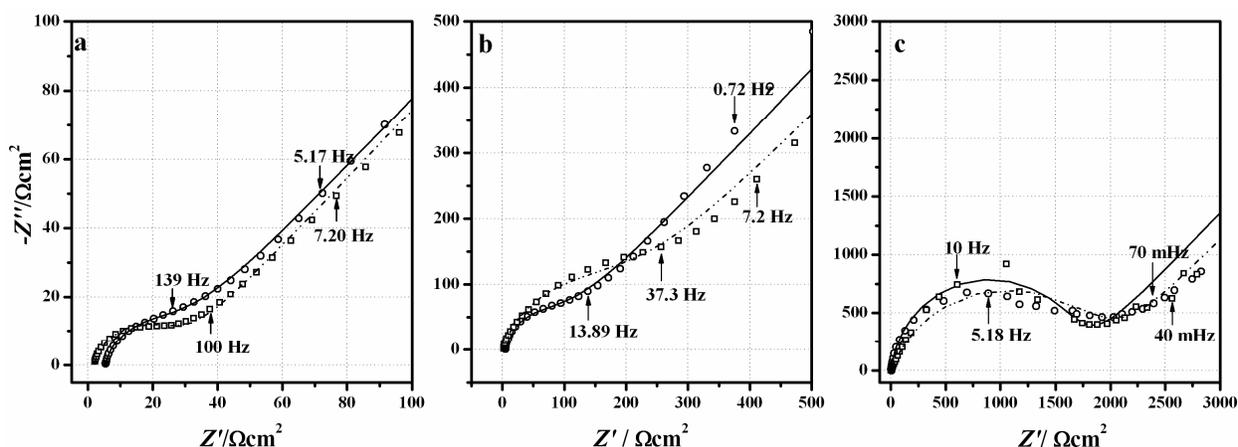


Figure. 3.7 Nyquist plots obtained for different electron-transfer reactions on BDD electrodes: a ferrocyanide, $10^{-3} \text{ mol L}^{-1}$; b hexamineruthenium, $10^{-3} \text{ mol L}^{-1}$; c dopamine, $10^{-3} \text{ mol L}^{-1}$. Supporting electrolyte was 1.0 mol L^{-1} KCl. $T=25^\circ \text{C}$. Legend: Type A, experimental (circles) and curve fitting (solid line); type B, experimental (squares) and curve fitting (dashed line)

Figure 3.7 shows Nyquist plots obtained for the type A and type B electrodes as a function of the solution composition.

Analyzing Figure 3.7 reveals that both electrode preparation conditions and the solution composition considerably affect the electrochemical response of BDD electrodes in the frequency domain. The behavior of the impedance spectra displayed in Figure 3.7a and b, which were obtained for the $\text{Fe}(\text{CN})_6^{3-/4-}$, $\text{Ru}(\text{NH}_3)_6^{3+/2+}$ redox couples, respectively, supports a typical diffusion-controlled process, mainly characterized by a slight deviation from the ideal reversible case, thus showing quasi-reversible behaviors. As a result, the analysis fit via CNLS revealed a good correspondence in these cases. In the special case of dopamine, the impedance data are considerably different from the ones already described. A

qualitative analysis of these findings indicated a considerable deviation from the ideal reversible case, as expected.

The heterogeneous electron-transfer rate constant k^0 at the equilibrium potential can be obtained from the R_{ct} values (available from the CNLS fit) according to Eq. 2:

$$k^0 = \frac{RT}{n^2 F^2 R_{ct} C} \quad 3.2$$

The parameters presented in Eq. 2 have their usual significance.

A summary of results obtained by EIS are presented in Table 3.2.

Table 3.2 The double-layer capacity C_{dl} and the heterogeneous electron-transfer rate constant k^0 for BDD type A and type B electrodes obtained from the EIS data for different redox systems

BDD type System	$k^0/\text{cm s}^{-1}$		$C_{dl}/\mu\text{F cm}^{-2}$		ϕ	
	A	B	A	B	A	B
$\text{Fe}(\text{CN})_6^{3-/4-}$	0.015	0.010	5.87	0.76	0.88	0.84
$\text{Ru}(\text{NH}_3)_6^{3+/2+}$	2.99×10^{-3}	1.31×10^{-3}	8.90	0.56	0.95	0.84
Dopamine	1.70×10^{-4}	1.46×10^{-4}	5.56	5.64	0.91	0.92

Supporting electrolyte: 1.0 mol L⁻¹ KCl. T=25 °C

Comparing the results listed in Tables 3.1 and 3.2, it is clearly observed that the EIS technique provided k^0 values slightly different. This discrepancy probably finds its origin in the different accuracies presented by the CV and EIS techniques [7].

The standard electrochemical behavior of high-quality BDD electrodes, in the presence of different redox systems, was reported by Granger et al. [4]. These authors found that the $\text{Fe}(\text{CN})_6^{3-/4-}$ redox couple undergoes electron transfer that is extremely sensitive to the surface chemistry of BDD, thus reporting k^0 in the range of 10^{-2} - 10^{-1} cm s⁻¹. According to the literature

[7] for BDD electrodes, the electrochemical reactivity for the $\text{Fe}(\text{CN})_6^{3-/4-}$ system is considerably affected by the p-type surface conductivity in which the hole generation process is correlated by the presence of an hydrogenated surface. Granger et al. [4] also found that the $\text{Ru}(\text{NH}_3)_6^{3+/2+}$ redox couple is mainly characterized by an electron transfer that is insensitive to the diamond surface

microstructure and chemistry with k^0 in the 10^{-2} - 10^{-1} cm s⁻¹ range. Besides, according to this report, dopamine undergoes sluggish electron transfer with k^0 between 10^{-4} and 10^{-5} cm s⁻¹.

A comparison of the findings discussed above with the results in Table 3.2 shows a rather good correspondence, thus revealing that the electrochemical properties of our BDD films are very similar to those pertaining to high quality BDD electrodes reported in the literature.

Table 3.2 shows double-layer capacitance values in the range of 0.56-8.90 $\mu\text{F cm}^{-2}$. These values are in agreement with the ones presented by Mahé et al. [7], who reported a capacitance of 2.5 $\mu\text{F cm}^{-2}$ in the case of $\text{Fe}(\text{CN})_6^{3-/4-}$. Therefore, based on the discussion presented by these authors

[7]

, it can be argued that both electrodes are adequate for analytical applications. However, it is worthwhile to mention that the type B presents better characteristics because of its lower capacitance.

Analysis of the capacitance data presented in Table 3.2 also reveals that the influence of the electrode preparation conditions is more pronounced in the case of the $\text{Fe}(\text{CN})_6^{3-/4-}$ redox couple, while in the case of $\text{Ru}(\text{NH}_3)_6^{3+/2+}$ and dopamine, it is observed that the value remained approximately constant. This behavior is in agreement with the literature [4], where it was found that the $\text{Ru}(\text{NH}_3)_6^{3+/2+}$ redox couple is less sensitive to the BDD surface microstructure than the $\text{Fe}(\text{CN})_6^{3-/4-}$ system.

Any possible influence on the electrode behavior promoted by changes in the semiconductor properties during the CVD process using ethanol as a carbon source would appear by the change in the electrochemical behavior of BDD for “standard” redox reactions (*e.g.*, outer and inner sphere redox processes). Thus the developed BDD electrodes present similar electronic properties when comparing with the BDD electrodes currently found in the literature.

Conclusions

BDD film electrodes grown from the ethanol carbon source presented improved electrochemical properties because ethanol can decrease significantly the

graphitic carbon impurities in situ, differently from the others sources. Hence, a great number of electrochemical technologies could benefit from the use of this electrode material such, as: electroanalysis, electrocatalysis, electrooxidation, spectroelectrochemistry, and bioelectrochemistry. The conclusions of this work allow emphasis on several points: (1) Ethanol is an excellent alternative source of carbon for growing BDD films for electrochemical applications; (2) a mild cathodic treatment caused significant increases in the electron transfer; (3) Raman reveals that only a very small amount of nondiamond carbon impurity was itself formed at the BDD film using ethanol as a source of carbon; (4) both type A and type B samples had similar morphologies, although their grain sizes are different because of the different deposition temperatures; (5) the background for type A and type B electrodes is low, featureless, and ideally polarizable indicating, again, that no significant level of nondiamond carbon impurities is present; and (6) according to investigations from several aqueous-based redox systems using CV and EIS, it could be noticed that the electrochemical parameters for the same technique attributed to both films are similar. However, small differences occur in the parameters between the techniques. From our point of view, EIS is a much more accurate technique, compared to CV, because of its more direct way in obtaining the system information; (7) the behavior of the EI spectra for $\text{Fe}(\text{CN})_6^{3-/4-}$ and $\text{Ru}(\text{NH}_3)_6^{3+/2+}$ redox couples supports a typical diffusion-controlled process. In the special case of dopamine, a considerable deviation from the ideal reversible case happened, possibly because of the absence of carbonyl surface functionalities on the hydrogen-terminated surface; (8) the resulting capacitance values for both types of electrodes indicate that these BDD electrodes are promising for analytical application.

References

1. Swain, G. Electrically conducting diamond thin films: advanced electrode materials for electrochemical technologies. In *Electroanalytical chemistry: a series of advances*, Bard, A. J. and Rubinstein, I., Eds. Marcel Dekker Inc: New York, 2004, Vol. 22, pp 181-277.
2. Pleskov, Y. V. Electrochemistry of diamond: A review. *Russ. J. Electrochem.* **2002**, 38, (12), 1275-1291.

3. Xu, J. S.; Granger, M. C.; Chen, Q. Y.; Strojek, J. W.; Lister, T. E.; Swain, G. M. Boron-doped diamond thin-film electrodes. *Anal. Chem.* **1997**, 69, (19), A591-A597.
4. Granger, M. C.; Witek, M.; Xu, J. S.; Wang, J.; Hupert, M.; Hanks, A.; Koppang, M. D.; Butler, J. E.; Lucazeau, G.; Mermoux, M.; Strojek, J. W.; Swain, G. M. Standard electrochemical behavior of high-quality, boron-doped polycrystalline diamond thin-film electrodes. *Anal. Chem.* **2000**, 72, (16), 3793-3804.
5. Ranganathan, S.; Kuo, T. C.; McCreery, R. L. Facile preparation of active glassy carbon electrodes with activated carbon and organic solvents. *Anal. Chem.* **1999**, 71, (16), 3574-3580.
6. Granger, M. C.; Swain, G. M. The influence of surface interactions on the reversibility of ferri/ferrocyanide at boron-doped diamond thin-film electrodes. *J. Electrochem. Soc.* **1999**, 146, (12), 4551-4558.
7. Mahe, E.; Devilliers, D.; Comninellis, C. Electrochemical reactivity at graphitic microdomains on polycrystalline boron doped diamond thin-films electrodes. *Electrochim. Acta* **2005**, 50, (11), 2263-2277.
8. Hupert, M.; Muck, A.; Wang, R.; Stotter, J.; Cvackova, Z.; Haymond, S.; Show, Y.; Swain, G. M. Conductive diamond thin-films in electrochemistry. *Diam. Relat. Mat.* **2003**, 12, (10-11), 1940-1949.
9. Ford, I. J. Boundaries of the diamond domain in the C-H-O diagram of carbon film deposition. *J. Phys. D-Appl. Phys.* **1996**, 29, (9), 2229-2234.
10. Shah, S. I.; Waite, M. M. Effect of oxygen on the nucleation and growth of diamond thin-films. *Appl. Phys. Lett.* **1992**, 61, (26), 3113-3115.
11. Ruan, J.; Choyke, W. J.; Kobashi, K. Oxygen-related centers in chemical vapor-deposition of diamond. *Appl. Phys. Lett.* **1993**, 62, (12), 1379-1381.
12. Baranauskas, V.; Tosin, M. C.; Peterlevitz, A. C.; Ceragioli, H.; Durrant, S. F. Enhancement of diamond nucleation using the solid-liquid-gas interface energy. *J. Appl. Phys.* **2000**, 88, (3), 1650-1654.
13. Baranauskas, V.; Li, B. B.; Peterlevitz, A.; Tosin, M. C.; Durrant, S. F. Nitrogen-doped diamond films. *J. Appl. Phys.* **1999**, 85, (10), 7455-7458.
14. Su, Q. F.; Lu, J. F.; Wang, L. J.; Liu, J. M.; Ruan, J. F.; Cui, J. T.; Shi, W. M.; Xia, Y. B. Electrical properties of [100]-oriented CVD diamond film. *Solid-State Electron.* **2005**, 49, (6), 1044-1048.
15. Ma, Z. B.; Wang, J. H.; Wang, C. X.; Man, W. D. Microwave plasma chemical vapor deposition of diamond films on silicon from ethanol and hydrogen. *Plasma Sci. Technol.* **2003**, 5, (2), 1735-1741.

16. Li, B. B.; Baranauskas, V.; Peterlevitz, A.; Chang, D. C.; Doi, I.; Trava-Airoldi, V. J.; Corat, E. J. Annealing-induced enhancement in the activation energy of heavily boron-doped polycrystalline diamond. *Diam. Relat. Mat.* **1998**, 7, (9), 1259-1262.
17. Honda, K.; Yoshimura, M.; Rao, T. N.; Tryk, D. A.; Fujishima, A.; Yasui, K.; Sakamoto, Y.; Nishio, K.; Masuda, H. Electrochemical properties of Pt-modified nano-honeycomb diamond electrodes. *J. Electroanal. Chem.* **2001**, 514, (1-2), 35-50.
18. Farabaugh, E. N.; Robins, L.; Feldman, A.; Johnson, C. E. Growth and oxidation of boron-doped diamond films. *J. Mater. Res.* **1995**, 10, (6), 1448-1454.
19. Baranauskas, V.; Peled, A.; Trava-Airoldi, V. J.; Lima, C.; Doi, I.; Corat, E. J. Morphological-studies of laser etching processes in self-sustained cvd diamond wafers. *Appl. Surf. Sci.* **1994**, 80, 129-135.
20. Marken, F.; Paddon, C. A.; Asogan, D. Direct cytochrome c electrochemistry at boron-doped diamond electrodes. *Electrochem. Commun.* **2002**, 4, (1), 62-66.
21. Boukamp, B. A. A nonlinear least-squares fit procedure for analysis of immittance data of electrochemical systems. **1986**, 20, (1), 31-44.
22. Ferro, S.; De Battisti, A. Electron transfer reactions at conductive diamond electrodes. *Electrochim. Acta* **2002**, 47, (10), 1641-1649.
23. Kawarada, H. Hydrogen-terminated diamond surfaces and interfaces. *Surf. Sci. Rep.* **1996**, 26, (7), 205-259.
24. Shinagawa, H.; Kido, G.; Takamasu, T.; Gamo, M. N.; Ando, T. High mobility holes on hydrogen-terminated diamond surface. *Superlattices Microstruct.* **2002**, 32, (4-6), 289-294.
25. Looi, H. J.; Jackman, R. B.; Foord, J. S. High carrier mobility in polycrystalline thin film diamond. *Appl. Phys. Lett.* **1998**, 72, (3), 353-355.
26. Ristein, J.; Maier, F.; Riedel, M.; Stammer, M.; Ley, L. Diamond surface conductivity experiments and photoelectron spectroscopy. *Diam. Relat. Mat.* **2001**, 10, (3-7 SI Sp. Iss. SI), 416-422.
27. Goss, J. P.; Jones, R.; Heggie, M. I.; Ewels, C. P.; Briddon, P. R.; Oberg, S. Theory of hydrogen in diamond. *Phys. Rev. B* **2002**, 65, (11 AR 115207 DI ARTN).
28. Fischer, A. E.; Show, Y.; Swain, G. M. Electrochemical performance of diamond thin-film electrodes from different commercial sources. *Anal. Chem.* **2004**, 76, (9), 2553-2560.
29. Yang, H. H.; McCreery, R. L. Effects of surface monolayers on the electron-transfer kinetics and adsorption of methyl viologen and phenothiazine derivatives on glassy carbon electrodes. *Anal. Chem.* **1999**, 71, (18), 4081-4087.

30. DuVall, S. H.; McCreery, R. L. Self-catalysis by catechols and quinones during heterogeneous electron transfer at carbon electrodes. *J. Am. Chem. Soc.* **2000**, 122, (28), 6759-6764.
31. Chen, P. H.; Fryling, M. A.; McCreery, R. L. Electron-transfer kinetics at modified carbon electrode surfaces - the role of specific surface sites. *Anal. Chem.* **1995**, 67, (18), 3115-3122.
32. Granger, M. C.; Xu, J. S.; Strojek, J. W.; Swain, G. M. Polycrystalline diamond electrodes: basic properties and applications as amperometric detectors in flow injection analysis and liquid chromatography. *Anal. Chim. Acta* **1999**, 397, (1-3), 145-161.
33. Duo, I.; Levy-Clement, C.; Fujishima, A.; Comninellis, C. Electron transfer kinetics on boron-doped diamond Part I: Influence of anodic treatment. *J. Appl. Electrochem.* **2004**, 34, (9), 935-943.
34. Duo, I.; Fujishima, A.; Comninellis, C. H. Electron transfer kinetics on composite diamond (sp³)-graphite (sp²) electrodes. *Electrochem. Commun.* **2003**, 5, (8), 695-700.
35. Goeting, C. H.; Foord, J. S.; Marken, F.; Compton, R. G. Sonoelectrochemistry at tungsten-supported boron-doped CVD diamond electrodes. *Diam. Relat. Mat.* **1999**, 8, (2-5), 824-829.
36. Chen, P. H.; McCreery, R. L. Control of electron transfer kinetics at glassy carbon electrodes by specific surface modification. *Anal. Chem.* **1996**, 68, (22), 3958-3965.
37. Fujishima, A.; Rao, T. N.; Popa, E.; Sarada, B. V.; Yagi, I.; Tryk, D. A. Electroanalysis of dopamine and NADH at conductive diamond electrodes. *J. Electroanal. Chem.* **1999**, 473, (1-2), 179-185.
38. Alehashem, S.; Chambers, F.; Strojek, J. W.; Swain, G. M.; Ramesham, R. Cyclic voltammetric studies of charge-transfer reactions at highly boron-doped polycrystalline diamond thin-film electrodes. *Anal. Chem.* **1995**, 67, (17), 2812-2821.
39. Nicholson, R. S. Theory and application of cyclic voltammetry for measurement of electrode reaction kinetics. *Anal. Chem.* **1965**, 37, (11), 1351-&.
40. Bard, A. J.; Faulkner, L. R. *Electrochemical methods: fundamentals and applications*. 2nd ed.; John Wiley & Sons: New York, 2001; p 833.
41. Brug, G. J.; Vandeneeden, A.; Sluytersrehabach, M.; Sluyters, J. H. The analysis of electrode impedances complicated by the presence of a constant phase element. **1984**, 176, (1-2), 275-295.

Capítulo 4

Estudo QSPR do Bloqueio Eletroquímico Provocado por Fenóis em Eletrodos de Platina e Diamante Dopado com Boro

“Toda força é fraca se não é unida”

La Fontaine

Neste capítulo, um estudo quantitativo relacionando a estrutura fenólica à propriedade de resistência ao bloqueio eletroquímico foi realizado sobre as superfícies do diamante dopado com boro (BDD) e da platina (Pt). O envenenamento destas superfícies eletródicas (Pt e BDD), provocado por vinte compostos fenólicos foi investigado por meio de cronoamperometria, métodos teóricos *i.e.*, QSPR (Relação quantitativa estrutura-propriedade) e métodos quimiométricos. A diferença entre as densidades de corrente depois de 15s e 90s de oxidação foi proposta pela primeira vez como medida quantitativa do bloqueio eletroquímico de eletrodos. As estruturas das moléculas fenólicas e de seus complexos formados via ligação de hidrogênio com o íon fluoreto foram modeladas e a geometria otimizada com os métodos B3LYP com o conjunto de bases 6-31G**. Vários descritores moleculares foram calculados e correlacionados com as medidas de bloqueio eletroquímico usando a regressão por quadrados mínimos parciais (PLS). Um modelo com uma variável latente (componente) de cinco descritores foi construído. Este modelo apresentou alta capacidade de previsão para o bloqueio eletroquímico no eletrodo de platina e foi validado externamente com 4 amostras. Os parâmetros estatísticos obtidos foram: $Q^2 = 0,786$, $R^2 = 0,851$, $RMSECV = 0,097$, and $Q^2_{ext} = 0,636$. Para o BDD um modelo com uma variável latente de quatro descritores também foi construído e validado da mesma maneira,

porém os parâmetros estatísticos ($Q^2 = 0,333$, $R^2 = 0,586$, $RMSECV = 0,159$ e $Q^2_{\text{ext}} = -0,031$) foram muito inferiores ao da Pt. Ambos os modelos foram utilizados para realizar previsões de dez novos compostos fenólicos e o modelo da platina mostrou ser apto para tal finalidade. Observou-se um bloqueio eletroquímico muito mais fraco sobre o eletrodo de BDD que sobre o de Pt e diferenças entre as interações eletrodo-fenol são as principais razões para as grandes diferenças entre os modelos. Análises exploratórias (PCA e HCA) foram também realizadas e interpretadas juntamente com os modelos de regressão em termos de conceitos químicos tais como reatividade fenólica, tamanho/forma, característica eletrônica e ligações de hidrogênio. Estes resultados podem ser úteis para explorar a possibilidade de prever o envenenamento e planejar experimentos eletroquímicos envolvendo diversos compostos fenólicos.

O estudo completo será apresentado em idioma inglês conforme artigo 2º da informação CCPG/002/06 da Unicamp, que permite a inclusão de artigo científico publicado ou submetido para publicação como capítulo ou apêndice da tese. No caso em questão, o artigo foi submetido à arbitragem em revista devidamente indexada..

QSPR study of passivation by phenolic compounds at platinum and boron-doped diamond electrodes

Reinaldo F. Teófilo^a, Rudolf Kiralj^a, Helder J. Ceragioli^b, Alfredo C. Peterlevitz^b, Vitor Baranauskas^b, Lauro T. Kubota^a, Márcia M.C. Ferreira^a

^aInstituto de Química, Universidade Estadual de Campinas, P.O. Box 6154, 13084 – 971

Campinas, SP, Brazil

^bFaculdade de Engenharia Elétrica e Computação, Universidade Estadual de Campinas, Av. Albert Einstein 400, 13083 – 852

Campinas, SP, Brazil

Introduction

It is well known that the oxidation of phenolic compounds at solid electrodes produces phenoxyl radicals, which covalently bind to form an unreactive and

impermeable polymeric film on the electrode surfaces, a phenomenon known as passivation, poisoning, blocking or fouling of the electrode. Such polymeric films represent a serious problem in electroanalysis (precision loss) and electrolysis (efficiency loss) of phenols since rapidly deactivate the electrode by blocking the electron transfer and slowing down further electron transfer [1-5]. However, electrochemical polymerization of organic films has received a great attention owing to some advantages in the use of aromatic amines to produce electroactive and conducting polymer films [6-8] and to protect electrodes against corrosion [9,10].

Despite the great relevance of the phenol passivation process, as much as in electroanalysis and electrooxidation, the formation and characteristics of the inhibitory polymeric film, and its dependence on the operational conditions are not well understood. A better understanding of the phenol electropolymerization mechanism is of great interest from the fundamental and practical viewpoints [4].

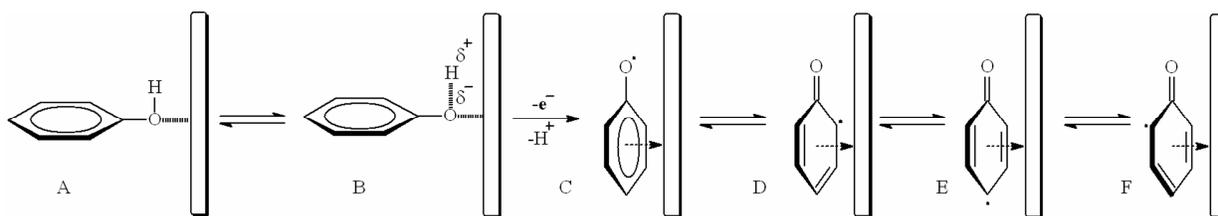


Figure 4.1. Mechanism of electrooxidation of phenol [11-15]. Adsorption of phenol on an electrode (A-B), electrooxidation (B-C) and resonance structures of the formed phenoxyl radical (C-F).

General pathways for the phenol electrooxidation have been described by several authors [2,16-18]. The behavior of phenol during electrooxidation may be considered as a general prototype for phenolics. Theoretical and experimental studies [1,11,15] have shown that the adsorption process occurs in two distinct and successive steps, as outlined in Figure 4.1. The first one is the physical adsorption of the phenol oxygen followed by heterolytic scission of the O-H bond [13,14] and anion oxidation [1,12], with the aromatic ring pointed towards the bulk of the solution (Figure 4.1, A-B). In the second step, the phenoxyl radical reorients from perpendicular (Figure 4.1 A-C) to a horizontal (Figure 4.1 D-E) configuration at the electrode surface, until the surface is fully covered by the phenolic radicals [11,15].

In fact, the formed phenoxyl radical is the key intermediate in the phenolic electropolymerization pathway. Its structure can be rationalized in terms of resonance forms, which explains how the radicals can combine to form different dimers. Such dimers may in turn be oxidized, in potentials lower than that of phenol, to produce radicals which can combine with another phenoxyl or dimer radicals to produce materials of high molecular weight, a passivating ultrathin insulating film [1,2]. This radical coupling occurs at the electrode surface or in its immediate vicinity but not in the bulk solution [19]. The polymerization occurs mainly *via ortho* and *para* positions due to the radical instability in these positions. [20]. The presence of substituents at the phenol ring determines whether the electropolymerization will be slower or faster relative to that for phenol and which intermediate phenoxyl radicals are more stable. Therefore, the reaction products will be formed and adsorbed at the electrode surface, because the products are usually kinetically and not thermodynamically controlled [21]. In fact, the pathway of this phenomenon is rather complex and may be affected by physical and chemical properties of the electrode material, solvent, pH, additives, concentration and structural features of phenolic compounds, electrode potential and current density, among others [1,5,22,23]. For instance, the formation of C-O bonds between radicals is favored in alkaline medium, while C-C bonds prevail in acid solutions [2, 17,24]. The polymeric tars binding will also depend on the nature of the active groups present on the electrode surface.

The majority of the studies prioritize the analysis of the obtained films, but little attention is dedicated to the initial phenolic structures *i.e.* the monomers. Besides that, quantitative data analyses on the passivation ability of different monomers are very scarce. The monomers structure as well as reactivity, defined by the substituents at various substitution positions of the ring, will be responsible for the adhesion and permeability of the resulting polymer on the electrode. The deactivation ability of the polymer is directly related to its adhesion and permeability. Hence, a quantitative study on the monomer structure and its passivation ability is desirable and can help understanding the electropolymerization of phenolic compounds.

Computational chemistry approaches can be useful not only in understanding the electropolymerization phenomena but also in the prediction of polymerization ability of a compound prior to experimental verification. Quantitative structure-activity relationship (QSAR) and quantitative structure-property relationship (QSPR) approaches are well-established methodologies that are useful in the prediction of various biological and physicochemical properties and chemical reactivity of substances for which there is no measured value yet of the properties in question [25]. QSPR methods have been successfully used to predict several physicochemical properties. For instance, in two recent QSPR studies, the O-H bond dissociation energy (BDE) of phenols was modeled using electronic properties of the substituent [26,27]. Furthermore, Nesmerak *et al.* [28] has presented a QSPR model to predict the substituent effect on half-wave potentials of benzoxazines.

In general, a QSPR analysis uses descriptors derived from the molecular structure. The advantage of this approach lies in the fact that the descriptors can be obtained computationally and are not dependent on carrying out experiments. Therefore, once a reliable regression model is established, it can be used to predict the studied property of new analogous compounds. Another advantage of QSPR studies is that it can show which structural features of the compounds play an important role in the modeled property. To the best of our knowledge, there is no QSPR work in the literature to predict electropassivation by phenols.

The purpose of this work is present a parsimonious QSPR models proposition correlating molecular descriptors to the electrode inactivation by phenolic compounds on platinum and boron-doped electrodes *via* Partial Least Squares (PLS) regression [29,30]. The models give can more insight into the phenolic polymer formation at molecular and mechanistic level.

Experimental

Boron-doped diamond film deposition

Boron-doped diamond (BDD) films were produced as previously described in an all-quartz cylindrical hot-filament CVD reactor. [31,32] In summary, a boron-

doping source was prepared by dissolving solid B₂O₃ in ethanol in a liquid reservoir before starting-up the reactor. Tungsten wires with diameter of 238 μm and length of 30 mm were used as substrates. The wires were mounted onto a polished silicon wafer placed below the hot filament, and parallel to the central axis of the filament coil. Deposition temperatures were measured by a thermocouple underside of the silicon wafer. Prior to the deposition, the tungsten wires were dipped in a colloidal mixture of diamond dust (0.25 μm in diameter) dispersed by ultrasonic vibration in *n*-hexane. This ‘seeding’ is a procedure used to increase the nucleation density of diamond films. The diamond films were deposited employing a mixture of ethanol vapor diluted in hydrogen gas (99.5% vol) at a volumetric flow rate of 100 standard cubic centimeters per minute (sccm), regulated by precision mass flow meters, and a total pressure kept at 20 torr. The boron dopant concentrations, estimated from Mott-Schottky plots were of $\sim 2.9 \times 10^{20}$ (5000 ppm or 0.5% B/C). Electric connections were placed on both types by wire-wrapping a nickel-chromium wire and sealing with a silver paint. The electrodes were insulated by a Teflon band and epoxy resin and then inserted into polypropylene tubes, exposing only the diamond area on the top of the electrodes for the electrochemical studies.

Voltammetry and chronoamperometry measurements

An AUTOLAB[®] PGSTAT-30 potentiostat was used to conduct electrochemical measurements. Polycrystalline platinum (Pt) wire (0.016 cm², geometric area) and a polycrystalline BDD (0.030 cm², geometric area) were used as working electrodes. All potentials were recorded against a saturated calomel reference electrode (SCE). The Pt electrode was polished before each measurement in 1 and 0.3 μm alumina slurred powders with ultra pure water (Milli-Q) on microcloth polishing followed by 5 min sonicating in alcoholic solution. The BDD electrode received before any measurement a mild cathodic treatment cycling the potential in the range from 0 to -4 V (vs. SCE) for at least three times in 1.0 M HNO₃ under vigorous stirring.

The experimental study was carried out in a cell of 5 ml using 0.05 M phosphate buffer solution at pH 6.5 and a final concentration of 5.0×10^{-4} M

phenolic compounds (from Sigma, Aldrich, Merck, Carlo Erba). All chemicals were of analytical grade and used without further purification. The solutions were prepared with ultra-high purity water (MilliPore, Milli-system) or ethanol (Merck). The electrode potential was scanned in the range defined for each phenolic compound. The potential sweep rate was 100 mV s^{-1} at step potential of 0.5 mV .

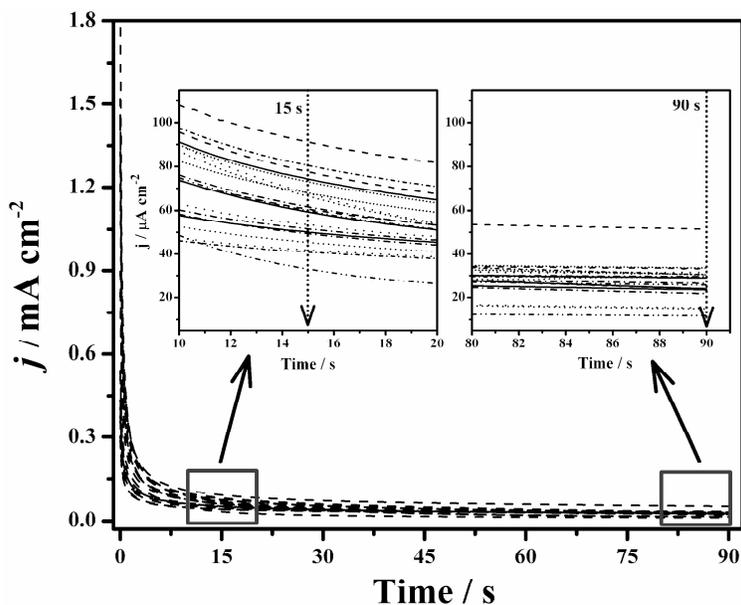


Figure 4.2. Passivation measured using chronoamperometric curves for Pt electrode.

The electrochemical passivation was monitored using the chronoamperometric method. This method was selected due to its facility to fix the oxidation potential with time. The potential was fixed in 50 mV more positive than the phenolic compounds oxidation peak. The oxidation was carried out during 90 s and the current was measured each 0.2 s (Figure 4.2). The difference between the current densities after 15 s and 90 s of oxidation time was defined for the first time as a parameter of passivation measurement on the Pt (Δj_{Pt}) and BDD (Δj_{BDD}) electrodes. This difference was chosen inasmuch as the polymerization is easily observed in the beginning of the process. The chronoamperometric curves for Pt electrode are presented in Figure 4.2 for illustration. The measurements were carried out in triplicate and the mean value was used for further data analyses. All experiments were performed at $25 \pm 2 \text{ }^\circ\text{C}$. The studied phenolic compounds were: catechol, chloroguaiacol, dopamine, guaiacol, hydroquinone, *L*-dopa, *o*-aminophenol, *o*-

nitrophenol, *p*-aminophenol, paracetamol, phenol, resorcinol, serotonine, 5-hydroxyindole, *o*-cresol, *p*-chloro-*m*-cresol, *m*-cresol, *p*-cresol, *o*-chlorophenol and *L*-tyrosine (Figure 4.3).

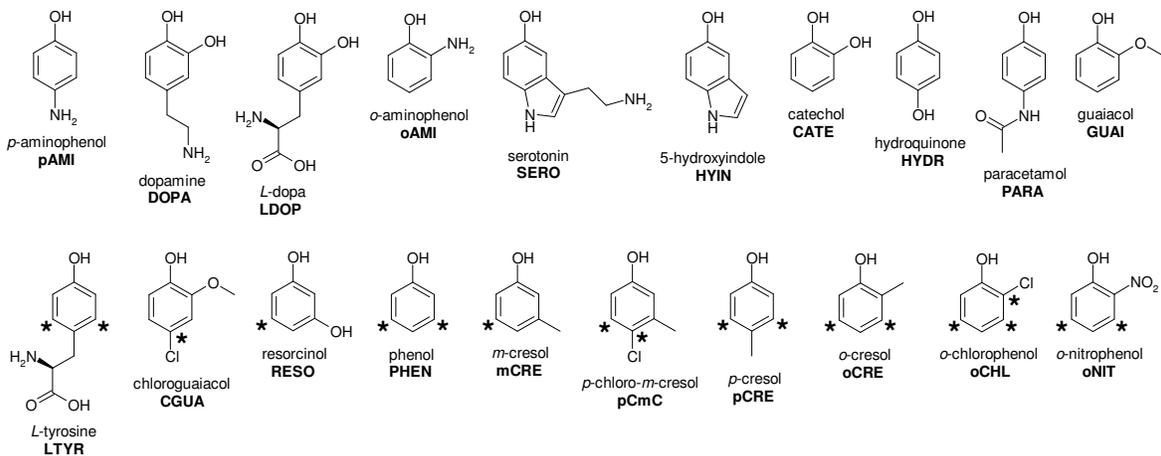


Figure 4.3. Structures of twenty phenolic compounds with measured passivation rates (training set), together with their names and used abbreviations, and reactivity markers (sites for nucleophilic aromatic substitution).

Molecular modeling

In order to model the molecular structures of the studied compounds, systematic searches were performed in the Cambridge Structural Database (CSD) [33] with January/May 2007 update [34] *via* ConQuest 1.9 [35,36] searching interface and Mercury [35,37] and WebLab View Pro [38] visualization software. Experimental geometries of phenolic compounds or their analogues were obtained with the most probable O-H bond orientations. Whenever possible, internal hydrogen bonds were considered as stabilizing factors. Geometry optimization was performed at the DFT (Density Functional Theory [39]) level with the B3LYP [40], [41] functional and 6-31G** basis set within the framework of the Titan molecular modeling software [42]. The resulted geometries of phenolic molecules were used in the next series of calculations: a quantum chemical probe (fluoride ion) was used to simulate strong hydrogen bond as equivalent to the interaction between the phenolic OH and electrode. All modeled hydrogen bonded complexes were geometry-optimized at the same computational level as the phenolic molecules. Dihydroxybenzenes were modeled differently than the other phenols: only one OH

group was hydrogen-bonded to the fluoride for the complexes calculations. Compounds **CATE**, **DOPA** and **LDOP** possess an internal hydrogen bond between the two neighboring OH groups. For such compounds, the OH group that act as the hydrogen bond donor was modified by turning the O-H bond in the opposite direction to the other OH group. When calculating the hydrogen bonding complexes, the formation of bifurcated hydrogen bond at the fluoride ion was avoided.

Various quantum chemical descriptors (independent variables) of steric and electronic nature were obtained from property calculations for phenolic molecules and their hydrogen complexes with the F⁻ ion by using the Titan software. Several compositional, topological and steric descriptors for phenolic molecules were also obtained by using the Dragon software [43] on the optimized geometries. Other compositional, topological, hydrogen bonding, electronic and geometrical descriptors were calculated and the total number of descriptors reached seven hundred.

Quantitative structure-property relationships (QSPR)

The calibration models were built using autoscaled molecular descriptors and the PLS regression method [29,30]. The input data for each PLS modeling were organized as a **X** matrix (descriptors data) where each row corresponded to a phenolic compound and each column to a molecular descriptor, and a **y** vector (dependent variable). The entries of the **y** vector were the experimentally measured passivation abilities of the studied phenolic compounds, defined as the logarithms of the means $\log(\Delta j_{Pt} / \mu\text{A cm}^{-2})$ and $\log(\Delta j_{BDD} / \mu\text{A cm}^{-2})$ for the Pt and BDD electrode, respectively. The variable selection for both models was performed in two steps. In the first step, the correlation coefficients of the **y** vectors with all the descriptors were inspected and a cut-offs of 0.60 and 0.40 were applied to the correlations relative to the Pt and BDD electrodes, respectively. In the second step, the variable selection was manually directed to satisfy several criteria: (1) satisfactory correlation with the passivation and acceptable PLS statistics, (2) selection of descriptors of different nature and origin, (3) computationally simple descriptors, and (4) chemically interpretable and understandable descriptors.

Two PLS models were further validated by leave- N -out crossvalidation, where N varied from 1 to 8. The robustness of the models was tested by performing ten Y-randomizations according to Wold and Eriksson [44]. The original data was randomized prior to leave- N -out crossvalidation and Y-randomization. The final validation procedure consisted of external validations, in which the original data sets were divided into as much as possible similar training (16 samples) and external (4 samples) validation sets (see Table 4.5) with the aid of Hierarchical Cluster Analysis (HCA).

HCA with incremental linkage and Principal Component Analysis (PCA) of the two data sets that had been used in the construction of the PLS models were performed in order to give deeper insight into relationships between electropolymerization and electronic structure of phenolic compounds. The PLS models were used to predict the passivation power of ten phenolics, for which there were no experimental passivation data measured in the same conditions as for the twenty phenolics in this work (Figure 4.4).

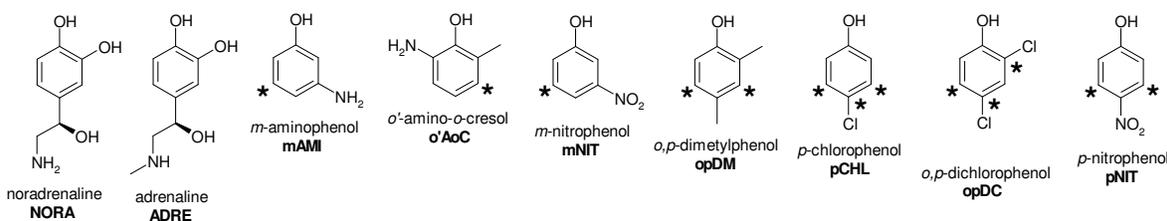


Figure 4.4. Structures of phenolic compounds without measured passivation rates (prediction set), together with their names and used abbreviations, and reactivity markers (sites for nucleophilic aromatic substitution).

All data analysis was performed using home-built Matlab-based functions on Matlab[®] 6.5 [45] and Pirouette package [46].

Results and Discussion

Cyclic voltammetry and chronoamperometry

Table 4.1 shows experimental results obtained for the phenolic compounds during their oxidation on the Pt and BDD electrodes using voltammetry and chronoamperometry.

Table 4.1. Experimental results obtained using cyclic voltammetry and chronoamperometry for the Pt and BDD electrodes.*

Compound	Pt					BDD				
	Ep	j_{15}	j_{90}	Δj	d Δj	Ep	j_{15}	j_{90}	Δj	d Δj
PHEN	0.90	59.31	23.45	35.86	3.31	1.38	52.90	41.25	11.64	1.24
CATE	0.38	45.35	29.14	16.21	2.05	0.91	40.25	29.46	10.79	1.59
RESO	0.90	67.63	15.49	52.14	2.44	1.32	50.79	36.56	14.23	1.32
HYDR	0.37	53.63	30.69	22.94	0.41	0.82	39.02	27.72	11.31	2.42
oAMI	0.44	50.02	28.95	21.08	1.59	0.59	43.57	25.21	18.36	1.77
pAMI	0.36	41.84	31.49	10.34	2.85	0.74	34.65	26.50	8.14	1.48
PARA	0.60	60.62	33.17	27.45	0.49	0.91	43.22	30.32	12.90	3.56
oCRE	0.91	72.99	30.42	42.56	0.52	1.31	45.02	30.68	14.34	2.17
mCRE	0.95	66.29	15.05	51.24	0.65	1.28	56.67	35.02	21.65	4.80
pCRE	0.84	74.19	23.89	50.30	3.62	1.19	56.26	36.68	19.58	4.65
pCmC	0.82	61.58	21.73	39.85	0.39	1.19	30.64	24.53	6.11	2.94
oCHL	0.88	77.58	23.80	53.78	1.72	1.25	26.75	18.25	8.50	1.37
oNIT	0.96	80.45	30.45	50.01	5.29	1.29	39.99	27.83	12.17	1.06
GUAI	0.79	88.51	53.15	35.36	3.85	1.08	61.54	46.08	15.47	2.52
CGUA	0.70	67.79	33.52	34.26	1.61	1.06	32.35	22.99	9.36	2.43
LYR	0.93	59.89	25.60	34.29	2.09	1.21	35.21	18.26	16.94	2.78
DOPA	0.32	40.82	29.20	11.62	1.13	0.62	33.55	30.56	2.99	0.50
LDOP	0.59	49.39	29.61	19.79	0.44	0.87	40.23	29.73	10.50	0.80
HYIN	0.52	33.00	11.80	21.21	2.10	0.78	58.55	39.27	19.28	3.07
SERO	0.46	51.57	26.37	25.20	1.91	0.63	26.77	12.92	13.85	1.28

*Ep (Volt): peak potential obtained using the cyclic voltammetry study; j_{15} and j_{90} ($\mu\text{A cm}^{-2}$): current densities obtained using the chronoamperometric time-current curve after 15 and 90 seconds, respectively; Δj ($\mu\text{A cm}^{-2}$): difference between j_{15} and j_{90} ; d Δj : standard deviation differences from triplicate measures.

The phenolic oxidation on the Pt electrode is significantly different from that on the BDD electrode (Table 4.1). The mean, standard deviation, maximum and minimum values of the peak potential for the Pt electrode are 0.68 V, 0.23, 0.96 V and 0.32 V, respectively. For the BDD electrode, the equivalent values are 1.02 V, 0.26, 1.38 V and 0.59 V, respectively. It can be noticed that all phenolic compounds are characterized by peak potentials higher for the BDD electrode than for the Pt electrode. The oxidation at higher potentials indicates low rate of electron transfer. This is in agreement with the literature, which reports relatively slow electrode kinetics as one of the reasons for lack of adsorption on diamond [31,47-49].

The difference of current density (Δj) obtained for the Pt electrode is significantly higher than that for the BDD electrode (Figure 4.5). The mean, standard deviation, maximum and minimum values of Δj for the Pt electrode are $32.77 \mu\text{A cm}^{-2}$, 14.10, $53.78 \mu\text{A cm}^{-2}$ and $10.34 \mu\text{A cm}^{-2}$, respectively. For the BDD electrode, these values are $12.91 \mu\text{A cm}^{-2}$, 4.77, $21.65 \mu\text{A cm}^{-2}$ and $2.99 \mu\text{A cm}^{-2}$, respectively. These results indicate weaker fouling on the boron-doped diamond electrode, which agrees with the literature on the boron-doped diamond electrode [49,50].

It is worth noting that although there is a current decrease at the BDD electrode with time, the formed phenolic polymer is easily removed by stirring or, in worst situations, by performing a mild cathodic treatment under stirring [31].

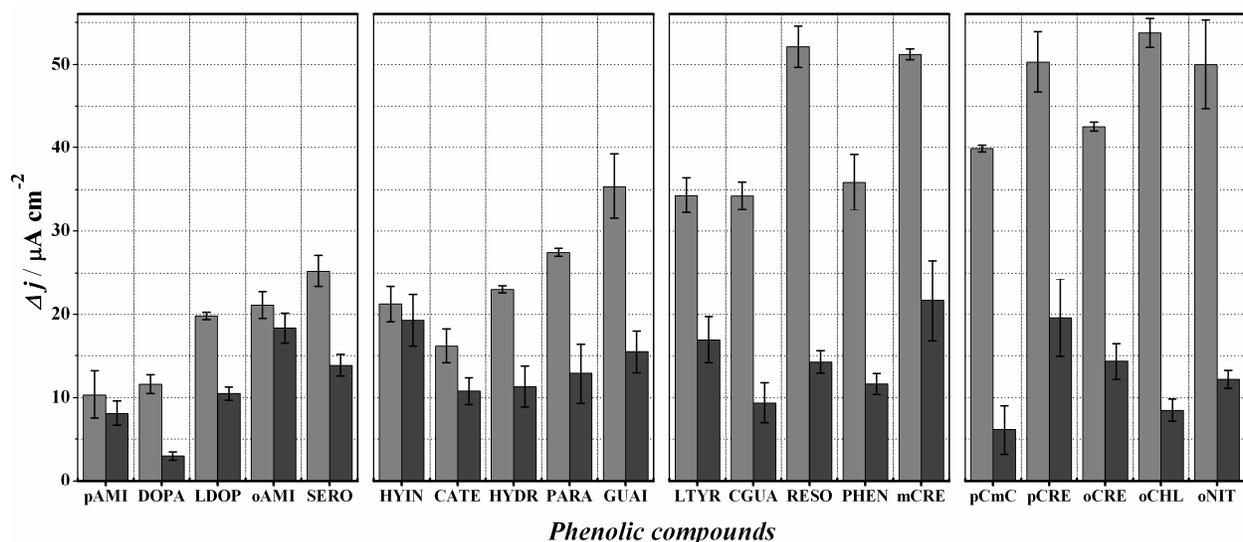


Figure 4.5. Graphic representation of the phenolic passivation for different substituents on the phenol ring obtained from chronoamperometric measurements. Gray: Pt electrode; Dark gray: BDD electrode. Bar error is the standard deviation.

Molecular descriptors

The final selected descriptors are presented in Table 4.2, which also includes the dependent variables. No descriptor is common for the two electrodes, indicating significant differences in electrochemical processes involving phenolic compounds at the electrodes. The descriptors are described in the text below.

a) Molecular descriptors related to the phenolic passivation on the Pt electrode:

HBD/N - the number of acidic hydrogen atoms (related to hydrogen bond donors, HBD) divided by the number of non-hydrogen atoms, as obtained from molecular formula;

Mor06u - an unweighted 3D MoRSe signal 06 descriptor, a steric descriptor obtained by using the Dragon software;

Qcnpa - the NPA (Natural Population Analysis) partial atomic charge of the phenolic carbon atom which is chemically bound to the phenolic OH group, obtained from DFT computations for the isolated molecule;

Ar - a modified Julg's aromaticity index [51] defined as $Ar = 255 - (\sigma_b / d_b)^2$, where: d_b is the average bond length and σ_b is the standard deviation of bond lengths of the benzene ring, as obtained from the DFT geometry of the isolated molecule;

QNUnpa - the sum of NPA atomic charges of phenolic carbon atoms which are considered as formal sites for nucleophilic aromatic substitution. The value for this descriptor was set to zero when no sites of these kind were identified. This descriptor is a reactivity index, which may be related, directly or indirectly, to the reactions between phenoxy radicals. Interestingly, analogue descriptors describing sites in electrophilic aromatic substitution were poorly correlated with the phenolic passivation on the Pt electrode.

b) Molecular descriptors related to the phenolic passivation on the BDD electrode:

D-dCO - C-O bond shortening after the H...F hydrogen bond is formed, defined as the difference between the phenolic C-O bond lengths in the isolated molecule and the hydrogen bonding complex, as obtained from DFT calculations;

DISPm - a dCOMMA2 value weighted by atomic masses, a conformationally dependent descriptor based on molecular geometry, as obtained by using the software Dragon;

C-026 - the number of R-CH-R fragments, belonging to the Ghose-Crippen counts of atom-centered fragments as obtained by using the software Dragon;

$L(nCaH)$ - linearized form of $nCaH$, defined as $L(nCaH) = (nCaH - 3.6)^2$, where $nCaH$ is the number of unsubstituted aromatic Csp^2 atoms, a count of functional groups as obtained by using the software Dragon.

Table 4.2. Molecular descriptors and passivations used in QSPR

Compound	HBD/N	Mor06u	Qcnpa	Ar	QNUnpa	$\log(\Delta j_{pi})$	D-dCO/Å	DISPm	C-026	$L(nCaH)$	$\log(\Delta j_{BDD})$
PHEN	0.1429	-2	0.327	0.9993	-0.444	1.55	0.072	4	1	0.36	1.07
CATE	0.2500	-1	0.270	0.9965	0	1.21	0.078	5	2	0.16	1.03
RESO	0.2500	-1	0.342	0.9985	-0.206	1.72	0.073	3	2	0.16	1.15
HYDR	0.2500	-2	0.300	0.9994	0	1.36	0.074	0	2	0.16	1.05
oAMI	0.3750	-2	0.285	0.9944	0	1.32	0.072	2	2	0.16	1.26
pAMI	0.3750	-2	0.295	0.9773	0	1.01	0.073	3	2	0.16	0.91
PARA	0.1818	-2	0.313	0.9975	0	1.44	0.075	6	2	0.16	1.11
oCRE	0.1250	-1	0.332	0.9975	-0.451	1.63	0.075	2	1	0.16	1.16
mCRE	0.1250	-1	0.334	0.9990	-0.216	1.71	0.072	6	1	0.16	1.34
pCRE	0.1250	-1	0.321	0.9983	-0.437	1.70	0.073	5	1	0.16	1.29
pCmC	0.1111	-1	0.332	0.9977	-0.292	1.60	0.074	9	2	1.96	0.79
oCHL	0.1250	-1	0.315	0.9976	-0.577	1.73	0.080	9	2	0.16	0.93
oNIT	0.1000	-1	0.368	0.9994	-0.387	1.70	0.074	11	2	0.16	1.09
GUAI	0.1111	-1	0.296	0.9936	0	1.55	0.061	5	2	0.16	1.19
CGUA	0.1000	0	0.294	0.9925	-0.053	1.53	0.077	9	3	1.96	0.97
LTYR	0.3077	-2	0.271	0.9966	-0.445	1.54	0.086	10	2	0.16	1.23
DOPA	0.4545	-2	0.276	0.9903	0	1.07	0.089	3	2	1.96	0.48
LDOP	0.3571	-1	0.308	0.9835	0	1.30	0.075	7	1	1.96	1.02
HYIN	0.2000	-2	0.311	0.9803	0	1.33	0.072	2	1	0.36	1.29
SERO	0.3846	-2	0.327	0.9975	0	1.40	0.063	2	1	0.04	1.14
NORA	0.4167	-2	0.253	0.9958	0	-	0.090	9	2	0.36	-
ADRE	0.3077	-1	0.251	0.9947	0	-	0.086	15	2	0.36	-
mAMI	0.3750	-2	0.345	0.9971	-0.206	-	0.072	4	2	0.16	-
o'AoC	0.3333	-1	0.314	0.9949	-0.233	-	0.054	2	2	0.36	-
TYRA	0.3000	-2	0.323	0.9979	-0.434	-	0.074	9	1	0.16	-
mNIT	0.1000	-1	0.332	0.9982	-0.217	-	0.075	9	2	0.16	-
opDM	0.1111	-1	0.326	0.9976	-0.444	-	0.075	5	1	0.36	-
pCHL	0.1250	-1	0.325	0.9989	-0.542	-	0.074	9	2	0.36	-
opDC	0.1111	0	0.313	0.9972	-0.404	-	0.072	0	3	0.36	-
pNIT	0.1000	-1	0.357	0.9940	-0.393	-	0.079	6	2	0.16	-

Charge analysis was performed for the phenolic carbon atoms to identify types of aromatic reactivity that would be related to phenolic polymerization. In general, positions occupied by hydrogen atoms at the benzene ring can suffer electrophilic

or nucleophilic aromatic substitution, while halogen atoms can be eliminated in a nucleophilic aromatic substitution. Phenols undergo radical-radical and even radical-anion reactions in which halogen atoms can be eliminated in radical-nucleophilic mechanism S_{NR1} [17]. On the other hand, *ortho*- and even *para*-, *para'*-positioned hydrogen atoms are substituted in phenolic radical-radical reactions [3,24,51]. Classical rules for strong, moderately strong and weak *ortho*-*para* and *meta* orientations for electrophilic and nucleophilic substitution at the benzene ring [21,52], aided in identifying the sites for the two aromatic substitution mechanisms. Cut-off value of -0.24 in NPA charges for the phenolic carbon atoms was established, what enabled identification of the reaction sites nature that could not be resolved by the classical rules. Nucleophilic substitution occurs when the carbon NPA charge is more positive than the cut-off, otherwise the site prefers electrophilic substitution. Figures 4.3 and 4.4 show the sites for radical-nucleophilic aromatic substitution in phenolics studied in this work.

Table 4.3. Correlation matrix for dependent and independent variables used for QSPR.

Correlations between molecular descriptors					
Descriptors for the Pt QSPR model					
	Mor06u	Qcnpa	Ar	QNUnpa	$\log(\Delta j_{Pt})$
HBD/N	-0.579	-0.582	-0.425	0.525	-0.767
Mor06u	1.00	0.262	0.196	-0.238	0.558
Qcnpa		1.00	0.347	-0.686	0.777
Ar			1.00	-0.512	0.580
QNUnpa				1.00	-0.725
Descriptors for the BDD QSPR model					
	DISPm	C-026	L(nCaH)		$\log(\Delta j_{BDD})$
D-dCO	0.305	0.316		0.579	-0.585
DISPm	1.00	0.348		0.408	-0.512
C-026		1.00		0.464	-0.509
L(nCaH)				1.00	-0.662

Correlation analysis (Table 4.3) shows rather fair correlations between descriptors and the dependent variables (absolute values of correlation coefficients vary from 0.51 to 0.78), whilst there are low to moderate correlations between the descriptors (absolute correlation coefficients are in the range from 0.20 to 0.69). In the case of the BDD electrode, the descriptor-passivation correlations are somewhat smaller. This difference between the two electrodes is perhaps related to the fact that the electropassivation of the Pt electrode has been observed more intensely than that of the BDD electrode. Variable selection reconfirms these differences

between the electrodes and perhaps between the corresponding electropassivation mechanisms (Table 4.2). While descriptors for the BDD electrode are predominantly steric/topological (only D-dCO contains some electronic effects), descriptors for the Pt electrode account for hydrogen bonding (HDB/N), electronic (Qcnpa, Ar and QNUnpa) and some steric (Mor06u) properties of the twenty phenols.

These facts can account for differences in electrode-phenolics interactions and chemical composition of the electroactive electrode surfaces. However, more detailed inspection of the descriptors role is necessary to consider, as follow. Negative contribution of HBD/N to $\log(\Delta j_{Pt})$ means that a small number of polar hydrogen atoms is preferred for high passivation. Groups with polar hydrogens (from -NH₂, -NH, -OH and -CO₂H groups *i.e.* hydrogen bonding groups) can play several roles in phenolic passivation. They can participate in chemical reactions in which branched and less dense polymers are formed. This includes formation of cavities filled with the solvent (water) which interacts with free hydrogen bonding groups. The hydrogen bonding groups can interact with the electrode surface or bulk solvent. These processes can slow down radical formation and polymerization reaction, resulting in lower passivation. The molecular shapes descriptor Mor06u is related to steric aspects of the phenolic polymerization. In general, more branched (less compact) molecules have smaller values of Mor06u, accounting for lower fouling. This situation is similar to that of the HBD/N, because more compact phenolic molecules will form denser polymers.

The increase of Qcnpa results in higher passivation. More positive atomic charge of the carbon from phenolic C-OH means less negative charge at the O (the two NPA charges are highly correlated for phenolic molecules, with correlation coefficient 0.83). The descriptor Ar is an average measure of the benzene ring's aromaticity. As higher the ring's aromaticity (Ar) higher is the passivation. Both Qcnpa and Ar are probably related to chemical reactivity of phenolics, and perhaps to their interactions with the electrode.

QNUnpa (Table 4.2), a quantum-chemical reactivity index, is directly related to the number of reaction sites for nucleophilic aromatic substitution (marked with asterisks in Figures 4.3 and 4.4). The more negative the QNUnpa, the higher the

phenolic passivation on the Pt electrode. This descriptor is related to the stability of intermediates in radical-mediated reactions. The first phenolic oxidation followed by radical formation is a crucial step for phenolic passivation. The radicals act like electrophiles by attacking substitution sites free for electrophilic aromatic substitution, and they also attack halogen-binding sites *via* the S_{NR}1 mechanism. Therefore, QNUnpa should be understood as an index showing how many sites undergo radical-electrophilic aromatic substitution (the sites not marked with asterisks in Figures 4.3 and 4.4) and how many halogen-binding sites can suffer radical-nucleophilic substitution (the sites are marked with asterisks in Figures 4.3 and 4.4).

Descriptors for the BDD electrode can be also rationalized in a way similar to the descriptors for the Pt electrode in the following text. Fluoride ion F⁻ was used as a quantum chemical probe to simulate hydrogen bonding and electron withdrawal/oxidation of phenolic compounds. Negative contribution of D-dCO to log(Δj_{BDD}) can be interpreted as an external interference into electron structure of the phenolic molecule by means of intermolecular interactions or chemical reaction. Hydrogen withdrawal causes geometrical changes in the C-O group, visible as the bond shortening due to the charge loss at O and stronger delocalization between the benzene ring and the CO group. Some phenolic molecules are more resistant to such changes due to the stability of the heteroaromatic system, so the corresponding D-dCO values are small.

Less flexible phenolics tend to form more homogeneous and denser polymers that are better absorbed at the electrode surface, as has been already noticed from the Pt data set. This fact is visible from three descriptors. DISPm describes conformational features of the samples. Low values of DISPm are expected for more rigid molecules that exhibit higher passivation due to the absence of long and flexible substituents and the presence of unsaturated bonds. C-026 shows that the absence of R-CX-R fragments *i.e.* branching in the molecular structure results in higher passivation. The descriptor nCaH shows that the absence of substituted Csp² atoms in the structure favors the passivation. Most compounds (16 from 20, with exception of **CGUA**, **DOPA**, **LDOP** and **SERO**) show negative correlation

between $nCaH$ and $\log(\Delta j_{BDD})$. After transforming $nCaH$ into $L(nCaH)$, this trend is preserved.

Exploratory analysis

Results of PCA and HCA for both data sets are shown in Figure 4.6. These analyses give deeper insight into the nature of the molecular descriptors and their correlations with the dependent variables at qualitative level.

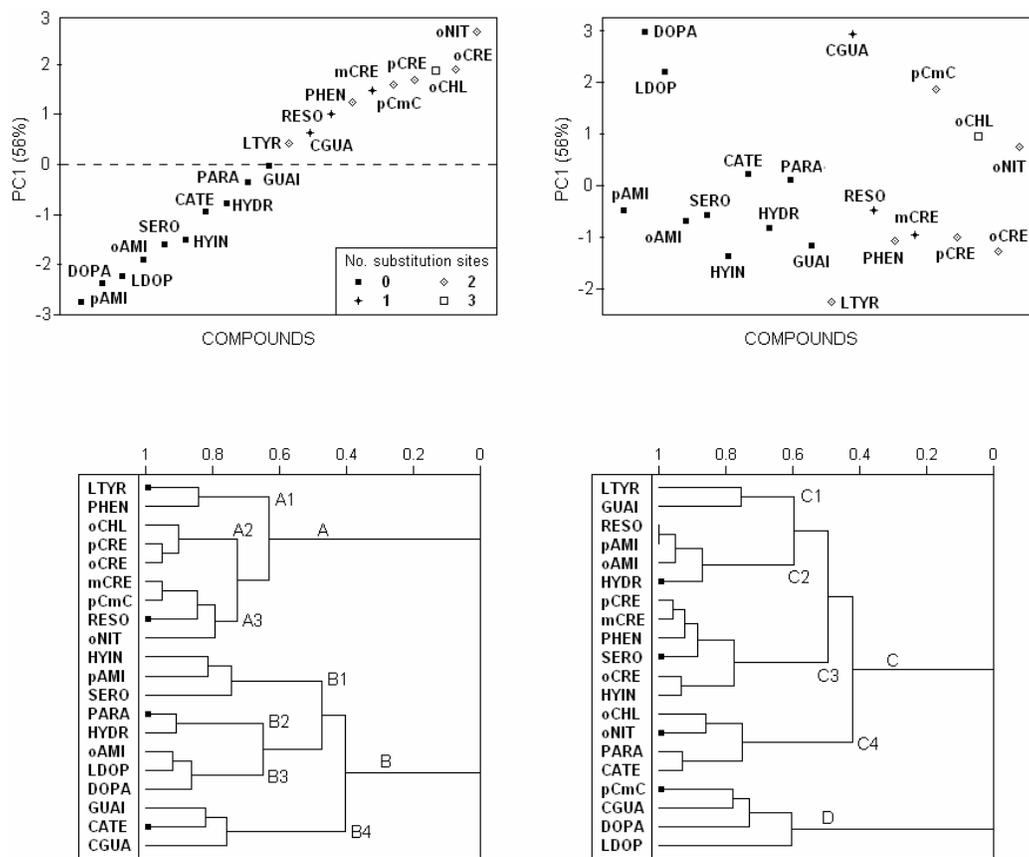


Figure 4.6. Scores plots (top figures) and HCA dendrograms (bottom figures) for molecular descriptors which were used to build the Pt model (left figures) and the BDD model (right figures). Compounds in the PCA plots are distinguished by the number of the most probable sites for nucleophilic aromatic substitution (zero and non-zero numbers are separated by the dashed line at $PC1 = 0$ for the Pt data set) and are arranged in the order of appearance along PC1 for the Pt data set. HCA sub-clusters are based on passivation and structural features of the compounds. Phenols which form external validation sets are marked with solid squares in the dendrograms.

Molecular structures of the twenty phenolics in Figure 4.3 are arranged exactly in the same order as the samples are positioned along PC1 (from left to right) in the PCA plot for the Pt data set (Figure 4.6 left top). This direction can be considered as increasing passivation, what can be easily verified by comparing the samples display and the experimental passivation data (Table 4.2). The number of sites for nucleophilic aromatic substitution (derived from Figure 4.3) can be zero (negative PC1) or can vary between 1 and 3 (PC1 is positive). These four groups of samples can be easily visualized in the scores plot, with elevated tendency for the nucleophilic substitution along PC1. When the same analysis of the scores plot for the BDD data (Figure 4.6 right top) is performed, rather different and irregular trends in the samples characterization can be observed. The passivation on the BDD electrode increases with the decrease in PC1, but the pattern of reactivity (number of substitution sites and corresponding charges) obviously does follow some other trend than in the PCA plot for the Pt data. This, once again, suggests rather different nature of phenolic-platinum and phenolic-diamond electrode interactions.

The PC1 for the Pt and BDD data sets can be expressed by the following equations, where $[]_{\text{au}}$ denotes autoscaled forms of the original descriptors:

$$PC1_{\text{Pt}} = -0.505 [\text{HBD/N}]_{\text{au}} + 0.343 [\text{Mor06u}]_{\text{au}} + 0.480 [\text{Qcnpa}]_{\text{au}} + 0.393 [\text{Ar}]_{\text{au}} - 0.492 [\text{QNUnpa}]_{\text{au}} \quad 4.1$$

$$PC1_{\text{BDD}} = 0.632 [\text{DdC-O}]_{\text{au}} + 0.625 [\text{DISPm}]_{\text{au}} + 0.382 [\text{C-026}]_{\text{au}} - 0.253 [\text{L(nCaH)}]_{\text{au}} \quad 4.2$$

According to the nature and contribution of the descriptors in these expressions, the passivation process on the Pt electrode accounts for interactions/reactions which have 25% of hydrogen bonding character (HBD/N), 27% steric (Mor06u and Ar) and 47% electronic/reactivity nature (QNUnpa). The same analysis for the BDD data points out mainly steric/geometrical (79%, DdC-O and DISPm) and compositional/topological (21%, C-026 and L(nCaH)) nature that determine phenol-electrode interactions.

The HCA dendrograms consists of clusters, sub-clusters and patterns that are rather well recognizable. The dendrogram for the Pt data (Figure 4.6 bottom left) consists of the clusters A and B which correspond to the non-zero and zero numbers of substitution sites (see the corresponding PCA plot), with exception of

CGUA. This compound belongs to the B instead of A cluster because it is the only sample with zero value of the descriptor Mor06u (Table 4.2). It seems that the reactivity of the phenolic rings in this dendrogram is the main criteria for classification. This reactivity can be expressed in terms of *ortho*-/*para*- and *meta*-positioned activators or deactivators in *nucleophilic aromatic substitution*, what depends on the nature of substituents, their position and mutual interaction. Here, it will be verified the influence of the substituent in the basic phenol structure.

All samples in the cluster A have *meta*- and/or *meta'*- and *ortho*-halogenated substitution sites (Figure 4.2). The sub-cluster A1 is characterized with moderate passivation ($\log(\Delta j_{Pt}) = 1.54-1.55$) and despite of the presence of bulky group (very weak activators) in the **LTYR** in regard to **PHEN** (absence of activators or deactivators), the reactivity in both compounds are very similar. The sub-clusters A2 and A3 consist of samples with high passivation (1.60-1.73). The samples in A2 posses halogens and/or methyl that act as weak activators/deactivators. The samples in A3 contain weak deactivators and/or moderately strong/strong activators (**oNIT**).

All the samples in the cluster B contain some deactivators. The sub-clusters B1-B3 consist of phenolics with very low to moderate passivation (1.01-1.44) and which have strong *ortho/para*-placed deactivators (OH, NH₂, NHR): B1 only with *para*-activators (amino groups), B2 with *para*- OH/NHR, and B3 with *ortho*-/*para*-OH or NH₂. The sub-cluster B4 consists of samples with low to moderate passivation (1.21-1.55). These samples have a weak *ortho*-deactivator, with additional weak *para*-deactivator in **CGUA**.

The dendrogram for the BDD data (Figure 4.6 right bottom) can be explained in terms of passivation ranges and some molecular features that are rather different from those for the Pt data set. Two simple properties can illustrate phenolics behavior at the BDD electrode: molecular topology expressed as the number of substituents on the phenolic ring (a fused heteroaromatic ring in **SERO** and **HYIN** is counted as one substituent), and average molecular polarity expressed as the C/X ratio (the number of carbon atoms divided by the number of X atoms, where X = O, N and Cl). While samples in the C cluster range from low ($\log(\Delta j_{BDD})=0.93$) to high (1.34) passivation and have at most one substituent, those in the D cluster

have very low to low passivation (0.48-1.02) and two substituents. The latter compounds have $C/X \leq 3.5$ what makes them rather polar species. Among the C1-C4 sub-clusters, C3 is the only one containing samples with one or none (**PHEN**) substituent. These samples have moderate to high passivation (1.07-1.34) and $C/X > 3$ meaning that they are rather amphiphilic. Smaller ratio ($C/X = 3$) and broad range of passivation (0.91-1.26) characterize the sub-cluster C2. While the C1 sub-cluster has high passivation (1.19-1.23) and $C/X \leq 3.5$, for C4 the samples show low passivation (0.93-1.11) and polarity relatively high ($C/X \leq 3$).

The BDD dendrogram points out possible interactions of rather steric nature between the phenolics and BDD electrode. There must be various types of putative interactions which depend on the structure of phenolic molecules and their size/shape properties. Lower branchness and amphiphilicity as preferred phenolic molecular features indicate the existence of limited domains or places of definite geometry to accommodate phenolic species at the BDD electrode, what agrees with findings of Chang et al [39].

Quantitative structure-property relationships (QSPR)

Statistics of the two PLS models is presented in Table 4.4. Prediction power of the models is demonstrated in Table 4.5 and Figure 4.7, whilst leave- N -out crossvalidation and Y-randomization of the models are presented in Figure 4.8. Some statistical conditions, as defined by Tropsha *et al.* [53,54] were applied to verify the prediction power of the models. These conditions include: a) all correlation coefficients are $R^2 > 0.6$ and $Q^2 > 0.5$ for the training sets; b) average Q^2 from leave- N -out crossvalidations must be high; c) the values R^2 and Q^2 from Y-randomizations must be small. The Pt model satisfies well all these conditions, whilst the BDD model shows to be weaker. Both models use only one component describing 56% of the total variance. Most statistical parameters show that the Pt model is more reliable than the BDD model. Pronounced differences between the models are obvious from the correlation coefficients, especially Q^2 from various validation approaches especially from leave- N -out crossvalidation (Figure 4.8 top). However, there was no chance correlation found for both models (Figure 4.8 bottom).

Table 4.4. Comparison of the Pt and BDD PLS models with basic statistics

Parameters	Pt model	BDD model
Componente (%Var) ^a	1 (55.6%)	1 (55.5%)
RMSECV	0.097	0.159
Q^{2b}	0.786	0.333
R^{2b}	0.851	0.586
Leave- <i>N</i> -out crossvalidation ^c		
$\langle Q^2_{LNO} \rangle$	0.795	0.350
y-randomization		
Q^2_{Yrand} (maximum)	-0.058	-0.051
R^2_{Yrand} (maximum)	0.281	0.326
External validation (ev)		
RMSECV _{ev}	0.095	0.177
RMSEC _{ev}	0.080	0.140
Q^2_{ev}	0.807	0.245
R^2_{ev}	0.881	0.586
Regression vector	HBD/N: -0.276 Mor06u: 0.201 Qcnpa: 0.279 Ar: 0.208 QNUnpa: -0.261	DdC-O: -0.263 DISPm: -0.230 C-026: -0.229 L(nCaH): -0.298

^aNumber of used components and the corresponding % variance of the **X** data matrix.

^bSquares of correlation coefficients of the PLS model: Q^2 - correlation coefficient of leave-one-out crossvalidation, R^2 - correlation coefficient of prediction (calibration).

^cAverage Q^2 correlation coefficient of leave-*N*-out crossvalidations where *N* = 1, 2, ..., 8.

Predictions obtained from the Pt model are rather fair, even when external validation is applied (Table 4.5). There are only three molecules with relative errors between 10 and 15% (Figure 4.7). As expected according to the parameters in Table 4.4, the BDD model has several samples with relative errors above 10%: four samples with errors between 10 and 15%, two samples with errors between 15 and 20%, one sample with error above 20% (**pAMI**), and one sample with large error of 60% (**DOPA**, Figure 4.7). Both models show very similar prediction trends when externally validated. It is interesting to note that, even when 4 samples were used in the training sets during the external validation, the prediction power of the BDD model did not alter significantly. It can be concluded that the Pt model is a parsimonius model with high predictive power, and though, recommended for practical purposes. However, the BDD model does not posses such predictive ability and cannot be recommended for prediction of phenolic passivation on the BDD electrode.

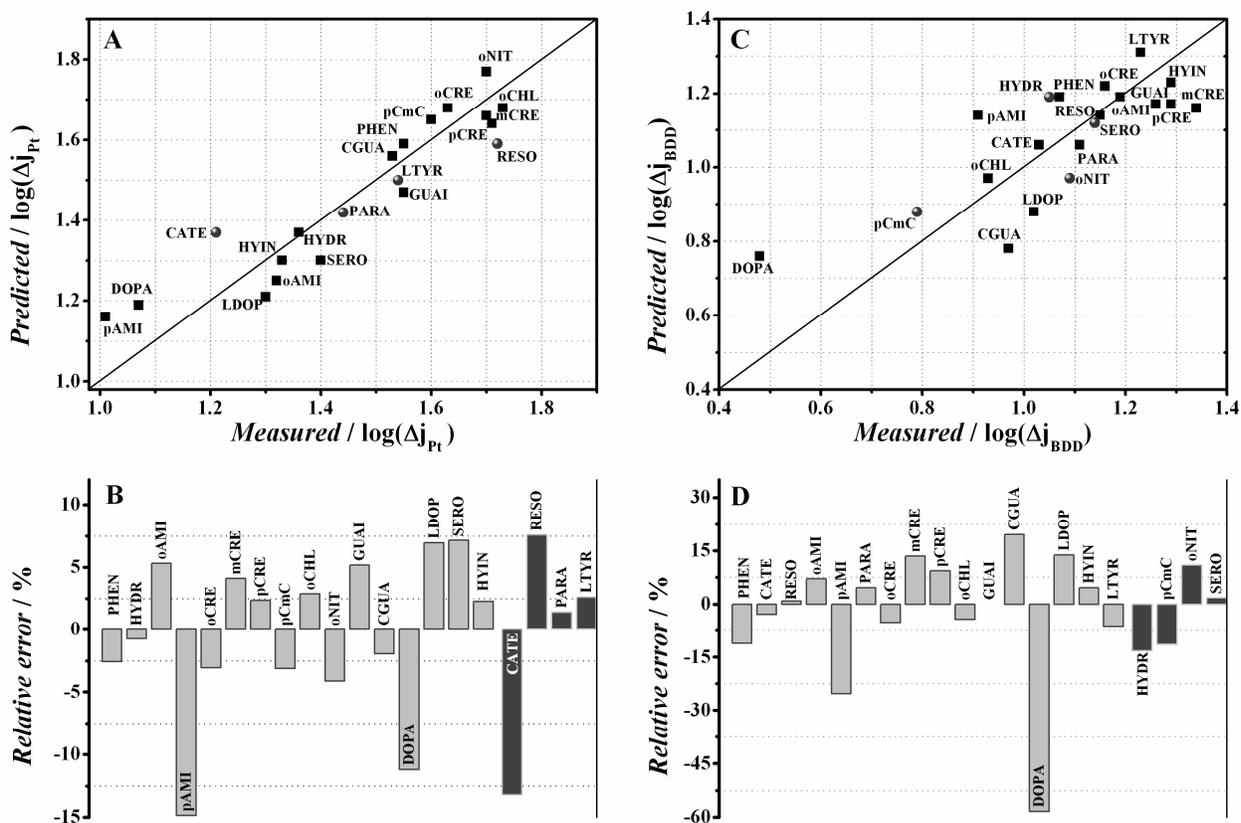


Figure 4.7. Prediction capacity using PLS models (A–C) and relative error (B–D) for Pt model and BDD model, respectively. Key for graphics A and C: Predictions for calibration compounds (■); predictions for external compounds (●). Key for graphics B and D: light gray - relative error for calibration compounds; dark gray - relative error for external compounds.

This model can be useful for better understanding phenolic electropassivation at the BDD electrode. Equations 4.3 and 4.4 present the regression vectors for the final models in deautoscaled form:

$$\log(\Delta j_{Pt}) = - 5.918 - 0.514 \text{ HBD/N} + 0.072 \text{ Mor06u} + 2.368 \text{ Qcnpa} + 6.860 \text{ Ar} - 0.269 \text{ QNUnpa} \quad 4.3$$

$$\log(\Delta j_{BDD}) = 1.955 - 8.404 \text{ DdC-O} - 0.015 \text{ DISPm} - 0.080 \text{ C-026} - 0.081 \text{ L(nCaH)} \quad 4.4$$

Table 4.5. Results for the QSPR models prediction capacity.

	Pt Model				BDD Model			
	Pred. ^a	Diff. ^a	EV Pred. ^b	EV Diff. ^b	Pred. ^a	Diff. ^a	EV Pred. ^b	EV Diff. ^b
PHEN	1.61	-0.06	1.59	-0.04	1.18	-0.11	1.19	-0.12
CATE ^c	1.36	-0.15	1.37	-0.16	1.05	-0.02	1.06	-0.03
RESO ^c	1.60	0.12	1.59	0.13	1.12	0.03	1.14	0.01
HYDR ^d	1.37	-0.01	1.37	-0.01	1.16	-0.11	1.19	-0.14
oAMI	1.24	0.08	1.25	0.07	1.15	0.11	1.17	0.09
pAMI	1.15	-0.14	1.16	-0.15	1.12	-0.21	1.14	-0.23
PARA ^c	1.43	0.01	1.42	0.02	1.06	0.05	1.06	0.05
oCRE	1.69	-0.06	1.68	-0.05	1.20	-0.04	1.22	-0.06
mCRE	1.65	0.06	1.64	0.07	1.17	0.17	1.16	0.18
pCRE	1.67	0.03	1.66	0.04	1.18	0.11	1.17	0.12
pCmC ^d	1.66	-0.06	1.65	-0.05	0.88	-0.09	0.88	-0.09
oCHL	1.69	0.04	1.68	0.05	0.98	-0.05	0.97	-0.04
oNIT ^d	1.79	-0.09	1.77	-0.07	1.00	0.09	0.97	0.12
GUAI	1.47	0.08	1.47	0.08	1.20	-0.01	1.19	0.00
CGUA	1.55	-0.02	1.56	-0.03	0.78	0.19	0.78	0.19
DOPA	1.18	-0.11	1.19	-0.12	0.77	-0.29	0.76	-0.28
LDOP	1.20	0.10	1.21	0.09	0.84	0.18	0.88	0.14
SERO ^d	1.29	0.11	1.30	0.10	1.13	0.01	1.12	0.02
HYIN	1.30	0.03	1.30	0.03	1.21	0.08	1.23	0.06
LTYR ^c	1.52	0.02	1.50	0.04	1.30	-0.07	1.31	-0.08
NORA ^e	1.15		1.17		0.88		0.87	
ADRE ^e	1.27		1.29		0.82		0.78	
mAMI ^e	1.46		1.44		1.12		1.13	
o'AoC ^e	1.47		1.47		-2.80		-2.48	
TYRA ^e	1.51		1.50		1.11		1.08	
mNIT ^e	1.65		1.64		1.02		1.00	
opDM ^e	1.69		1.68		1.14		1.14	
pCHL ^e	1.71		1.70		1.03		1.01	
opDC ^e	1.71		1.72		1.08		1.12	
pNIT ^e	1.73		1.71		1.03		1.03	

^aPredictions obtained from the PLS models based on training sets with 20 samples: Pred. - prediction values of passivation, Diff. - difference between experimental and predicted values of passivation.

^bPredictions obtained from the externally validated PLS models based on training sets with 4 samples: EV Pred. - prediction values of passivation, EV Diff. - difference between experimental and predicted values of passivation.

^cSamples forming the external validation set for the passivation on the Pt electrode. Corresponding predicted values and the differences are typed bold.

^dSamples forming the external validation set for the passivation on the BDD electrode. Corresponding predicted values and the differences are typed bold.

^eSamples of the prediction set with corresponding values of passivation.

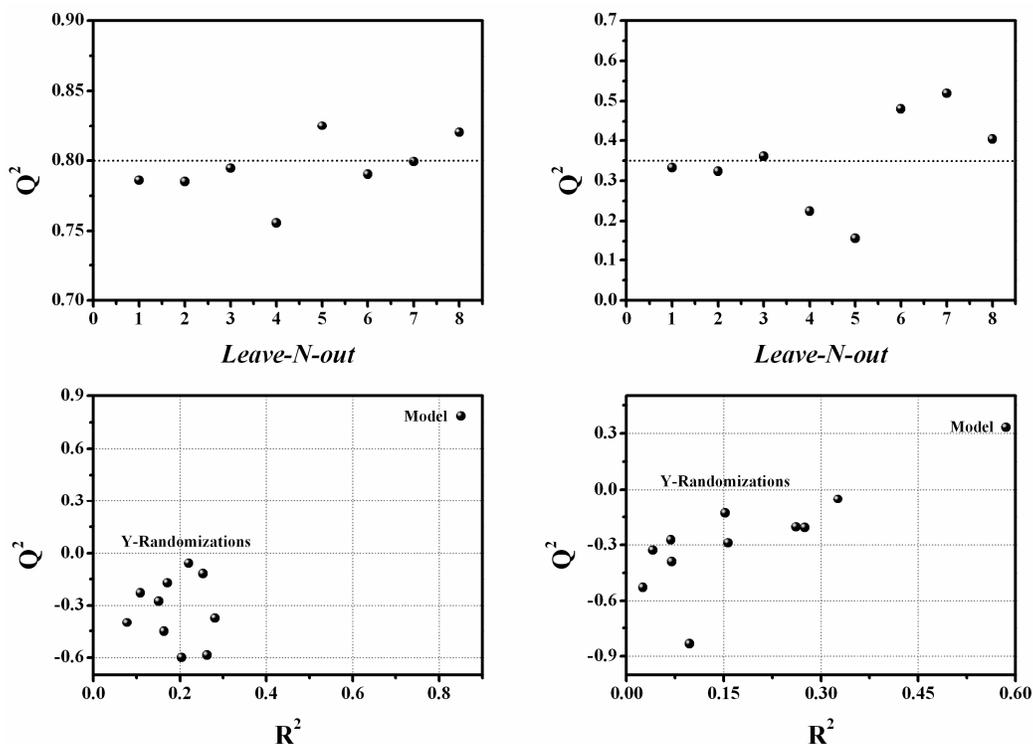


Figure 4.8. Validation of the Pt model (left figures) and the BDD model (right figures) by means of leave- N -out crossvalidation (upper figures) and Y-randomization (lower figures). Mean Q^2 values are shown by dashed lines for leave- N -out crossvalidation. Models resulting from Y-randomizations are shown by solid squares to be distinct from the validated PLS models (white squares).

Predictions for ten phenolics (Figure 4.4) are in Table 4.5. Both PLS models provide reasonable predictions even when externally validated, with exception of the prediction for **o'AoC** for which the BDD model seriously fails. The validity of these predictions can be verified by comparing predictions and corresponding deviations for similar phenolics from the training and/or external validation sets. With respect to some phenolics that are under study in this work also, the comparisons additionally confirm the reliability of the predictions in Table 4.5.

Molecular descriptors (Table 4.2) make positive or negative contributions to the passivation at the Pt or BDD electrodes, which is well noticeable from descriptor-passivation correlations (Table 4.3), regression vectors (Table 4.4), regression equations (equations 4.3 and 4.4), and especially from the exploratory analysis. Due to this fact, interpretation of the regression vectors was already presented in the correlation and exploratory analysis. It can be concluded that both

PLS models indicate more rigid and compact phenolic molecules which form polymers with elevated adsorption power with respect to the electrodes. This fact can be useful in predicting behavior of phenolics during electrooxidation, both at qualitative and quantitative level.

Conclusions

Substituent effects largely determine the electropassivation processes at the Pt and BDD electrodes. The preference of radical attack to *ortho*, *ortho'*, *meta*, *meta'* or *para* positions which are occupied with hydrogen or halogen atoms can be visualized by means of NPA partial atomic charges of the corresponding carbon atoms. The charges show whether the substitution site is electron-rich (attracting radicals) or electron-poor (not attracting radicals), and consequently, the site would undergo electrophilic or nucleophilic aromatic substitution, respectively. Classical organic chemistry approach in terms of resonance structures that show inductive and resonance effects of the substituents is a qualitative view on the NPA charges distribution. In the absence of steric effects, this charge classification is a good aid in predicting phenoxyl radical formation.

Oxidation of phenolics into phenoxyl radicals results in higher reactivity of the new species. Phenolic electropassivation at the Pt electrode is satisfactorily described by the PLS model which can be used for practical purposes. Although the PLS model for the phenolic passivation at the BDD electrode is far from being statistically relevant, it gives some idea about this phenomenon.

This is the first time that phenolic passivation is treated at a quantitative level by means of QSPR analysis. Such studies could be expanded to diverse phenolic compounds and electrodes. Such QSPR analyses may become routine in analytical laboratories which use phenols in electrochemical assays.

References

1. Gattrell, M.; Kirk, D. W. A fourier-transform infrared-spectroscopy study of the passive film produced during aqueous acidic phenol electrooxidation. *J. Electrochem. Soc.* **1992**, 139, (10), 2736-2744.

2. Gattrell, M.; Kirk, D. W. A study of electrode passivation during aqueous phenol electrolysis. *J. Electrochem. Soc.* **1993**, 140, (4), 903-911.
3. Gattrell, M.; Kirk, D. W. A study of the oxidation of phenol at platinum and preoxidized platinum surfaces. *J. Electrochem. Soc.* **1993**, 140, (6), 1534-1540.
4. Wang, J.; Jiang, M.; Lu, F. Electrochemical quartz crystal microbalance investigation of surface fouling due to phenol oxidation. *J. Electroanal. Chem.* **1998**, 444, (1), 127-132.
5. Ežerskis, Z.; Jusys, Z. Electropolymerization of chlorinated phenols on a Pt electrode in alkaline solution Part - I: A cyclic voltammetry study. *J. Appl. Electrochem.* **2001**, 31, (10), 1117-1124.
6. Salavagione, H. J.; Arias, J.; Garces, P.; Morallon, E.; Barbero, C.; Vazquez, J. L. Spectroelectrochemical study of the oxidation of aminophenols on platinum electrode in acid medium. *J. Electroanal. Chem.* **2004**, 565, (2), 375-383.
7. Gonçalves, D.; Faria, R. C.; Yonashiro, M.; Bulhoes, L. Electrochemical oxidation of o-aminophenol in aqueous acidic medium: formation of film and soluble products. *J. Electroanal. Chem.* **2000**, 487, (2), 90-99.
8. Guenbour, A.; Kacemi, A.; Benbachir, A.; Aries, L. Electropolymerization of 2-aminophenol - Electrochemical and spectroscopic studies. *Prog. Org. Coat.* **2000**, 38, (2), 121-126.
9. Guenbour, A.; Kacemi, A.; Benbachir, A. Corrosion protection of copper by polyaminophenol films. *Prog. Org. Coat.* **2000**, 39, (2-4), 151-155.
10. Mengoli, G.; Musiani, M. M. An overview of phenol electropolymerization for metal protection. *J. Electrochem. Soc.* **1987**, 134, (12), C643-C652.
11. Neves, R. S.; Motheo, A. J.; Fernandes, F.; Fartaria, R. Monte Carlo simulation of the adsorption of phenol on gold electrodes. A simple model. *J. Braz. Chem. Soc.* **2004**, 15, (2), 224-231.
12. Bartlett, P. N.; Cooper, J. M. A review of the immobilization of enzymes in electropolymerized films. *J. Electroanal. Chem.* **1993**, 362, (1-2), 1-12.
13. Chia, V.; Soriaga, M. P.; Hubbard, A. T. Kinetics of oriented adsorption - hydroquinone on platinum. *J. Phys. Chem.* **1987**, 91, (1), 78-82.
14. Lu, F.; Salaita, G. N.; Lagurendavidson, L.; Stern, D. A.; Wellner, E.; Frank, D. G.; Batina, N.; Zapien, D. C.; Walton, N.; Hubbard, A. T. Characterization of hydroquinone and related-compounds adsorbed at pt(111) from aqueous-solutions - electron energy-loss spectroscopy, auger-spectroscopy, low-energy electron-diffraction, and cyclic voltammetry. *Langmuir* **1988**, 4, (3), 637-646.
15. Soriaga, M. P.; Hubbard, A. T. Determination of the orientation of adsorbed molecules at solid-liquid interfaces by thin-layer electrochemistry - aromatic-compounds at platinum-electrodes. *J. Am. Chem. Soc.* **1982**, 104, (10), 2735-2742.

16. Iotov, P. I.; Kalcheva, S. V. Mechanistic approach to the oxidation of phenol at a platinum/gold electrode in an acid medium. *J. Electroanal. Chem.* **1998**, 442, (1-2), 19-26.
17. Ežerskis, Z.; Jusys, Z. Electropolymerization of chlorinated phenols on a Pt electrode in alkaline solution. Part IV: A gas chromatography mass spectrometry study. *J. Appl. Electrochem.* **2002**, 32, (5), 543-550.
18. Karhu, M. Formation of diphenyl ethers from cyclohexa-2,5-dienones via 4-phenoxy-4-(1-alkoxy)cyclohexa-2,5-dienones as probable intermediates. *J. Chem. Soc.-Perkin Trans. 1* **1981**, (1), 303-306.
19. Vasileva, L. V.; Shchukin, I. V.; Babkin, V. A. Investigation of electrochemical and chemical guaiacol oxidation-products. *Russ. Electrochem.* 29, (4), 609-615.
20. Nishiki, Y.; Furuta, T.; Kurosu, T. Diamond electrode, its characteristics and potential applications. *Electrochemistry* **2004**, 72, (7), 521-528.
21. Smith, M. B.; March, J. *March's advanced organic chemistry: reactions, mechanisms and structure*. 5 ed.; John Wiley & Sons Inc: New York, 2001; Vol. 1, p 2083.
22. De Carvalho, R. M.; Kubota, L. T.; Rath, S. Influence of EDTA on the electrochemical behavior of phenols. *J. Electroanal. Chem.* **2003**, 548, 19-26.
23. Ežerskis, Z.; Stalnionis, G.; Jusys, Z. Electropolymerization of chlorinated phenols on a Pt electrode in alkaline solution. Part II: An electrochemical quartz crystal microbalance study. *J. Appl. Electrochem.* **2002**, 32, (1), 49-55.
24. Armstrong, D. R.; Cameron, C.; Nonhebel, D. C.; Perkins, P. G. Oxidative coupling of phenols .8. a theoretical-study of the coupling of phenoxy radicals. *J. Chem. Soc.-Perkin Trans. 2* **1983**, (5), 575-579.
25. Ferreira, M. M. C. Multivariate QSAR. *J. Braz. Chem. Soc.* **2002**, 13, (6), 742-753.
26. Xue, C. X.; Zhang, R. S.; Liu, H. X.; Yao, X. J.; Liu, M. C.; Hu, Z. D.; Fan, B. T. An accurate QSPR study of O-H bond dissociation energy in substituted phenols based on support vector machines. *J. Chem. Inf. Comput. Sci.* **2004**, 44, (2), 669-677.
27. Bosque, R.; Sales, J. A QSPR study of O-H bond dissociation energy in phenols. *J. Chem. Inf. Comput. Sci.* **2003**, 43, (2), 637-642.
28. Nesmerak, K.; Nernec, I.; Sticha, M.; Waisser, K.; Palat, K. Quantitative structure-property relationships of new benzoxazines and their electrooxidation as a model of metabolic degradation. *Electrochim. Acta* **2005**, 50, (6), 1431-1437.
29. Ferreira, M. M. C.; Antunes, A. M.; Melgo, M. S.; Volpe, P. L. O. Chemometrics I: Multivariate calibration, a tutorial. *Quim. Nova* **1999**, 22, (5), 724-731.
30. Martens, H.; Naes, T. *Multivariate Calibration*. John Wiley & Sons: New York, 1989; p 419.

31. Teófilo, R. F.; Ceragioli, H. J.; Peterlevitz, A. C.; Da Silva, L. M.; Damos, F. S.; Ferreira, M. M. C.; Baranauskas, V.; Kubota, L. T. Improvement of the electrochemical properties of an "as-grown" boron-doped polycrystalline diamond electrodes deposited on tungsten wires using ethanol. *J. Solid State Electrochem.* **2007**, 11, (10), 1449-1457.
32. Baranauskas, V.; Peterlevitz, A. C.; Ceragioli, H. J.; Kubota, L. T.; Teófilo, R. F. Processo de fabricação de eletrodos de diamante dopado para eletroquímica. PI0600897-6, 03/16/2006.
33. Allen, F. H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr. Sect. B-Struct. Sci.* **2002**, 58 PN Part 3 Sp. Iss, 380-388.
34. Cambridge Structural Database, version 5.28, with January and May 2007 updates; The Cambridge Structural Data Centre, University of Cambridge: Cambridge, UK, November 2006.
35. Bruno, I. J.; Cole, J. C.; Edgington, P. R.; Kessler, M.; Macrae, C. F.; McCabe, P.; Pearson, J.; Taylor, R. New software for searching the Cambridge Structural Database and visualizing crystal structures. *Acta Crystallogr. Sect. B-Struct. Sci.* **2002**, 58 PN Part 3 Sp. Iss, 389-397.
36. ConQuest, version 1.9, CSD searching software; The Cambridge Structural Data Centre; University of Cambridge: Cambridge, UK, 2007..
37. Mercury, version 1.5., crystal structure viewer; The Cambridge Structural Data Centre; University of Cambridge: Cambridge, UK, 2006.
38. WebLab ViewerPro, version 4.0, high-quality molecular visualization software; Accelrys, Inc.: Burlington, MA, 2000.
39. Ziegler, T. Approximate density functional theory as a practical tool in molecular energetics and dynamics. *Chem. Rev.* **1991**, 91, (5), 651-667.
40. Becke, A. D. Density-functional thermochemistry .3. the role of exact exchange. *J. Chem. Phys.* **1993**, 98, (7), 5648-5652.
41. Lee, C. T.; Yang, W. T.; Parr, R. G. Development of the colle-salvetti correlation-energy formula into a functional of the electron-density. *Phys. Rev. B* **1988**, 37, (2), 785-789.
42. Titan, version 1.0.8, molecular modeling software; Wavefunction, Inc.: Irvine, CA, 2001..
43. Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M. *Dragon Web version, software for calculation of molecular descriptors*, 3.0; University of Milano - Bicocca: Milano, 2003.

44. Wold, S.; Eriksson, L. Statistical Validation of QSAR Results. In *Chemometric Methods in Molecular Design*, Waterbeemd, V. d., Ed. VCH: Weinheim, 1995, pp 309-318.
45. Matlab – The Language of Technical Computing, version 7.0; The Mathworks, Inc: Natick, MA, 2004.
46. Pirouette, version 3.02, multivariate data analysis software; Infometrix, Inc.: Woodinville, WA, 2001.
47. Alehashem, S.; Chambers, F.; Strojek, J. W.; Swain, G. M.; Ramesham, R. Cyclic voltammetric studies of charge-transfer reactions at highly boron-doped polycrystalline diamond thin-film electrodes. *Anal. Chem.* **1995**, 67, (17), 2812-2821.
48. Granger, M. C.; Witek, M.; Xu, J. S.; Wang, J.; Hupert, M.; Hanks, A.; Koppang, M. D.; Butler, J. E.; Lucazeau, G.; Mermoux, M.; Strojek, J. W.; Swain, G. M. Standard electrochemical behavior of high-quality, boron-doped polycrystalline diamond thin-film electrodes. *Anal. Chem.* **2000**, 72, (16), 3793-3804.
49. Muna, G. W.; Tasheva, N.; Swain, G. M. Electro-oxidation and amperometric detection of chlorinated phenols at boron-doped diamond electrodes: A comparison of microcrystalline and nanocrystalline thin films. *Environ. Sci. Technol.* **2004**, 38, (13), 3674-3682.
50. Sarada, B. V.; Rao, T. N.; Tryk, D. A.; Fujishima, A. Electrochemical oxidation of histamine and serotonin at highly boron doped diamond electrodes. *Anal. Chem.* **2000**, 72, (7), 1632-1638.
51. Kiralj, R.; Ferreira, M. Predicting bond lengths in planar benzenoid polycyclic aromatic hydrocarbons: A chemometric approach. *J. Chem. Inf. Comput. Sci.* **2002**, 42, (3), 508-523.
52. Morrison, R. T.; Boyd, R. N. *Organic Chemistry*. 3rd ed.; Allyn and Bacon, Inc: Boston, 1981; .
53. Tropsha, A.; Gramatica, P.; Gombar, V. K. The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb. Sci.* **2003**, 22, (1), 69-77.
54. Golbraikh, A.; Tropsha, A. Beware of q(2)! *J. Mol. Graph.* **2002**, 20, (4), 269-276.

Capítulo 5

Determinação Simultânea de DA, AU e AA com Diamante Dopado com Boro usando Voltametria e Métodos Quimiométricos

“Entre as dificuldades se esconde a oportunidade”

Albert Einstein

Introdução

A detecção eletroquímica é um dos métodos mais elegantes da química analítica [1]. As metodologias que empregam técnicas baseadas neste tipo de detecção satisfazem muitas das exigências para as tarefas a que são confiadas devido a sua inerente especificidade, velocidade de resposta, sensibilidade e simplicidade. Além disso, as técnicas eletroanalíticas são versáteis, de baixo custo e tem a capacidade de atingir baixos limites de detecção.

Atualmente, muitas pesquisas em eletroanalítica tem se dedicado à determinação seletiva de dopamina (DA) e ácido úrico (AU) na presença de interferentes como o ácido ascórbico (AA) [2-12].

A dopamina e outras catecolaminas exercem um papel importante nos sistemas nervoso central, cardiovascular, renal e hormonal [13]. Uma baixa produção de dopamina está intimamente relacionada a desordem neurológica conhecida como mal de Parkinson e, por outro lado, uma alta produção deste neurotransmissor pode ser relacionada à desordens psicológicas, tais como a esquizofrenia [14]. O ácido úrico é o produto final do catabolismo das bases purínicas [14]. Níveis anormais de AU são sintomas de várias doenças tais como hiperuricemia, gota, síndrome de Lesch-Nyhan e também estão relacionados à imunodeficiência e ao infarto [14-16]. O ácido ascórbico (vitamina C) é um

componente vital na dieta humana e tem sido usado para a prevenção e tratamento de resfriados, indisposições mentais, infertilidade, câncer e em algumas manifestações clínicas de infecções provocadas pelo vírus HIV [17].

Faz-se, portanto, extremamente necessário o desenvolvimento de métodos analíticos rápidos, simples, sensíveis, seletivos e estáveis para a determinação simultânea destes compostos em fluidos biológicos, principalmente para análises de rotina. Neste sentido, os métodos eletroanalíticos apresentam características versáteis para a análise destes compostos eletroativos, como por exemplo, para análise direta, análise *in vivo* [18], detecção eletroquímica em métodos cromatográficos ou eletroforéticos [19,20] e possibilidade de acoplamentos com métodos espectroscópicos [21].

Entretanto, alguns problemas da detecção eletroquímica simultânea destes analitos em matrizes biológicas usando eletrodos sólidos convencionais (*e.g.* carbono vítreo) [5] limitam o uso desta metodologia para a análise de rotina. São eles: (1) A DA, o AA e o AU muitas vezes coexistem na maioria dos fluidos biológicos e a oxidação destes analitos está próxima de um mesmo potencial para a maioria dos eletrodos sólidos, o que resulta em sobreposição da resposta voltamétrica, fazendo com que a discriminação de cada composto seja extremamente difícil [4,5,6,10,22]; (2) geralmente, as oxidações diretas são irreversíveis e requerem altos potenciais [5]; (3) eletrodos sólidos convencionais sofrem o efeito de bloqueio eletroquímico, o que compromete a precisão, a sensibilidade e a viabilidade do método e; (4) o AA e o AU, normalmente, são encontrados em concentrações relativamente muito maiores que a da DA, o que compromete ainda mais a falta de seletividade. As concentrações de DA em fluidos biológico em nível basal está na faixa de 0,01 a 1,0 $\mu\text{mol L}^{-1}$; as concentrações de AU podem ser encontradas na faixa de 240 a 520 $\mu\text{mol L}^{-1}$ em sangue e de 1,4 a 4,4 mmol L^{-1} em urina e as concentrações de AA podem ser encontradas na faixa de 0,05 a 0,1 mmol L^{-1} em sangue e de 100 a 500 $\mu\text{mol L}^{-1}$ em fluido extracelular do sistema nervoso central [4,22-24].

Embora haja muitas dificuldades, diversos trabalhos divulgados na literatura propõem modificações físico-químicas na superfície do eletrodo no esforço de contornar os problemas relacionados à determinação simultânea de DA, AU e AA,

empregando métodos eletroquímicos [4,5,10,22,24]. Alguns exemplos são o uso de microeletrodo modificado com DNA [25], aplicações com monocamadas auto-organizadas [26-28], eletrodo de grafite esfoliado [29], eletrodos de argila ou de cerâmica (modificadas) [5,24] e eletrodos modificados com complexos [1,22,30]. A maioria destes eletrodos modificados, infelizmente, apresenta algum tipo de desvantagem que limitam seu uso, tais como: lixiviação do mediador, pobre estabilidade para uso por longo período de tempo, algumas preparações são de alto custo e/ou difíceis e; além disso, adsorções irreversíveis podem acontecer e a necessidade de renovação da superfície após cada medida pode dificultar as análises de rotina [5,31].

Assim, um eletrodo ideal para os eletroanalíticos, além de todas as suas características inerentes, deve principalmente ser estável por um longo período de tempo (*e.g.* meses) e resistente ao bloqueio eletroquímico. Recentemente, o diamante dopado com boro tem apresentado estas duas principais características, entre outras desejáveis do ponto de vista eletroquímico [32-36], como está descrito nos Capítulos 2 e 3.

Porém, para compostos que se oxidam próximos a um mesmo potencial, uma baixa seletividade será ainda observada. Isto ocorre, pois, a complexidade de sistemas naturais pode contribuir com sinais interferentes, o que dificulta a discriminação das respostas e assim, a análise quantitativa se torna complicada. Neste sentido, o uso de ferramentas quimiométricas em eletroanálise, tais como a calibração multivariada [37], aliada à robustez do diamante dopado com boro, apresenta-se como uma excelente alternativa para contornar o problema de falta de seletividade.

As ferramentas quimiométricas são extremamente úteis em química analítica quando problemas de sobreposições de sinais são observados, uma vez que informações de múltiplos potenciais são consideradas e assim, modelos seletivos podem ser obtidos e quantificações podem ser realizadas mesmo na presença de interferentes, desde que estes sejam modelados [38-41].

No presente trabalho é apresentado um estudo de simulação de misturas para a determinação simultânea de DA, AU e AA usando voltametria de onda quadrada (SWV), eletrodo de BDD e ferramentas quimiométricas.

Planejamentos fatoriais e planejamentos para superfície de resposta foram empregados para encontrar as melhores condições operacionais da análise [42]. Quadrados mínimos parciais (PLS) foi empregado para construir os modelos de calibração multivariada [38,41,43-45] e cálculo de parâmetros analítico tais como seletividade, sensibilidade e limite de detecção foram calculados para os modelos de calibração [46,47].

Portanto, um estudo preliminar para o desenvolvimento de métodos eletroanalíticos diretos, que permitem a determinação simultânea, sem prévia separação da DA, AU e AA é o objetivo desta pesquisa.

Experimental

Reagentes e soluções

Os reagentes, dopamina, ácido úrico e ácido ascórbico foram adquiridos da SigmaTM e o fosfato de sódio foi adquirido da SynthTM. Todos os reagentes são de grau analítico sem purificação adicional. Todas as soluções foram preparadas com água ultra pura de um sistema de purificação Milli QTM (>18 MΩ cm). As soluções estoque de cada analito foram preparadas com uma concentração de 0,5 mol L⁻¹ em água e as misturas foram preparadas do estoque por diluições em tampão como necessário.

Medidas Eletroanalíticas

As medidas eletroquímicas foram executadas em uma célula com três eletrodos: o eletrodo de calomelano saturado (SCE) foi usado como referência, uma espiral de platina foi usada como contra eletrodo e o BDD foi utilizado como eletrodo de trabalho. Todas as medidas foram realizadas em um potenciostato AUTOLABTM PGSTAT 30 gerenciado pelo software GPES 4.9 a partir de um microcomputador.

Para evitar o acúmulo de possíveis filmes poliméricos formados sobre a superfície do eletrodo de trabalho durante as análises, após cada ensaio a solução foi agitada vigorosamente por aproximadamente 60 segundos. Para manter a superfície hidrogenada e assim garantir medidas reprodutivas por longo período de

tempo, após aproximadamente 15 ensaios uma varredura de 0 a -3 V sob agitação em meio ácido foi realizada. Todos os experimentos foram executados em ordem aleatória.

Parâmetros Voltamétricos

A técnica voltamétrica empregada foi a voltametria de onda quadrada (SWV). Esta técnica foi escolhida por ser geralmente mais rápida e mais sensível que outras técnicas voltamétricas de pulso. As condições foram definidas usando planejamentos experimentais e metodologia de superfície de resposta com o objetivo de encontrar as melhores condições de análise. Os estudos foram realizados para cada analito separadamente e serão detalhados nos resultados. A concentração dos analitos na célula foi fixada em $100 \mu\text{mol L}^{-1}$. A faixa de potencial investigada foi de 0,01 V a 1,05 V. A linha de base de todos os voltamogramas foi corrigida usando uma janela de média móvel (Capítulo 1). A resposta usada neste estudo foi definida de acordo com Equação 5.1.

$$r_n = \frac{i_p}{l_p \sqrt{v}} \quad 5.1$$

em que i_p é a corrente de pico em ampere (A), l_p é a largura do pico à meia altura e v é a velocidade de varredura em V s^{-1} . A divisão pela largura do pico foi adicionada na equação com o objetivo de se obter melhor resolução. A divisão pela raiz quadrada da velocidade de varredura é necessária, pois, de acordo com Swain (1993) [48] no BDD de alta qualidade a corrente de pico de oxidação varia linearmente com a raiz da velocidade de varredura para todos os analitos. Isto indica uma corrente controlada por difusão linear semi-infinita do analito na zona de reação interfacial. Assim, nem sempre um aumento na corrente devido a altas velocidades de varredura significa um ganho real de corrente. A divisão por este fator minimiza a influência da velocidade de varredura durante o estudo.

Conjunto de dados

Variáveis Dependentes

Um planejamento fatorial parcial para mistura com cinco níveis foi elaborado de acordo com Munoz (1998) [49] para obter as variáveis dependentes (concentrações), o que garantiu a ortogonalidade destas variáveis. Isto significa que este planejamento extrai e separa os efeitos de diferentes fatores (componentes da mistura) igualmente. As misturas de DA, AU e AA foram elaboradas sinteticamente a partir de padrões conforme descrito pelo planejamento parcial.

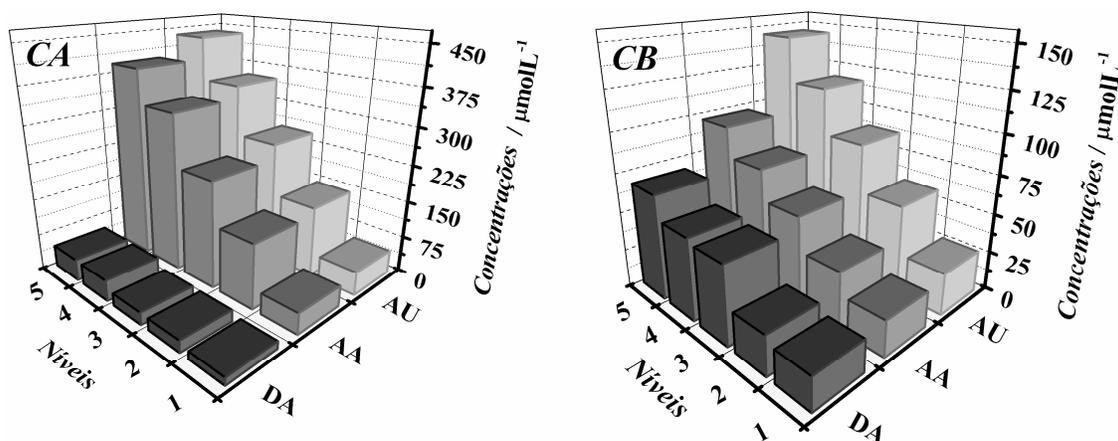


Figura 5.1. Níveis e concentrações estudadas para os três analitos. (CA) concentrações altas de AA e AU em relação à DA, (CB) concentrações próximas de uma mesma faixa de concentração.

Dois conjuntos de dados foram analisados: um deles as concentrações médias de AA e AU são aproximadamente 8 vezes maiores que a de DA (CA) e outro as concentrações médias de AA e AU são aproximadamente 2 vezes maiores que a de DA (CB), conforme apresentado na Figura 5.1. Um total de 37 amostras foi obtido para o conjunto CA e de 32 amostras para o conjunto CB.

Variáveis independentes

Na faixa de potencial investigada (0,01 a 1,05 V), 525 variáveis independentes foram obtidas. As colunas da matriz de dados foram centradas na média, corrigidas em relação à linha de base usando a média móvel e os potenciais foram selecionados com o objetivo de se obter as melhores previsões. O número de

fatores para o modelo foi determinado por validação cruzada aplicando o método *leave-n-out* e se baseou no cálculo do RMSECV, conforme equação 1.45. A habilidade de previsão do modelo foi avaliada pelo erro médio quadrático de previsão (RMSEP) e pelo coeficiente de correlação conforme equações 1.47 e 1.48, respectivamente. Os parâmetros analíticos, tais como seletividade, sensibilidade e limite de detecção, foram estimados conforme metodologias apresentadas por Lorber et al. (1997) [50], Bro e Andersen [46,47], Boqué e Rius [51] e Ortiz et al. 2003 [52].

Seleção de variáveis e amostras

As variáveis independentes foram selecionadas usando o método de seleção das variáveis ordenadas (OPS) [53,54] e o algoritmo genético (GA). Os parâmetros usados como entrada do algoritmo GA foram investigados empregando uma matriz saturada de Plackett-Burman [55-57], com sete variáveis sendo uma fictícia conforme mostrado na Tabela 5.1. O objetivo é encontrar a melhor configuração deste algoritmo.

Tabela 5.1. Variáveis e níveis investigados do algoritmo genético

Parâmetros	Níveis		
	-1	0	+1
População	52	100	152
Largura da janela	2	1	3
Taxa de Mutação	0,001	0,003	0,005
Inicialização	20	35	50
<i>Cross over</i> (variável fictícia)	2	2	2
Gerações	50	100	150
Convergência	20	45	70

Para cada conjunto e analito, dez ensaios foram executados em triplicatas, usando o algoritmo PLS com uma validação cruzada aleatória com *split* de 5 amostras em 5 iterações.

Em ambos os métodos, as variáveis com melhor capacidade de previsão na validação cruzada *i.e.*, com menor valor de RMSECV (equação 1.45), foram usadas para comparação.

A seleção de variáveis foi executada usando todas as amostras. Após a seleção ter sido realizada, os conjuntos de calibração e validação foram separados. Este procedimento é realizado por dois motivos: (1) o modelo deve ser representativo para todas as amostras; (2) deve-se evitar a especificidade do modelo apenas para a previsão de um conjunto que pode ser não representativo da população com todas as amostras, principalmente para pequenos conjuntos de dados.

As amostras retiradas para validação externa, além de serem representativas, não devem promover nenhuma grande modificação no vetor de regressão que é obtido quando todas as amostras estão presentes. Para isto, o método de Kennard e Stone (1969) [58] foi empregado para selecionar as amostras para previsão. Este método consiste na construção de um planejamento para superfície de resposta com a matriz de variáveis independentes transposta. O método é sequencial e cada etapa cobre a região experimental uniformemente [58]. As amostras a serem selecionadas são aquelas que são primeiramente usadas como padrões para a ortogonalidade e normalmente orientam outras amostras a ficarem dentro do planejamento.

Análise dos dados

As análises de dados relativas aos planejamentos experimentais foram realizadas empregando planilhas eletrônicas do software Microsoft Excel™ 2003 [42]. Para a construção dos modelos de calibração, cálculos do NAS e dos parâmetros analíticos foram realizados usando funções *.m* para o software Matlab™ 7 (The MathWorks, Co., Natick, MA, USA). As funções foram programadas no Laboratório de Quimiometria Teórica e Aplicada (LQTA). Os cálculos usando o GA, foram realizados usando o pacote PLS_Toolbox 3.01 para Matlab [59].

Resultados e Discussão

Otimização das condições voltamétricas

A SWV é uma técnica diferencial em que um sinal de excitação formado por uma onda quadrada simétrica (Figura 5.2B), superposta na base de uma rampa de potencial (Figura 5.2A), é aplicado no eletrodo de trabalho [60,61] (Figura 5.2).

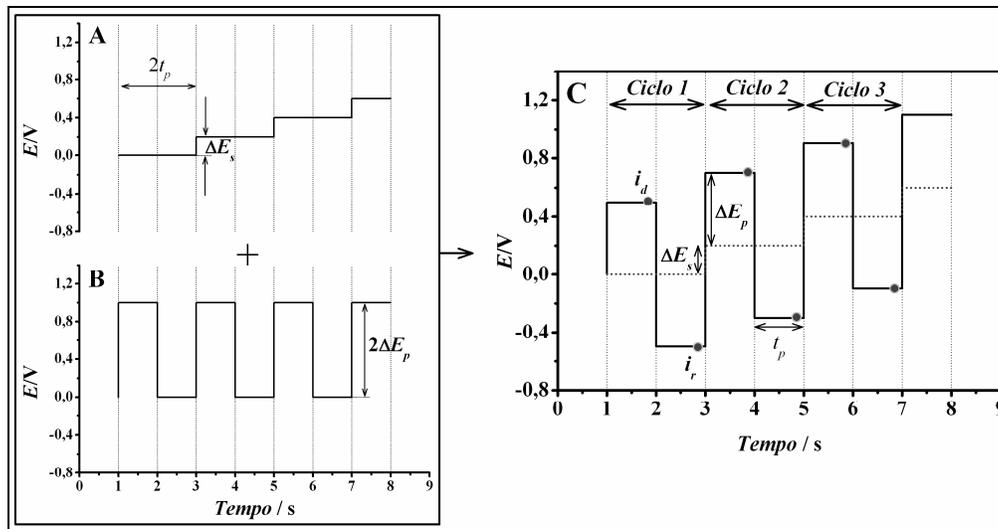


Figura 5.2. Forma de aplicação do potencial na SWV e esquema de medida. (A) varredura do potencial em rampa, (B) varredura em pulsos constantes, (C) voltametria de onda quadrada.

A onda quadrada é caracterizada por uma altura do pulso, ΔE_p medida em relação ao correspondente degrau da rampa e por uma largura do pulso t_p . Alternativamente, a largura do pulso pode ser expressa em termos da frequência da onda quadrada, $f = 1/2t_p$. O incremento por ΔE_s na rampa é o início de cada ciclo; assim, a velocidade de varredura é definida por $v = \Delta E_s/2t_p = f\Delta E_s$.

Conforme apresentado na Figura 5.2, as amostragens das correntes são realizadas duas vezes por ciclo e no fim de cada pulso. A amostragem da corrente direta (i_d) origina-se no primeiro pulso do ciclo, que está na direção do incremento na rampa. A amostragem da corrente reversa (i_r) é obtida no fim do segundo pulso, que está na direção oposta ao incremento na rampa. A diferença das correntes Δi é calculada como $i_r - i_d$. Há valores diagnósticos nas correntes direta e reversa; consequentemente elas são preservadas separadamente. Finalmente, o resultado de uma simples varredura usando a SWV é composto de três voltamogramas *i.e.*, o direto, o reverso e a diferença entre as correntes destes voltamogramas no correspondente degrau da rampa [62].

Uma das principais vantagens da SWV é a que ela é geralmente mais sensível que a voltametria de pulso diferencial para sistemas eletroquímicos reversíveis. Além disso, na forma em que as correntes são medidas prioriza-se a minimização da corrente capacitiva e os picos obtidos são mais simétricos e melhor resolvidos

em comparação com outras técnicas [63]. Porém, estas vantagens somente são obtidas desde que a SWV esteja otimizada. Neste sentido, planejamentos experimentais *i.e.*, triagem e metodologia de superfície de resposta (RSM) [42] foram empregados para encontrar as melhores condições operacionais de análise para o uso da voltametria de onda quadrada. Neste estudo, apenas a DA e o AU foram otimizados, visto que, a princípio, o AA será considerado apenas como um interferente.

Um planejamento fracionário 2^{4-1} com ponto central foi elaborado para investigar as variáveis (ou fatores) que apresentam influência significativa sobre o sistema. As variáveis estudadas foram: concentração do tampão fosfato de sódio ($[TP]$), pH do tampão (pH), amplitude ($Ampl.$) e frequência ($Freq.$). O incremento de potencial foi fixado em $0,002 \text{ V s}^{-1}$. A Tabela 5.2 mostra os fatores, níveis estudados e respostas obtidas. Observe na Tabela 5.2. que a resposta obtida para a DA está com valores que correspondem a aproximadamente a metade dos valores para o AU. Isto ocorreu principalmente devido a largura a meia altura para a DA, que apresentou valores aproximadamente duas vezes maiores que para o AU.

Tabela 5.2. Fatores, níveis investigados codificados e decodificados e respostas obtidas para a DA e AU.

Ensaio	Fatores				Respostas (r_n)	
	$[TP]/\text{mol L}^{-1}$	pH	Ampl./V	Freq./Hz	DA	AU
12	0,075 (0)	6,5 (0)	0,075 (0)	75 (0)	3,89	7,41
9	0,075 (0)	6,5 (0)	0,075 (0)	75 (0)	3,76	7,84
6	0,1 (+)	5,8 (-)	0,1 (+)	50 (-)	4,38	9,87
7	0,05 (-)	7,2 (+)	0,1 (+)	50 (-)	4,05	6,74
1	0,05 (-)	5,8 (-)	0,05 (-)	50 (-)	2,34	6,99
11	0,075 (0)	6,5 (0)	0,075 (0)	75 (0)	3,81	7,82
3	0,05 (-)	7,2 (+)	0,05 (-)	100 (+)	1,98	3,80
4	0,1 (+)	7,2 (+)	0,05 (-)	50 (-)	2,63	6,96
8	0,1 (+)	7,2 (+)	0,1 (+)	100 (+)	3,73	6,87
10	0,075 (0)	6,5 (0)	0,075 (0)	75 (0)	3,78	7,98
2	0,1 (+)	5,8 (-)	0,05 (-)	100 (+)	2,38	6,95
5	0,05 (-)	5,8 (-)	0,1 (+)	100 (+)	3,33	6,94
13	0,075 (0)	6,5 (0)	0,075 (0)	75 (0)	3,89	7,74

Os ensaios estão colocados na ordem em que foram executados. Os níveis codificados estão entre parênteses e os decodificados, fora. A unidade da resposta é dada por: $A \cdot s^{1/2} \cdot V^{-3/2}$

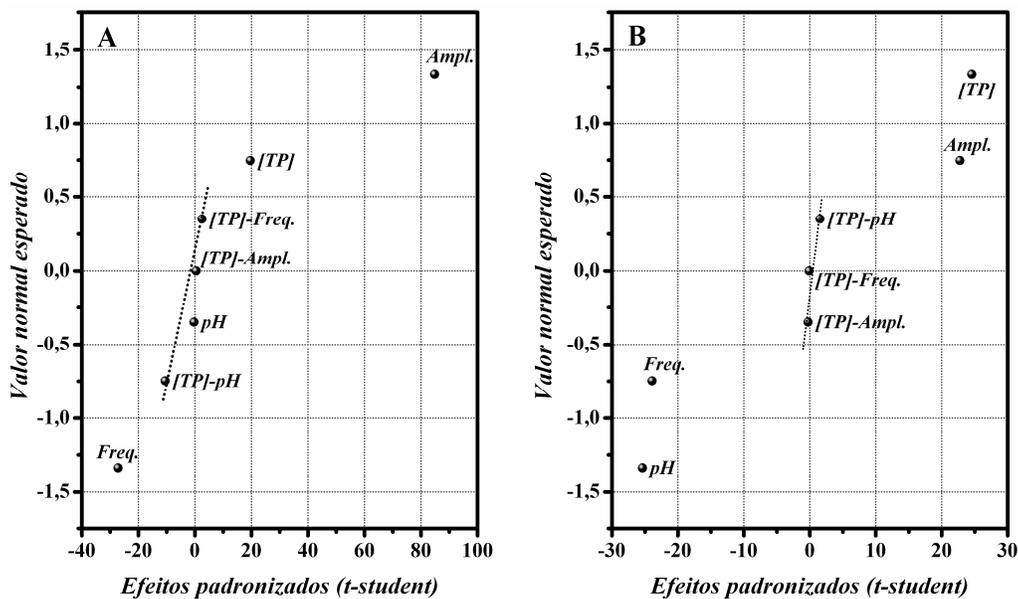


Figura 5.3. Gráfico de probabilidade normal para a DA (A) e AU (B). Fatores fora da linha tracejada são significativos.

Os resultados mostrados na Figura 5.3 e na Tabela 5.3 indicam que as variáveis *Ampl.* e *Freq.* necessitam de um estudo mais detalhado visto que elas foram de grande importância para os dois compostos. Outra variável significativa e positiva para ambos os compostos foi a concentração do tampão. Embora ela tenha sido importante, optou-se por fixá-la em $0,1 \text{ mol L}^{-1}$, pois se observou com experimentos independentes pouco ganho na resposta para concentrações maiores que $0,1 \text{ mol L}^{-1}$. Esta conclusão foi possível, pois não foi notado nenhum efeito de interação desta variável com outra. Além disso, muitos trabalhos na literatura usaram esta concentração [4,5,23,24], o que contribuiu para esta escolha.

Tabela 5.3. Efeitos principais do planejamento fracionário 2^{4-1} para cada composto e seus respectivos valor *p* calculados a partir do erro puro.

	DA				AU			
	Efeitos	Erro puro	t	<i>p</i>	Efeitos	Erro puro	t	<i>p</i>
Intercepto	3,380*	0,017	194,13	0	7,22*	0,060	120,75	0
[TP]	0,356*	0,044	8,03	$1,3 \times 10^{-3}$	1,55*	0,153	10,14	$5,3 \times 10^{-4}$
pH	-0,005	0,044	-0,11	0,92	-1,59*	0,153	-10,46	$4,7 \times 10^{-4}$
Ampl.	1,540*	0,044	34,70	$4,0 \times 10^{-6}$	1,43*	0,153	9,39	$7,2 \times 10^{-4}$
Freq.	-0,494*	0,044	-11,13	$3,7 \times 10^{-4}$	-1,50*	0,153	-9,85	$6,0 \times 10^{-4}$

*efeitos significativos a um nível de significância de 0.05 e 4 graus de liberdade pelo teste t.

O efeito da variável pH não foi significativo para a DA, porém foi o mais significativo e negativo para o AU. Como o tampão fosfato tem como limite inferior um pH de 5,8 optou-se por fixá-lo em 6,0.

Um planejamento Doehlert para superfície de resposta [42,64,65] foi elaborado para investigar as variáveis amplitude e frequência conforme apresentado na Tabela 5.4.

Tabela 5.4 Planejamento Doehlert com os níveis codificados e experimentais e as respectivas respostas observadas para as variáveis amplitude e frequência.

Ensaio	Níveis				Respostas (r_n)	
	Codificados		Experimentais		DA	AU
	x_1	x_2	Ampl.	Freq.		
11	0	0	0,075	16,5	4,57	14,03
4	-0,5	-0,866	0,0625	8	4,24	15,82
9	0	0	0,075	16,5	4,68	14,47
10	0	0	0,075	16,5	5,02	14,29
3	-1	0	0,05	16,5	3,67	12,87
1	1	0	0,1	16,5	6,00	13,68
5	0,5	-0,866	0,0875	8	4,95	15,20
7	0	0	0,075	16,5	4,69	14,37
8	0	0	0,075	16,5	4,76	14,29
6	-0,5	0,866	0,0625	25	4,05	12,87
2	0,5	0,866	0,0875	25	5,00	13,41

As análises de variância (ANOVA) para ambos os modelos quadráticos indicaram que as regressões foram significativas e não houve falta de ajuste, conforme é apresentado na Tabela 5.5. A variação explicada foi de 94,85 % e 97,14 % para o modelo da DA e do AU, respectivamente.

A Figura 5.4. mostra os valores medidos vs. previstos pelos modelos bem como os valores medidos vs. os resíduos. Observe que os gráficos para ambos compostos indicam um bom ajuste do modelo concordando com os resultados da análise de variância. Os resíduos se distribuíram aleatoriamente, novamente indicando que o modelo quadrático descreve bem o comportamento dos dados.

Tabela 5.5. Análises de variância para os modelos DA e AU

Fonte de variação	ANOVA DA				
	SQ	nGL	MQ	F _{calc.}	<i>p</i>
Regressão	3,48*	5	0,70	18,40	3,10x10 ⁻³
Resíduos	0,19	5	0,04		
Falta de Ajuste	0,08 [†]	1	0,08	2,69	0,18
Erro Puro	0,11	4	0,03		
Total	3,67	10			

Fonte de variação	ANOVA AU				
	SQ	nGL	MQ	F _{calc.}	<i>p</i>
Regressão	7,86*	5	1,57	33,98	7,28x10 ⁻⁴
Resíduos	0,23	5	0,05		
Falta de Ajuste	0,13 [†]	1	0,13	4,99	0,09
Erro Puro	0,10	4	0,03		
Total	8,09	10			

SQ. Soma quadrática, nGL. número de graus de liberdade, MQ. Média quadrática, F_{calc.} Valor calculado da distribuição *F*. *valores significativos a um nível de significância de 0,05 pelo teste *F*. [†]valores não significativos a um nível de significância de 0,05 pelo teste *F*.

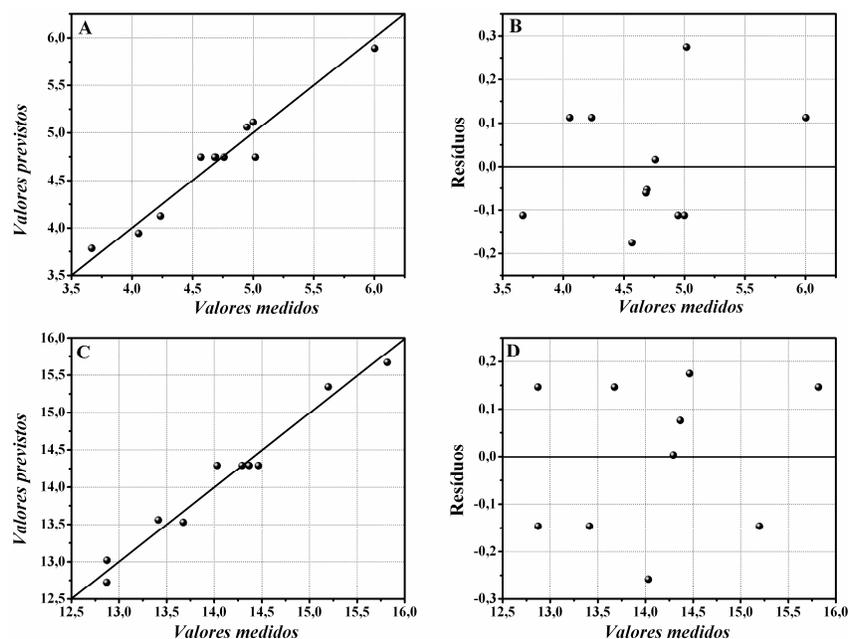


Figura 5.4. Valores medidos vs. previstos pelo modelo da DA (A) e do AU (C), respectivamente e valores medidos vs. resíduos para os modelos da DA (B) e AU (D), respectivamente.

As equações 5.2 e 5.3 mostram os modelos construídos para a DA e AU, respectivamente. As superfícies de resposta para estes modelos são apresentadas na

Figura 5.5. Observe que as respostas ótimas para ambos os analitos não coincidem, o que sugere escolher os níveis de maneira compromissada, para se obter respostas satisfatórias para ambos os compostos. Assim, pode-se concluir que os níveis que satisfazem os dois sistemas para as variáveis amplitude e frequência são 0,075 V e 22,5 Hz, respectivamente.

$$r_{DA} = 2,37 + 11,08Ampl. + 0,05Freq. + 147,40Ampl.^2 - 0,003Freq.^2 + 0,54Ampl. \times Freq. \quad 5.2$$

$$r_{AU} = 11,16 + 208,60Ampl. - 0,47Freq. - 1.623,60Ampl.^2 + 0,004Freq.^2 + 2,74Ampl. \times Freq. \quad 5.3$$

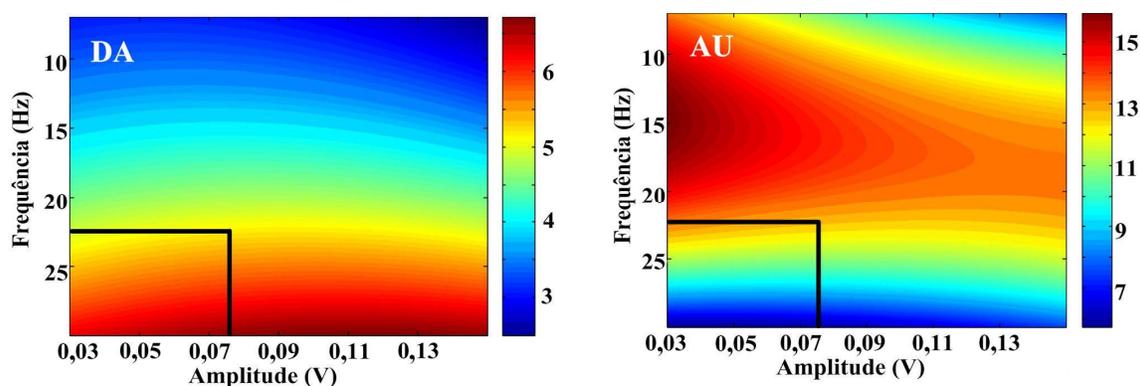


Figura 5.5. Superfícies de resposta para os modelos quadráticos construídos para a DA e para o AU.

Utilizando as condições ótimas obtidas *i.e.*, tampão fosfato 0,1 M pH 6,0; amplitude 0,075 V e frequência 22,5 Hz, modelos de calibração multivariada foram construídos para realizar previsões da mistura DA, AU e AA.

Modelos PLS

As Figuras 5.6 e 5.7 apresentam os voltamogramas originais e com correção de linha de base nas condições ótimas.

As Figuras 5.6A,C e 5.7A,C mostram a alta sobreposição dos três picos, indicando uma baixa seletividade da análise. Os voltamogramas das misturas apresentados nas Figuras 5.6B,D e 5.7B,D não apresentaram nenhum tipo de estrutura que pudesse contribuir para identificar um ou outro componente, entretanto, o deslocamento dos picos no eixo das abscissas certamente contém informações importantes das variações das concentrações nas misturas.

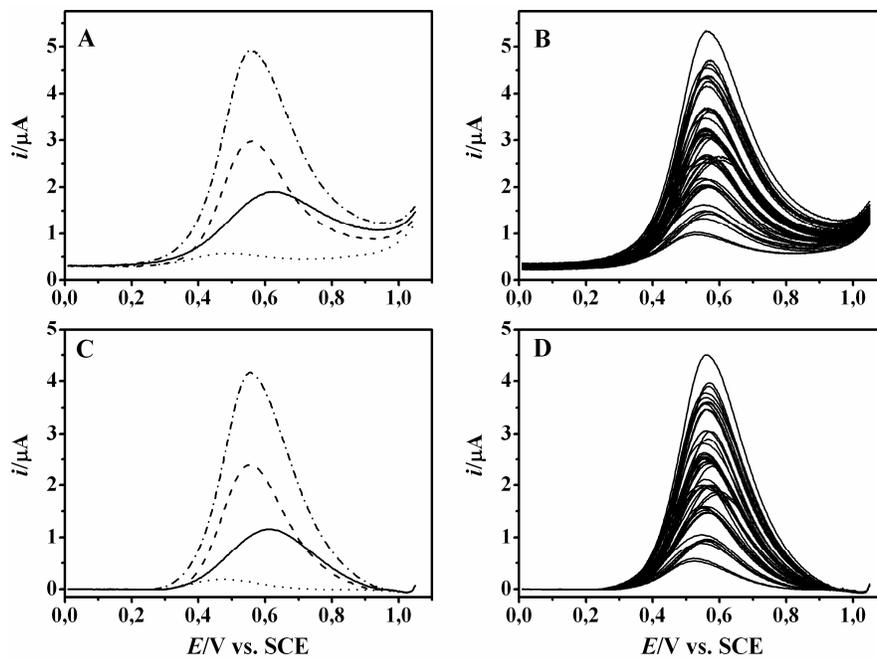


Figura 5.6. Conjunto **CA**: Voltamogramas originais (A) e com correção de linha base (C) para DA(⋯), AU(---), AA(—) e mistura(-·-); Voltamogramas originais das misturas (B) e com correção de linha de base com janela móvel (D). A largura da janela na correção de linha de base foi igual a 0,015 V.

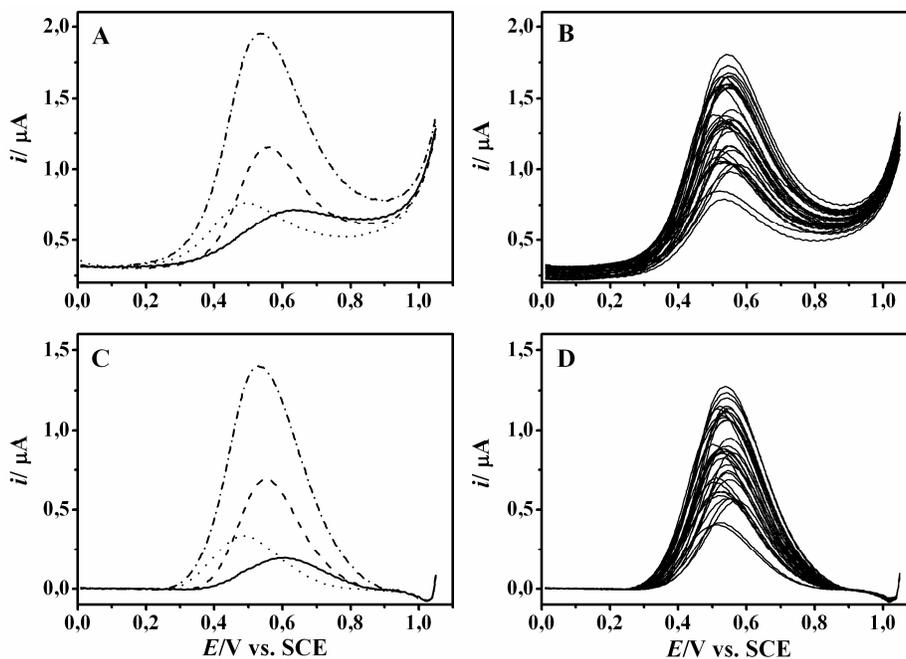


Figura 5.7 Conjunto **CB**: Voltamogramas originais (A) e com correção de linha base (C) para DA(⋯), AU(---), AA(—) e mistura(-·-); Voltamogramas originais das misturas (B) e com correção de linha de base (D). A largura da janela na correção de linha de base foi igual a 0,02 V.

Nota-se pela Tabela 5.6 que as concentrações para conjunto **CA** são maiores para AU e AA em relação à DA.

Tabela 5.6. Concentrações de cada componente nas misturas usadas nos conjuntos de calibração **CA** e **CB**, além das concentrações médias, mínimas e máximas, para cada componente em cada conjunto.

Misturas	Conjunto CA			Conjunto CB		
	DA	AU	AA	DA	AU	AA
1	25,1	251	227	37,9	87,2	58,3
2	26,4	48,1	43,3	24,8	146	63,8
3	44,7	145	402	68,8	90,7	36,7
4	45,3	249	133	20,2	117	94
5	15,1	151	324	50,3	145	78,6
6	14,6	347	403	52,5	87,9	98,4
7	33,9	436	304	69,9	143	24
8	23,9	437	393	69,9	143	24
9	45,7	45,8	319	53,9	31,5	54,7
10	34,9	350	132	39,4	117	81,1
11	36,4	152	42,2	55,4	58,1	23,4
12	15,3	259	42,5	23,7	29,7	39,4
13	25,1	251	227	65,2	146	100
14	25,1	251	227	37,9	87,2	58,3
15	25,1	252	227	37,9	87,2	58,3
16	34,5	244	401	37,9	87,2	58,3
17	4,7	364	42,3	37,9	87,2	58,3
18	36,3	46,7	232	36,8	57,6	42,1
19	24,5	347	311	23	57,6	78,3
20	25,1	252	227	67,7	114	56,7
21	14,6	449	223	36,9	146	94,1
22	15,6	48,2	141	54,7	60,2	21,5
23	14,6	449	223			
24	14,6	347	403			
25	25,1	252	227			
26	25,1	251	227			
27	25,7	153	137			
<i>Média</i>	25,8	252,9	231,1	45,6	96,7	59,2
<i>Mínima</i>	4,7	45,8	42,2	20,2	29,7	21,5
<i>Máxima</i>	45,7	449,0	403,0	69,9	146,0	100,0

As concentrações são dadas em $\mu\text{mol L}^{-1}$.

Em análise multivariada e principalmente em eletroanalítica, o efeito da concentração de um analito sobre outro pode influenciar significativamente a

análise e parâmetros analíticos dos modelos como seletividade, sensibilidade e limite de detecção. Para verificar esta influência o conjunto **CB** foi elaborado com concentrações menores em relação ao conjunto **CA** para AU e AA, porém, ainda maiores em relação à DA.

Seleção de Variáveis

Como discutido no Capítulo 1, melhores modelos de calibração são normalmente obtidos quando variáveis são selecionadas. Neste sentido, dois algoritmos foram empregados para realizar a tarefa de encontrar as variáveis com maior potencial para realizar boas previsões. O primeiro deles foi idealizado durante a realização desta tese e foi nomeado de OPS[®] (Capítulo 1) [53,54]. O segundo foi o GA (Capítulo 1), que é um algoritmo bastante usado devido à sua estratégia intensiva de busca. O GA foi empregado para comparação com os resultados do algoritmo proposto.

Os resultados do planejamento Plackett-Burman que forneceram os menores valores de RMSECV são mostrados na Tabela 5.7.

Tabela 5.7. Parâmetros usados para executar o algoritmo genético.

	CA					CB				
	Originais			Correção		Originais			Correção	
	AA	DA	AU	AA	DA/AU	AA	DA	AU	AA	DA/AU
População	52	52	100	152	152	52	152	152	152	152
Largura da janela	3	1	2	1	1	3	3	1	3	3
Taxa de Mutação	0,005	0,005	0,003	0,001	0,005	0,005	0,005	0,005	0,001	0,005
Inicialização	20	50	35	50	20	20	50	20	20	50
<i>Cross over</i>	2	2	2	2	2	2	2	2	2	2
Gerações	50	150	100	50	50	50	150	50	150	150
Convergência	20	20	45	20	70	20	70	70	20	70

Durante a seleção de variáveis empregando o método OPS, os seguintes vetores informativos foram utilizados (ver Capítulo 1): o vetor de regressão do modelo PLS (R); o vetor de correlações entre cada variável independente e o vetor de variável dependente (C); o vetor que contém o inverso da soma quadrática das colunas da matriz de resíduos, que é obtida subtraindo a matriz de dados da matriz reconstruída com *h* componentes (S) e; a combinações entre estes vetores *i.e.* RC,

RS, CS e RCS. A Figura 5.8 mostra os dados originais para o conjunto **CA** e os vetores informativos calculados para a dopamina. Observe que os vetores que trazem mais informação estruturada na região de oxidação são o **R** e o **C** e a combinação **RC**. O vetor **S** basicamente indica uma região de potenciais baixos, onde praticamente não há oxidação dos analitos da mistura. Observou-se que para todos os compostos e para ambos os conjuntos, os vetores que mais traziam informações importantes e que realmente apresentaram maior poder de previsão foram os vetores **R** e a combinação **RC**. Assim, durante este trabalho, o uso do OPS se restringirá à aplicação destes vetores prognósticos para a seleção das variáveis.

A Figura 5.9 mostra os resultados obtidos quando o método OPS foi executado para os conjuntos **CA** e **CB**, usando todas as amostras com e sem correção de linha de base. As investigações foram realizadas com janela e incremento iguais a cinco variáveis. O valor do RMSECV foi coletado para cada conjunto com um valor fixo de componentes conforme as Tabelas 5.8 e 5.9.

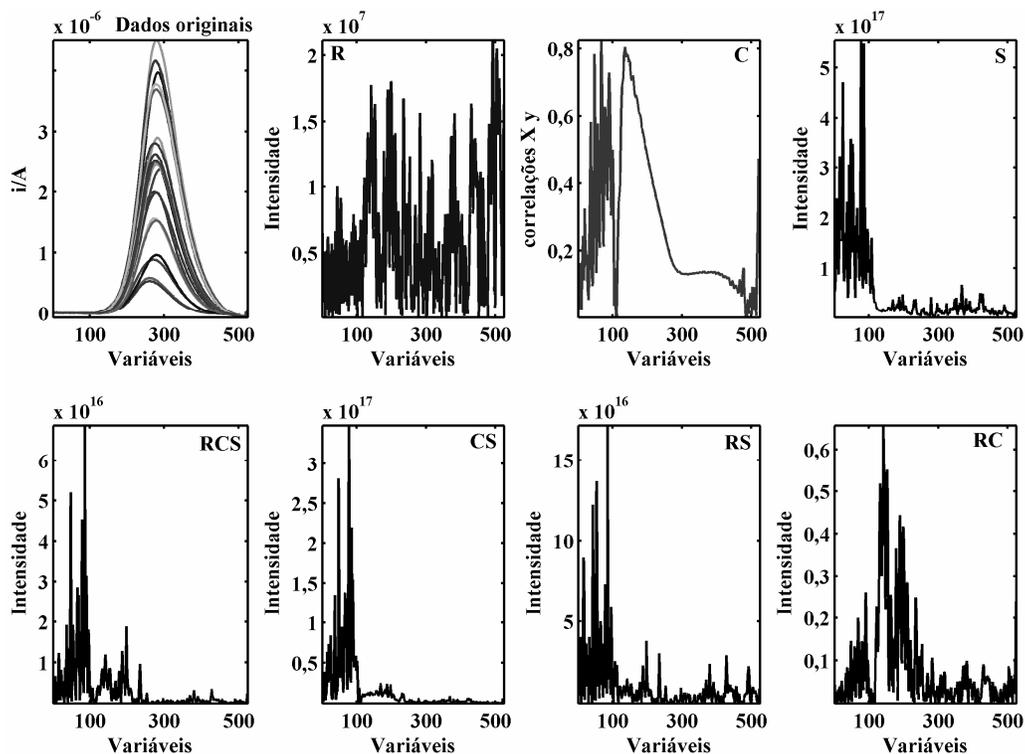


Figura 5.8. Dados originais e vetores prognósticos usados para selecionar variáveis para a dopamina usando o método OPS. No eixo das variáveis são indicados os seus números e não os potenciais.

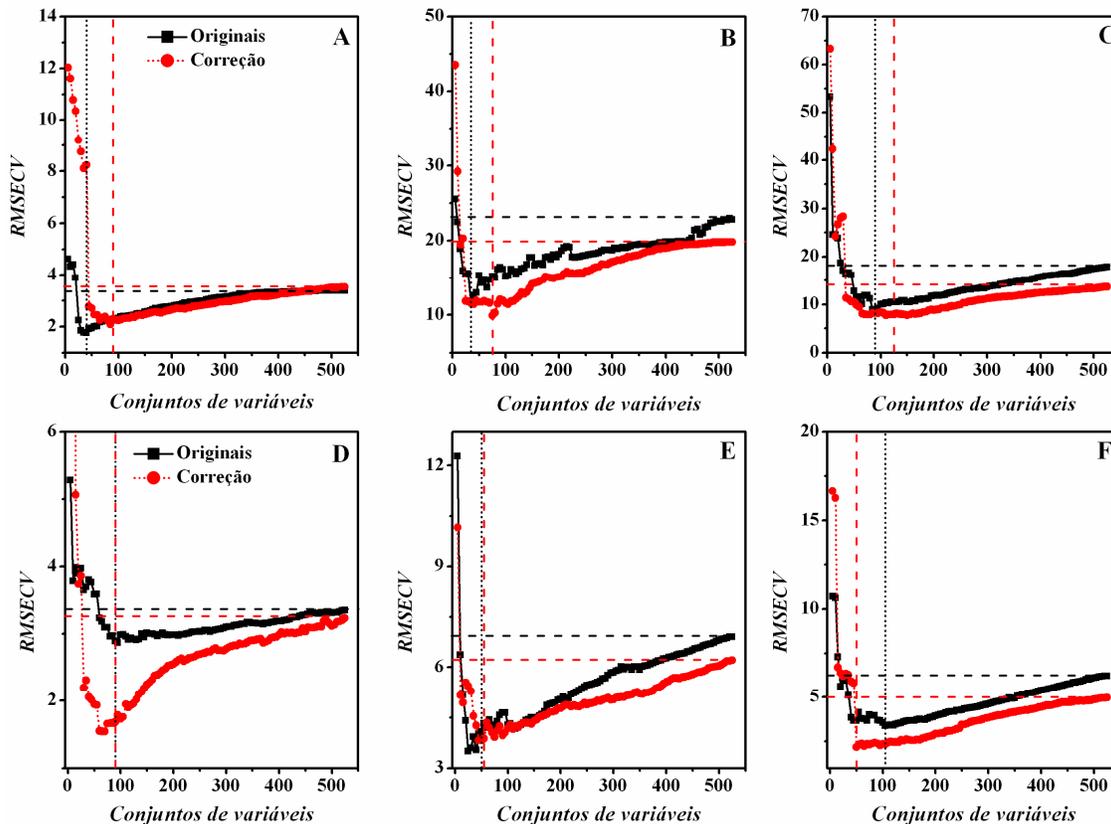


Figura 5.9. Gráficos característicos obtidos usando o método OPS. As linhas horizontais indicam o valor de RMSECV com todas as variáveis e as linhas verticais indicam o conjunto escolhido. A, B, C são os resultados para DA, AU e AA para o conjunto **CA**, e D, E, F para o conjunto **CB**, respectivamente. Os símbolos ■ e ● indicam os voltamogramas sem e com correção de linha de base, respectivamente.

Os parâmetros estatísticos dos modelos construídos para os conjuntos **CA** e **CB** são apresentados nas Tabelas 5.8 e 5.9. Após ter sido definido um número ótimo de componentes para cada analito, este número foi usado para os diferentes modelos.

Observa-se que os maiores erros (RMSECV e RMSEP) foram obtidos para os modelos que utilizaram todas as variáveis. Isto confirma a necessidade de se realizar uma seleção de variáveis em calibração multivariada com dados voltamétricos com o objetivo de se melhorar o poder de previsão dos modelos. Além disso, como observado, tanto na Figura 5.9 como nas Tabelas 5.8 e 5.9, os modelos relativos aos dados que receberam correção de linha de base tiveram, na maioria dos casos, melhor desempenho em relação àqueles que não receberam nenhuma correção.

Tabela 5.8. Diferentes modelos PLS construídos para a quantificação simultânea de DA, AU e AA para o conjunto CA.

		Originais			Correção		
		Completo	Seleção		Completo	Seleção	
			OPS [®]	GA		OPS [®]	GA
DA	nº variáveis	525	40	30	525	90	35
	nº componentes		8			8	
	RMSECV	5,2	2,1	3,2	4,9	3,1	3,6
	r_{cv}	0,886	0,979	0,954	0,895	0,956	0,941
	RMSEP	1,6	1,0	2,6	2,2	1,0	1,3
	r_p	0,992	0,997	0,985	0,985	0,997	0,994
	<hr/>						
AU	nº variáveis	525	35	100	525	75	85
	nº componentes		7			7	
	RMSECV	26,2	14,3	14,0	21,2	12,0	15,5
	r_{cv}	0,977	0,993	0,993	0,985	0,995	0,992
	RMSEP	21,3	10,1	13,2	17,3	8,1	12,1
	r_p	0,991	0,998	0,996	0,993	0,998	0,998
	<hr/>						
AA	nº variáveis	525	90	85	525	125	145
	nº componentes		9			9	
	RMSECV	19,8	11,4	11,6	17,4	11,1	19,4
	r_{cv}	0,985	0,995	0,995	0,988	0,995	0,986
	RMSEP	20,7	12,5	11,3	13,2	9,3	11,6
	r_p	0,982	0,993	0,995	0,993	0,996	0,995

Em relação aos algoritmos para seleção de variáveis observe na Figura 5.9 e nas Tabelas 5.8 e 5.9 que o algoritmo proposto (OPS) se mostrou bastante eficiente para encontrar o conjunto de variáveis com melhor capacidade de previsão se comparado com o conjunto completo (linha horizontal na Figura 5.9). Uma comparação entre os algoritmos mostra que o OPS, na maioria dos casos, apresentou valores de RMSEP e RMSECV menores que o GA para ambos os conjuntos. Isto indica a eficiência do OPS e o torna apto a ser empregado em seleção de variáveis.

Tabela 5.9. Diferentes modelos PLS construídos para a quantificação simultânea de DA, AU e AA para o conjunto **CB**

		Originais			Correção		
		Completo	Seleção		Completo	Seleção	
			OPS [®]	GA		OPS [®]	GA
DA	n° variáveis	525	90	130	525	90	110
	n° componentes		5			5	
	RMSECV	4,1	3,4	2,3	3,5	2,3	4,3
	r_{cv}	0,968	0,977	0,989	0,975	0,990	0,964
	RMSEP	3,8	4,3	2,9	4,1	2,0	3,5
	r_p	0,982	0,981	0,996	0,984	0,997	0,996
	AU	n° variáveis	525	50	30	525	55
n° componentes			8			8	
RMSECV		7,7	3,3	3,5	7,3	5,6	6,2
r_{cv}		0,979	0,996	0,996	0,981	0,989	0,986
RMSEP		8,9	7,4	5,5	7,00	4,0	7,5
r_p		0,987	0,996	0,998	0,989	0,997	0,994
AA		n° variáveis	525	105	78	525	50
	n° componentes		8			8	
	RMSECV	7,7	4,7	6,7	7,5	3,3	5,6
	r_{cv}	0,951	0,982	0,963	0,953	0,991	0,974
	RMSEP	8,2	5,6	6,9	6,0	2,7	7,8
	r_p	0,975	0,990	0,985	0,983	0,999	0,977

É importante destacar aqui o tempo computacional de ambos os métodos. Enquanto o OPS realizou os cálculos para estes conjuntos em frações de segundos, o GA demorou aproximadamente 2 horas para cada ensaio. Outro fato interessante é que os resultados obtidos com o método OPS são os mesmos se o algoritmo for executado várias vezes nas mesmas condições, por outro lado, nas mesmas condições os resultados obtidos com o método GA raramente se repetem.

Olhando agora para os dois conjuntos, observa-se que não houve grandes melhorias do modelo para a DA quando as concentrações de AU e AA foram diminuídas no conjunto **CB**. Isto ocorreu possivelmente devido ao potencial de oxidação da DA que ocorre antes dos outros analitos. Porém, os modelos para AU e AA tiveram uma leve piora se forem considerados os valores dos coeficientes de correlação. Isto certamente aconteceu devido à presença mais ativa da DA no

conjunto **CB** em relação aos outros analitos. A DA proporcionou uma maior influência na competição pela superfície do eletrodo e assim, as espécies AU e AA neste conjunto, estando em menor concentração, sentiram mais significativamente a presença da DA. Em eletroanalítica, ao contrário da espectroscopia, após o analito começar a responder, este processo continua até que toda a faixa seja varrida. Isto significa que mesmo após o sinal de pico diminuir em um voltamograma de pulso, a oxidação ou redução continua. Portanto, em um sistema de três analitos como este estudado aqui, quando o AA começa a se oxidar, a DA e o AU já estão disputando ao máximo a superfície do eletrodo. No potencial de 0,8 V, por exemplo, todos os três analitos estão se oxidando simultaneamente. Além disso, o coeficiente de difusão e a cinética de transferência eletrônica de cada analito são importantes nestas considerações, visto que estes parâmetros influenciam a competitividade pela superfície e assim, a qualidade do modelo final. Esta discussão justifica ainda, a grande quantidade de componentes (fatores) necessários para extrair as informações dos voltamogramas.

Em ambas as Tabelas nota-se que reduções/aumentos no RMSECV nem sempre são acompanhados por reduções/aumentos nos valores de RMSEP, principalmente para as seleções empregando GA e quando se utilizou as variáveis originais. Isto possivelmente aconteceu devido ao pequeno número de amostras nos diferentes conjuntos e certamente devido a baixa robustez das variáveis selecionadas, que não foram representativas do conjunto completo.

Finalmente, com o objetivo padronizar a análise, os modelos definidos como aptos para realizar previsões de ambos os compostos e para ambos os conjuntos de dados são aqueles nos quais se aplicou a correção de linha de base e que as variáveis foram selecionadas com o algoritmo OPS.

As Figuras 5.10 e 5.11 mostram as variáveis com correção de linha de base e selecionadas pelos algoritmos GA e OPS para os conjuntos **CA** e **CB**, respectivamente. Nota-se que em ambos os conjuntos, o GA se mostrou bastante disperso, não reconhecendo regiões específicas do voltamograma e principalmente, selecionando potenciais antes do início da oxidação da DA em todos os conjuntos. O algoritmo OPS, por outro lado, mostrou ser bem mais objetivo, selecionando regiões que dão maiores condições para interpretações.

De acordo com as regiões selecionadas pelo método OPS, pode-se notar que os níveis de concentrações definiram as regiões com informações analíticas nestes sistemas de baixa seletividade. As variáveis para a DA no conjunto **CA** ficaram mais distribuídas após o seu potencial de pico (0,47 V). No conjunto **CB**, o comportamento inverso foi observado e menos componentes foram necessárias. Isto certamente ocorreu devido ao nível de concentração dos interferentes. O AU, o mais concentrado no conjunto **CA**, exigiu as variáveis distribuídas próximas ao seu potencial de pico (0,55 V). Quando a concentração do AU foi diminuída no conjunto **CB**, observou-se a exigência de mais uma componente e as variáveis selecionadas foram escolhidas em potenciais maiores que o potencial de pico. Isto justifica a influência da DA sobre o AU, nestas concentrações.

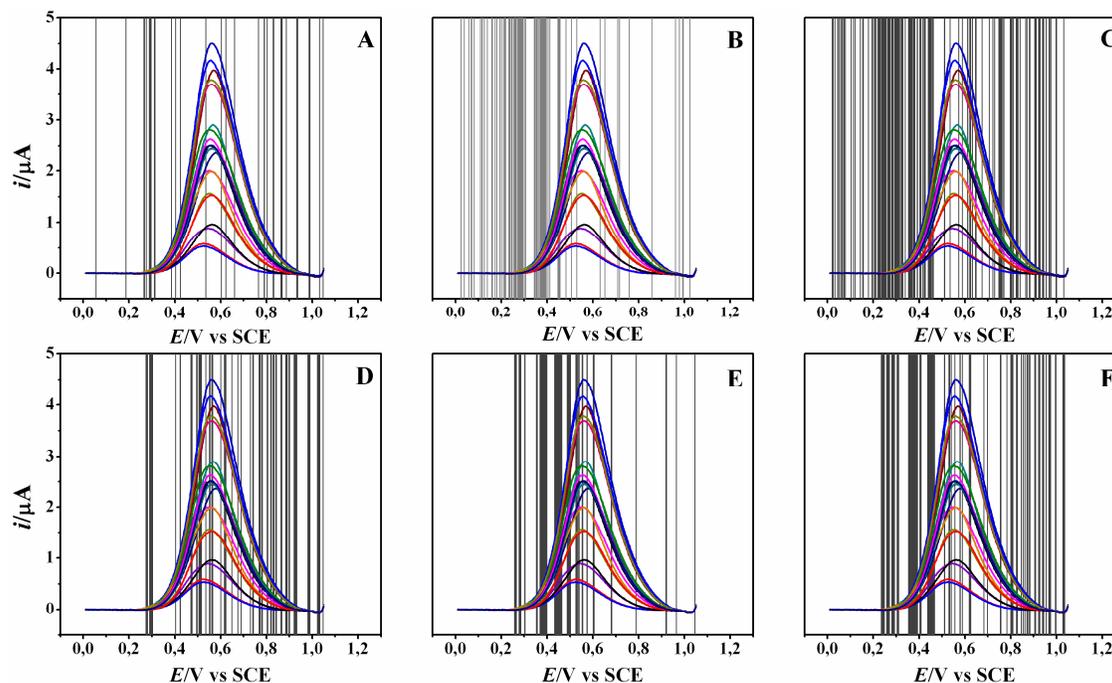


Figura 5.10. Conjunto **CA**: Variáveis selecionadas pelo GA para os compostos DA (A), AU(B) e AA(C) e variáveis selecionadas pelo método OPS para os compostos DA(D), AU(E), AA(F).

O AA no conjunto **CA** exigiu muitas variáveis distribuídas em toda a região, mesmo antes do potencial de pico (0,61 V), o que indica a dificuldade do AA em se oxidar após a DA e o AU, além disso, muitas componentes foram necessárias (9). Quando as concentrações de AU e AA foram diminuídas no conjunto **CB**, as

variáveis se deslocaram para potenciais mais próximos em relação ao potencial de pico e, além disso, menos variáveis e componentes (8) foram necessárias. Uma possível explicação para este comportamento é a menor influência do AU que se oxida antes e estava em altas concentrações.

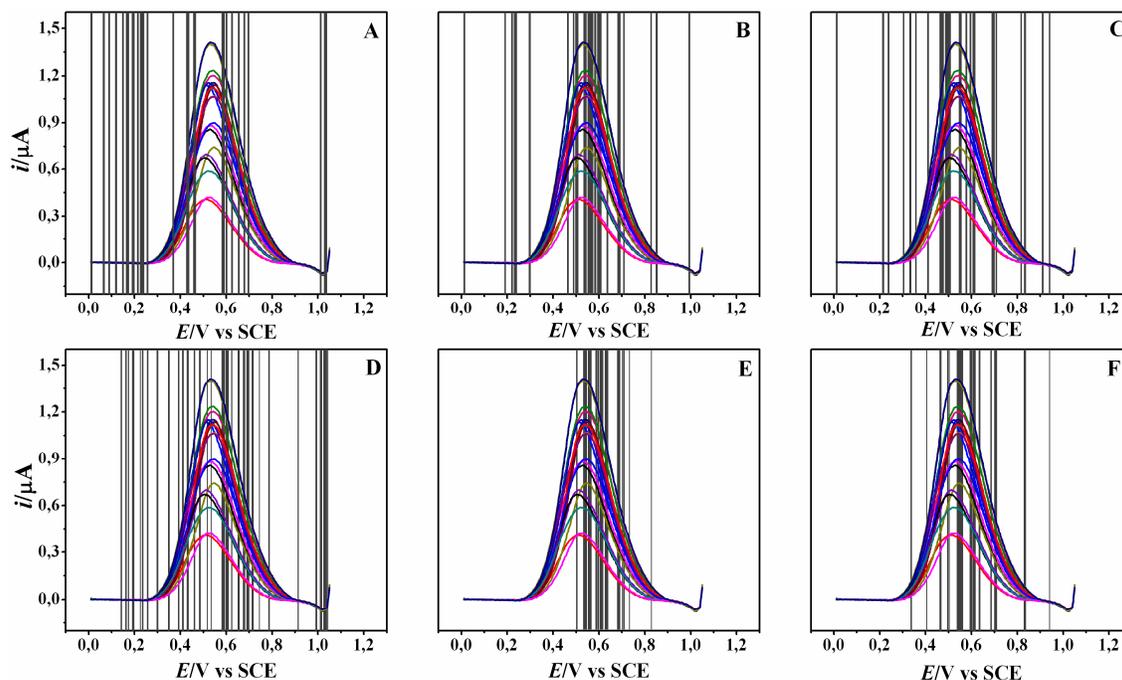


Figura 5.11. Conjunto **CB**: Variáveis selecionadas pelo GA para os compostos DA (A), AU(B) e AA(C) e variáveis selecionadas pelo OPS para os compostos DA(D), AU(E), AA(F).

Finalmente, conclui-se que os modelos com correção de linha de base e com variáveis selecionadas, na maioria dos casos, se mostraram aptos a realizar previsões para novas amostras como pode ser observado nas Tabelas 5.10 e 5.11 e nas Figuras 5.12 e 5.13. Estes modelos certamente não são capazes de realizar previsões dos três compostos em, por exemplo, fluidos biológicos, visto que para isto modelos deveriam ser construídos usando diretamente o ambiente da amostra de maneira a modelar todos os possíveis interferentes, além de considerar o efeito da matriz. Porém, estes resultados sugerem para a possibilidade de explorar o BDD e a quimiometria para a análise simultânea destes analitos em fluidos biológicos.

Tabela 5.10. Concentrações previstas pelos modelos do conjunto **CA** e seus respectivos erros relativos

	DA			AU			AA		
	Ref.	Prev.	Erro relat.	Ref.	Prev.	Erro relat.	Ref.	Prev.	Erro relat.
1	25,1	25,0	0,4	251	253,4	-1,0	227	236,9	-4,4
2	26,4	25,8	2,2	48,1	49,7	-3,3	43	40,1	7,4
3	25,7	25,8	-0,4	152	150,1	1,3	137	137,5	-0,4
4	24,5	25,1	-2,3	347	351,4	-1,3	311	305,4	1,8
5	45,3	47,4	-4,5	249	260,4	-4,6	133	127,1	4,4
6	33,8	31,9	5,8	437	418,9	4,1	304	314,2	-3,4
7	44,3	43,2	2,5	343	338,9	1,2	219	200,8	8,3
8	15,6	15,0	3,7	48,2	50,7	-5,2	141	140,5	0,4
9	42,9	43,4	-1,1	428	439,7	-2,7	386	370,0	4,1
10	4,75	5,2	-9,8	154	156,3	-1,5	236	233,8	0,9

Ref. e Prev. são as concentrações de referência e previstas dadas em $\mu\text{mol L}^{-1}$, Erros relat. são os erros relativos dados em porcentagem.

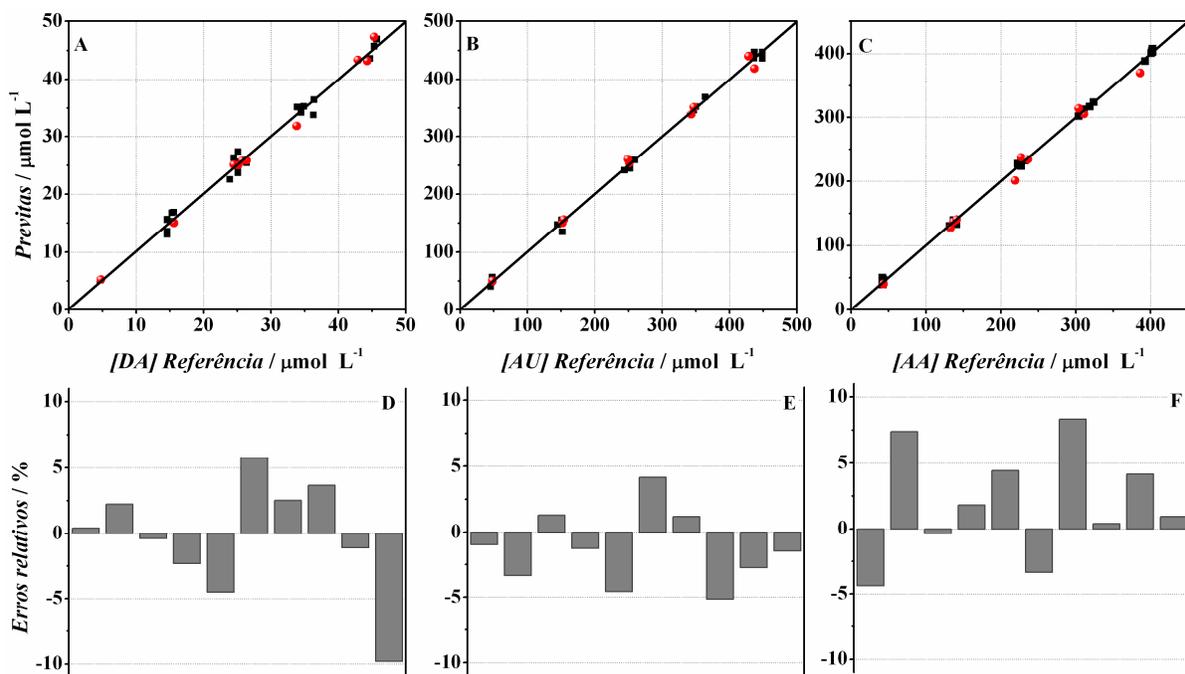


Figura 5.12. Conjunto **CA**: Concentrações adicionadas vs. previstas para o modelo (■) e para o conjunto externo de previsão (●) para DA (A), AA (B) e AU(C). As barras verticais são os erros relativos para as amostras externas para DA (D), AA (E) e AU(F)

Considerando as previsões para DA, observa-se que ambos os modelos **CA** e **CB** foram capazes de realizar boas previsões para o conjunto externo, mesmo em situação de baixa seletividade.

Em relação ao AU e AA para o conjunto **CB**, os modelos também se mostraram robustos e com boa capacidade de previsão. Na maioria dos casos, as previsões tiveram erros menores que 10%, exceto para uma amostra para AU.

Tabela 5.11. Concentrações previstas pelos modelos do conjunto **CB** e seus respectivos erros relativos

	DA			AU			AA		
	Ref.	Prev.	Erro relat.	Ref.	Prev.	Erro relat.	Ref.	Prev.	Erro relat.
1	51,1	48,2	5,6	116	120,0	-3,5	40,6	38,4	5,5
2	40,9	42,9	-4,8	26,9	29,4	-9,4	22,6	22,8	-1,0
3	65,2	62,7	3,8	146	149,1	-2,1	100	95,2	4,8
4	53,5	50,0	6,6	28,7	31,4	-9,2	60,6	60,4	0,4
5	49,9	48,1	3,5	144	142,6	1,0	77,7	76,0	2,1
6	67,6	65,0	3,8	58,2	60,8	-4,4	96,5	94,5	2,0
7	66,1	65,2	1,4	29,9	30,7	-2,5	80,6	79,1	1,9
8	20,9	20,6	1,4	118	122,3	-3,6	95,9	92,0	4,1
9	23	23,0	0,2	57,6	66,6	-15,6	78,3	73,8	5,8
10	24,3	24,34	-0,2	86,5	82,6	4,5	19,6	21,1	-7,5

Ref. e Prev. são as concentrações de referência e previstas dadas em $\mu\text{mol L}^{-1}$, Erros relat. são os erros relativos dados em percentagem.

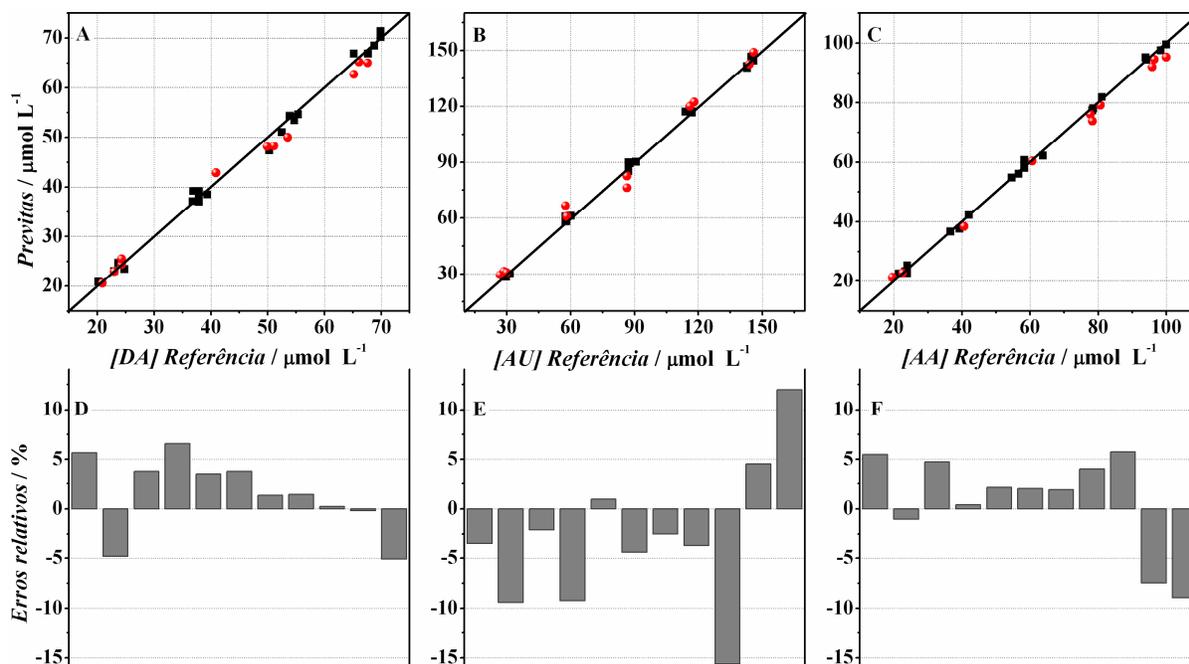


Figura 5.13. Conjunto **CB**: Concentrações adicionadas vs. previstas para o modelo (■) e para o conjunto externo de previsão (●) para DA (A), AA (B) e AU(C). As barras verticais são os erros relativos para as amostras externas para DA (D), AA (E) e AU(F)

Alguns parâmetros analíticos foram calculados para os modelos de maneira a verificar a qualidade da análise. A Tabela 5.12 mostra os valores de seletividade, sensibilidade e limite de detecção (LOD), calculados para os dois modelos finais para os três analitos e para ambos os conjuntos. A sensibilidade e a seletividade foram estimadas empregando o cálculo do NAS (Capítulo 1) [47,50,66]. Para o limite de detecção, quatro diferentes métodos foram utilizados visto que ainda não há um método bem definido para realizar este cálculo para modelos de calibração multivariada inversa. Os métodos empregados estão discutidos no Capítulo 1.

Tabela 5.12. Parâmetros analíticos para os modelos finais de ambos os conjuntos

CA						
	Seletividade	Sensibilidade	LOD ^a _{ort}	LOD ^b _{br}	LOD ^c _σ	LOD ^d _{mod}
DA	0,0103	2,36	2,5	2,3	5,2	2,5
AU	0,0216	0,67	12,0	2,2	8,7	2,5
AA	0,0221	0,62	6,2	2,7	11,4	2,0
CB						
DA	0,0682	3,84	3,4	2,4	6,5	3,0
AU	0,0126	0,49	4,2	3,3	6,9	2,6
AA	0,0124	0,68	2,5	3,0	7,5	2,4

^aMétodo de Ortiz, ^bMétodo de Boqué e Rius, ^cMétodo do desvio padrão, ^dMétodo proposto. A sensibilidade é dada por $\mu\text{A}/\text{mmol L}^{-1}$ e o LOD em $\mu\text{mol L}^{-1}$.

Os baixos valores de seletividade confirmam o que era esperado, devido à alta sobreposição dos picos. Observe que a seletividade da DA no conjunto **CB** foi maior que no **CA**. Isto confirma a perda em seletividade na determinação de DA na presença de interferentes em altas concentrações. A maior seletividade do AU e AA no conjunto **CA** certamente ocorreu devido às suas altas concentrações. No conjunto **CB** nota-se uma diminuição da seletividade para o AU e AA, confirmando a grande influência da concentração na determinação voltamétrica destes compostos empregando calibração multivariada.

A Tabela 5.13 mostra a faixa de valores encontrados para os parâmetros analíticos sensibilidade e LOD na literatura. Fazendo uma comparação entre as duas tabelas (5.12 e 5.13), pode se concluir que as sensibilidades do AU e AA estão 10 vezes menores que os menores valores encontrados na literatura, enquanto a DA está dentro da ordem de grandeza esperada. As possíveis explicações para isto é

certamente devido à competição na superfície do eletrodo, visto que a DA apresenta uma alta sensibilidade e se oxida em potencial menor, assim ela dificulta o acesso dos outros analitos, mesmo estando em baixas concentrações. A Figura 5.14 mostra os gráficos pseudo-univariados inversos que usam os valores do NAS vs os valores de referência. Como o PLS é um modelo inverso, observe que as inclinações são inversas em relação sensibilidade. Observe que a DA mostrou ter alta sensibilidade o que é facilmente notado nos gráficos (Figura 5.14 A e D) com suas baixas inclinações.

Tabela 5.13. Faixa de parâmetros analíticos encontrada na literatura para determinação simultânea DA, AU e AA usando univariados em diferentes tipos de eletrodos modificados usando técnicas de pulso (SWV ou DPV).

	Sensibilidade $\mu\text{A}/\mu\text{mol L}^{-1}$			LOD $\mu\text{mol L}^{-1}$		
	Mínima	Média	Máxima	Mínimo	Médio	Máximo
DA	0,0074	0,198	0,75	0,0027	0,295	1,40
AU	0,0071	2,386	6,47	0,0010	0,700	2,40
AA	0,0022	0,101	0,36	0,0052	1,354	5,00

*Valores obtidos das referências: [1,4,5,12,23,24,67].

A partir de uma comparação entre as sensibilidades dos dois conjuntos nota-se que leves variações ocorreram, porém não consideradas significativas, exceto para o AU que mostrou ser mais sensível em altas concentrações.

Finalmente, em relação ao limite de detecção, o método de Ortiz [52] é falho para modelos construídos longe da região próxima ao limite de detecção, o que justifica seus altos valores estimados no conjunto **CA** e a significativa diminuição no conjunto **CB**, porém, ainda altos. O método do desvio padrão baseado na extensão do método univariado desconsidera a hipótese falsa positiva e falsa negativa e por isso é criticada e muitas vezes não apresentam valores coerentes. Os métodos de Boqué e Rius [51] e o proposto nesta tese são os que chegaram mais próximos dos valores de LOD observados na literatura para ambos os conjuntos. Assim, os valores de LOD que serão considerados para ambos os modelos será uma média de ambas as estimativas destes métodos para os conjuntos **CA** e **CB**. Portanto, os limites de detecção para DA, AU e AA foram respectivamente de 2,37; 2,34 e 2,33 para o conjunto **CA** e de 2,68; 2,98 e 2,72 para o conjunto **CB**.

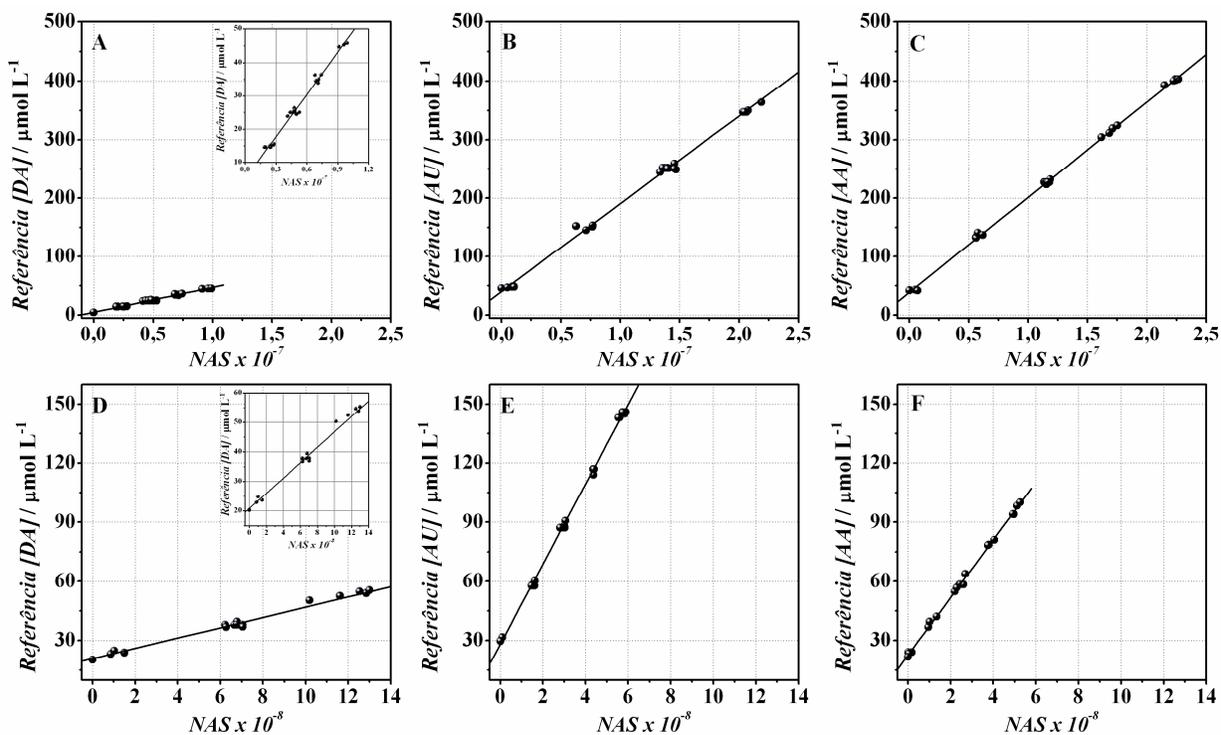


Figura 5.14. Gráficos pseudo-univariados (referência vs. NAS) para o conjunto **CA**: DA(A), AU(B) e AA(C) e pra o conjunto **CB**: DA(D), AU(E), AA(F).

Considerando estes valores, nota-se um leve aumento nos valores de LOD para o conjunto **CB**, mesmo para a DA que estava em concentrações baixas em ambos os conjuntos. Isto sugere que as diferenças não podem ser consideradas como devido à variação nas concentrações, mas sim devido a algum erro aleatório ou sistemático entre um experimento e outro. Com esta hipótese conclui-se que não há variações nos limites de detecção devido a variações nas concentrações dos analitos em situações de baixa seletividade.

Conclusões

Neste trabalho, as condições voltamétricas foram otimizadas empregando planejamentos experimentais (triagem e metodologia de superfície de resposta) e modelos de calibração multivariada foram construídos usando PLS.

O uso do diamante dopado com boro foi fundamental para este estudo, visto que os analitos quantificados apresentam grande poder de bloqueio eletroquímico sobre eletrodos sólidos convencionais.

Os modelos que foram construídos com variáveis selecionadas pelo método OPS mostraram ter maior capacidade de predição que aqueles construídos com todas as variáveis. Em comparação com o GA, os modelos foram melhores na maioria das vezes e, além disso, o método OPS selecionou variáveis mais interpretativas.

Os parâmetros analíticos mostraram que o efeito da concentração na determinação simultânea de DA, AU e AA no eletrodo de BDD interfere significativamente na seletividade. A sensibilidade e o limite de detecção tiveram leves variações com as concentrações, porém, o AU mostrou ser mais sensível em altas concentrações. Os resultados sugerem melhorias do eletrodo para ganhos em sensibilidade e no limite de detecção. Modificações do ambiente de análise de modo a aumentar a separação dos potenciais de oxidação destes analitos com o objetivo de aumentar a seletividade, é também desejável. Contudo, avanços foram alcançados na aplicação de ferramentas quimiométricas neste tipo de sistema para a determinação simultânea destes analitos.

A unificação eletrodo de diamante dopado com boro – eletroanalítica – quimiometria é certamente muito promissora para a determinação de misturas de compostos eletroativos que se interferem e apresentam alto poder de bloqueio eletroquímico em eletrodos sólidos convencionais. Por um lado se ganha em precisão, robustez e estabilidade, por outro se ganha em seletividade, confiabilidade e aplicabilidade.

Referências

1. Milczarek, G.; Ciszewski, A. 2,2-bis(3-amino-4-hydroxyphenyl)hexafluoropropane modified glassy carbon electrodes as selective and sensitive voltammetric sensors. Selective detection of dopamine and uric acid. *Electroanalysis* **2004**, 16, (23), 1977-1983.
2. Kumar, S. S.; Mathiyarasu, J.; Phani, K.; Yegnaraman, V. Simultaneous determination of dopamine and ascorbic acid on poly (3,4-ethylenedioxythiophene) modified glassy carbon electrode. *J. Solid State Electrochem.* **2006**, 10, (11), 905-913.
3. Ciszewski, A.; Milczarek, G. Poly Eugenol-modified platinum electrode for selective detection of dopamine in the presence of ascorbic acid. *Anal. Chem.* **1999**, 71, (5), 1055-1061.

4. Zare, H. R.; Rajabzadeh, N.; Nasirizadeh, N.; Ardakani, M. M. Voltammetric studies of an oracet blue modified glassy carbon electrode and its application for the simultaneous determination of dopamine, ascorbic acid and uric acid. *J. Electroanal. Chem.* **2006**, 589, (1), 60-69.
5. Salimi, A.; MamKhezri, H.; Hallaj, R. Simultaneous determination of ascorbic acid, uric acid and neurotransmitters with a carbon ceramic electrode prepared by sol-gel technique. *Talanta* **2006**, 70, (4), 823-832.
6. Wang, C. Y.; Liu, Q. X.; Shao, X. Q.; Hu, X. Y. Voltammetric determination of dopamine in human serum and urine at a glassy carbon electrode modified by cysteic acid based on electrochemical oxidation of L-cysteine. *Anal. Lett.* **2007**, 40, (4), 689-704.
7. Dos Reis, A. P.; Tarley, C.; Maniasso, N.; Kubota, L. T. Exploiting micellar environment for simultaneous electrochemical determination of ascorbic acid and dopamine. *Talanta* **2005**, 67, (4), 829-835.
8. Luz, R.; Damos, F. S.; de Oliveira, B.; Beck, J.; Kubota, L. T. Development of a sensor based on tetracyanoethylene (LiTCNE)/poly-L-lysine (FILL) for dopamine determination. *Electrochim. Acta* **2005**, 50, (13), 2675-2683.
9. Rouhollahi, A.; Rajabzadeh, R.; Ghasemi, J. Simultaneous determination of dopamine and ascorbic acid by linear sweep voltammetry along with chemometrics using a glassy carbon electrode. *Microchim. Acta* **2007**, 157, (3-4), 139-147.
10. Safavi, A.; Maleki, N.; Moradlou, O.; Tajabadi, F. Simultaneous determination of dopamine, ascorbic acid, and uric acid using carbon ionic liquid electrode. *Anal. Biochem.* **2006**, 359, (2), 224-229.
11. Santos, P. M.; Sandrino, B.; Moreira, T. F.; Wohnrath, K.; Nagata, N.; Pessoa, C. A. Simultaneous voltammetric determination of dopamine and ascorbic acid using multivariate calibration methodology performed on a carbon paste electrode modified by a mer-[RuC₁₃(dppb)(4-pic)] complex. *J. Braz. Chem. Soc.* **2007**, 18, (1), 93-99.
12. Dutt, J.; Cardosi, M. F.; Livingstone, C.; Davis, J. Diagnostic implications of uric acid in electroanalytical measurements. *Electroanalysis* **2005**, 17, (14), 1233-1243.
13. Wightman, R. M.; May, L. J.; Michael, A. C. Detection of dopamine dynamics in the brain. *Anal. Chem.* **1988**, 60, 769A-755A.
14. Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry*. 4 ed.; W. H. Freeman: New York, 2004; p 1100.
15. Eswaradu, V. V.; Mottola, H. A. Determination of uric-acid at microgram level by a kinetic procedure based on a pseudo-induction period. *Anal. Chem.* **1974**, 46, (12), 1777-1781.

16. John, S. A. Simultaneous determination of uric acid and ascorbic acid using glassy carbon electrodes in acetate buffer solution. *J. Electroanal. Chem.* **2005**, 579, (2), 249-256.
17. Arrigoni, O.; De Tullio, M. C. Ascorbic acid: much more than just an antioxidant. *Biochim. Biophys. Acta-Gen. Subj.* **2002**, 1569, (1-3), 1-9.
18. Robinson, D. L.; Volz, T. J.; Schenk, J. O.; Wightman, R. M. Acute ethanol decreases dopamine transporter velocity in rat striatum: In vivo and in vitro electrochemical measurements. *Alcoholism (NY)* **2005**, 29, (5), 746-755.
19. Kehr, J.; Hu, X. J.; Yoshitake, T.; Scheller, D. Determination of the dopamine agonist rotigotine in microdialysates from the rat brain by microbore column liquid chromatography with electrochemical detection. *J. Chromatogr. B* **2007**, 845, (1), 109-113.
20. Chu, Q. C.; Lin, M.; Geng, C. H.; Ye, J. N. Determination of uric acid in human saliva and urine using miniaturized capillary electrophoresis with amperometric detection. *Chromatographia* **2007**, 65, (3-4), 179-184.
21. Shamsipur, M.; Hemmateenejad, B.; Babaei, A.; Faraj-Sharabiani, L. Use of multivariate curve resolution analysis in the spectroelectrochemistry of 9,10-anthraquinone reduction in dimethylformamide solution. *J. Electroanal. Chem.* **2004**, 570, (2), 227-234.
22. Zare, H. R.; Nasirizadeh, N.; Ardakani, M. M. Electrochemical properties of a tetrabromo-p-benzoquinone modified carbon paste electrode. Application to the simultaneous determination of ascorbic acid, dopamine and uric acid. *J. Electroanal. Chem.* **2005**, 577, (1), 25-33.
23. Zhang, L.; J., J.; Zou, X.; Dong, S. Simultaneous determination of dopamine and ascorbic acid at an In-Site functionalized self-assembled monolayer on gold electrode. *Electroanalysis* **2004**, 16, (17), 1413-1418.
24. Zen, J. M.; Chen, P. J. A selective voltammetric method for uric acid and dopamine detection using clay modified electrodes. *Anal. Chem.* **1997**, 69, (24), 5087-5093.
25. Luo, J. W.; Zhang, M.; Pang, D. W. Selective and sensitive determination of uric acid at DNA-modified graphite powder microelectrodes. *Sens. Actuator B-Chem.* **2005**, 106, (1), 358-362.
26. Liu, T.; Li, M. X.; Li, Q. Y. Electroanalysis of dopamine at a gold electrode modified with N-acetylcysteine self-assembled monolayer. *Talanta* **2004**, 63, (4), 1053-1059.
27. Raj, C. R.; Ohsaka, T. Voltammetric detection of uric acid in the presence of ascorbic acid at a gold electrode modified with a self-assembled monolayer of heteroaromatic thiol. *J. Electroanal. Chem.* **2003**, 540, 69-77.

28. Raj, C. R.; Tokuda, K.; Ohsaka, T. Electroanalytical applications of cationic self-assembled monolayers, square-wave voltammetric determination of dopamine and ascorbate. *Bioelectrochemistry* **2001**, 53, (2), 183-191.
29. Ramesh, P.; Suresh, G. S.; Sampath, S. Selective determination of dopamine using unmodified, exfoliated graphite electrodes. *J. Electroanal. Chem.* **2004**, 561, (1-2), 173-180.
30. Oni, J.; Westbroek, P.; Nyokong, T. Electrochemical behavior and detection of dopamine and ascorbic acid at an iron(II)tetrakisulfophthalocyanine modified carbon paste microelectrode. *Electroanalysis* **2003**, 15, (10), 847-854.
31. Cai, X. H.; Kalcher, K.; Neuhold, C.; Ogorevc, B. An improved voltammetric method for the determination of trace amounts of uric-acid with electrochemically pretreated carbon-paste electrodes. *Talanta* **1994**, 41, (3), 407-413.
32. Granger, M. C.; Xu, J. S.; Strojek, J. W.; Swain, G. M. Polycrystalline diamond electrodes: basic properties and applications as amperometric detectors in flow injection analysis and liquid chromatography. *Anal. Chim. Acta* **1999**, 397, (1-3), 145-161.
33. Teófilo, R. F.; Ceragioli, H. J.; Peterlevitz, A. C.; Da Silva, L. M.; Damos, F. S.; Ferreira, M. M. C.; Baranauskas, V.; Kubota, L. T. Improvement of the electrochemical properties of "as-grown" boron-doped polycrystalline diamond electrodes deposited on tungsten wires using ethanol. *J. Solid State Electrochem.* **2007**, 11, 1449-1457.
34. Prado, C.; Murcott, G. G.; Marken, F.; Foord, J. S.; Compton, R. G. Detection of chlorophenols in aqueous solution via hydrodynamic channel flow cell voltammetry using a boron-doped diamond electrode. *Electroanalysis* **2002**, 14, (14), 975-979.
35. Granger, M. C.; Witek, M.; Xu, J. S.; Wang, J.; Hupert, M.; Hanks, A.; Koppang, M. D.; Butler, J. E.; Lucazeau, G.; Mermoux, M.; Strojek, J. W.; Swain, G. M. Standard electrochemical behavior of high-quality, boron-doped polycrystalline diamond thin-film electrodes. *Anal. Chem.* **2000**, 72, (16), 3793-3804.
36. Muna, G. W.; Tasheva, N.; Swain, G. M. Electro-oxidation and amperometric detection of chlorinated phenols at boron-doped diamond electrodes: A comparison of microcrystalline and nanocrystalline thin films. *Environ. Sci. Technol.* **2004**, 38, (13), 3674-3682.
37. De Carvalho, R. M.; Mello, C.; Kubota, L. T. Simultaneous determination of phenol isomers in binary mixtures by differential pulse voltammetry using carbon fibre electrode and neural network with pruning as a multivariate calibration tool. *Anal. Chim. Acta* **2000**, 420, (1), 109-121.
38. Martens, H.; Naes, T. *Multivariate Calibration*. John Wiley & Sons: New York, 1989; p 419.

39. Naes, T.; Isaksson, T.; Fearn, T.; Davies, T. *Multivariate calibration and classification*. Nir publications: Chichester, 2002; p 344.
40. Workman, J. J.; Mobley, P. R.; Kowalski, B. R.; Bro, R. Review of chemometrics applied to spectroscopy: 1985-95 .1. *Appl. Spectrosc. Rev.* **1996**, 31, (1-2), 73-124.
41. Brereton, R. G. Introduction to multivariate calibration in analytical chemistry. *Analyst* **2000**, 125, (11), 2125-2154.
42. Teófilo, R. F.; Ferreira, M. M. C. Chemometrics II: Spreadsheets for experimental design calculations, a tutorial. *Quim. Nova* **2006**, 29, (2), 338-350.
43. Brereton, R. G. *Chemometrics: Data Analysis for the laboratory and chemical plant*. JOHN WILEY & SONS INC: Chichester, 2003; p 407.
44. Ferreira, M. M. C.; Antunes, A. M.; Melgo, M. S.; Volpe, P. L. O. Chemometrics I: Multivariate calibration, a tutorial. *Quim. Nova* **1999**, 22, (5), 724-731.
45. Beebe, K. R.; Pell, R. J.; Seasholtz, M. B. *Chemometrics: A practical guide*. 1 ed.; John Wiley & Sons Inc: New York, 1998; p 348.
46. Booksh, K. S.; Kowalski, B. R. Theory of analytical chemistry. *Anal. Chem.* **1994**, 66, (15), 782A-791A.
47. Bro, R.; Andersen, C. M. Theory of net analyte signal vectors in inverse regression. *J. Chemometr.* **2003**, 17, (12), 646-652.
48. Swain, G. M.; Ramesham, R. The electrochemical activity of boron-doped polycrystalline diamond thin-film electrodes. *Anal. Chem.* **1993**, 65, (4), 345-351.
49. Munoz, J. A.; Brereton, R. G. Partial factorial designs for multivariate calibration: extension to seven levels and comparison of strategy. *Chemometrics Intell. Lab. Syst.* **1998**, 43, (1-2), 89-105.
50. Lorber, A.; Faber, K.; Kowalski, B. R. Net analyte signal calculation in multivariate calibration. *Anal. Chem.* **1997**, 69, (8), 1620-1626.
51. Boque, R.; Rius, F. X. Computing detection limits in multicomponent spectroscopic analysis. *Trac-Trends Anal. Chem.* **1997**, 16, (8), 432-436.
52. Ortiz, M. C.; Sarabia, L. A.; Herrero, A.; Sanchez, M. S.; Sanz, M. B.; Rueda, M. E.; Gimenez, D.; Melendez, M. E. Capability of detection of an analytical method evaluating false positive and false negative (ISO 11843) with partial least squares. *Chemometrics Intell. Lab. Syst.* **2003**, 69, (1-2), 21-33.
53. Teófilo, R. F.; Martins, J. P. A.; Ferreira, M. M. C. In *Ordered predictors Selection: an intuitive method to find the most relevant variable in multivariate calibration*, 10th International Conference on Chemometrics in Analytical Chemistry, Águas de Lindóia, Brazil, Águas de Lindóia, Brazil, 2006, p ^pp P066.
54. Teófilo, R. F.; Martins, J. P. A.; Ferreira, M. M. C. *OPS Toolbox 1.0*, 1.0; INPI - 0000270703255138: Brasil, 2007.

55. Plackett, R. L.; Burman, J. P. The design of optimum multifactorial experiments. *Biometrika* **1946**, 33, (4), 305-325.
56. StatSoft, *STATISTICA (data analysis software system)*: Version 6; StatSoft.: Tulsa, 2001.
57. Neto, B. B.; Scarminio, I. S.; Bruns, R. E. *Como fazer experimentos: pesquisa e desenvolvimento na ciência e na indústria*. Editora Unicamp: Campinas, 2001; p 401.
58. Kennard, R. W.; Stone, L. A. Computer aided design of experiments. *Technometrics* **1969**, 11, (1), 137-148.
59. Wise, B. M.; Gallagher, N. B.; Bro, R.; Shaver, J. M. *PLS_Toolbox*: Version 3.01; Eigenvector Research, Inc.: Manson, 2003.
60. Bard, A. J.; Faulkner, L. R. *Electrochemical methods: fundamentals and applications*. 2nd ed.; John Wiley & Sons: New York, 2001; p 833.
61. De Souza, D.; Machado, S. A. S. Avaca, L. A. Square wave voltammetry. Part I: Theoretical aspects. *Quim. Nova* **2003**, 26, (1), 81-89.
62. Brett, C. M. A.; Brett, A. M. O. *Electrochemistry: principles, methods, and applications*. 1 ed.; Oxford University Press: Oxford, 1993; Vol. 1, p 427.
63. Diaz-Cruz, M. S.; Esteban, M.; Rodriguez, A. R. Square wave voltammetry data analysis by multivariate curve resolution: application to the mixed-metal system Cd-Zn-Lys-Cys-Thr-Cys-Cys-Ala. *Anal. Chim. Acta* **2001**, 428, (2), 285-299.
64. Doehlert, D. H. Uniform shell designs. *Appl. Stat.* **1970**, 19, 231-239.
65. Ferreira, S.; dos Santos, W.; Quintella, C. M.; Neto, B. B.; Bosque-Sendra, J. A. Doehlert matrix: a chemometric tool for analytical chemistry - review. *Talanta* **2004**, 63, (4), 1061-1067.
66. Faber, N. M. Efficient computation of net analyte signal vector in inverse multivariate calibration models. *Anal. Chem.* **1998**, 70, (23), 5108-5110.
67. Roy, P. R.; Okajima, T.; Ohsaka, T. Simultaneous electrochemical detection of uric acid and ascorbic acid at a poly(N,N-dimethylaniline) film-coated GC electrode. *J. Electroanal. Chem.* **2004**, 561, (1-2), 75-82.

Capítulo 6

Determinação Simultânea de Guaiacol e Cloroguaicol com Diamante Dopado com Boro usando Voltametria e Métodos Quimiométricos

“As águas são sempre puras em suas nascentes”

Blaise Pascal

Introdução

Compostos fenólicos e seus derivados são amplamente usados em processos industriais, tais como fabricação de polímeros, medicamentos, aditivos conservantes, corantes, etc [1-7]. Esta classe de composto pode ser também encontrada na decomposição química de alguns pesticidas e como subprodutos da indústria de polpa de papel, além de estarem presentes nos lixiviados provenientes dos aterros sanitários, entre outros resíduos [1,2,6]. Assim, os tipos e abundância destes compostos variam com a fonte ou processo de fabricação [6,8] e, conseqüentemente, um número considerável de poluentes orgânicos, amplamente distribuídos em toda parte do ambiente, possuem uma estrutura fenólica [9].

Os compostos fenólicos com alta toxicidade pertencem a uma classe de poluentes químicos que são facilmente absorvidos por animais e humanos através da pele e mucosas [10,11]. Além disso, estes compostos apresentam uma ampla faixa de níveis de toxicidade mesmo a baixas concentrações e, devido a sua alta periculosidade e persistência no ambiente, estão presentes na lista de substâncias perigosas e poluentes prioritários de diversas agências de proteção ambiental tais como: do Conselho Nacional do Meio Ambiente (CONAMA) [12], da Comissão

Européia de Proteção Ambiental (EC) [13] e da Agência de Proteção Ambiental Norte Americana (EPA) [14].

Estes compostos são também muitas vezes responsáveis por propriedades organolépticas de alimentos, da água potável, bebidas, flores e frutos [15].

Devido a todas estas características, uma grande importância tem sido dada na triagem, monitoramento e controle dos compostos fenólicos [4,16,17] e o desenvolvimento de procedimentos para a detecção destas espécies químicas em diferentes matrizes é de grande interesse para diversas áreas. Além disso, em amostras reais, normalmente o que se encontra são misturas de compostos fenólicos, assim, as técnicas analíticas devem levar em consideração a coexistência de diversos compostos fenólicos, muitas vezes interferentes entre si.

Os métodos analíticos para detecção e quantificação de misturas de compostos fenólicos são usualmente baseados nas técnicas analíticas de separação, que permitem a identificação e quantificação de constituintes individuais, tais como extração líquido-líquido ou líquido-sólido seguida por cromatografia gasosa ou cromatografia líquida usando diferentes dispositivos de detecção [4]. Entretanto, estas técnicas não são apropriadas para o monitoramento contínuo *on site*, são relativamente de alto custo, consomem muito tempo de análise e um operador especialista é necessário [18].

Por outro lado, dentre os métodos analíticos mais sensíveis para detectar e quantificar fenóis, os eletroanalíticos, tais como voltametria e amperometria, têm atraído pouca atenção dos laboratórios de rotina, apesar do fato de apresentarem baixo custo e empregarem uma análise relativamente rápida [19,20]. Uma primeira razão para este uso limitado é a baixa seletividade quando se analisam compostos que oxidam em potenciais muito próximos [21]. Para minimizar este problema, vários trabalhos na literatura apresentam diferentes tipos de modificações de eletrodos que têm como objetivo determinar simultaneamente compostos que se interferem [18,19]. Estas modificações de fato tornam as técnicas voltamétricas mais seletivas e muitas vezes mais sensíveis. Entretanto, mesmo quando boa seletividade e sensibilidade são alcançadas com estas modificações, há frequentes complicações relacionadas com os fenômenos de adsorção, necessidade de renovação da superfície após cada medida, instabilidade e baixa reprodutibilidade

[22,23]. Estas são outras razões que muitas vezes limitam o uso das técnicas voltamétricas e/ou amperométricas.

Recentemente, o diamante dopado com boro demonstra ser um material atrativo para aplicações eletroquímicas, tendo em vista suas propriedades únicas, principalmente em relação à sua estabilidade, robustez e inércia química (Capítulo 2) [17,24-26]. Porém, para compostos que se oxidam próximos a um mesmo potencial, uma baixa seletividade será ainda observada. Neste sentido, o uso de ferramentas quimiométricas em eletroanálise [21], tais como a calibração multivariada, aliada à robustez do diamante dopado com boro, apresenta-se como uma alternativa excelente para contornar este problema. As ferramentas quimiométricas são extremamente úteis em química analítica quando problemas de sobreposições de sinais são observados [27-29].

Nesta aplicação, uma metodologia eletroanalítica é apresentada para a determinação simultânea de GUA e CGUA diretamente em água de ribeirão contaminado com esgoto doméstico empregando o eletrodo de BDD.

Os compostos guaiacol (GUA) e cloroguaicol (CGUA) são espécies químicas que podem estar presentes em água de rios, ribeirões e lagoas devido ao resultado do descarte sem tratamento de rejeitos gerados por atividades industriais, principalmente por fábricas de papel, entre outras fontes antrópicas [30-32]. Os guaiacóis, tais como o GUA e CGUA são motivo de preocupação ambiental, visto que tais compostos, principalmente os clorados *e.g.* CGUA são tóxicos para espécies aquáticas, podem se acumular em animais bentônicos [33], são recalcitrantes e apresentam propensão à bioacumulação [31]. Desde modo, há um considerável interesse em suas medidas em matrizes ambientais [34]. A discriminação destes compostos a nível traço é importante devido às suas diferenças em toxicidade e impacto sobre organismos vivos. Porém, esta determinação seletiva é complicada visto que suas propriedades físico-químicas são muito similares [35].

Do ponto de vista eletroanalítico, a oxidação destes compostos GUA/CGUA em eletrodo de BDD está muito próxima de um mesmo potencial, o que resulta em uma alta sobreposição da resposta voltamétrica fazendo com que a discriminação de cada composto seja extremamente difícil. Neste sentido, o uso de ferramentas

quimiométricas *i.e.* calibração multivariada, é de fundamental importância para realizar a quantificação individual destes analitos.

Neste trabalho, o método de regressão por quadrados mínimos parciais (PLS) foi empregado para construir os modelos de calibração multivariada [27,36-38] diretamente em água contaminada por esgoto doméstico. Variáveis foram selecionadas usando o método OPS e após os modelos terem sido construídos e validados os parâmetros analíticos (seletividade, sensibilidade e limite de detecção) foram estimados [39,40].

Assim, o desenvolvimento de um método eletroanalítico direto, que permita a determinação simultânea de GUA e CGUA, sem prévia separação destes compostos, é o objetivo desta aplicação.

Experimental

Reagentes e soluções

Os reagentes, guaiacol e cloroguaiacol foram adquiridos da SigmaTM e AldrichTM, respectivamente. O fosfato de sódio e o cloreto de potássio foram adquiridos da SynthTM e o ácido cítrico da SigmaTM. Todos os reagentes são de grau analítico sem purificação adicional. Todas as soluções foram preparadas com água ultra pura de um sistema de purificação Milli QTM (>18 MΩ cm). As soluções estoque de cada analito foram preparadas com uma concentração de 2 mmol L⁻¹ em água e as misturas foram preparadas do estoque por diluições diretas em água contaminada com esgoto doméstico como necessário.

Amostra

Água do Rio das Pedras (água esta que recebe esgoto doméstico) foi coletada para simular uma matriz complexa. A coleta foi realizada em frascos plásticos e para cada litro de amostra 1 ml de HNO₃ concentrado foi adicionado para sua conservação. As amostras foram armazenadas em refrigerador na temperatura de aproximadamente 4 °C por um prazo máximo de cinco dias. Todas as medidas para a construção do modelo de calibração foram realizadas diretamente nesta amostra juntamente com o tampão, visto que modelos de calibração de primeira ordem

exigem que todos os possíveis interferentes que possam ser encontrados na amostra sejam modelados.

Medidas Eletroanalíticas

As medidas eletroquímicas foram executadas usando o eletrodo de calomelano saturado (SCE) como referência, uma espiral de platina como contra eletrodo e o BDD como eletrodo de trabalho. Todas as medidas foram realizadas em um potenciostato AUTOLAB™ PGSTAT 30 gerenciado pelo software GPES 4.9 a partir de um micro computador. Para evitar o acúmulo de possíveis filmes poliméricos formados durante as análises, após cada ensaio a solução foi agitada vigorosamente por aproximadamente 60 segundos. Para manter a superfície hidrogenada e assim garantir medidas reprodutivas por longo período de tempo, após aproximadamente 15 ensaios uma varredura de 0 a -3 V sob agitação em meio ácido foi realizada. Todos os experimentos foram executados em ordem aleatória.

Parâmetros Voltamétricos

A técnica voltamétrica empregada foi a voltametria de onda quadrada (SWV). As condições voltamétricas foram definidas usando metodologias de superfície de resposta com o objetivo de encontrar as melhores condições de análise. Os estudos foram realizados para cada analito separadamente e serão detalhados nos resultados. A concentração dos compostos na célula foi fixada em $100 \mu\text{mol L}^{-1}$. A faixa de potencial investigada foi de 0,5 V a 1,2 V. A linha de base de todos os voltamogramas foi corrigida usando uma janela de média móvel (Capítulo 1). A resposta usada neste estudo foi definida de acordo com a equação 5.1.

Conjunto de dados

Variáveis Dependentes

Um planejamento fatorial parcial para mistura com cinco níveis foi elaborado para obter as variáveis dependentes (concentrações), o que garantiu a ortogonalidade destas variáveis. As misturas de GUA e CGUA foram elaboradas sinteticamente a partir de padrões conforme descrito pelo planejamento diretamente

na mistura tampão – amostra, conforme indicado. Um total de 45 amostras foi utilizado neste estudo.

Variáveis independentes

Na faixa de potencial investigada 353 variáveis independentes foram obtidas. As colunas da matriz de dados foram centradas na média, corridas em relação à linha de base usando a média móvel (ver Capítulo 1) e potenciais foram selecionados com o objetivo de se obter as melhores previsões. O número de fatores selecionados para o modelo foi determinado por validação cruzada aplicando o método ‘*leave-n-out*’ e se baseou no cálculo do RMSECV, conforme equação 1.45. A habilidade de previsão do modelo foi avaliada pelo erro médio quadrático de previsão (RMSEP) e pelos coeficientes de correlação (r_p) conforme equações 1.47 e 1.48, respectivamente. Os parâmetros analíticos tais como seletividade, sensibilidade e limite de detecção, foram estimados conforme metodologias apresentadas por Lorber et al. (1997) [41], Bro e Andersen [40], Boqué e Rius [42] e Ortiz et al. 2003 [43].

Seleção de variáveis e amostras

As variáveis independentes foram selecionadas sobre o conjunto com todas as amostras usando o método de seleção das variáveis ordenadas (OPS[®]) [44,45] e também com o algoritmo genético (GA) para comparação [46]. Os parâmetros usados no GA foram investigados empregando a matriz saturada de Plackett-Burman [47-49].

Em ambos os métodos, as variáveis com melhor capacidade de previsão na validação cruzada *i.e.*, com menor valor de RMSECV, foram usadas.

As amostras de calibração e previsão foram separadas empregando o algoritmo de Kennard e Stone (1969) [50] após a realização das seleções de variáveis como descrito no Capítulo 5.

Análise dos dados

As análises de dados relativas aos planejamentos experimentais foram realizadas empregando planilhas eletrônicas do software Microsoft Excel[™] 2003

[51] e para a construção dos modelos de calibração as análises foram executadas usando funções *.m* para o software Matlab™ 7 (The MathWorks, Co., Natick, MA, USA). Os programas foram implementados no Laboratório de Quimiometria Teórica e Aplicada (LQTA) exceto aqueles empregando o GA, que foram realizados no pacote PLS_Toolbox 3.01 para Matlab.

Resultados e Discussão

Otimização das condições voltamétricas

Metodologia de superfície de resposta (RSM) [51] foram empregadas para encontrar as melhores condições operacionais de análise para a voltametria de onda quadrada (Capítulo 5 Figura 5.2). O planejamento composto central (CCD) com três variáveis foi aplicado inicialmente sobre os dois compostos individualmente com o objetivo de atingir mais rapidamente a região ótima. O incremento de potencial foi fixado em 0,002 mV. As medidas foram executadas em 5 ml de tampão McIlvaine 0,05 M. A resposta analisada foi definida na parte experimental do Capítulo 5. As variáveis investigadas foram: frequência da onda quadrada (*Freq.*), amplitude (*Ampl.*) e pH (*pH*).

A Figura 6.1 apresenta o planejamento CCD usado e a Tabela 6.1 mostra os fatores, níveis estudados e respostas obtidas.

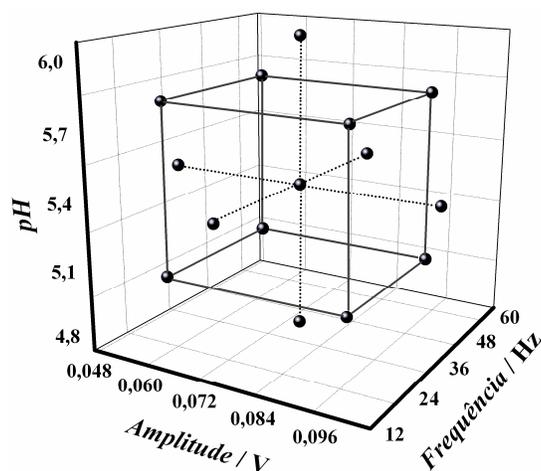


Figura 6.1. Representação geométrica do planejamento composto central usado para o estudo de três variáveis do sistema voltamétrico para a determinação de GUA e CGUA.

Tabela 6.1. Fatores, níveis investigados codificados e decodificados e respostas obtidas para a GUA e CGUA.

Ensaio	Fatores			Resposta (r_n)	
	Ampl./V	Freq./Hz	pH	GUA	CGUA
2	0,09 (+1)	20 (-1)	5,0 (-1)	21,8	13,0
3	0,05 (-1)	50 (+1)	5,0 (-1)	15,2	11,8
4	0,09 (+1)	50 (+1)	5,0 (-1)	16,4	14,5
7	0,05 (-1)	50 (+1)	5,8 (+1)	10,8	5,0
1	0,05 (-1)	20 (-1)	5,0 (-1)	17,2	12,2
16	0,07 (0)	35 (0)	5,4 (0)	15,6	9,3
14	0,07 (0)	35 (0)	6,07 (+1,682)	11,6	4,9
8	0,09 (+1)	50 (+1)	5,8 (+1)	13,4	6,4
5	0,05 (-1)	20 (-1)	5,8 (+1)	12,3	5,5
11	0,07 (0)	9,8 (- 1,682)	5,4 (0)	12,8	9,4
10	0,104 (+1,682)	35 (0)	5,4 (0)	17,8	10,2
9	0,036 (- 1,682)	35 (0)	5,4 (0)	10,9	5,8
12	0,07 (0)	60,2 (+1,682)	5,4 (0)	14,0	8,9
15	0,07 (0)	35 (0)	5,4 (0)	15,3	9,2
6	0,09 (+1)	20 (-1)	5,8 (+1)	17,3	7,6
13	0,07 (0)	35 (0)	4,73 (- 1,682)	20,3	14,6
18	0,07 (0)	35 (0)	5,4 (0)	16,7	9,4
17	0,07 (0)	35 (0)	5,4 (0)	12,46	8,30

Os ensaios estão colocados na ordem em que foram executados. Os níveis codificados estão entre parênteses e os decodificados, fora. A unidade da resposta é dada por: $A \cdot s^{1/2} \cdot V^{-3/2}$.

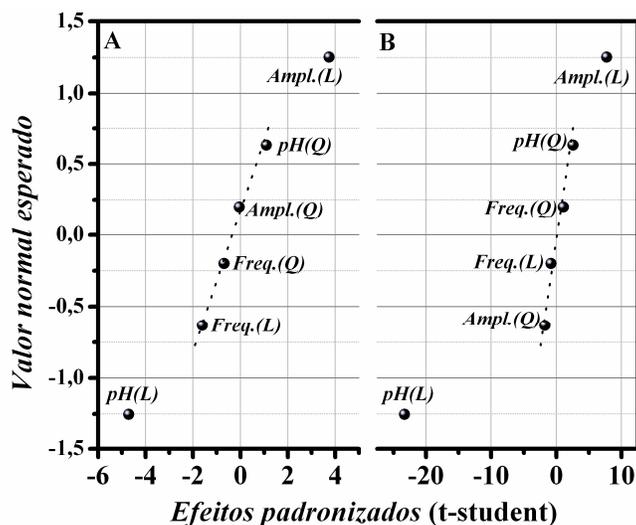


Figura 6.2. Gráfico de probabilidade normal para a GUA (A) e CGUA (B). Fatores fora da linha tracejada são significativos.

Os resultados mostrados na Figura 6.2 e na Tabela 6.2 sugerem que as variáveis *Ampl.* e *pH* são as mais importantes para ambos os compostos e

necessitam de um estudo mais detalhado. Os resultados mostram que um aumento da amplitude e/ou uma diminuição do pH provocam um aumento na resposta. O efeito da variável *Freq.* não foi significativo e foi fixada em 35 Hz.

Tabela 6.2. Efeitos principais do planejamento fracionário CCD para cada composto, e seus respectivos valores *p* calculados a partir do erro puro.

	GUA				CGUA			
	Efeitos	Erro puro	t	<i>p</i>	Efeitos	Erro puro	t	<i>p</i>
Intercepto	14,97*	0,91	16,53	0,0005	9,02	0,25	35,94	0,0000
<i>Ampl.</i>	3,68*	0,98	3,75	0,0332	2,10	0,27	7,73	0,0045
<i>Ampl.</i> ²	-0,04	1,02	-0,04	0,9686	-0,50	0,28	-1,75	0,1779
<i>Freq.</i>	-1,57	0,98	-1,60	0,2083	-0,22	0,27	-0,82	0,4742
<i>Freq.</i> ²	-0,70	1,02	-0,69	0,5413	0,30	0,28	1,08	0,3598
<i>pH</i>	-4,62*	0,98	-4,71	0,0181	-6,34	0,27	-23,30	0,0002
<i>pH</i> ²	1,12	1,02	1,10	0,3514	0,73	0,28	2,57	0,0827

*efeitos significativos a um nível de significância de 0,05 e 3 graus de liberdade pelo teste t.

As análises de variância (ANOVA) para o GUA e CGUA indicam, para ambos compostos, uma regressão significativa e uma falta de ajuste não significativa (Tabela 6.3). As variâncias explicadas foram de 83,65 e 97,44 % para GUA e CGUA, respectivamente.

Tabela 6.3. ANOVA para os modelos quadráticos obtidos com o planejamento CCD para GUA e CGUA.

GUA						
FV	SQ	nGL	MQ	<i>F</i> calc.	<i>p</i>	
Regressão	139,62*	9	15,51	4,54	0,0221	
Resíduos	27,29	8	3,41			
F. Ajuste	17,43 [†]	5	3,48	1,06	0,5143	
Erro Puro	9,87	3	3,29			
Total	166,91	17				
CGUA						
Regressão	156,93*	9	17,44	33,793	2,0x10 ⁻⁵	
Resíduos	4,13	8	0,52			
F. Ajuste	3,37 [†]	5	0,67	2,669	0,2244	
Erro Puro	0,76	3	0,25			
Total	161,06	17				

SQ. Soma quadrática, nGL. número de graus de liberdade, MQ. Média quadrática, *F*calc. Valor calculado da distribuição *F*. *valores significativos a um nível de significância de 0,05 pelo teste *F*. [†]valores não significativos a um nível de significância de 0,05 pelo teste *F*.

Os resultados da ANOVA tornam as interpretações das superfícies de resposta estatisticamente viáveis (Figuras 6.3).

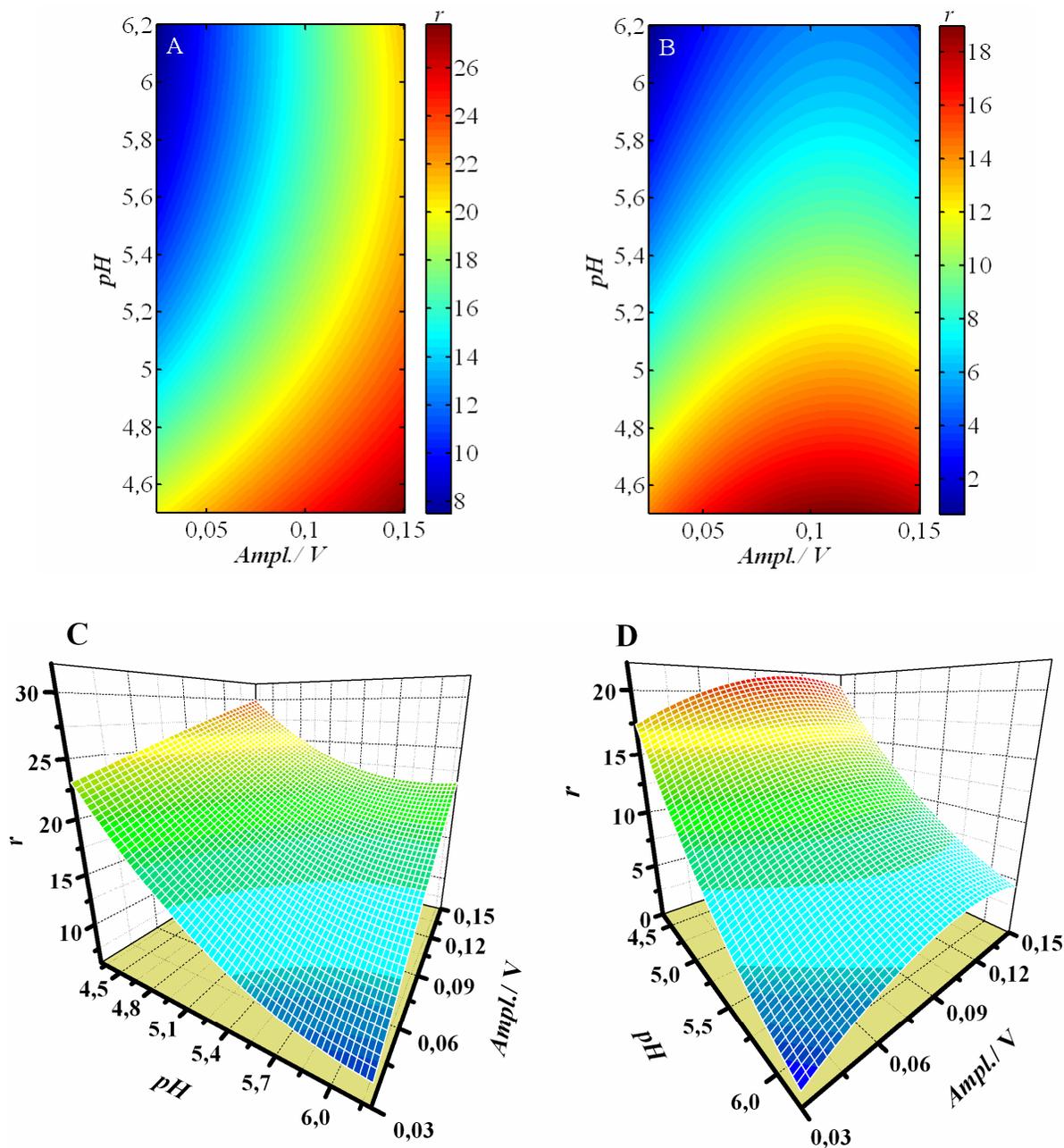


Figura 6.3. Superfícies de resposta extrapoladas para o GUA(A-C) e CGUA(B-D) obtidas através do planejamento CCD com a *Freq.* fixada em 35 Hz.

Como esperado, as superfícies confirmam que a amplitude deve ser aumentada e o pH deve ser diminuído para se obter maiores respostas. A amplitude certamente está aumentando a resposta devido à maior polarizabilidade do eletrodo

durante cada pulso, o que estimula a oxidação e, portanto, a obtenção de uma maior corrente. A diminuição do pH está seguramente relacionada à estabilidade da molécula e à facilidade de realizar a transferência de elétrons em meio levemente ácido além de contribuir para picos menos largos. O fato de a frequência da onda quadrada não ter sido significativa foi, possivelmente, devido à utilização da equação usada para definir a resposta (equação 5.1), que minimizou a influência da velocidade sobre o aumento da corrente.

Estes resultados recomendam para um deslocamento na região estudada, uma vez que uma condição ótima não foi obtida. Assim, um planejamento Doehlert [51-53] com duas variáveis *i.e* *pH* e *Ampl.* foi escolhido para investigar a nova região. Neste estudo apenas o GUA foi investigado devido às semelhanças eletroanalíticas com o CGUA e também devido à tendência de ambos de atingirem uma região ótima praticamente comum, conforme observado na Figura 6.3.

A Tabela 6.4 mostra a matriz Doehlert para duas variáveis, os valores experimentais empregados e as respostas obtidas em ensaios aleatórios. A Figura 6.4 apresenta a representação geométrica da matriz Doehlert.

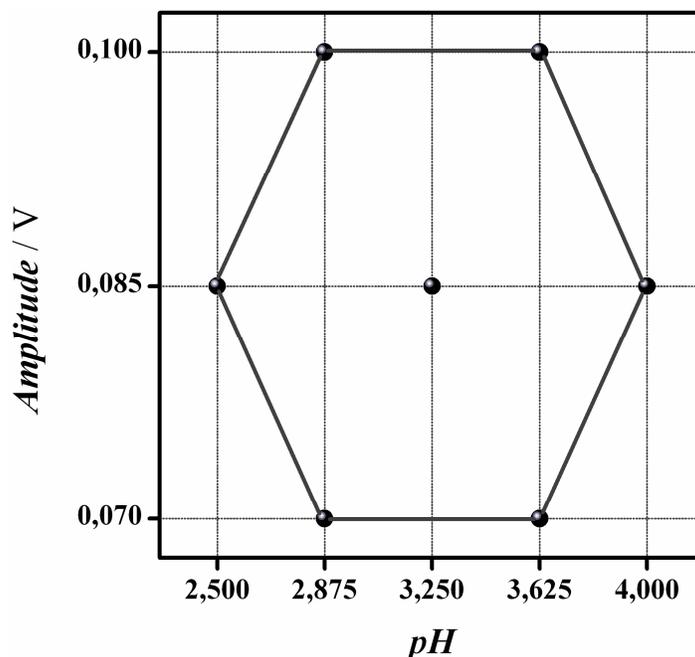


Figura 6.4. Pontos experimentais da matriz Doehlert representados geometricamente conforme as variáveis invetigadas.

Tabela 6.4. Planejamento Doehlert para duas variáveis, em ordem aleatória, para guaiacol.

Ordem Exp.	Níveis codificados		Níveis experimentais		Respostas
	Var.1	Var.2	pH	Ampl.	GUA
1	1	0	4	0,085	18,4
10	0	0	3,25	0,085	21,9
4	-0,5	-0,866	2,875	0,07	18,8
8	0	0	3,25	0,085	21,0
3	-1	0	2,5	0,085	21,5
5	0,5	-0,866	3,625	0,07	19,7
6	-0,5	0,866	2,875	0,1	21,4
11	0	0	3,25	0,085	21,1
7	0	0	3,25	0,085	20,7
2	0,5	0,866	3,625	0,1	20,0
9	0	0	3,25	0,085	21,6

A avaliação estatística dos coeficientes do modelo construído com a matriz Doehlert (níveis codificados), indicada na Tabela 6.5 ratificam a importância das variáveis *Ampl.* e *pH* para o sistema estudado. Entretanto, a interação entre estas variáveis (*pH x Ampl.*), não é significativa.

Tabela 6.5. Análise estatística dos coeficientes do modelo quadrático usando a matriz Doehlert codificada.

	Coef.	Erro puro	t	<i>p</i>
Intercepto*	21,25	0,23	90,51	0,0000
<i>pH</i> *	-1,13	0,30	3,74	0,0201
<i>Ampl.</i> *	0,84	0,30	2,77	0,0499
<i>pH</i> ² *	-1,30	0,44	2,97	0,0412
<i>Ampl.</i> ² *	-1,27	0,44	2,90	0,0441
<i>pH x Ampl.</i>	-1,38	0,61	2,27	0,0856

*Coeficientes significativos com 4 graus de liberdade pelo teste t a um nível de significância de 0,05.

Tabela 6.6. Análises de variância para o modelo do GUAI

FV	SQ	nGL	MQ	Fcalc.	<i>p</i>
Regressão	11,93*	5	2,39	5,10	0,049
Resíduos	2,34	5	0,47		
F. Ajuste	1,24 [†]	1	1,24	4,49	0,101
Erro Puro	1,10	4	0,28		
Total	14,27	10			

SQ. Soma quadrática, nGL. número de graus de liberdade, MQ. Média quadrática, Fcalc. Valor calculado da distribuição F. *valores significativos a um nível de significância de 0,05 pelo teste F. †valores não significativos a um nível de significância de 0,05 pelo teste F.

O modelo foi significativo para a regressão e não significativo para a falta de ajuste dentro de um nível de significância de 0,05 como indicado pela Tabela 6.6. A variação explicada foi de 83,60 %. O modelo quadrático estimado, considerando os coeficientes decodificados, ou seja, a matriz de níveis experimentais é representado na equação 6.1.

$$r = -62,40 + 22,56pH + 1114,80Ampl. - 2,32pH^2 - 4.246,00Ampl.^2 - 105,98pH \times Ampl. \quad 6.1$$

A análise de superfície de resposta (Figura 6.5) sugere que os melhores níveis das variáveis estudadas, para se obter as melhores respostas estão na faixa de pH 2,2 a 3,0 e amplitude de 0,09 a 1,10 V. Aqui foi escolhido as seguintes condições: pH 2,6 e Ampl. 0,1 V.

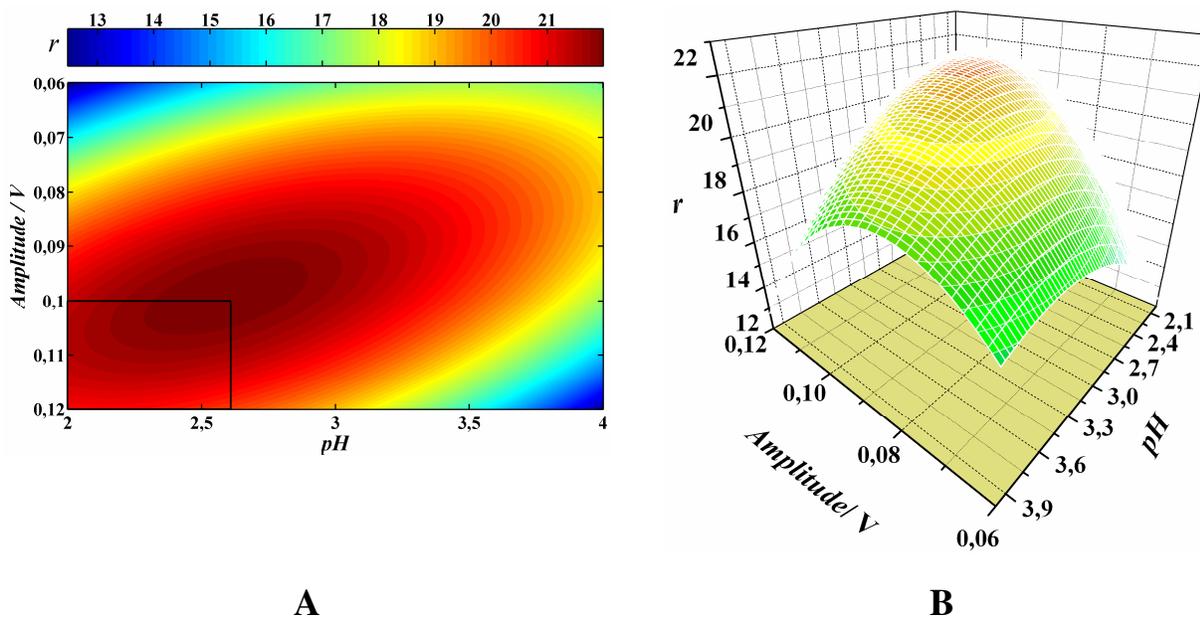


Figura 6.5. Superfícies de resposta obtidas através do modelo Doehlert para duas variáveis. Superfícies em 2 dimensões (A) e em 3 dimensões (B). As linhas pretas internas à superfície A indicam a região ótima usada.

Apesar de o primeiro planejamento ter atingido uma resposta muito próxima do valor ótimo obtido, é preciso ressaltar que naquele planejamento, a maior resposta foi obtida com uma frequência de 20 Hz enquanto no segundo, todas as respostas foram obtidas com uma frequência de 35 Hz. Assim, embora os valores serem parecidos é preciso levar em consideração a equação para se obter a resposta, que é inversamente dependente da raiz da velocidade de varredura. A menor

velocidade no primeiro planejamento contribuiu, portanto para uma maior resposta. Entretanto, foi observado que houve um aumento de 26 % no valor da corrente para a maior resposta do segundo planejamento em relação à maior no primeiro.

Modelos PLS

Utilizando as condições ótimas obtidas *i.e.*, tampão McIlvaine 0,05 M pH 2,6; amplitude 1,1 V e frequência de 35 Hz, um modelo de calibração multivariada foi construído para realizar previsões da mistura GUA e CGUA. A Figura 6.6 mostra os voltamogramas obtidos nas condições otimizadas.

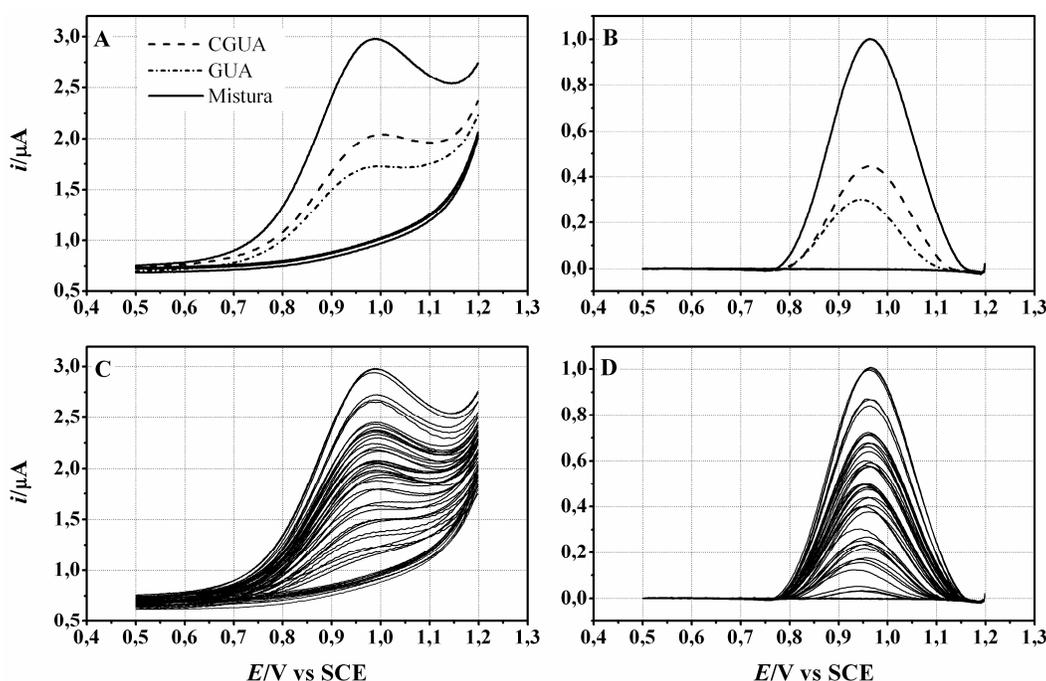


Figura 6.6. Voltamogramas de onda quadrada obtidos nas condições otimizadas. Voltamogramas puros e para a mistura antes (A) e após correção da linha de base com janela móvel (B) e voltamogramas das misturas para a construção do modelo antes (C) e após correção de linha de base (D). Os voltamogramas da água do rio no tampão sem os guaiacóis são mostrados abaixo das curvas. A largura da janela na correção de linha de base foi igual a 0,005 V.

Na Figura 6.6 A e B uma alta sobreposição dos três picos é observada o que indica uma baixa seletividade para ambos os componentes. Os potenciais de pico do GUA e CGUA são, para estes sistema de 0,946 e 0,962 V, respectivamente. Como observado também no Capítulo 5 os voltamogramas das misturas não apresentaram

nenhum tipo de estrutura que possa contribuir para identificar um ou outro componente, entretanto, o deslocamento dos picos no eixo das abscissas certamente traz informações importantes das variações das concentrações nas misturas.

Seleção de Variáveis

O GA [54] e o método OPS [44,45], ambos descritos no Capítulo 1, foram empregados para realizar a tarefa de encontrar as variáveis com maior potencial para realizar boas previsões.

Durante a seleção de variáveis empregando o método OPS, os seguintes vetores informativos foram empregados: o vetor de regressão do modelo PLS (R); o vetor de correlações entre cada variável independente e o vetor de variável dependente (C) e combinações entre estes vetores (RC). A Figura 6.7 mostra os vetores informativos calculados para GUA e CGUA com correção de linha de base. Ambos os vetores apresentam maior intensidade na região onde ocorre a oxidação dos analitos. Observou-se que para GUA e CGUA os vetores que apresentaram melhores resultados foram o R e o RC.

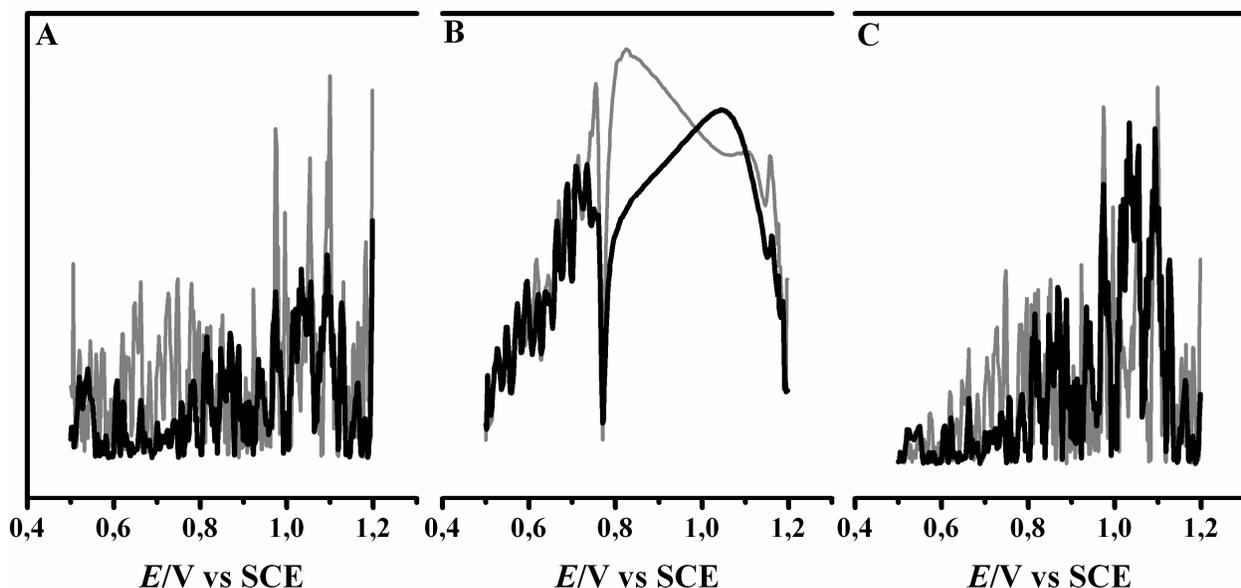


Figura 6.7. Módulo dos vetores informativos para GUA (linha cinza) e CGUA (linha preta) usados pelo OPS na seleção de variáveis. (A) vetor de regressão, (B) vetor de correlações X - y e (C) combinações entre estes vetores. Estes vetores estão ou normalizados e/ou ponderados e as unidades de suas intensidades são arbitrárias.

Uma análise dos três vetores mostrados na Figura 6.7 indica que as intensidades, na maioria das vezes, mostraram serem maiores em todos os vetores para o GUA e, para regiões de menores potenciais, os vetores para CGUA mostraram ter menor influência. Para ambos os compostos os vetores R e RC indicaram para a maior influência dos potenciais acima de aproximadamente 1,0 V, região onde ambos já atingiram o potencial de pico. O vetor de correlações, por outro lado, indicou regiões de menores potenciais para GUA e de maiores para CGUA o que está coerente com os potenciais de pico do GUA e CGUA que são de 0,946 e 0,962 V, respectivamente.

Para o algoritmo genético, as variáveis e níveis investigados são mostrados na Tabela 5.6 no Capítulo 5. As configurações que apresentaram melhores resultados de um conjunto de dez ensaios, definidos pela matriz Plackett-Burman [47-49], são mostradas na Tabela 6.7.

Para cada conjunto e analito, dez ensaios foram executados em triplicatas, usando o algoritmo PLS com uma validação cruzada aleatória com *split* de 5 amostras em 5 iterações. Foi usada a matriz contendo todas as amostras. A resposta obtida foi o valor de RMSECV e a replicatas que apresentou o menor valor de RMSECV foi utilizada para comparação.

Tabela 6.7. Parâmetros selecionados para executar o algoritmo genético.

	Originais		Correção	
	GUA	CGUA	GUA	CGUA
População	152	152	152	100
Largura da janela	3	3	3	2
Taxa de Mutação	0,001	0,005	0,001	0,003
Inicialização	20	50	20	35
<i>Cross over</i>	2	2	2	2
Gerações	150	150	150	100
Convergência	20	70	20	45

A Figura 6.8 mostra os resultados obtidos quando o método OPS foi executado para GUA e CGUA com e sem correção de linha de base. As variáveis foram investigadas com janela e incremento iguais a cinco. O valor de RMSECV obtido usando todas as amostras foi coletado para cada conjunto com um valor fixo de componentes. Note que há diminuições significativas no valor de RMSECV em todas as seleções, além disso, os erros para o CGUA foram maiores em todos os

conjuntos. Observa-se também que os valores de RMSECV são maiores para o conjunto original, o que confirma a necessidade de correção de linha de base para se atingir melhores previsões. Possivelmente, isto é um indicativo que a corrente responsável pela inclinação da linha de base ao ser removida, aumenta a relação sinal ruído e, portanto, uma maior informação analítica é obtida, o que melhora a capacidade de previsão dos modelos.

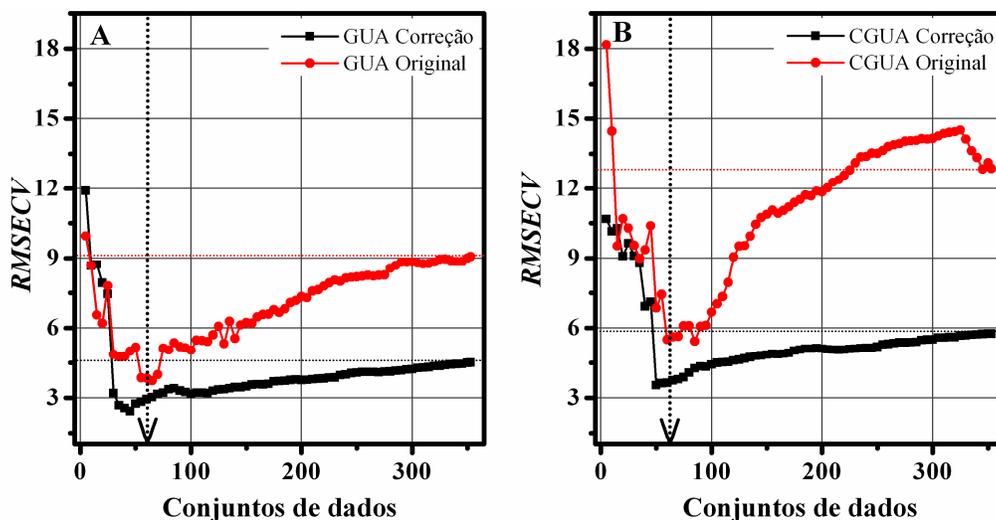


Figura 6.8. Gráficos característicos usados para a seleção de variáveis com o método OPS. As linhas horizontais indicam o valor de RMSECV com todas as variáveis e as linhas verticais indicam o conjunto escolhido para seleção com correção. (A) seleções realizadas para o GUA e (B) seleções realizadas para o CGUA. As 45 amostras foram usadas nestas seleções e o número de componentes foram fixados em 6 para A e 5 para B. Os símbolos ● e ■ indicam os voltamogramas sem e com correção de linha de base, respectivamente.

Após a seleção de variáveis os conjuntos de calibração e o conjunto de validação externa foram separados usando o Algoritmo de Kennard-Stone [50] (ver Capítulo 5). As concentrações média, mínima e máxima para GUA no conjunto de calibração foram respectivamente de 37,79; 3,02 e 74,46 $\mu\text{mol L}^{-1}$ e para CGUA foram respectivamente de 48,68; 3,19 e 85,22 $\mu\text{mol L}^{-1}$.

Os parâmetros estatísticos dos modelos construídos para os conjuntos originais e com correção de linha de base são apresentados na Tabela 6.8. Observe que novamente os maiores erros (RMSECV e RMSEP) foram obtidos para os modelos que utilizaram todas as variáveis. Isto confirma a importância de se

realizar uma seleção de variáveis em calibração multivariada com dados voltamétricos.

Em relação aos algoritmos para seleção de variáveis note na Tabela 6.8 que ambos mostraram ter resultados similares e os melhores resultados foram obtidos quando se realizou seleção sobre os dados com correção de linha de base.

Tabela 6.8. Parâmetros estatísticos de diferentes modelos PLS construídos para a quantificação simultânea de GUA e CGUA.

		Originais			Correção		
		Completo	Seleção		Completo	Seleção	
			OPS [®]	GA		OPS [®]	GA
GUA	n° variáveis	353	55	95	353	60	65
	n° componentes		6			6	
	RMSECV	10,3	4,1	4,4	5,1	3,7	3,6
	r_{cv}	0,924	0,985	0,983	0,978	0,988	0,989
	RMSEP	6,6	4,1	3,6	5,5	2,6	2,6
	r_p	0,959	0,984	0,987	0,976	0,995	0,995
CGUA	n° variáveis	353	70	85	353	60	90
	n° componentes		5			5	
	RMSECV	16,7	7,3	7,2	6,8	4,4	4,5
	r_{cv}	0,803	0,960	0,961	0,965	0,986	0,985
	RMSEP	9,6	4,8	4,7	6,8	2,6	2,8
	r_p	0,928	0,983	0,986	0,973	0,996	0,997

A Figura 6.9 mostra as variáveis selecionadas pelo GA e pelo método OPS para os conjuntos com correção de linha de base. Nota-se que em ambos os conjuntos, o GA se mostrou bastante disperso, não reconhecendo regiões específicas do voltamograma e selecionando linha de base. O OPS, por outro lado, mostrou ser bem mais objetivo, selecionando regiões mais definidas e que dão condições para interpretações. Observa-se também, que muitas variáveis selecionadas pelo GA, foram também selecionadas pelo OPS.

De acordo com as regiões selecionadas pelo OPS, sugere-se que os potenciais de pico definiram as regiões com informações analíticas nestes sistemas de baixa seletividade. Embora o GUA e CGUA tenham exigido o mesmo número de variáveis, aquelas selecionadas pelo OPS para o GUA, ficaram distribuídas por toda região de potencial, antes e após do potencial de pico (0,946 V). Porém, para o CGUA, as variáveis selecionadas ficaram agrupadas principalmente após o

potencial de pico do CGUA (0,962 V). Isto sugere que antes do potencial de pico do CGUA as informações para GUA estão mais disponíveis e após o potencial de 0,9 V, onde ambos estão oxidando, é possível obter informações para CGUA.

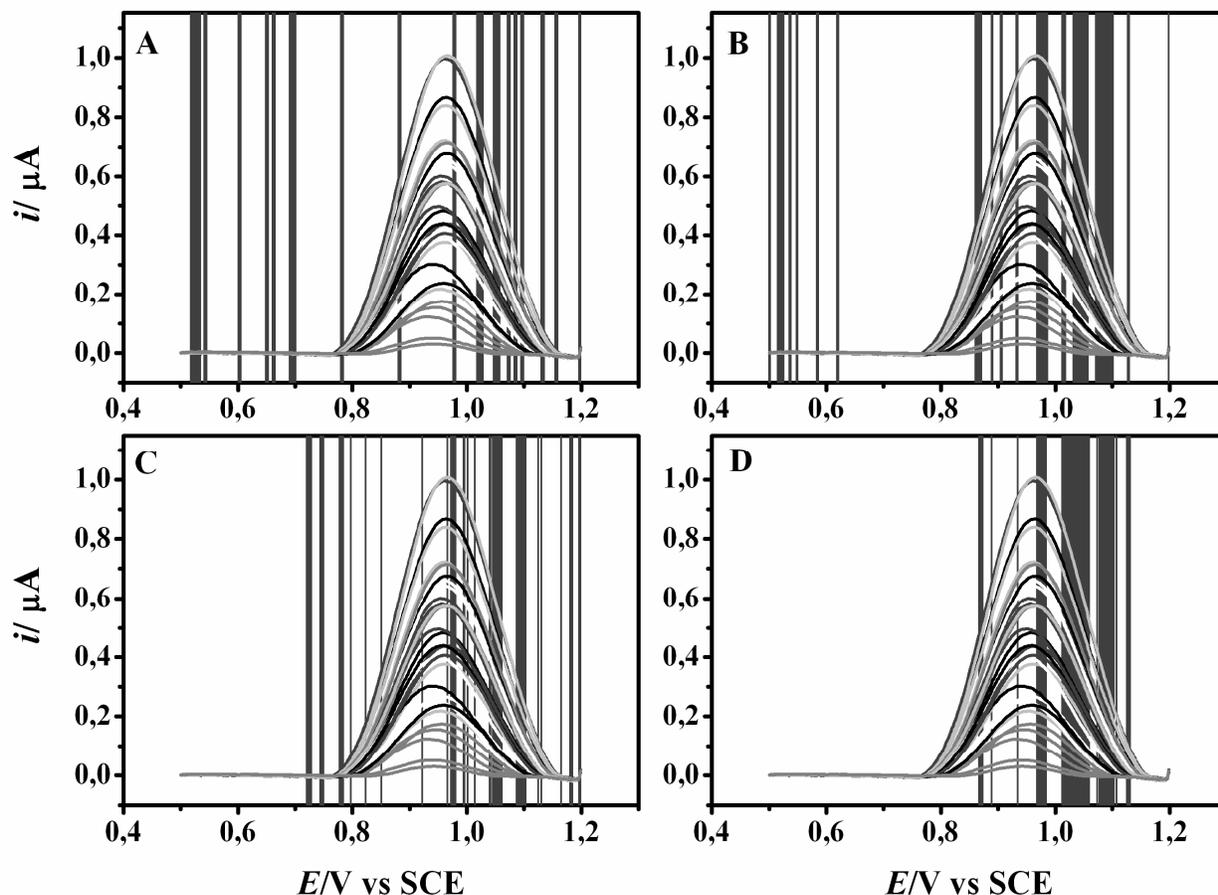


Figura 6.9. Variáveis selecionadas pelo GA para os compostos GUA (A), CGUA(B) e variáveis selecionadas pelo OPS para os compostos GUA(C) e CGUA (D). As linhas verticais indicam as variáveis selecionadas.

Tendo em vista o tempo de cálculo para os algoritmos GA e OPS, as variáveis selecionadas e os resultados estatísticos obtidos; fica bastante claro que o OPS é o método mais viável e coerente para se selecionar variáveis. Além disso, como pode ser observado na Tabela 6.9 e na Figura 6.10, os modelos construídos com voltamogramas que receberam correção de linha de base e tiveram variáveis selecionadas pelo OPS se mostraram aptos a realizar previsões para novas amostras. Como todo o modelo foi construído diretamente em água contaminada com esgoto doméstico, eles são capazes de realizar previsões de GUA e CGUA em

amostras de água nas condições em que foram coletadas sem necessidade de separação prévia.

Das quinze amostras previstas pelo modelo construído para o GUA observa-se apenas um erro maior que 10%, indicando que o modelo está com excelente capacidade de previsão. Para o modelo construído para o CGUA, notam-se três erros acima de 10% e ambos abaixo de 15%. Apesar destes erros, o modelo mostra-se apto para realizar previsões com qualidade se considerarmos a amostra e as concentrações estudadas.

Tabela 6.9. Concentrações previstas para GUA e CGUA e seus respectivos erros relativos

	GUA			CGUA		
	Ref.	Prev.	Erro relat.	Ref.	Prev.	Erro relat.
1	40,9	39,2	4,3	45,9	47,8	-4,1
2	40,9	43,7	-6,8	45,9	45,9	0,1
3	40,9	40,8	0,3	45,9	45,1	1,8
4	23,5	22,3	5,0	84,4	81,4	3,6
5	74,4	70,7	5,1	45,9	47,3	-3,1
6	42,5	40,9	3,7	27,6	31,0	-12,5
7	78,1	73,4	6,0	8,6	9,9	-14,9
8	60,8	58,5	3,8	28,7	29,0	-1,1
9	70,7	67,9	3,9	44,5	41,1	7,7
10	19,7	20,4	-3,5	22,3	20,3	9,2
11	8,9	8,6	4,0	8,9	8,0	10,5
12	56,2	57,8	-2,9	83,7	79,3	5,2
13	75,1	73,3	2,4	84,5	87,0	-3,0
14	10,7	9,8	8,5	42,2	41,8	1,0
15	39,5	45,1	-14,2	81,6	76,0	6,9

Ref. e Prev. são as concentrações de referência e previstas dadas em $\mu\text{mol L}^{-1}$, Erros relat. são os erros relativos dados em percentagem.

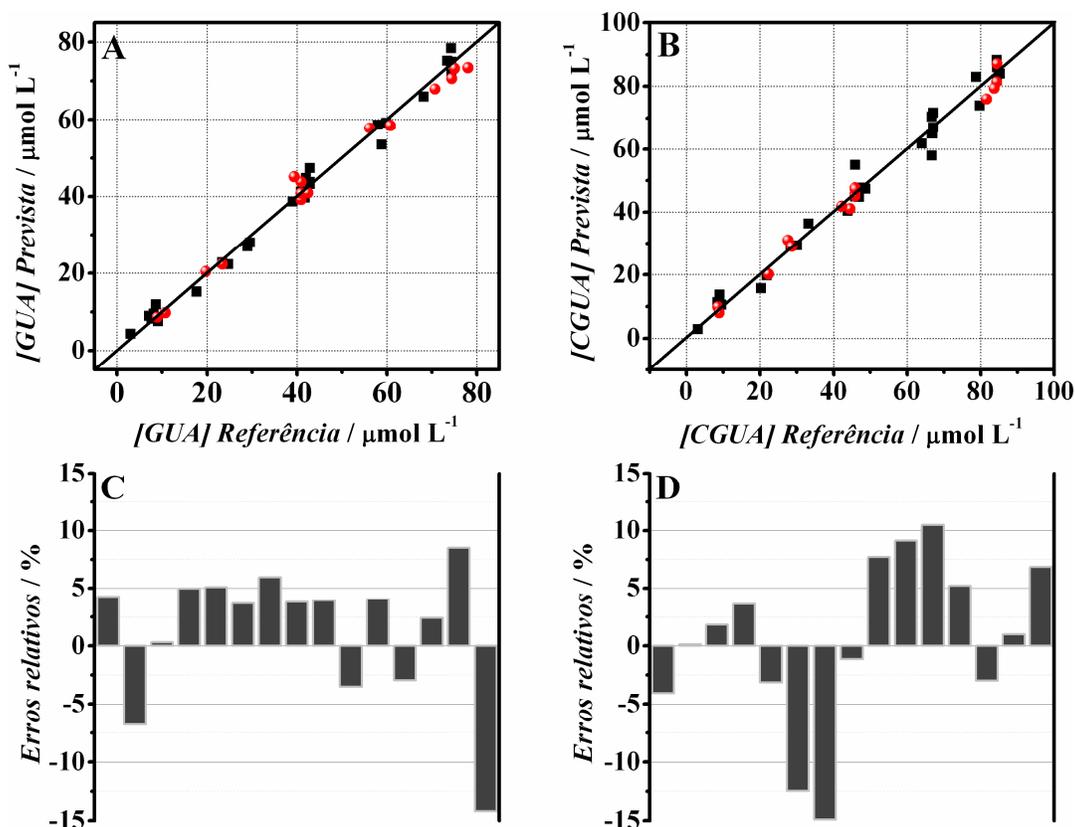


Figura 6.10. Concentrações referência vs. previstas para o modelo (■) e para o conjunto externo de previsão (●) para GUA (A) e CGUA (B). As barras verticais são os erros relativos para as amostras externas para GUA (C), CGUA (D).

Alguns parâmetros analíticos foram calculados para os modelos de maneira a verificar a qualidade da análise. A Tabela 6.10 mostra os valores de seletividade, sensibilidade e limite de detecção (LOD) calculados para os modelos finais. A sensibilidade e a seletividade foram estimadas empregando o cálculo do NAS (Capítulo 1) [40,41,55]. Para o limite de detecção, quatro diferentes métodos foram utilizados uma vez que ainda não há um método bem definido para realizar este cálculo para modelos de calibração multivariada. Os métodos empregados estão discutidos no Capítulo 1.

Tabela 6.10. Parâmetros analíticos para os modelos finais de ambos os conjuntos

	Seletividade	Sensibilidade	LOD ^a _{ort}	LOD ^b _{br}	LOD ^c _σ	LOD ^d _{mod}
GUA	0,0197	0,43	3,1	1,3	4,5	1,7
CGUA	0,0186	0,55	5,8	3,0	12,8	2,5

^aMétodo de Ortiz, ^bMétodo de Boqué e Rius, ^cMétodo do desvio padrão, ^dMétodo proposto. A sensibilidade é dada em $\mu\text{A}/\text{mmol L}^{-1}$ e o LOD em $\mu\text{mol L}^{-1}$.

Os baixos valores de seletividade confirmam o que era esperado, visto que uma alta sobreposição dos picos foi obtida. O fato de a seletividade do GUA ser levemente maior que a do CGUA talvez esteja relacionado com o potencial de pico, pois o GUA apresenta potencial de pico menor que a do CGUA. Assim, como o GUA difunde primeiramente para a superfície do eletrodo, ele dificulta o acesso do CGUA o que certamente influencia a seletividade.

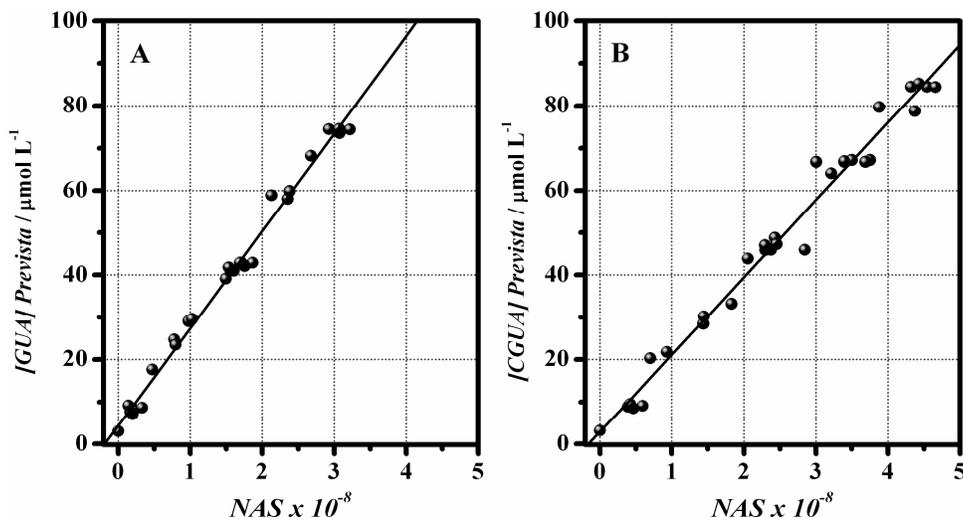


Figura 6.11. Gráficos pseudo-univariados (NAS vs. referência) para o GUA(A), CGUA(B).

A sensibilidade do CGUA é levemente maior que a do GUA. Na Figura 6.11 este parâmetro analítico é mostrado pelos gráficos pseudo-univariados que usam os valores do NAS vs os valores de referência. Como o PLS é um modelo inverso, observe que as inclinações são inversas em relação às sensibilidade calculadas.

Para discutir sobre os valores estimados para o limite de detecção é necessário lembrar que a concentração mínima do GUA no conjunto de calibração foi $3,02 \mu\text{mol L}^{-1}$ e para CGUA foi de $3,19 \mu\text{mol L}^{-1}$, espera-se portanto, valores menores que estes, que foram usados para a construção do modelo. Em relação aos valores estimados pelo método de Ortiz et al. [43], note que para ambos analitos as estimativas estão acima do valor mínimo encontrado no modelo o que torna este método questionável. O mesmo ocorre com o método do desvio padrão [39], que faz estimativas para valores ainda maiores que o método de Ortiz. Os métodos que estão mais coerentes são aquele apresentado por Boqué e Rius [42] e o método

proposto. Assim, pode-se considerar que o limite de detecção para a determinação de GUA e CGUA são uma média dos valores encontrados por estes dois métodos.

Tabela 6.11. Limites máximos definidos pelo CONAMA para alguns compostos fenólicos e fenóis totais.

	Águas doces	Águas salinas	Águas salobras	Lançamento de Efluentes
	<i>Classe 1</i>	<i>Classe 1</i>	<i>Classe 1</i>	
Parâmetros	Valor máximo	Valor máximo	Valor máximo	Valor máximo
2-clorofenol	0,1 µg L ⁻¹			
Fenóis totais	3,0 µg L ⁻¹	60 µg L ⁻¹	3,0 µg L ⁻¹	500 µg L ⁻¹
Pentaclorofenol	9,0 µg L ⁻¹	7,9 µg L ⁻¹	7,9 µg L ⁻¹	
	<i>Classe 1 (d)</i>	<i>Classe 1 (c)</i>	<i>Classe 1 (c)</i>	
Pentaclorofenol	3 µg L ⁻¹	3,0 µg L ⁻¹	3,0 µg L ⁻¹	
2,4,6-triclorofenol	2,4 µg L ⁻¹	2,4 µg L ⁻¹	2,4 µg L ⁻¹	
2-clorofenol		150 µg L ⁻¹	150 µg L ⁻¹	
2,4-Diclorofenol		290 µg L ⁻¹	290 µg L ⁻¹	
	<i>Classe 3</i>			
Fenóis totais	10,0 µg L ⁻¹			
Pentaclorofenol	9,0 µg L ⁻¹			

(d), (c) padrões para corpos de água onde haja pesca ou cultivo de organismos para fins de consumo intensivo.

Segundo a resolução CONAMA nº 357 de 2005, Art. 3º [12], as águas doces, salobras e salinas do Território Nacional são classificadas segundo a qualidade requerida para os seus usos preponderantes em treze classes. Para algumas destas classes o CONAMA definiu alguns padrões a partir de concentrações máximas de determinados parâmetros. Entre estes parâmetros há alguns tipos de fenóis e concentração máxima de fenóis totais, que são mostrados na Tabela 6.11 para suas respectivas classes de águas.

Da Tabela 6.10 os limites de detecção do GUA e CGUA em massa/volume são iguais a 186,2 µg L⁻¹ e 435,30 µg L⁻¹, respectivamente. Com estes limites de detecção, note que este método estaria apto a determinar fenóis totais em efluentes para lançamento em corpos de água se tal efluente contiver unicamente GUA e CGUA. Isto torna vantajoso o uso desta metodologia para este tipo de monitoramento devido às suas características robustas e não necessidade de uso de técnicas de separação. Porém, na literatura, encontra-se limite de detecção para análise de fenóis e clorofenóis em BDD na faixa de 1 a 110 µg L⁻¹ [56-59],

entretanto, nenhum trabalho descreve a análise simultânea de GUA e CGUA simultaneamente usando BDD. Portanto, de uma maneira geral, os resultados encontrados neste trabalho indicam que, apesar de o limite de detecção atender uma das classes de água definida pelo CONAMA, é necessário melhorar este método, principalmente as características eletroanalíticas do BDD, para atingir valores de detecção ainda menores.

Conclusões

Neste trabalho as condições voltamétricas foram otimizadas empregando metodologia de superfície de resposta com os planejamentos CCD e Doehlert. Utilizando as condições otimizadas modelos de calibração multivariada foram construídos usando o método PLS diretamente em água contaminada com esgoto doméstico. Modelos foram construídos com os voltamogramas originais e com correção de linha de base, além de modelos com todas e variáveis e com potenciais selecionados usando os métodos OPS e GA. Uma comparação entre os resultados sugeriu que os voltamogramas deveriam ser pré-tratados com correção de linha de base e os potenciais deveriam ser selecionados empregando o método OPS. Os modelos que foram construídos com variáveis selecionadas pelo método OPS mostraram ter maior capacidade de predição que aqueles construídos com todas as variáveis. Os resultados dos métodos de seleção foram similares, porém o método OPS mostrou ser mais eficiente computacionalmente além selecionar variáveis interpretáveis.

O uso do diamante dopado com boro foi fundamental para definir esta metodologia, uma vez que os analitos quantificados apresentam grande poder de bloqueio eletroquímico sobre eletrodos sólidos convencionais, tais como platina e carbono vítreo.

Os parâmetros analíticos mostraram que a determinação simultânea de GUA e CGUA indicaram para a possibilidade de empregar o eletrodo de BDD para o monitoramento de fenóis totais em lançamento de efluentes. Apesar disso, melhorias do eletrodo para se obter ganhos em sensibilidade e menores limites de detecção são necessárias. Porém, avanços foram alcançados na aplicação de

ferramentas quimiométricas neste tipo de sistema para a determinação simultânea destes analitos.

Estes resultados reafirmam a promissora união entre o eletrodo de diamante dopado com boro, a eletroanalítica e a quimiometria para a determinação de misturas de compostos eletroativos que apresentam alto poder de bloqueio eletroquímico em eletrodos sólidos tais como platina, carbono vítreo, ouro, etc.

Referências

1. Farre, M.; Barcelo, D. Toxicity testing of wastewater and sewage sludge by biosensors, bioassays and chemical analysis. *Trac-Trends Anal. Chem.* **2003**, 22, (5), 299-310.
2. Benfenati, E.; Barcelo, D.; Johnson, I.; Galassi, S.; Levsen, K. Emerging organic contaminants in leachates from industrial waste landfills and industrial effluent. *Trac-Trends Anal. Chem.* **2003**, 22, (10), 757-765.
3. Panizza, M.; Michaud, P. A.; Cerisola, G.; Comninellis, C. Anodic oxidation of 2-naphthol at boron-doped diamond electrodes. *J. Electroanal. Chem.* **2001**, 507, (1-2 SI Sp. Iss. SI), 206-214.
4. Puig, D.; Barcelo, D. Determination of phenolic compounds in water and waste water. *Trac-Trends Anal. Chem.* **1996**, 15, (8), 362-375.
5. Canizares, P.; Lobato, J.; Paz, R.; Rodrigo, M. A.; Saez, C. Electrochemical oxidation of phenolic wastes with boron-doped diamond anodes. *Water Res.* **2005**, 39, (12), 2687-2703.
6. Freire, R. S.; Duran, N.; Kubota, L. T. Development of a laccase-based flow injection electrochemical biosensor for the determination of phenolic compounds and its application for monitoring remediation of Kraft E1 paper mill effluent. *Anal. Chim. Acta* **2002**, 463, (2), 229-238.
7. Rosatto, S. S.; Freire, R. S.; Duran, N.; Kubota, L. T. Amperometric biosensors for phenolic compounds determination in the environmental interest samples. *Quim. Nova* **2001**, 24, (1), 77-86.
8. Lacorte, S.; Latorre, A.; Barcelo, D.; Rigol, A.; Malmqvist, A.; Welander, T. Organic compounds in paper-mill process waters and effluents. *Trac-Trends Anal. Chem.* **2003**, 22, (10), 725-737.
9. Ortega, F.; Dominguez, E.; Burestedt, E.; Emneus, J.; Gorton, L.; Markovarga, G. Phenol oxidase-based biosensors as selective detection units in column liquid-chromatography for the determination of phenolic-compounds. *J. Chromatogr. A* **1994**, 675, (1-2), 65-78.

10. Asakawa, E.; Hirose, H.; Hagiwara, A.; Takahashi, S.; Ito, N. Carcinogenicity of 4-methoxyphenol and 4-methoxycatechol in F344 rats. *Int. J. Cancer* **1994**, 56, 146.
11. Garg, R.; Kurup, A.; Hansch, C. Comparative QSAR: On the toxicology of the phenolic OH moiety. *Crit. Rev. Toxicol.* **2001**, 31, (2), 223-245.
12. MMA resoluções, *Resoluções do Conama*. Ministério do Meio Ambiente: Brasília, pp 255-283.
13. Vincent, G. Organic micropollutants in the aquatic environment. Angeletti, G. and Bjorseth, A., Eds. Kluwer Dordrecht: 1991, pp 285-292.
14. EPA method 604, *Phenols in federal register*. Part VIII, 40 CFR, Environmental Protection Agency: New York, Friday October 26, pp 58-66.
15. Campanella, L.; Beone, T.; Sammartino, M. P.; Tomassetti, M. Determination of phenol in wastes and water using an enzyme sensor. *Analyst* **1993**, 118, (8), 979-986.
16. Muna, G. W.; Quaiserova-Mocko, V.; Swain, G. M. The analysis of chlorinated phenol solutions by capillary electrophoresis coupled with direct and indirect amperometric detection using a boron-doped diamond microelectrode. *Electroanalysis* **2005**, 17, (13), 1160-1170.
17. Prado, C.; Murcott, G. G.; Marken, F.; Foord, J. S.; Compton, R. G. Detection of chlorophenols in aqueous solution via hydrodynamic channel flow cell voltammetry using a boron-doped diamond electrode. *Electroanalysis* **2002**, 14, (14), 975-979.
18. Wang, J. Remote electrochemical sensors for monitoring inorganic and organic pollutants. *Trac-Trends Anal. Chem.* **1997**, 16, (2), 84-88.
19. Jin, G. P.; Lin, X. Q.; Ding, Y. F. Glassy carbon electrodes modified with mixed covalent monolayers of choline, glycine, and glutamic acid for the determination of phenolic compounds. *J. Solid State Electrochem.* **2006**, 10, (12), 987-994.
20. Gutes, A.; Ibanez, A.; Cespedes, F.; Alegret, S.; del Valle, M. Simultaneous determination of phenolic compounds by means of an automated voltammetric "electronic tongue". *Anal. Bioanal. Chem.* **2005**, 382, (2), 471-476.
21. De Carvalho, R. M.; Mello, C.; Kubota, L. T. Simultaneous determination of phenol isomers in binary mixtures by differential pulse voltammetry using carbon fibre electrode and neural network with pruning as a multivariate calibration tool. *Anal. Chim. Acta* **2000**, 420, (1), 109-121.
22. Strein, T. G.; Ximba, B. J.; Hamad, A. H. Steady-state voltammetry of catechol and guaiacol analogues at carbon fiber microdisk electrodes following laser and electrochemical activation procedures. *Electroanalysis* **1999**, 11, (1), 37-46.
23. Zen, J. M.; Chen, P. J. A selective voltammetric method for uric acid and dopamine detection using clay modified electrodes. *Anal. Chem.* **1997**, 69, (24), 5087-5093.

24. Teófilo, R. F.; Ceragioli, H. J.; Peterlevitz, A. C.; Da Silva, L. M.; Damos, F. S.; Ferreira, M. M. C.; Baranauskas, V.; Kubota, L. T. Improvement of the electrochemical properties of "as-grown" boron-doped polycrystalline diamond electrodes deposited on tungsten wires using ethanol. *J. Solid State Electrochem.* **2007**, 11, 1449-1457.
25. Granger, M. C.; Witek, M.; Xu, J. S.; Wang, J.; Hupert, M.; Hanks, A.; Koppang, M. D.; Butler, J. E.; Lucazeau, G.; Mermoux, M.; Strojek, J. W.; Swain, G. M. Standard electrochemical behavior of high-quality, boron-doped polycrystalline diamond thin-film electrodes. *Anal. Chem.* **2000**, 72, (16), 3793-3804.
26. Muna, G. W.; Tasheva, N.; Swain, G. M. Electro-oxidation and amperometric detection of chlorinated phenols at boron-doped diamond electrodes: A comparison of microcrystalline and nanocrystalline thin films. *Environ. Sci. Technol.* **2004**, 38, (13), 3674-3682.
27. Martens, H.; Naes, T. *Multivariate Calibration*. John Wiley & Sons: New York, 1989; p 419.
28. Naes, T.; Isaksson, T.; Fearn, T.; Davies, T. *Multivariate calibration and classification*. Nir publications: Chinchester, 2002; p 344.
29. Workman, J. J.; Mobley, P. R.; Kowalski, B. R.; Bro, R. Review of chemometrics applied to spectroscopy: 1985-95 .1. *Appl. Spectrosc. Rev.* **1996**, 31, (1-2), 73-124.
30. Michalowicz, J.; Duda, R. Analysis of chlorophenols, chlorocatechols, chlorinated methoxyphenols and monoterpenes in communal sewage of Lodz and in the Ner River in 1999-2000. *Water Air Soil Pollut.* **2005**, 164, (1-4), 205-222.
31. Gundersen, J. L.; MacIntyre, W. G. Dissociation constants of chloroguaiacols in water: A comparison of measured and predicted values. *Environ. Toxicol. Chem.* **1996**, 15, (6), 809-813.
32. Paasivirta, J.; Heinola, K.; Humppi, T.; Karjalainen, A.; Knuutinen, J.; Mantykoski, K.; Pauku, R.; Piilola, T.; Surmaaho, K.; Tarhanen, J.; Welling, L.; Vihonen, H.; Sarkka, J. Polychlorinated phenols, guaiacols and catechols in environment. *Chemosphere* **1985**, 14, (5), 469-491.
33. Kovacs, T. G.; Martel, P. H.; Voss, R. H.; Wrist, P. E.; Willes, R. F. Aquatic toxicity equivalency factors for chlorinated phenolic-compounds present in pulp-mill effluents. *Environ. Toxicol. Chem.* **1993**, 12, (2), 281-289.
34. Freire, R. S.; Thongngamdee, S.; Duran, N.; Wang, J.; Kubota, L. T. Mixed enzyme (laccase/tyrosinase)-based remote electrochemical biosensor for monitoring phenolic compounds. *Analyst* **2002**, 127, (2), 258-261.
35. Korenman, Y. I.; Alyмова, A. T.; Vatutina, I. V. Determination of phenol and guaiacol in aqueous solutions using extraction-chromatographic preconcentration. **2002**, 57, (1), 31-34.

36. Ferreira, M. M. C.; Antunes, A. M.; Melgo, M. S.; Volpe, P. L. O. Chemometrics I: Multivariate calibration, a tutorial. *Quim. Nova* **1999**, 22, (5), 724-731.
37. Beebe, K. R.; Pell, R. J.; Seasholtz, M. B. *Chemometrics: A practical guide*. 1 ed.; John Wiley & Sons Inc: New York, 1998; p 348.
38. Brereton, R. G. *Chemometrics: Data Analysis for the laboratory and chemical plant*. John Wiley & Sons Inc: Chinchester, 2003; p 407.
39. Booksh, K. S.; Kowalski, B. R. Theory of analytical chemistry. *Anal. Chem.* **1994**, 66, (15), 782A-791A.
40. Bro, R.; Andersen, C. M. Theory of net analyte signal vectors in inverse regression. *J. Chemometr.* **2003**, 17, (12), 646-652.
41. Lorber, A.; Faber, K.; Kowalski, B. R. Net analyte signal calculation in multivariate calibration. *Anal. Chem.* **1997**, 69, (8), 1620-1626.
42. Boque, R.; Rius, F. X. Computing detection limits in multicomponent spectroscopic analysis. *Trac-Trends Anal. Chem.* **1997**, 16, (8), 432-436.
43. Ortiz, M. C.; Sarabia, L. A.; Herrero, A.; Sanchez, M. S.; Sanz, M. B.; Rueda, M. E.; Gimenez, D.; Melendez, M. E. Capability of detection of an analytical method evaluating false positive and false negative (ISO 11843) with partial least squares. *Chemometrics Intell. Lab. Syst.* **2003**, 69, (1-2), 21-33.
44. Teófilo, R. F.; Martins, J. P. A.; Ferreira, M. M. C. In *Ordered predictors Selection: an intuitive method to find the most relevant variable in multivariate calibration*, 10th International Conference on Chemometrics in Analytical Chemistry, Águas de Lindóia, Brazil, Águas de Lindóia, Brazil, 2006, p ^pp P066.
45. Teófilo, R. F.; Martins, J. P. A.; Ferreira, M. M. C. *OPS Toolbox*, Versão 1.0; INPI - 0000270703255138: Brasil, 2007.
46. Majidi, M. R.; Jouyban, A.; Asadpour-Zeynali, K. Genetic algorithm based potential selection in simultaneous voltammetric determination of isoniazid and hydrazine by using partial least squares (PLS) and artificial neural networks (ANNs). *Electroanalysis* **2005**, 17, (10), 915-918.
47. Plackett, R. L.; Burman, J. P. The design of optimum multifactorial experiments. *Biometrika* **1946**, 33, (4), 305-325.
48. StatSoft. *STATISTICA (data analysis software system)*, Version 6; Statsoft: Tulsa, 2001.
49. Neto, B. B.; Scarminio, I. S.; Bruns, R. E. *Como fazer experimentos: pesquisa e desenvolvimento na ciência e na indústria*. Editora Unicamp: Campinas, 2001; p 401.
50. Kennard, R. W.; Stone, L. A. Computer aided design of experiments. *Technometrics* **1969**, 11, (1), 137-148.

51. Teófilo, R. F.; Ferreira, M. M. C. Chemometrics II: Spreadsheets for experimental design calculations, a tutorial. *Quim. Nova* **2006**, 29, (2), 338-350.
52. Doehlert, D. H. Uniform shell designs. *Appl. Stat.* **1970**, 19, 231-239.
53. Ferreira, S.; dos Santos, W.; Quintella, C. M.; Neto, B. B.; Bosque-Sendra, J. A. Doehlert matrix: a chemometric tool for analytical chemistry - review. *Talanta* **2004**, 63, (4), 1061-1067.
54. Wise, B. M.; Gallagher, N. B.; Bro, R.; Shaver, J. M. *PLS_Toolbox*, Version 3.01; Eigenvector Research, Inc.: Manson, 2003.
55. Faber, N. M. Efficient computation of net analyte signal vector in inverse multivariate calibration models. *Anal. Chem.* **1998**, 70, (23), 5108-5110.
56. Notsu, H.; Tatsuma, T.; Fujishima, A. Tyrosinase-modified boron-doped diamond electrodes for the determination of phenol derivatives. *J. Electroanal. Chem.* **2002**, 523, (1-2), 86-92.
57. Pedrosa, V. D.; Codognoto, L.; Avaca, L. A. Electroanalytical determination of 4-nitrophenol by square wave voltammetry on diamond electrodes. *J. Braz. Chem. Soc.* **2003**, 14, (4), 530-535.
58. Wang, J.; Chen, G.; Chatrathi, M. P.; Fujishima, A.; Tryk, D. A.; Shin, D. Microchip capillary electrophoresis coupled with a boron-doped diamond electrode-based electrochemical detector. *Anal. Chem.* **2003**, 75, (4), 935-939.
59. Saterlay, A. J.; Foord, J. S.; Compton, R. G. An ultrasonically facilitated boron-doped diamond voltammetric sensor for analysis of the priority pollutant 4-chlorophenol. *Electroanalysis* **2001**, 13, (13), 1065-1070.

Capítulo 7

Monitoramento da Degradação Eletroquímica de Fenóis no Eletrodo de Diamante Dopado com Boro empregando Fluorescência e Métodos Multimodos

*“Ambiente limpo não é o que mais se limpa,
mas o que menos se suja”*

Chico Xavier

Introdução

Processos industriais geram resíduos contendo compostos orgânicos recalcitrantes tais como os compostos fenólicos de alta toxicidade que precisam ser recuperados ou eliminados antes do descarte ou reuso dos resíduos [1,2]. A incineração de compostos fenólicos requer oxidantes muito fortes, tais como o radical hidroxila (OH^\bullet). As duas principais rotas para a geração de OH^\bullet são a irradiação na presença de TiO_2 , que oxida a água formando OH^\bullet [3-6], e a geração do radical na superfície de um eletrodo [2,7-10].

A seleção de um tratamento considerado economicamente ótimo dependerá do controle, confiança, custo-benefício e eficiência no tratamento. Destas tecnologias, a oxidação eletroquímica é considerada uma das mais promissoras [2,8,11]. Muitas investigações tem sido executadas em diversos tipos de materiais, tais como RuO_2 , PbO_2 e SnO_2 , que são empregados como anodos para a degradação de poluentes orgânicos [9,12-14]. Na superfície destes materiais os produtos da oxidação da água, tais como, OH^\bullet , H_2O_2 , O_2 e O_3 , são gerados e oxidam os poluentes orgânicos. Entretanto, estes anodos se corroem com facilidade e tendem a se desativar com facilidade, diminuindo sua eficiência e vida útil [15].

Recentemente, eletrodos de diamante dopado com boro fabricados pelo processo CVD (Capítulo 2) estão sendo aplicados com sucesso para o tratamento de resíduos aquosos contendo compostos orgânicos não biodegradáveis tais como fenol, dihidroxibenzeno, guaiacóis, clorofenóis, nitrofenóis e misturas de compostos fenólicos, entre outras classes de compostos [2,8,10]. Diversos trabalhos na literatura têm divulgado as excelentes características do BDD para tratamento de resíduos orgânicos em água aplicando oxidação eletroquímica [16,17]. O BDD é um material com uma extraordinária estabilidade anódica e com uma condutividade aceitável [18]. Além disso, o BDD apresenta uma característica de inércia química e, devido a isto, uma excelente resistência à adsorção de compostos orgânicos em sua superfície. Isto sugere que este material é um eletrodo não ativo ideal, no qual a oxidação de compostos orgânicos e a evolução de oxigênio podem ser obtidas via a formação de radicais hidroxilas [18].

Em eletrodos não ativos há uma fraca interação entre os radicais hidroxila e a superfície do eletrodo. A oxidação de compostos orgânicos mediada por radicais hidroxila (equação 7.1) pode resultar em produtos de uma oxidação completa, tais como CO₂.



Nesta equação da reação esquemática, R é a fração do composto orgânico a ser transformado completamente em produtos oxidados e D representa a superfície do diamante. Os valores de *m* e *n* dependem da composição elementar do composto a ser oxidado, R. Esta reação está em competição com o lado da reação da descarga dos radicais hidroxila para O₂ sem nenhuma participação da superfície do anodo (equação 7.2) [18].

Dentre os poluentes que podem ser degradados eletroquimicamente, os compostos fenólicos são provavelmente os mais extensivamente estudados no campo do tratamento de águas contaminadas, devido a sua persistência e alta toxicidade. Além disso, estes compostos são gerados e descartados por um considerável número de indústrias, visto que são frequentemente usados como

reagentes industriais na produção de borracha, tintas, plásticos, fármacos, papel, etc.

Através do processo de degradação, os fenóis se decompõem a ácidos orgânicos de cadeias curtas tais como os ácidos maléico, fumárico, acético, succínico, malônico, glioxálico, oxálico e posteriormente a CO_2 [9,10,19]. Conhecer a cinética individual de degradação dos compostos que se degradam em uma mistura de diferentes moléculas é importante, pois como foi descrito por Canizares et al. (2005) [10], a eficiência do processo depende fortemente da concentração dos poluentes orgânicos e de sua natureza, e não da densidade de corrente. Além disso, devido à diferença de toxicidade entre os compostos, a degradação do mais tóxico é prioritária.

O monitoramento da degradação de compostos orgânicos pode ser realizado usando medidas de parâmetros globais, como conteúdo de carbono orgânico dissolvido (DOC), carbono orgânico total (TOC) e demanda química de oxigênio [20]. Métodos mais seletivos como os espectroscópicos (ultravioleta ou fluorescência) ou eletroquímicos também podem ser empregados, mas normalmente com separação prévia usando técnicas cromatográficas [21,22]. A separação cromatográfica é muitas vezes necessária devido à complexidade do sistema químico, principalmente após o início da degradação, visto que muitos compostos gerados podem interferir nas respostas dos analitos degradados, além de muitas vezes os próprios compostos na mistura interferirem mutuamente na análise empregada. As desvantagens da separação cromatográfica para este tipo de monitoramento estão no grande consumo de tempo, no alto consumo de reagentes e no relativo alto custo da análise como um todo. Por outro lado, há métodos quimiométricos disponíveis na literatura para resolver e identificar espécies presentes em misturas que apresentam sobreposição de sinais analíticos [23]. Entre estes métodos destacam-se as calibrações multivariadas de primeira e segunda ordem, tais como PLS e *N*-PLS, respectivamente e aqueles que aplicam a resolução de curvas multivariadas que se baseiam em métodos multimodos (Capítulo 1), tais como a análise dos fatores paralelos (PARAFAC) [24-29]. Para aplicar os métodos multimodos é necessário obter uma matriz de respostas para cada amostra e não mais um vetor. Espectrofluorímetros e cromatógrafos a líquido de alto desempenho

com detector de arranjo de diodos (HPLC-DAD) são fontes para este tipo de dados. Aplicando métodos multimodos sobre um conjunto de amostras é possível resolver misturas de compostos em seus perfis originais, além de conseguir o perfil de concentração destes analitos nas amostras [24,30]. Esta metodologia, porém, é ainda pouco empregada em monitoramento de reações, mas é uma alternativa para resolver sinais com baixa seletividade.

Neste estudo, duas misturas de compostos fenólicos foram degradadas empregando o eletrodo de BDD. O foco deste trabalho é monitorar individualmente a cinética de degradação destes compostos na mistura. Normalmente, este tipo de monitoramento é realizado por HPLC empregado diferentes tipos de detectores. Porém, uma alternativa é o monitoramento da degradação dos compostos fenólicos usando métodos espectroscópicos com tratamento dos dados usando métodos quimiométricos de resolução de curvas ou de calibração multivariada. Este tipo de metodologia (espectroscopia e quimiometria) apresenta menor custo, tem potencialidade para monitoramento *on-line* e pode fornecer possibilidades de realizar correlações com métodos não específicos, como o TOC, utilizando informações contidas nos espectros.

Assim, neste trabalho, a eletrólise de compostos fenólicos utilizando o eletrodo de BDD foi monitorada globalmente com espectroscopia de fluorescência e TOC e o monitoramento individual foi realizado com dados de fluorescência tratados com métodos PARAFAC, PLS do tensor matriciado (*M*-PLS) e PLS multilinear (*N*-PLS) [24,27,31].

Experimental

Reator para eletrólise

A eletrólise da solução foi realizada em uma célula de nylon cilíndrica e não dividida. O ânodo utilizado foi o BDD crescido conforme descrito previamente no Capítulo 3, usando como substrato para este estudo, um *wafer* quadrado de silício tipo p de 1,15 cm² de área e 1 mm de espessura ou dois deste *wafer* (Figura 7.1). O filme de BDD foi crescido sobre apenas um dos lados do *wafer*. Para cada ânodo, uma placa de aço inox (AISI 304) com 1,5 cm² de área e 1,5 mm de espessura foi

utilizada como cátodo. A área não ativa do ânodo e uma das áreas do aço inox foram isoladas utilizando resina epóxi.

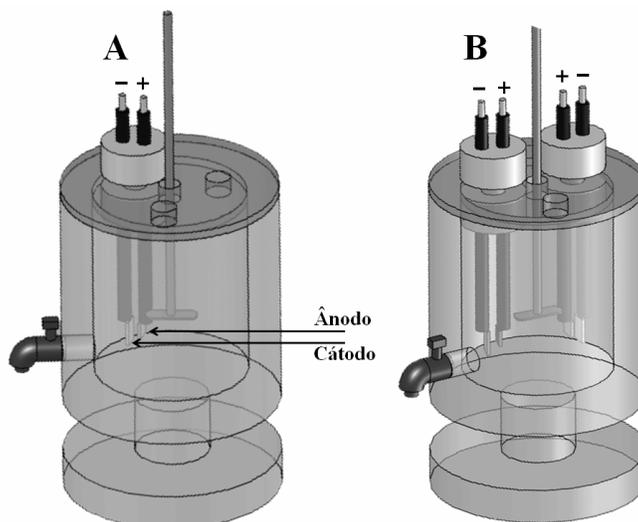


Figura 7.1. Esquema do reator, na escala utilizada. (A) reator com dois eletrodos, (B) reator com quatro eletrodos.

Uma fonte dc de corrente-voltagem foi empregada para fornecer a energia elétrica durante a eletrólise da solução. O contato elétrico do ânodo e do cátodo à fonte foi realizado utilizando um fio de cobre. As áreas ativas do ânodo e cátodo foram colocadas em paralelo a uma distância de 5 mm. O conjunto ânodo-cátodo foi colocado em paralelo à parede da célula a uma distância de 1,5 cm e a 3 cm do centro do cilindro. No centro do cilindro foi colocado um bastão de aço inox com hélices acoplado a um motor agitador. Nesta disposição a solução agitada passa através do ânodo e do cátodo continuamente. A Figura 7.1, mostra o esquema do reator utilizado na escala em que for construído. A densidade de corrente aplicada foi de 30 mA cm^{-2} .

Reagentes

Os reagentes, hidroquinona, guaiacol, p-cresol, m-cresol, catecol e fenol foram adquiridos da SigmaTM, o Na_2SO_4 foi adquirido da SynthTM e o H_2SO_4 foi adquirido da MerckTM. Todos os reagentes são de grau analítico sem purificação adicional. Todas as soluções foram preparadas com água ultra pura de um sistema de purificação Milli QTM ($>18 \text{ M}\Omega \text{ cm}$). As soluções estoque de cada analito foram

preparadas com uma concentração de $0,5 \text{ mmol L}^{-1}$ em água e as misturas foram preparadas do estoque por diluições em tampão, como necessário.

Soluções

Uma solução 5 g L^{-1} de Na_2SO_4 a um pH 2 (obtido com adição de H_2SO_4) foi utilizada como eletrólito suporte. Os compostos fenólicos degradados foram: hidroquinona (HYDR); guaiacol (GUAI); p-cresol (pCRE); m-cresol (mCRE); catecol (CATE) e fenol (PHEN).

Duas misturas foram elaboradas, *i.e.* (1) PHEN, CATE, HYDR, pCRE, mCRE e GUAI que foi denominada *MS* e (2) HYDR, PHEN e CATE que foi denominada *MT*. Os níveis estudados de concentrações iniciais destes compostos foram: nível A(*MS*) - $80 \text{ } \mu\text{mol L}^{-1}$ para todos os compostos; nível B(*MS*) - $1,75 \text{ } \mu\text{mol L}^{-1}$ para HYDR, GUAI e PHEN; $0,63 \text{ } \mu\text{mol L}^{-1}$ para pCRE e mCRE e $3,5 \text{ } \mu\text{mol L}^{-1}$ para CATE e nível C(*MT*) - $5,0 \text{ } \mu\text{mol L}^{-1}$ para ambos HYDR e PHEN e $10 \text{ } \mu\text{mol L}^{-1}$ para CATE. O volume de solução foi de 500 mL.

Monitoramento global

O tempo máximo de monitoramento da degradação foi de 45 h para o nível A e de 25 h para o nível B e de 25 h para o nível C. Alíquotas de 10 mL do eletrólito foram coletadas durante a degradação em intervalos de tempo definidos. Para cada alíquota, a degradação foi monitorada pela concentração de TOC e espectroscopia de fluorescência.

Para a determinação do TOC, um analisador de carbono orgânico total Shimadzu modelo TOC-5000A foi empregado. As medidas foram realizadas em duplicata tanto para a análise do carbono inorgânico (*IC*), quanto para a análise do carbono total (*TC*). A diferença ($\overline{TC} - \overline{IC}$) das médias foi utilizada.

Os espectros de fluorescência foram adquiridos em um espectrofluorímetro Perkin Elmer modelo LS55 e em uma cubeta de quartzo de 1 cm de caminho ótico. As aberturas das fendas foram de 5 nm. Para o monitoramento global, apenas as intensidades coletadas nos comprimentos de onda (λ) de 275 nm para a excitação e 320 nm para a emissão foram utilizadas.

Monitoramento individual: conjunto de dados para os modelos

Variáveis Dependentes

Um planejamento fatorial parcial para mistura com cinco níveis foi elaborado para obter as variáveis dependentes (concentrações), o que garantiu a ortogonalidade destas variáveis. As misturas foram elaboradas a partir de padrões conforme descrito pelo planejamento. A Tabela 7.1 mostra os níveis investigados. Um total de 30 misturas foi elaborado para o conjunto *MS* e de 38 misturas para o conjunto *MT*.

Tabela 7.1. Níveis usados para a elaboração do planejamento fatorial parcial para as duas misturas *MS* e *MT*

Níveis	Concentrações / $\mu\text{mol L}^{-1}$ - <i>MS</i>					
	MCRE	GUAI	PCRE	HYDR	PHEN	CATE
-2	0,1	0,2	0,1	0,2	0,3	0,5
-1	0,3	0,9	0,3	1,0	1,0	2,0
0	0,5	1,5	0,5	1,7	2,0	3,5
1	0,7	2,2	0,7	2,5	3,0	5,0
2	1,0	3,0	1,0	3,2	3,5	6,0
Níveis	Concentrações / $\mu\text{mol L}^{-1}$ - <i>MT</i>					
	-2			0,5	0,5	1,0
	-1			1,0	1,0	2,0
	0			2,5	3,5	5,0
	1			3,0	4,0	6,0
	2			5,0	5,0	8,0

Variáveis independentes

Para o monitoramento individual os espectros foram obtidos no mesmo instrumento descrito anteriormente e nas mesmas configurações, exceto para a coleta dos espectros. A excitação foi realizada na faixa de 245 a 300 nm com incrementos de 2 nm. Para cada comprimento de onda de excitação foi obtido um espectro de emissão na faixa de 270 a 380 nm a uma velocidade de varredura de 120 nm s^{-1} . Os espectros obtidos desta maneira são conhecidos como espectros excitação-emissão (EEM).

O espalhamento Raman foi removido por subtração do espectro branco e o espalhamento Rayleigh foi substituído por dados faltantes e posteriormente a região faltante gerada foi preenchida usando a interpolação dos dados [32].

Para os modelos PARAFAC nenhum tratamento dos dados foi realizado e nenhuma restrição foi empregada. A escolha do número de componentes foi realizada por avaliação do modelo para cada componente após diferentes inicializações (Capítulo 1). A detecção de amostras anômalas foi realizada empregando o método *Jack-knife-leave-one-out* (Capítulo 1) e validação dos modelos foi realizada empregando o método *split-half*.

Para a construção dos modelos de calibração multivariada, tais como o *M*-PLS e o *N*-PLS, as colunas da matriz de dados foram centradas na média. O número de fatores para o modelo foi determinado por validação cruzada aplicando o método '*leave-n-out*' e se baseou no cálculo do RMSECV, conforme equação 1.45. As variáveis independentes foram selecionadas usando o método de seleção das variáveis ordenadas para segunda ordem, o nOPSel (Capítulo 1) e as variáveis com melhor capacidade de previsão na validação cruzada *i.e.*, com menor valor de RMSECV, foram usadas.

A habilidade de previsão do modelo foi avaliada pelo erro médio quadrático de previsão (RMSEP) e pelos coeficientes de correlação (r_p) conforme equações 1.47 e 1.48, respectivamente.

As amostras de calibração e previsão foram separadas empregando o algoritmo de Kennard e Stone (1969) [33], após a seleção de variáveis.

Análise dos dados

As análises de dados foram executada usando funções *.m* para o software Matlab™ 7 (The MathWorks, Co., Natick, MA, USA). Para a remoção do espalhamento Rayleigh e para a interpolação dos dados foram empregados os algoritmos disponibilizados por Bro no *web site* www.models.kvl.dk/source/. e divulgado no trabalho de Bahran et al. (2006) [32]. Os algoritmos para *Jack-Knife* [34] e *Split-half* [35] foram obtidos do *web site* www.models.kvl.dk/source/. Os modelos PARAFAC foram calculados usando os algoritmos disponíveis no pacote *N-Way Toolbox 3.0* [36] disponíveis no *web site* www.models.kvl.dk/source/. Os algoritmos para *N*-PLS, PLS e OPS e nOPS foram desenvolvidos no Laboratório de Quimiometria Teórica e Aplicada (LQTA).

Resultados e Discussão

Caracterizações usando microscopia eletrônica de varredura e espectroscopia Raman, conforme descrito na metodologia do Capítulo 3, foram realizadas aqui para um dos eletrodos crescidos. A Figura 7.2 mostra os resultados destas caracterizações e novamente observa-se grãos bem definidos e o espectro Raman indica uma baixa quantidade de fase gráfitica sobre a superfície do eletrodo.

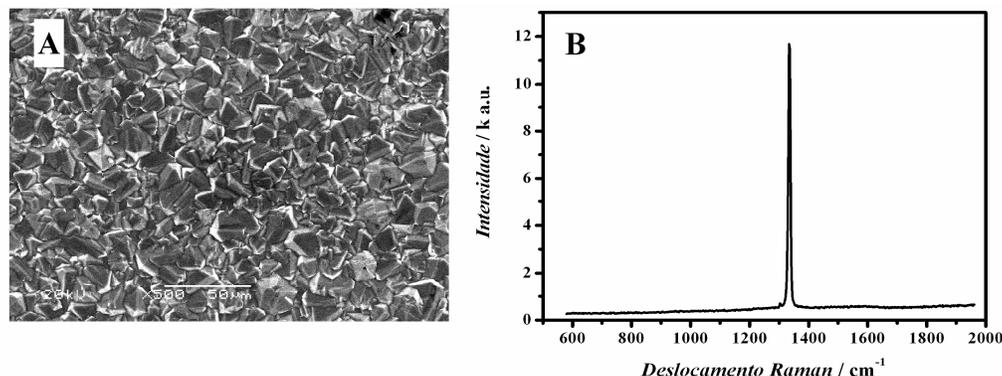


Figura 7.2. Imagem de microscopia eletrônica de varredura (A) e espectro Raman (B) para um dos eletrodos usados.

Monitoramento global

A Figura 7.3 mostra um perfil típico de remoção de TOC com a carga elétrica específica passada ($A \cdot h \cdot L^{-1}$). A mesma figura apresenta o comportamento da remoção de intensidades de fluorescência com a carga elétrica específica. A intensidade de fluorescência foi denominada aqui como sinal de aromaticidade, visto que na região de emissão investigada, além da maioria dos monofenóis, o anel aromático também emite e os produtos não aromáticos conhecidos da eletrodegradação dos fenóis não emitem nesta região, tais como o CO_2 e ácidos de cadeias curtas. Assim, a resposta é extremamente seletiva para a aromaticidade.

Nota-se pelos perfis que a remoção de aromaticidade é significativamente maior que a remoção de TOC. Estes resultados indicam, como esperado, que a ruptura do anel aromático ocorre rapidamente durante a degradação enquanto a remoção de TOC ocorre mais lentamente, visto que após a ruptura do anel várias etapas são necessárias até a mineralização completa dos compostos orgânicos gerados. Observa-se pelas Figuras 7.3A e B, que para maiores concentrações de

compostos fenólicos (Nível A), uma maior necessidade de carga é necessária. Estes resultados estão em concordância com aqueles obtidos por [10], que mostrou a grande influência da concentração dos poluentes orgânicos sobre a eficiência do processo. Por outro lado, para concentrações baixas (Nível B) a degradação via BDD é extremamente eficiente, pois nas concentrações investigadas outros métodos, tais como aqueles que empregam bactérias ou reagente de Fenton, não apresentam a sensibilidade obtida com este material. Esta vantagem do BDD o torna uma excelente alternativa para realizar a desinfecção de águas destinadas para o consumo humano [37], *i.e.* águas doces das classes especiais e classes 1, 2 e 3, conforme resolução nº 357 (17/03/05) do CONAMA [38].

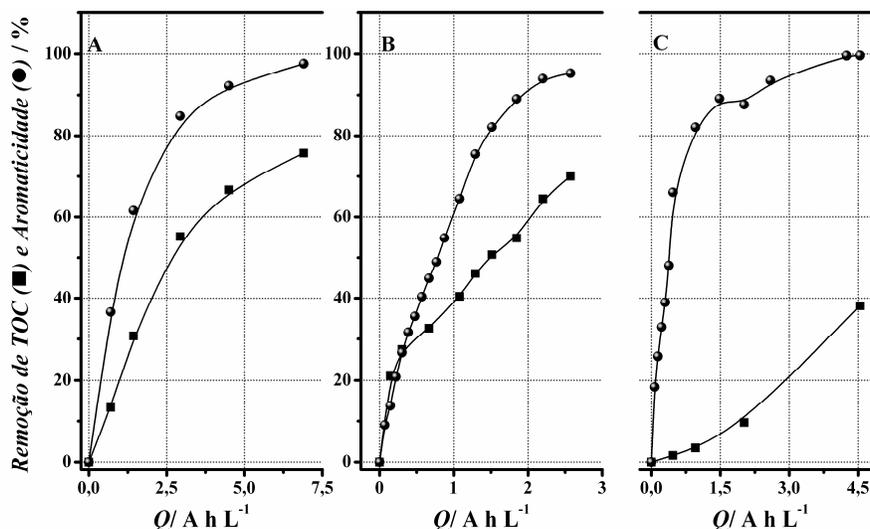


Figura 7.3. Remoções do sinal de emissão coletado a 320 nm (●) e do valor de TOC (■) obtidos durante a degradação para os conjuntos *MS* no nível A e B e para o conjunto *MT* no nível C. As medidas em cada ponto foram realizadas na mesma solução.

A Figura 7.3C apresenta a remoção da mistura *MT* em concentrações relativamente baixas. Este resultado (Figura 7.3C) mostra que a remoção da aromaticidade foi realizada de maneira extremamente eficiente, entretanto, a remoção de TOC ocorreu de modo contrário se for comparado com os resultados mostrados nos níveis A e B. Nasr et al. (2005) [39] realizando a degradação individual de hidroquinona, resorcinol e catecol no BDD concluíram que há formação de um grande número de intermediários principalmente na oxidação do

catecol, porém, os autores detectaram uma baixa geração de CO₂ no início da degradação. Neste mesmo trabalho, os autores detectaram uma baixa concentração de intermediários aromáticos mesmo quando uma baixa carga foi passada, o que está em concordância com os resultados obtidos aqui. Além disso, há indícios de oxidação direta neste experimento da mistura *MT*, pois a geração de radicais OH[•] na superfície do eletrodo certamente foi menor em relação aos outros experimentos, uma vez que o eletrodo empregado neste experimento já havia passado por extensivos ciclos de degradação por longos períodos de tempo. Outra importante conclusão é que os compostos da mistura *MT* apresentam fácil rompimento do anel em relação aos outros utilizados nos experimentos anteriores como o GUAI, MCRE e PCRE, que são extremamente recalcitrantes [4,40].

O custo energético para ambas as degradações foi calculado conforme equação 7.3.

$$CE = \frac{Eit_f}{V_{ol}} \quad 7.3$$

em que *E* é a média do potencial (em V), *i* é a corrente aplicada (em Ampére), *t_f* é o tempo necessário para a mineralização total (em h), *V_{ol}* é o volume inicial da solução (em L) e *CE* é o custo energético dado em kWh m⁻³.

Tabela 7.2. Custo energético para remoção de TOC e aromaticidade

Níveis	Área/cm ²	Tempo/h	<i>i</i> /A	<i>E</i> /V	Remoção TOC/%	Remoção Aromat.1/%	CE/kW h m ⁻³
A	2,30	45	0,069	7	75,71	97,66	43,47
B	1,15	25	0,035	6	70,02	95,36	10,50
C	2,30	25	0,069	7	38,10	99,61	24,15

Os resultados da Tabela 7.2 mostram que economicamente o uso do BDD para degradação de compostos fenólicos é viável devido ao baixo custo energético encontrado em ambos os experimentos. Estes baixos custos estão de acordo com resultados divulgados na literatura para degradação de fenóis [4,41]. Observe que para o nível A, o custo foi de 43,57 kW h m⁻³ para remover 75,71 % de TOC e 97,66 % de aromaticidade. Para o nível B, o custo foi de 10,50 kW h m⁻³ para remover 70,02 % de TOC e 95,36 % de aromaticidade. Para o nível C, o custo foi

24,15 kW h m⁻³ para remover 38,10 % de TOC e 99,61 % de aromaticidade. O resultado do nível C confirma a hipótese de perda de atividade do ânodo devido ao uso extensivo.

Como estes experimentos foram realizados em concentrações diferentes de compostos fenólicos, uma comparação entre eles não pode ser realizada.

O monitoramento global mostrou que o BDD crescido do etanol apresenta alta viabilidade econômica para ser empregado na eletrodegradação dos compostos fenólicos tanto para remoção de TOC quanto de aromaticidade. O monitoramento via fluorescência pode fornecer informações importantes em relação à toxicidade do rejeito, pois os compostos gerados são na sua maioria menos tóxicos que os compostos originais. Portanto, a eletrólise pode ser finalizada antes da remoção de altos níveis de TOC com garantia de uma solução com a toxicidade praticamente eliminada e evitando gastos excessivos de energia.

Monitoramento individual

Dois planejamentos de misturas foram elaborados para realizar a construção de modelos de calibração multivariada. Os dados coletados para cada mistura foram espectros de EEM, que torna possível explorar as metodologias quimiométricas para dados de ordem superior como PARAFAC, *N*-PLS e PLS do tensor matriciado (*M*-PLS). O objetivo destes modelos é tornar possível realizar previsões da degradação individual de cada composto na mistura. Neste sentido, espectros de EEM foram obtidos para todas as coletas durante a degradação dos compostos fenólicos para ambas as misturas. O modelo PARAFAC apresenta sobre as outras duas metodologias a vantagem de segunda ordem, que infelizmente não será explorada neste estudo. O *N*-PLS e o *M*-PLS são métodos que tendem a obter os mesmos resultados, porém, o *N*-PLS explora a estrutura multimodos dos dados e o vetor de regressão se torna mais simples de ser explorado para interpretação [27,31]. O método nOPS proposto nesta tese (Capítulo 1) para selecionar variáveis em modelos *N*-PLS explora a característica informativa do vetor de regressão e portanto a possibilidade de interpretação dos resultados a partir deste vetor.

A Figura 7.4 mostra as correções realizadas nos espectros EEM antes do tratamento dos dados. Observe que com o tratamento realizado, todos os dados

faltantes são removidos através da interpolação, que praticamente reconstrói o espectro. Este tratamento contribui para uma análise mais estável e precisa dos dados [32,42,43].

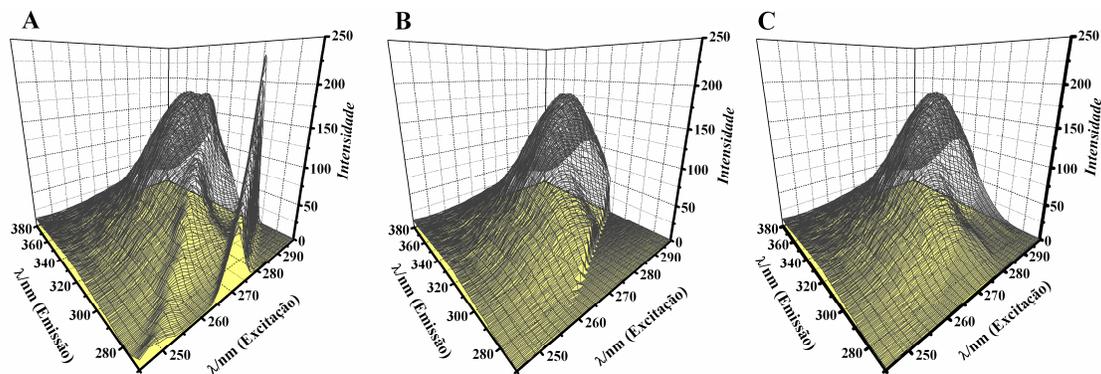


Figura 7.4. Comportamento típico dos espectros EEM durante as correções. Espectros originais (A), espectro com remoção dos espalhamentos Raman e Rayleigh (B) e espectro com interpolação sobre os dados faltantes (C).

Apesar de conhecer a mistura *MS* e saber que seis compostos estão presentes, a análise diagnóstica empregando vários ajustes com diferentes inicializações foi realizada para verificar se o número de componentes realmente é seis. A Figura 7.5 mostra gráficos obtidos para esta análise.

Como em modelos bilineares, as características de um modelo multimodo podem ser julgadas pelos resíduos: se uma variação sistemática é deixada nos resíduos, é um indicativo que mais componentes podem ser extraídas. Para analisar a Figura 7.5A em relação ao modelo PARAFAC, considere que um inesperado aumento no número de mínimos locais é frequentemente uma boa indicação que o ruído está sendo modelado e variações para um mesmo número de componentes podem ser observadas *i.e.* as repetições não são idênticas. Observe na Figura 7.5A (Figura interna) que na quarta componente há variações bruscas nos resíduos, o que indica que o modelo pode se ajustar com três componentes. A Figura 7.5B com os valores de CORCODIA (Capítulo 1) confirma o resultado anterior para três componentes. A Figura 7.5C mostra o número de iterações para cada parte e em cada componente. Normalmente, se houver um inesperado aumento no número de iterações, isto pode indicar que muitas componentes estão sendo usadas, mas também pode ser devido a um difícil problema a ser ajustado. Este gráfico deixa

dúvidas entre o uso de três componentes e não é muito informativo, mas com as informações prévias pode-se dizer que este modelo se ajusta com três componentes.

Entretanto, como dito anteriormente, eram esperados seis componentes, pois seis compostos estão na mistura! Estes resultados indicam que o modelo PARAFAC está tendo dificuldades em se ajustar e, possivelmente, combinações dos *loadings* foram realizadas e por isso, as análises diagnósticas empregadas indicam três componentes erroneamente.

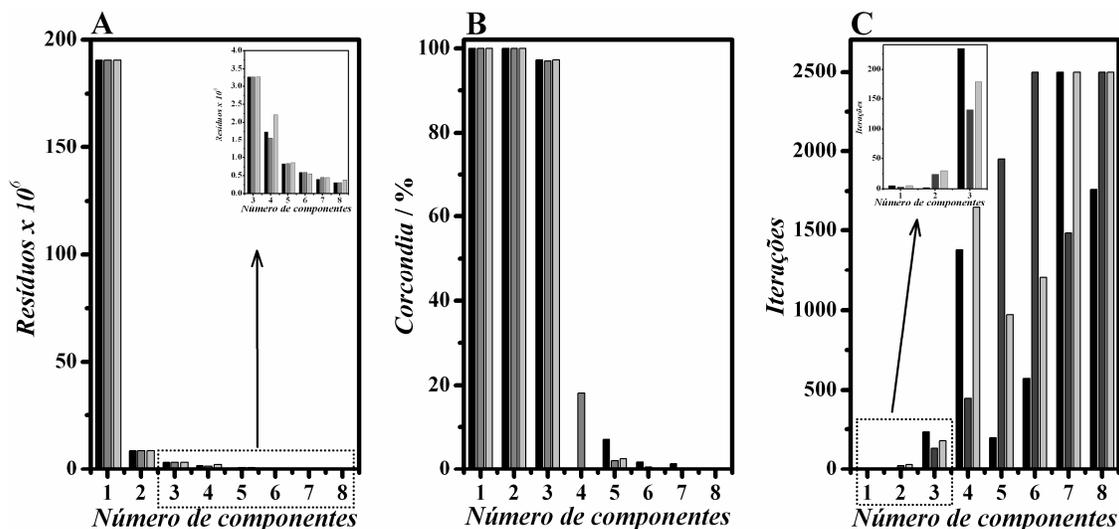


Figura 7.5. Resultados da análise com três diferentes inicializações (SVD, DTLD, aleatória) e oito componentes.

Apesar de as análises indicarem para três componentes, o que é quimicamente indesejável, uma observação no comportamento dos *loadings* para seis componentes pode indicar o motivo pelo qual o modelo PARAFAC não conseguiu se ajustar com o número quimicamente desejável de componentes.

A Figura 7.6 mostra o gráfico RIP (Capítulo 2) para os modos (excitação e emissão). Este gráfico indica que os *loadings* para cada amostra não apresentam um comportamento com grandes diferenças em relação ao modelo completo (todas as amostras). Este resultado é confirmado pela Figura 7.7, que mostra os *loadings* ajustados e os espectros puros experimentais de cada composto. Nesta figura, apenas os *loadings* para o catecol e fenol tiveram um comportamento significativamente diferente dos espectros puros.

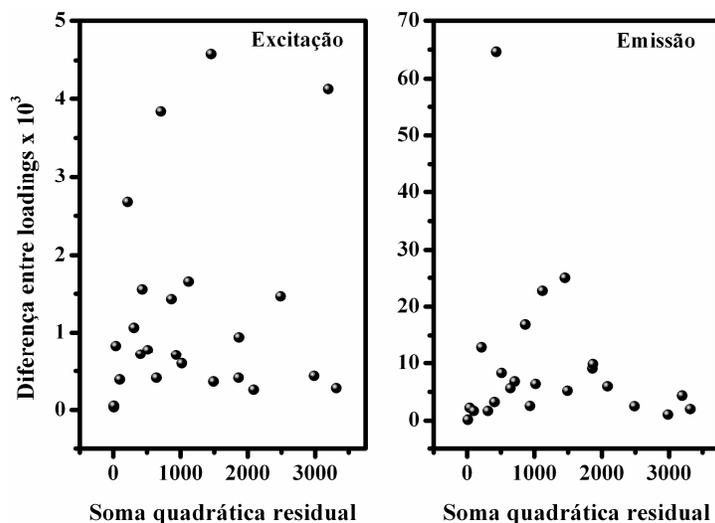


Figura 7.6. Gráficos RIP para os modos excitação e emissão. Os pontos nos gráficos representam os comportamentos das amostras e cada modo.

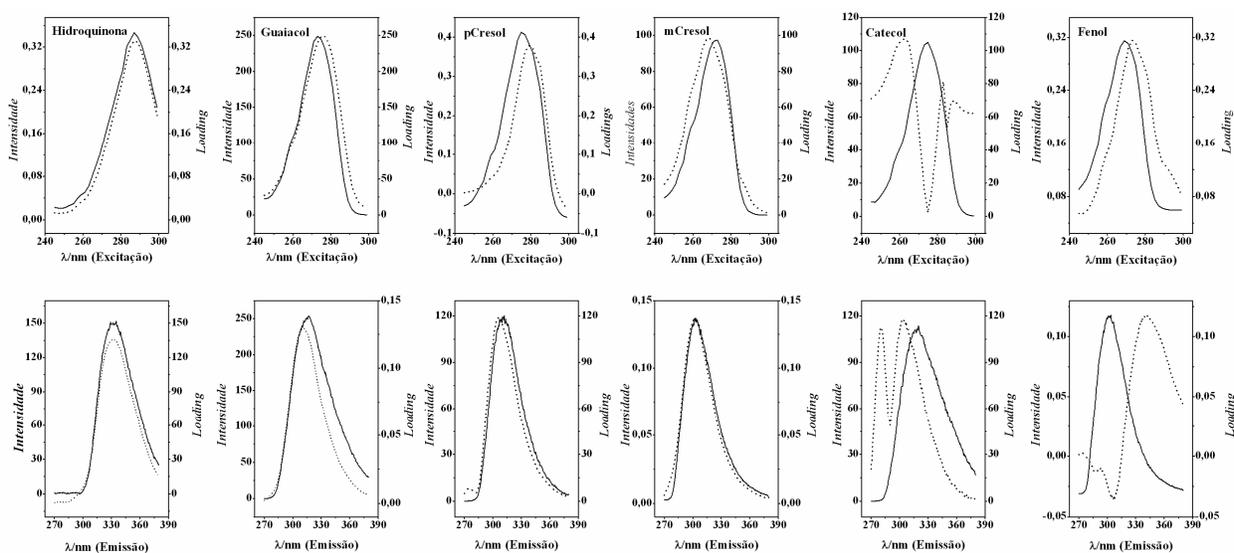


Figura 7.7. Espectros puros (—) de excitação (acima) e emissão (abaixo) e respectivos loadings (····) obtidos pelo modelo PARAFAC.

Entretanto, o gráfico IMP (Capítulo 1) apresentado na Figura 7.8, indica uma expressiva diferença nos escores. Praticamente todas as previsões foram realizadas erroneamente. Estes resultados confirmam a dificuldade em se ajustar o modelo PARAFAC com 6 componentes e aponta as amostras como o modo problemático do sistema químico. Neste sentido, o gráfico IMP, apesar de ser uma ferramenta apresentada para detecção de amostras anômalas, pôde aqui ser usado para detectar

não trilinearidade nos dados o que realmente torna o modelo PARAFAC não apto para ajustar o comportamento do sistema.

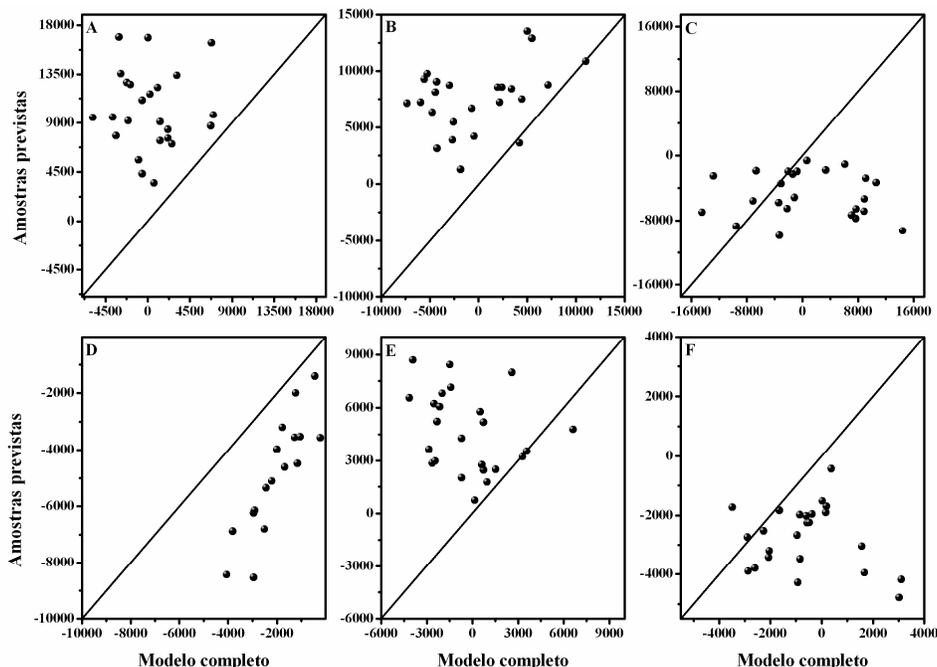


Figura 7.8. Gráfico IMP para HYDR (A), GUAI (B), PCRE (C), MCRE (D), CATE (E), PHEN (F).

A indicação do modo das amostras como problemático no modelo PARAFAC sugere que um possível fenômeno não desejado aconteceu durante a obtenção dos espectros EEM. Um processo foto-físico conhecido para este sistema é a supressão de fluorescência por transferência eletrônica ou por transferência de próton, dependendo da acidez, potencial de oxidação e polaridade do solvente [44-46]. Biczók et al. (1997) [45] mostraram que este processo foto-físico de supressão ocorre com álcoois e fenóis devido principalmente à ligação de hidrogênio e à polaridade do solvente. Nas condições de análise deste trabalho, o meio aquoso ácido certamente contribuiu para a supressão por transferência eletrônica. Três fatores, entretanto, contribuíram para aumentar a supressão: (1) quantidade de espécies fenólicas na mistura, (2) característica estrutural das espécies fenólicas e (3) concentração dos fenóis na mistura. Além destes fatores que influenciam a supressão, soma-se ainda a alta sobreposição dos sinais, que dificulta

significativamente o ajuste do modelo. Porém, para a confirmação de todas estas hipóteses, estudos posteriores são necessários.

Na literatura, del Olmo et al. (1996) [47] mostrou a possibilidade de analisar fenol, *o*-cresol, *m*-cresol e *p*-cresol em uma mistura via espectros de excitação. Neste trabalho, apesar de os autores não comentarem o possível efeito da supressão, observa-se pelos resultados um alto número de componentes necessários para ajustar o modelo PLS (6 para fenol e *m*-cresol e 7 para *o*-cresol e *p*-cresol). Os resultados destes autores sugerem para uma possível não linearidade do modelo e a necessidade de muitas componentes para realizar o ajuste. Esta não linearidade ocorreu certamente devido ao efeito de supressão. Outros dois trabalhos publicados por Bosco et al. (2006 e 2007) [30,48] mostram a análise de quatro fenóis *i.e.* fenol, hidroquinona, catecol e resorcinol. Em um dos trabalhos [30] os autores aplicaram PARAFAC e obtiveram bons resultados no ajuste dos *loadings*, entretanto, os resultados dos escores ficaram aquém para a realização de previsões quantitativas. No segundo trabalho [48] os autores aplicaram MCR-ALS e obtiveram resultados ainda piores que os anteriores para a realização de previsões quantitativas. Nestes dois trabalhos os autores realizaram degradação dos fenóis usando fotocatalise com TiO₂ e aplicaram os modelos para quantificações de dois dos fenóis durante a degradação. Em ambos os trabalhos destes autores os altos erros obtidos são novamente apontados como extensão da supressão de fluorescência.

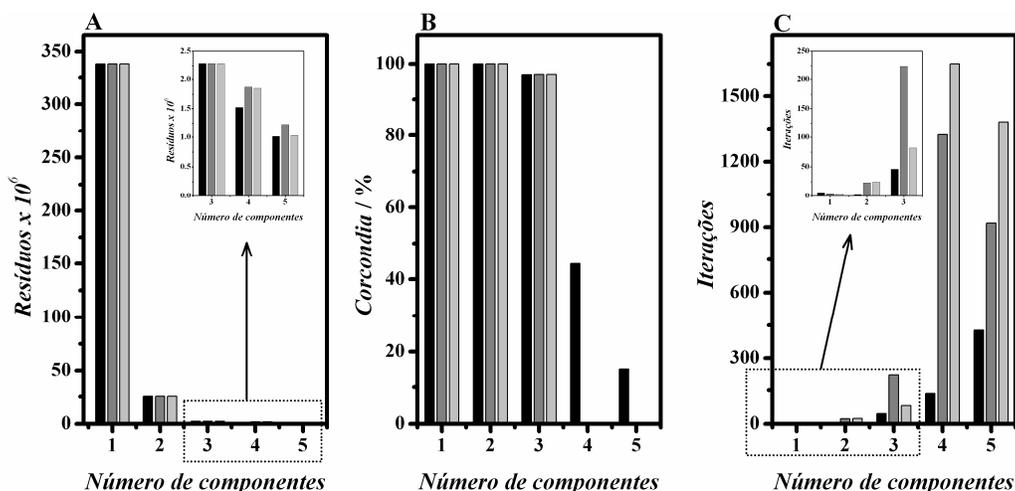


Figura 7.9. Análise diagnóstica para três diferentes inicializações (SVD, DTL, aleatória) e 5 componentes: (A) gráfico dos resíduos, (B) percentagem da análise Corcondia e (C) número de iterações

Aqui, nenhum modelo foi construído para a quantificação individual dos fenóis na mistura *MS* devido à complexidade do sistema estudado. Entretanto, para a mistura *MT* um sistema menos complexo foi obtido e modelos foram construídos para a quantificação individual dos fenóis nesta mistura.

A Figura 7.9 mostra os resultados da análise diagnóstica para três diferentes inicializações e cinco componentes. Os resultados sugerem que três componentes ajustam bem o modelo PARAFAC. Isto indica que a mistura realizada com menos espécies e as características das espécies apresenta menor probabilidade de supressão. Esta mistura mostra ainda, menor sobreposição que a mistura anterior, o que certamente aumenta a seletividade.

A Figura 7.10 mostra os gráficos RIP para os modos excitação e emissão da mistura *MT*.

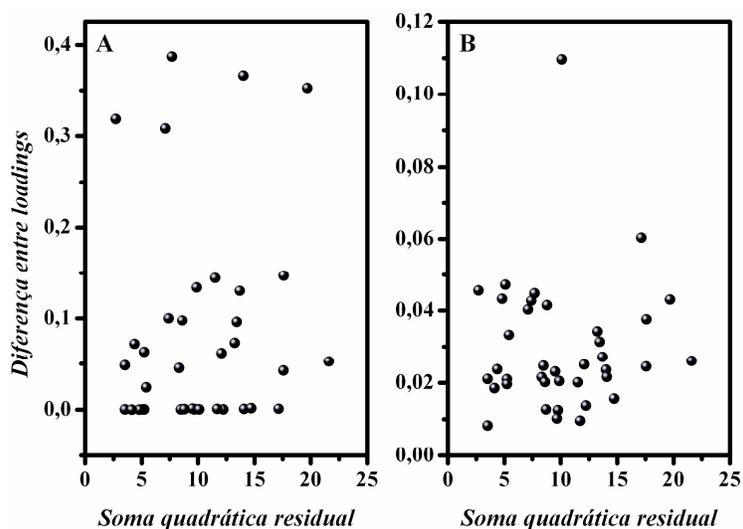


Figura 7.10. Gráficos RIP para o modo da excitação (A) e emissão (B).

Os resultados sugerem que os *loadings* também nesta mistura não apresentam grande diferenças em relação aos *loadings* da mistura com todas as amostras. Note ainda que os valores das diferenças entre os *loadings* vs. a soma quadrática residual são significativamente menores que aqueles observados na Figura 7.6.

A semelhança entre os *loadings* são confirmadas na Figura 7.11 que usa as médias dos *loadings* obtidos em 38 segmentos na análise *Jack-knife* e também na Figura 7.12, que mostra o resultado da análise *split-half* com 8 divisões. Em ambas as figuras os *loadings* foram obtidos de diferentes segmentos dos dados, o que

ratifica a unicidade da solução. Observe que um excelente ajuste foi obtido em relação aos espectros experimentais.

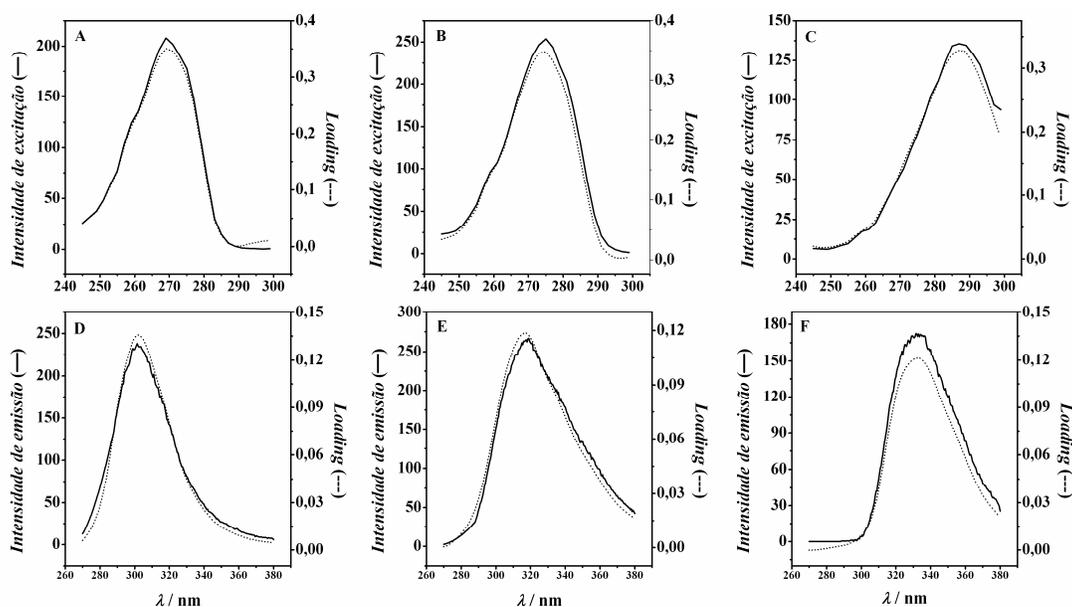


Figura 7.11. Espectros puros (—) para HYDR (A,D), PHEN (B,E) e CATE (C,F) e respectivas médias das estimativas dos *loadings* (····) para 38 segmentos *Jack-knife*.

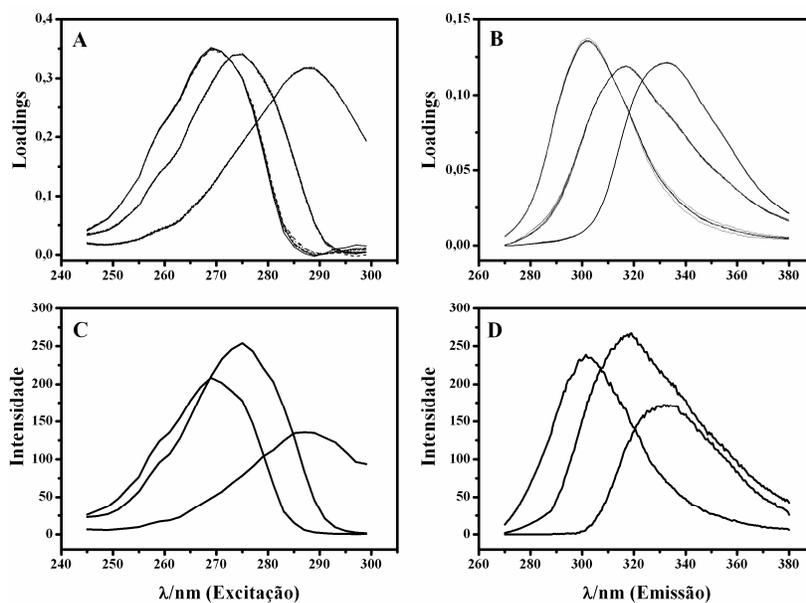


Figura 7.12. Análise *Split-half* para três componentes: *Loadings* dos modos excitação (A) e emissão (B) e espectros puros experimentais da excitação (C) e emissão (D). Os comprimentos de onda máximos de excitação e emissão são respectivamente em nm de: PHEN (270/300); CATE(275/320); HYDR(287/331).

O modo das amostras, que realmente garante a possibilidade de modelar os analitos da mistura, apresenta um bom indicativo da possibilidade de aplicar o modelo para realização de previsões com precisão (Figura 7.13).

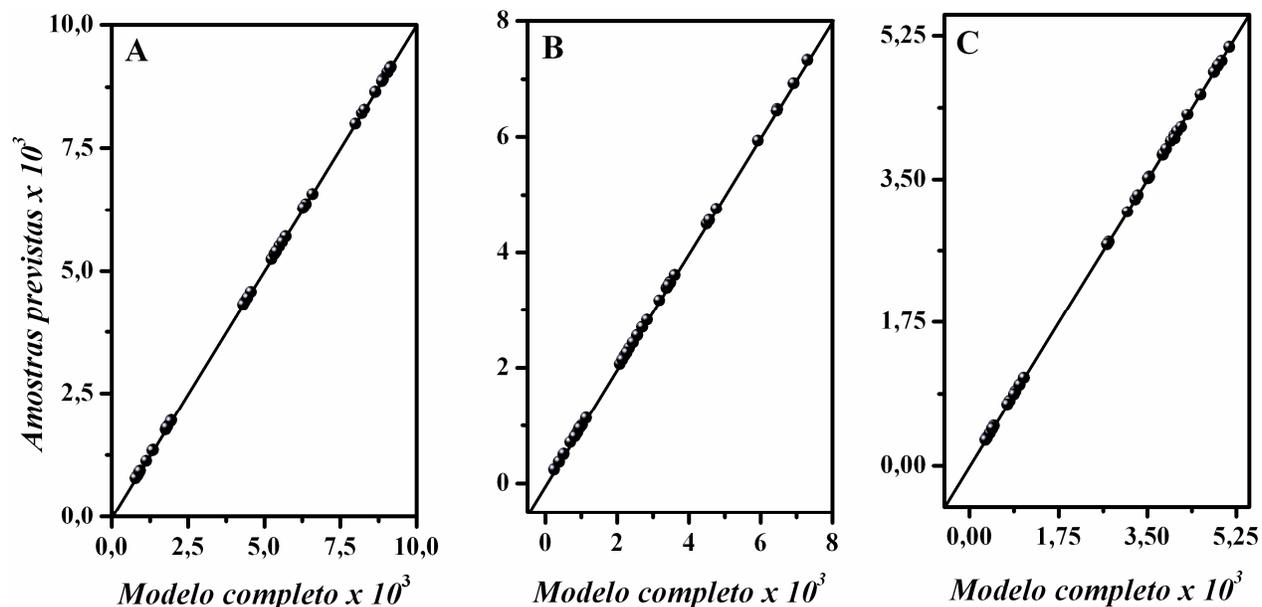


Figura 7.13. Gráficos IMP para HYDR (A), PHEN (B) e CATE (C).

Todos estes resultados orientam para a utilização do modelo PARAFAC para realização de previsões durante a degradação. Entretanto, modelos *M*-PLS e *N*-PLS também foram construídos e comparações foram realizadas para observar quais destes modelos apresentam maior capacidade de previsão *i.e.* quais ajustam melhor aos dados. Para isto, modelos PARAFAC foram previamente construídos com o conjunto de degradação para observar a possibilidade da existência de algum intermediário fluorescente na mistura, diferente dos analitos. A Figura 7.14 mostra os resultados da análise diagnóstica que confirma três componentes e, portanto a não existência de interferentes não modelados.

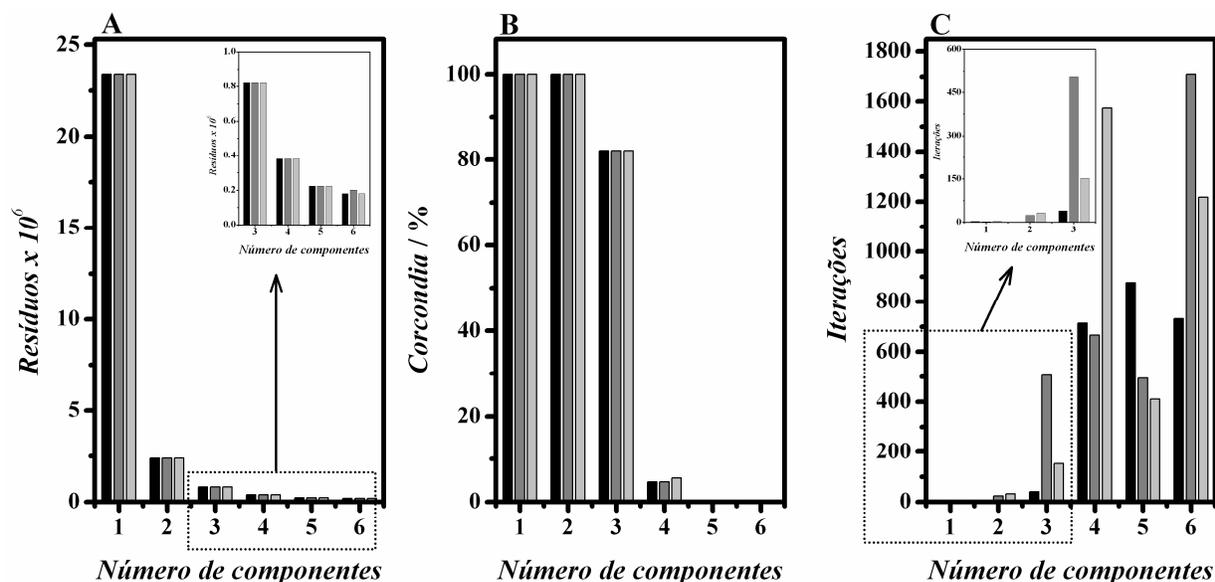


Figura 7.14. Análise diagnóstica do conjunto de degradação para três diferentes inicializações e seis componentes: (A) gráfico dos resíduos, (B) percentagem da análise Corcondia e (C) número de iterações

Tabela 7.3. Diferentes modelos multimodos para realizar a previsão das concentrações durante a degradação individual.

	PARAFAC		M-PLS		N-PLS	
			Completo	OPS	Completo	nOPS
HYDR						
Dimensão	28x221x28		28x6188	28x310	28x221x28	28x70x7
RMSECV	0,048*		0,084	0,088	0,088	0,077
r_{cv}	0,9987*		0,9985	0,9984	0,9984	0,9988
RMSEP	0,048		0,050	0,069	0,048	0,048
r_p	0,9995		0,9996	0,9996	0,9996	0,9997
PHEN						
Dimensão	28x221x28		28x6188	28x2925	28x221x28	28x35x17
RMSECV	0,189*		0,075	0,067	0,083	0,057
r_{cv}	0,9957*		0,999	0,9992	0,9987	0,9994
RMSEP	0,189		0,077	0,077	0,077	0,080
r_p	0,9962		0,9991	0,9992	0,9991	0,9991
CATE						
Dimensão	28x221x28		28x6188	28x2470	28x221x28	28x95x8
RMSECV	0,636*		0,438	0,443	0,423	0,384
r_{cv}	0,9958*		0,9943	0,9941	0,9946	0,9956
RMSEP	0,636		0,583	0,540	0,573	0,474
r_p	0,9922		0,9951	0,9956	0,9954	0,9969

*Valores de erros e correlações relacionados ao modelo de calibração *i.e.* RMSEC e r_c .

Previsões foram realizadas e os resultados apresentados na Tabela 7.3 indicam que os modelos PARAFAC para PHEN e CATE apresentaram piores previsões em relação aos modelos *M*-PLS e *N*-PLS. Uma comparação entre os modelos *M*-PLS e *N*-PLS com os dados completos, indica que tais modelos apresentaram capacidades equivalentes para previsão. Com a seleção de variáveis usando os algoritmos OPS e nOPS, melhorias foram obtidas tanto na previsão quanto na diminuição da dimensão das matrizes e tensores, respectivamente. Além disso, o modelo *N*-PLS/nOPS mostrou ser levemente melhor nas previsões que o modelo *M*-PLS/OPS. O modelo *N*-PLS/nOPS, devido à sua melhor capacidade de previsão e também devido à sua maior parsimônia, foi selecionado como o melhor modelo.

A seleção de variáveis empregando o algoritmo nOPS fornece melhorias significativas em relação à seleção de variáveis usando o OPS, exceto para PHEN, que apresentou uma leve piora. Entretanto, nota-se que as dimensões da seleção nOPS são, na maioria dos casos, significativamente menores que as seleção OPS, exceto para HYDR, com uma diferença de 180 variáveis.

Os métodos para seleção de variáveis mostram ser eficientes para encontrarem modelos aptos a realizarem boas previsões com dimensões significativamente menores. O método nOPS mostrou ser capaz a selecionar variáveis em ambos os modos diretamente no cubo. Este método diminui expressivamente o tempo de cálculo visto que os tensores são comprimidos para menores dimensões com ganhos, na maioria dos casos, na qualidade das previsões.

As variáveis selecionadas pelo algoritmo nOPS são apresentadas na Figura 7.15. Somente as variáveis selecionadas em ambos os modos e que se sobrepuseram, foram consideradas. O método baseado em vetores informativos e neste caso, o vetor de regressão, mostrou ser capaz de indicar regiões onde realmente estão as informações de cada analito de interesse como observado na Figura 7.15 D-F, onde estão os espectros EEM puros de cada analito e as respectivas variáveis selecionadas. Observe que na maioria dos casos as variáveis selecionadas destacam as regiões específicas de excitação–emissão do analito. Isto confirma que o vetor de regressão do *N*-PLS traz informações de ambos os modos.

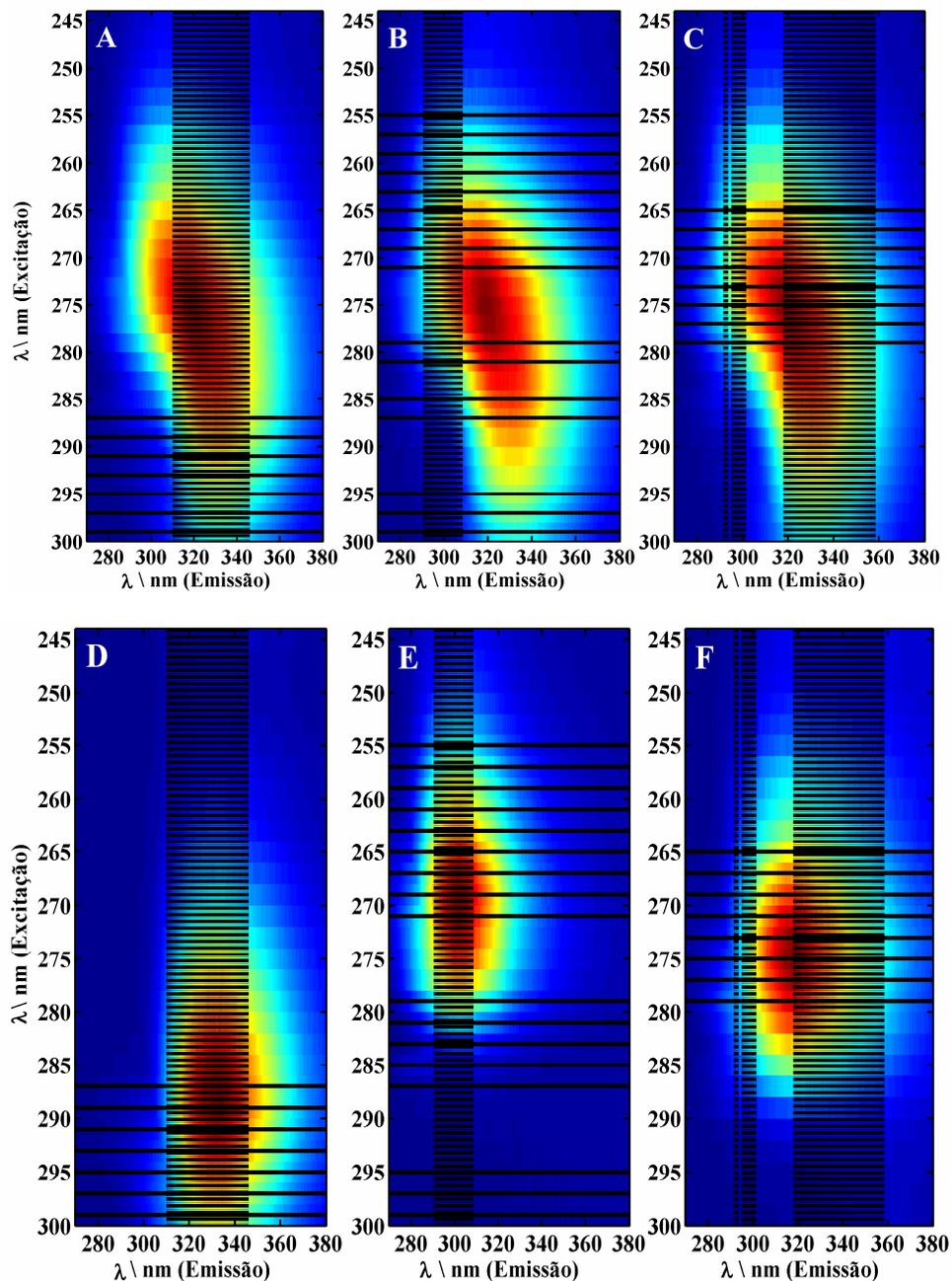


Figura 7.15. Variáveis selecionadas pelo nOPS, as linhas indicam as variáveis selecionadas em cada modo. HYDR(A,D); PHEN(B,E) e CATE (C,F), as imagens D-E mostram os espectros EEM puros de cada analito e as respectivas variáveis selecionadas.

Os modelos de calibração e as previsões para o conjunto externo realizadas pelo método *N*-PLS com as variáveis selecionadas podem ser observados na Figura 7.16 e os resultados numéricos das previsões são mostrados na Tabela 7.4.

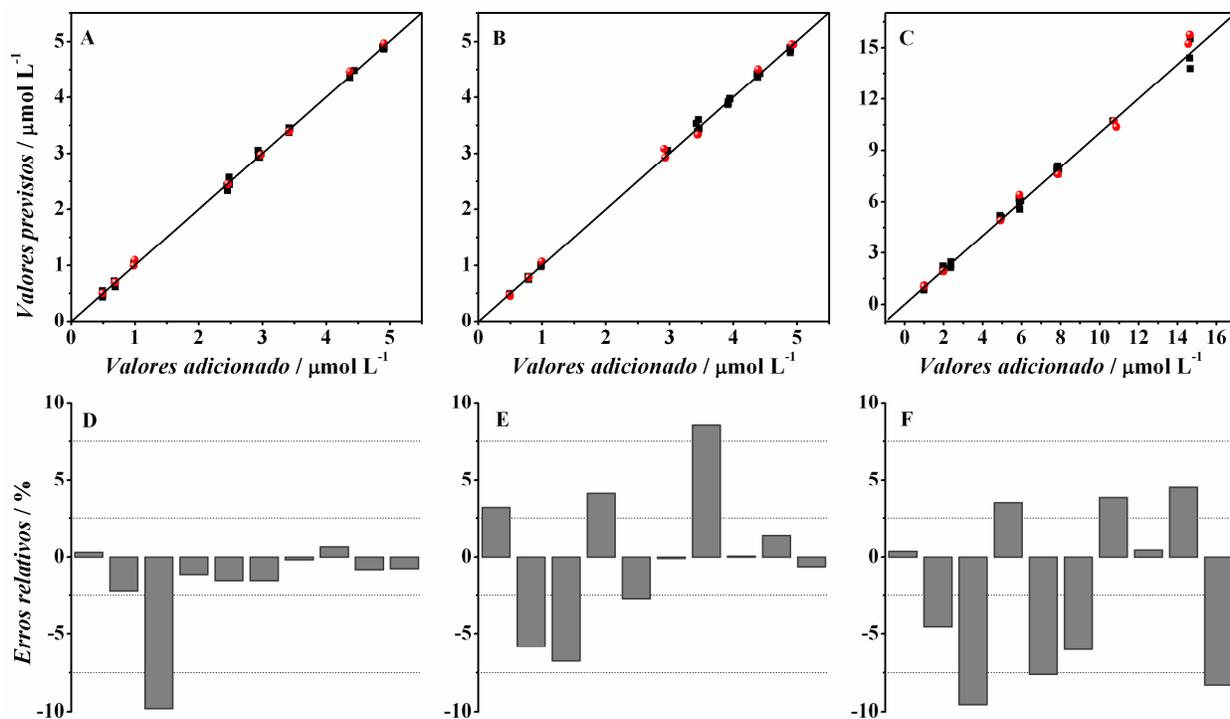


Figura 7.16. Valores adicionados vs. previstos para os modelos de calibração N-PLS/nOPS (■) e amostras previstas (●) e erros relativos das amostras previstas para HYDR (A,D), PHEN (B,E) e CATE (C,F).

Tabela 7.4. Valores adicionados (Adic.) vs. previstos (Prev.) e erros relativos (Erros relat.) para as amostras externas usando o modelo N-PLS com seleção nOPS.

	HYDR*			PHEN*			CATE*		
	Adic.	Prev.	Erros relat.	Adic.	Prev.	Erros relat.	Adic.	Prev.	Erros relat.
1	2,46	2,45	0,31	3,44	3,33	3,18	4,92	4,90	0,37
2	4,37	4,47	-2,25	2,91	3,08	-5,85	14,56	15,22	-4,54
3	1,00	1,09	-9,79	1,00	1,06	-6,77	1,00	1,09	-9,55
4	4,91	4,96	-1,18	0,49	0,47	4,09	7,85	7,58	3,50
5	0,68	0,69	-1,58	4,39	4,51	-2,73	14,63	15,74	-7,61
6	0,49	0,50	-1,58	4,94	4,95	-0,09	0,99	1,05	-6,03
7	2,97	2,98	-0,16	0,50	0,45	8,53	1,98	1,91	3,83
8	3,42	3,40	0,66	2,93	2,93	0,06	10,74	10,69	0,44
9	0,69	0,70	-0,88	0,79	0,78	1,38	10,85	10,36	4,49
10	0,98	0,99	-0,81	4,91	4,95	-0,70	5,89	6,38	-8,30

*Os valores Adic. e Prev. estão em mol L⁻¹ e os Erros relativ. estão em %.

Os resultados das previsões para os modelos N-PLS/nOPS apresentados na Figura 7.16 e Tabela 7.4 sugerem que tais modelos estão aptos a realizar previsões com alta precisão. Note que não ocorreu erro relativo maior que 10 %.

Finalmente, as previsões dos três analitos durante a degradação, para os três modelos construídos, usando o BDD são apresentadas na Figura 7.17. Observe que para HYDR e CATE as previsões de ambos os modelos são muito parecidas, entretanto, para PHEN o modelo PARAFAC para o início da degradação apresenta previsões diferentes dos outros dois modelos. As previsões com os modelos *N*-PLS/*n*OPS são apresentadas separadamente para ambos compostos na Figura 7.17D.

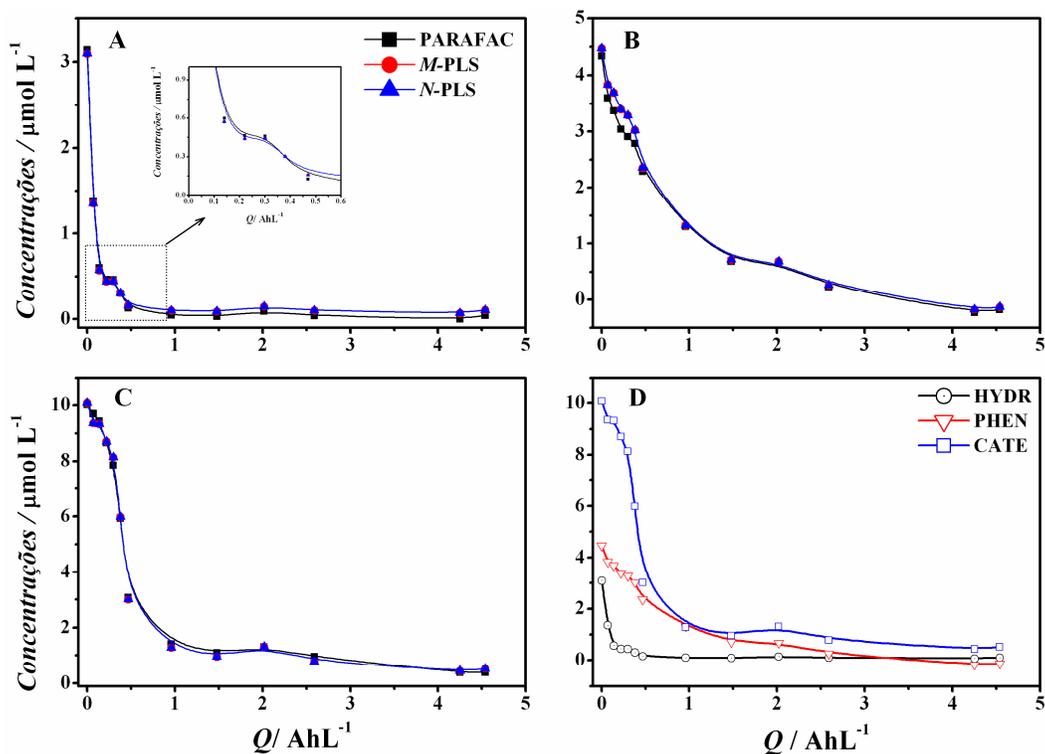


Figura 7.17. Estimativas das degradações da HYDR(A), PHEN(B) e CATE (C) para os modelos PARAFAC, *M*-PLS e *N*-PLS/*n*OPS e ambas as estimativas para o modelo *N*-PLS/*n*OPS (D).

Na Figura 7.17A nota-se que a HYDR é o composto que se degrada mais rapidamente, entretanto, entre 0,1 a 0,5 Ah L⁻¹ (gráfico interno) um leve ombro é observado, o que pode indicar o aparecimento de possíveis intermediários neste momento da degradação o que concorda com alguns autores [10,19,40], que têm encontrado o par hidroquinona/benzoquinona como intermediário na degradação de fenóis. No início da degradação, um leve ombro é observado nos perfis tanto para o PHEN quanto o CATE (Figura 7.17B,C), o que possivelmente indica que neste

momento, as reações ocorrem rapidamente e há geração de intermediários destes compostos [10,19]. Os perfis para HYDR e PHEN obtidos aqui apresentam comportamento similar aos mostrados por Canizares et al. (2005) em análises individuais [10]. O perfil do PHEN indica que este composto é o mais recalcitrante, visto que mesmo em concentrações bem menores que a do CATE, exigiu praticamente a mesma carga para se degradar. Traços de CATE são observados após $3,0 \text{ Ah L}^{-1}$, o que sugere que a aromaticidade ainda detectada após este momento é exclusivamente devido ao CATE. O ombro observado em $2,0 \text{ Ah L}^{-1}$ na Figura 7.17 e também notado na Figura 7.3, pode ser atribuído a algum erro sistemático na obtenção dos espectros ou contaminação da amostra durante a coleta, visto que tal ombro é notado para todos os compostos e também na remoção de aromaticidade.

Assim, é possível monitorar quantitativamente com precisão, usando espectros de fluorescência, a degradação de compostos fenólicos ou aromáticos que fluoresçam, desde que respeitadas as limitações exigidas pela supressão.

Conclusões

Neste trabalho, degradações de misturas de compostos fenólicos foram realizadas e monitoradas por TOC e fluorescência. De acordo com o monitoramento global o BDD mostrou ser eficiente para a degradação das misturas e viável economicamente. O uso da fluorescência mostrou ser uma ferramenta alternativa para este tipo de monitoramento devido a sua seletividade para a aromaticidade. O monitoramento individual realizado pelos espectros EEM e com uso do método PARAFAC indicaram que misturas complexas de compostos fenólicos inviabilizam a análise usando modelos estritamente lineares. Porém, a complexidade do sistema devido à ocorrência de supressão não é um problema de modelagem quimiométrica, mas sim químico. A diminuição da complexidade do sistema usando poucos compostos na mistura possibilitou a análise empregando métodos multimodos como PARAFAC, *M*-PLS e *N*-PLS. As metodologias para seleção de variáveis OPS e nOPS foram aplicadas nos métodos *M*-PLS e *N*-PLS respectivamente, mostrando serem capazes de diminuir significativamente a dimensão dos dados com melhorias, na maioria dos casos, nas previsões e com

ganhos na interpretação dos resultados. Os modelos apresentaram previsões similares, porém os modelos *N-PLS/nOPS* mostraram ser os mais precisos. Na análise dos perfis previstos para cada composto foi possível fazer atribuições para possíveis regiões de aparecimento de intermediários, recalcitrância dos compostos e composto responsável pela permanência da aromaticidade.

Devido às excelentes previsões obtidas pelos modelos construídos, a degradação eletroquímica de misturas de compostos fenólicos no BDD pode, portanto, ser estudada com auxílio da espectroscopia de fluorescência e com metodologias quimiométricas com ganhos significativos de informações para a interpretação dos resultados. Além disso, com uma única medida analítica é possível inferir sobre a degradação global e sobre a degradação individual de cada composto.

Referências

1. Ribordy, P.; Pulgarin, C.; Kiwi, J.; Peringer, P. Electrochemical versus photochemical pretreatment of industrial wastewaters. *Water Sci. Technol.* **1997**, 35, (4), 293-302.
2. Canizares, P.; Saez, C.; Lobato, J.; Rodrigo, M. A. Detoxification of synthetic industrial waste-waters using electrochemical oxidation with boron-doped diamond anodes. *J. Chem. Technol. Biotechnol.* **2006**, 81, (3), 352-358.
3. Kwan, C. Y.; Chu, W. A study of the reaction mechanisms of the degradation of 2,4-dichlorophenoxyacetic acid by oxalate-mediated photooxidation. *Water Res.* **2004**, 38, (19), 4213-4221.
4. Nava, J. L.; Nunez, F.; Gonzalez, I. Electrochemical incineration of p-cresol and o-cresol in the filter-press-type FM01-LC electrochemical cell using BDD electrodes in sulfate media at pH 0. *Electrochim. Acta* **2007**, 52, (9), 3229-3235.
5. Coleman, H. M.; Eggins, B. R.; Byrne, J. A.; Palmer, F. L.; King, E. Photocatalytic degradation of 17-beta-oestradiol on immobilised TiO₂. *Appl. Catal. B-Environ.* **2000**, 24, (1), L1-L5.
6. Fujishima, A.; Rao, T. N.; Tryk, D. A. TiO₂ photocatalysts and diamond electrodes. *Electrochim. Acta* **2000**, 45, (28), 4683-4690.
7. Michaud, P. A.; Panizza, M.; Ouattara, L.; Diaco, T.; Foti, G.; Comninellis, C. Electrochemical oxidation of water on synthetic boron-doped diamond thin film anodes. *J. Appl. Electrochem.* **2003**, 33, (2), 151-154.
8. Alfaro, M.; Ferro, S.; Martinez-Huitle, C. A.; Vong, Y. M. Boron doped diamond electrode for the wastewater treatment. *J. Braz. Chem. Soc.* **2006**, 17, (2), 227-236.

9. Li, X. Y.; Cui, Y. H.; Feng, Y. J.; Xie, Z. M.; Gu, J. D. Reaction pathways and mechanisms of the electrochemical degradation of phenol on different electrodes. *Water Res.* **2005**, 39, (10), 1972-1981.
10. Canizares, P.; Lobato, J.; Paz, R.; Rodrigo, M. A.; Saez, C. Electrochemical oxidation of phenolic wastes with boron-doped diamond anodes. *Water Res.* **2005**, 39, (12), 2687-2703.
11. Canizares, P.; Paz, R.; Lobato, J.; Saez, C.; Rodrigo, M. A. Electrochemical treatment of the effluent of a fine chemical manufacturing plant. *J. Hazard. Mater.* **2006**, 138, (1), 173-181.
12. Coteiro, R. D.; Teruel, F. S.; Ribeiro, J.; de Andrade, A. R. Effect of solvent on the preparation and characterization of DSA((R))-type anodes containing RuO₂-TiO₂-SnO₂. *J. Braz. Chem. Soc.* **2006**, 17, (4), 771-779.
13. Rajkumar, D.; Kim, J. G.; Palanivelu, K. Indirect electrochemical oxidation of phenol in the presence of chloride for wastewater treatment. *Chem. Eng. Technol.* **2005**, 28, (1), 98-105.
14. Chen, G. H. Electrochemical technologies in wastewater treatment. *Sep. Purif. Technol.* **2004**, 38, (1), 11-41.
15. Bonfatti, F.; Ferro, S.; Lavezzo, F.; Malacarne, M.; Lodi, G.; De Battisti, A. Electrochemical incineration of glucose as a model organic substrate I. Role of the electrode material. *J. Electrochem. Soc.* **1999**, 146, (6), 2175-2179.
16. Panizza, M.; Michaud, P. A.; Cerisola, G.; Comninellis, C. Anodic oxidation of 2-naphthol at boron-doped diamond electrodes. *J. Electroanal. Chem.* **2001**, 507, (1-2 SI Sp. Iss. SI), 206-214.
17. Boye, B.; Brillas, E.; Marselli, B.; Michaud, P. A.; Comninellis, C.; Farnia, G.; Sandona, G. Electrochemical incineration of chloromethylphenoxy herbicides in acid medium by anodic oxidation with boron-doped diamond electrode. **2006**, 51, (14), 2872-2880.
18. Marselli, B.; Garcia-Gomez, J.; Michaud, P. A.; Rodrigo, M. A.; Comninellis, C. Electrogeneration of hydroxyl radicals on boron-doped diamond electrodes. *J. Electrochem. Soc.* **2003**, 150, (3), D79-D83.
19. Zhi, J. F.; Wang, H. B.; Nakashima, T.; Rao, T. N.; Fujishima, A. Electrochemical incineration of organic pollutants on boron-doped diamond electrode. Evidence for direct electrochemical oxidation pathway. *J. Phys. Chem. B* **2003**, 107, (48), 13389-13395.
20. Farre, M.; Barcelo, D. Toxicity testing of wastewater and sewage sludge by biosensors, bioassays and chemical analysis. *Trac-Trends Anal. Chem.* **2003**, 22, (5), 299-310.

21. Suliman, F.; Al-Kindi, S. S.; Al-Kindy, S.; Al-Lawati, H. Analysis of phenols in water by high-performance liquid chromatography using coumarin-6-sulfonyl chloride as a fluorogenic precolumn label. *J. Chromatogr. A* **2006**, 1101, (1-2), 179-184.
22. Muna, G. W.; Tasheva, N.; Swain, G. M. Electro-oxidation and amperometric detection of chlorinated phenols at boron-doped diamond electrodes: A comparison of microcrystalline and nanocrystalline thin films. *Environ. Sci. Technol.* **2004**, 38, (13), 3674-3682.
23. Bro, R.; Workman, J. J.; Mobley, P. R.; Kowalski, B. R. Review of chemometrics applied to spectroscopy: 1985-95 .3. Multi-way analysis. *Appl. Spectrosc. Rev.* **1997**, 32, (3), 237-261.
24. Bro, R. PARAFAC. Tutorial and applications. *Chemometrics Intell. Lab. Syst.* **1997**, 38, (2), 149-171.
25. Ferreira, M. M. C.; Antunes, A. M.; Melgo, M. S.; Volpe, P. L. O. Chemometrics I: Multivariate calibration, a tutorial. *Quim. Nova* **1999**, 22, (5), 724-731.
26. Naes, T.; Isaksson, T.; Fearn, T.; Davies, T. *Multivariate calibration and classification*. Nir publications: Chinchester, 2002; p 344.
27. Smilde, A.; Bro, R.; Geladi, P. *Multi-way analysis with applications in the chemical science*. 1 ed.; John Wiley & Sons Inc: Chinchester, 2004; p 379.
28. Brereton, R. G. *Chemometrics: Data Analysis for the laboratory and chemical plant*. John Wiley & Sons Inc: Chinchester, 2003; p 407.
29. Martens, H.; Naes, T. *Multivariate Calibration*. John Wiley & Sons: New York, 1989; p 419.
30. Bosco, M. V.; Garrido, M.; Larrechi, M. S. Determination of phenol in the presence of its principal degradation products in water during a TiO₂-photocatalytic degradation process by three-dimensional excitation-emission matrix fluorescence and parallel factor analysis. *Anal. Chim. Acta* **2006**, 559, (2), 240-247.
31. Bro, R. Multiway calibration. Multilinear PLS. *J. Chemometr.* **1996**, 10, (1), 47-61.
32. Bahram, M.; Bro, R.; Stedmon, C.; Afkhami, A. Handling of Rayleigh and Raman scatter for PARAFAC modeling of fluorescence data using interpolation. *J. Chemometr.* **2006**, 20, (3-4), 99-105.
33. Kennard, R. W.; Stone, L. A. Computer aided design of experiments. *Technometrics* **1969**, 11, (1), 137-148.
34. Riu, J.; Bro, R. Jack-knife technique for outlier detection and estimation of standard errors in PARAFAC models. *Chemometrics Intell. Lab. Syst.* **2003**, 65, (1), 35-49.
35. Bro, R. *N-Way tutorial: curso on-line*. Royal Veterinary and Agricultural University: Frederiksberg, pp 1-72.

36. Andersson, C. A.; Bro, R. The N-way Toolbox for MATLAB. *Chemometrics Intell. Lab. Syst.* **2000**, 52, (1), 1-4.
37. Troster, I.; Schafer, L.; Fryda, M.; Matthee, T. Electrochemical advanced oxidation process using DiaChem (R) electrodes. *Water Sci. Technol.* **2004**, 49, (4), 207-212.
38. MMA, Resoluções. *Resoluções do Conama*. Ministério do Meio Ambiente: Brasília, pp 255-283.
39. Nasr, B.; Abdellatif, G.; Canizares, P.; Saez, C.; Lobato, J.; Rodrigo, M. A. Electrochemical oxidation of hydroquinone, resorcinol, and catechol on boron-doped diamond anodes. *Environ. Sci. Technol.* **2005**, 39, (18), 7234-7239.
40. Canizares, P.; Garcia-Gomez, J.; Saez, C.; Rodrigo, M. A. Electrochemical oxidation of several chlorophenols on diamond electrodes - Part I. Reaction mechanism. *J. Appl. Electrochem.* **2003**, 33, (10), 917-927.
41. Spitzer, M. Eletroxidação de cloranfenicol e outros poluentes orgânicos utilizando reatores eletroquímicos, constituídos de anodos tipo DSA ou de diamante dopado com boro (DDB). PhD, Universidade Estadual de Campinas, Campinas, 2005.
42. Rinnan, A.; Booksh, K. S.; Bro, R. First order Rayleigh scatter as a separate component in the decomposition of fluorescence landscapes. *Anal. Chim. Acta* **2005**, 537, (1-2), 349-358.
43. Rinnan, A.; Andersen, C. M. Handling of first-order Rayleigh scatter in PARAFAC modelling of fluorescence excitation-emission data. *Chemometrics Intell. Lab. Syst.* **2005**, 76, (1), 91-99.
44. Valeur, B. *Molecular Fluorescence: Principles and Applications.*; John Wiley & Sons: New York, 2001; Vol. 1, p 383.
45. Biczok, L.; Berces, T.; Linschitz, H. Quenching processes in hydrogen-bonded pairs: Interactions of excited fluorenone with alcohols and phenols. *J. Am. Chem. Soc.* **1997**, 119, (45), 11071-11077.
46. Avigal, I.; J, F.; M, O. Dissociation constants of excited phenols from fluorescence quenching data. *J. Chem. Phys.* **1969**, 50, (6), 2614-&.
47. Del Olmo, M.; Diez, C.; Molina, A.; deOrbe, I.; Vilchez, J. L. Resolution of phenol, o-cresol, m-cresol and p-cresol mixtures by excitation fluorescence using partial least-squares (PLS) multivariate calibration. *Anal. Chim. Acta* **1996**, 335, (1-2), 23-33.
48. Bosco, M.; Callao, M. P.; Larrechi, M. S. Resolution of phenol, and its dihydroxyderivative mixtures by excitation-emission fluorescence using MCR-ALS - Application to the quantitative monitoring of phenol photodegradation. *Talanta* **2007**, 72, (2), 800-807.

Conclusão Geral e Perspectivas

“Não é possível criar os acontecimentos. Pode-se navegar com eles e dar-lhes uma direção”

Otto von Bismarck

Embora esta tese tenha atingido o objetivo para a qual ela se propôs a defender, durante todo o seu processo de construção, modificações tanto no alicerce quanto nas paredes que a cercavam foram realizadas. O diamante dopado com boro foi uma das principais ferramentas desta construção e se mostrou extremamente eficiente para o que foi confiado a ele. Os diamantes crescidos para serem usados nesta tese apresentaram boas qualidades eletroquímicas e confirmou uma característica fundamental do ponto de vista eletroanalítico: a resistência à passivação. Portanto, este tipo material mostrou ser muito promissor nesta área, ratificando os diversos estudos, disponível na literatura, já realizados sobre este material.

Os métodos quimiométricos foram responsáveis pela engenharia e arquitetura para a construção desta tese. Um capítulo inteiro foi dedicado à apresentação de metodologias quimiométricas já conhecidas e de novas, propostas neste trabalho. Desde o primeiro estudo empregando QSPR, tais metodologias se mostraram fundamentais para a obtenção de informações extremamente importantes para se entender o sistema estudado. Além de ser possível concluir sobre a resistência do diamante à passivação, foi possível mostrar que há uma relação quantitativa entre a estrutura fenólica e o poder de bloqueio eletroquímica da superfície do eletrodo. Foi possível concluir ainda que este bloqueio é influenciado por características eletrônicas e estéricas das moléculas e que tal bloqueio pode ser facilmente relacionado com a estrutura fenólica em eletrodos que não apresentam grande resistência ao fenômeno de passivação, como é o caso do eletrodo de platina, usado juntamente ao BDD neste estudo. Os resultados do estudo QSPR abriram o

horizonte para a utilização do BDD para fins eletroanalíticos de maneira a empregá-lo em análise direta *i.e.*, sem modificações, de misturas de compostos fenólicos.

As duas metodologias analíticas apresentadas empregando o BDD como eletrodo de trabalho, SWV, e métodos quimiométricos mostraram o grande potencial desta parceria. Desde a otimização do sistema eletroanalítico, passando pela construção do modelo de calibração multivariada até os cálculos dos parâmetros analíticos dos modelos, as metodologias quimiométricas mostraram ser fundamentais. Um novo método de seleção de variáveis denominado OPS, usado para obter modelos confiáveis de calibração multivariada para previsão de novas amostras, mostrou ser uma excelente alternativa para este fim. Melhores previsões foram obtidas após a aplicação desta metodologia e uma comparação com o método GA indicou o grande potencial desta nova metodologia quimiométrica.

Bons resultados foram obtidos para ambas as metodologias de determinações simultâneas apresentadas usando o método PLS para calibração multivariada aliado ao método OPS para seleção de variáveis. Boas previsões para DA, AU e AA foram obtidas mesmo com a alta interferência existente entre estes três compostos no sinal voltamétrico. O mesmo pode ser dito para a determinação simultânea de GUA e CGUA diretamente em água de esgoto.

Porém, melhorias do eletrodo de BDD são ainda necessárias para se obter ganhos em sensibilidade e no limite de detecção. Modificações do ambiente de análise de modo a aumentar a separação dos potenciais de oxidação destes analitos com o objetivo de aumentar a seletividade é também desejável. Contudo, avanços foram alcançados na aplicação de ferramentas quimiométricas neste tipo de sistema para a determinação simultânea destes analitos.

A última aplicação mostrou como o BDD pode ser usado com eficiência e versatilidade na eletrólise de misturas de compostos fenólicos tendo como objetivo a degradação destes compostos. Nesta aplicação foi possível mostrar que o monitoramento global da degradação da mistura de compostos fenólicos pode ser realizado usando a espectroscopia de fluorescência. Foi possível observar que usando a fluorescência alternativamente à técnica TOC, é possível fazer um monitoramento seletivo para a aromaticidade. Desta maneira, a fluorescência

mostrou ser muito eficiente, pois o processo de degradação pode ser finalizado antes de todo carbono orgânico ter sido mineralizado, mas com minimização da toxicidade. De outro modo, usando metodologias quimiométricas de ordem superior tais como PARAFAC, *N*-PLS e *M*-PLS, foi possível realizar o monitoramento individual de cada composto na mistura durante a degradação. Nesta aplicação foi também empregada uma nova metodologia para seleção de variáveis em dados de ordem superior denominada nOPS. Esta nova metodologia apresentou melhorias nas previsões do modelo *N*-PLS aplicado ao conjunto completo diminuindo significativamente a dimensão dos dados, além disso, as variáveis selecionadas foram interpretativas. O modelo *N*-PLS/nOPS foi o modelo que apresentou melhor capacidade de previsão e um perfil cinético de degradação foi construído usando as previsões fornecidas por este modelo.

Embora muitos estudos tenham sido realizados, boas perspectivas podem ser apresentadas para estudos futuros.

Em relação ao BDD, modificações neste material e na confecção do eletrodo podem fornecer um material mais sensível à concentração de analitos; com alto poder de detecção em concentrações ainda menores que aquelas atingidas nesta tese e com maior eficiência para degradação de compostos persistentes. Para isso pode-se trabalhar com filmes menos espessos; estudar a influência da concentração de dopantes no diamante e sua influência sobre os parâmetros analíticos e sobre parâmetros de eficiência de degradação; adicionar gases inertes à atmosfera de hidrogênio como, por exemplo, nitrogênio, argônio, etc, para se obter diamantes nanocristalinos e verificar as propriedades eletroquímicas e eletroanalíticas destes materiais; otimizar o contato elétrico do diamante com algum metal, de maneira a diminuir a resistência ôhmica deste sistema; estudar mais detalhadamente a influência dos tratamentos eletroquímicos sobre as propriedades do BDD para fins eletroanalíticos e para a eletrólise.

Em relação às metodologias quimiométricas há muitos estudos a serem realizados. As metodologias propostas nesta tese ainda precisam ser detalhadamente investigadas e aplicadas para confirmar os seus potenciais. Os métodos OPS e nOPS funcionaram muito bem, mas merecem um estudo cuidadoso de suas características estatísticas e algébricas. É preciso verificar para cada uma

destas metodologias, a influência do número de componentes na construção do vetor de regressão usado para a seleção de variáveis. Um estudo detalhado sobre esta influência poderá tornar estes métodos ainda mais promissores.

O método apresentado para o cálculo do limite de detecção em calibração multivariada inversa também é outra metodologia que merece ser estudada e aplicada para verificação de sua exatidão na realização do cálculo para o qual ela foi proposta. Pelas poucas aplicações realizadas, pôde-se verificar o quão promissora é tal metodologia.

Apêndice

Tópicos de Álgebra Linear

*“O conhecimento pertence à humanidade
e é a tocha que ilumina o mundo”*

Louis Pasteur

Como muitas das operações e definições algébricas foram utilizadas nesta tese, é apropriado que os principais tópicos sejam descritos para orientar o leitor interessado. Assim, tópicos foram selecionados de alguns livros texto e serão descritos a seguir [1-6].

1. Sistema linear

Todo sistema linear de I equações e J incógnitas podem ser escrito como uma simples equação matricial $\mathbf{X}\mathbf{b} = \mathbf{y}$ em que

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1J} \\ x_{21} & x_{22} & \cdots & x_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ x_{I1} & x_{I2} & \cdots & x_{IJ} \end{pmatrix}, \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_J \end{pmatrix}, \text{ e } \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_I \end{pmatrix}$$

Portanto, toda equação matricial da forma $\mathbf{X}_{I \times J} \mathbf{b}_{J \times 1} = \mathbf{y}_{I \times 1}$ representa um sistema de I equações lineares em J incógnitas.

2. Posto de uma matriz

O posto de uma matriz é definido como número de linhas ou colunas linearmente independentes da matriz completa. Pode ser definido mais

especificamente como número de pivôs da matriz escalonada quando se emprega a eliminação de Gauss [1,2].

3. Adição de matrizes

Se \mathbf{X} e \mathbf{V} são matrizes $I \times J$, a soma de \mathbf{X} e \mathbf{V} é definida ser uma matriz $I \times J$, obtida por adição das entradas correspondentes, *i.e.*

$$[\mathbf{X} + \mathbf{V}]_{ij} = [\mathbf{X}]_{ij} + [\mathbf{V}]_{ij} \text{ para cada } i \text{ e } j.$$

4. Multiplicação de escalar

O produto de um escalar α , vezes uma matriz \mathbf{X} , indicado por $\alpha\mathbf{X}$, é definido ser a matriz obtida por multiplicação de cada entrada de \mathbf{X} por α , *i.e.*

$$[\alpha\mathbf{X}]_{ij} = \alpha[\mathbf{X}]_{ij} \text{ para cada } i \text{ e } j.$$

5. Multiplicação matricial

Matrizes \mathbf{X} e \mathbf{V} são ditas ser semelhantes para multiplicação na ordem \mathbf{XV} quando o número de colunas de \mathbf{X} é exatamente igual ao número de linhas de \mathbf{V} , *i.e.*, \mathbf{X} é $I \times P$ e \mathbf{V} é $P \times J$.

Para matrizes compatíveis $\mathbf{X}_{I \times P} = [x_{ip}]$ e $\mathbf{V}_{P \times J} = [v_{pj}]$, o produto matricial \mathbf{XV} é definido ser a matriz $I \times J$ cujas (i,j) entradas é o produto interno da i -ésima linha de \mathbf{X} com a j -ésima coluna em \mathbf{V} , *i.e.*,

$$[\mathbf{XV}]_{ij} = \mathbf{X}_{i*} \cdot \mathbf{V}_{*j} = x_{i1}v_{1j} + x_{i2}v_{2j} + \cdots + x_{ip}v_{pj} = \sum_{p=1}^P x_{ip}v_{pj}$$

No caso de \mathbf{X} e \mathbf{V} falharem na compatibilidade *i.e.*, \mathbf{X} ser $I \times P$ e \mathbf{V} ser $F \times J$ com $P \neq F$, então nenhum produto \mathbf{XV} é definido.

A multiplicação de matrizes é uma operação não comutativa *i.e.*, $\mathbf{XV} \neq \mathbf{VX}$, mesmo quando ambos os produtos existem e tem a mesma forma. Porém, as seguintes leis distributivas e associativa existem: $\mathbf{X}(\mathbf{V} + \mathbf{G}) = \mathbf{XV} + \mathbf{XG}$ (lei distributiva esquerda); $(\mathbf{D} + \mathbf{E})\mathbf{F} = \mathbf{DF} + \mathbf{EF}$ (lei distributiva direita); $\mathbf{X}(\mathbf{VG}) = (\mathbf{XV})\mathbf{G}$ (lei associativa).

6. Produtos Kronecker e Khatri-Rao

Para descrever modelos multimodos, o usual produto matricial não é suficiente. Dois outros produtos serão introduzidos aqui: o de Kronecker (\otimes) e o de Khatri-Rao (\odot).

Considere duas matrizes $\mathbf{A}(I \times J)$ e $\mathbf{B}(K \times M)$. O produto Kronecker destas matrizes é definido como

$$\mathbf{A}_{I \times J} \otimes \mathbf{B}_{K \times M} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1J}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{I1}\mathbf{B} & \cdots & a_{IJ}\mathbf{B} \end{bmatrix}_{IK \times JM}$$

Note que o produto Kronecker é também definido para duas matrizes as quais o produto matricial usual não existe (se $J \neq K$)

O produto Khatri-Rao é definido como segue

$$\mathbf{A} = [\mathbf{a}_1 \dots \mathbf{a}_J] \text{ e } \mathbf{C} = [\mathbf{c}_1 \dots \mathbf{c}_J]$$

$$\mathbf{A} \odot \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{c}_1 \dots \mathbf{a}_J \otimes \mathbf{c}_J]$$

Note que o produto Khatri-Rao exige que as matrizes a serem processadas tenham o mesmo número de colunas.

7. Matriz transposta

A transposta de $\mathbf{X}_{I \times J}$ é definida como a matriz $J \times I$ \mathbf{X}^t obtida pela troca de linhas e colunas em \mathbf{X} . Mais precisamente, se $\mathbf{X} = [a_{ij}]$, então $[\mathbf{X}^t]_{ij} = a_{ji}$. Por exemplo:

$$\begin{pmatrix} 3 & 5 \\ 7 & 2 \\ 8 & 1 \end{pmatrix}^t = \begin{pmatrix} 3 & 7 & 8 \\ 5 & 2 & 1 \end{pmatrix}$$

É, portanto evidente que para todas as matrizes $(\mathbf{X}^t)^t = \mathbf{X}$.

A lei da ordem reversa da transposição para matrizes compatíveis \mathbf{X} e \mathbf{V} é: $(\mathbf{XV})^t = \mathbf{V}^t \mathbf{X}^t$.

A transposição não muda o posto de uma matriz *i.e.*, $\text{posto}(\mathbf{X}) = \text{posto}(\mathbf{X}^t)$.

8. Combinações lineares

Para escalares α_j e matrizes \mathbf{X}_j , a expressão

$$\alpha_1 \mathbf{X}_1 + \alpha_2 \mathbf{X}_2 + \cdots + \alpha_J \mathbf{X}_J = \sum_{j=1}^J \alpha_j \mathbf{X}_j$$

É chamada uma combinação linear de \mathbf{X}_j 's.

Quando for referido que há uma combinação linear de uma matriz, significa que há uma combinação linear das colunas da matriz. Por exemplo, se \mathbf{X} é uma matriz $I \times J$, então uma combinação linear de \mathbf{X} é

$$\mathbf{c} = \mathbf{X}\mathbf{b} = b_1 \mathbf{x}_1 + b_2 \mathbf{x}_2 + \cdots + b_J \mathbf{x}_J$$

Em que $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_J$ representa as J colunas de \mathbf{X} e $\mathbf{b} = (b_1, b_2, \dots, b_J)^t$ é um vetor $J \times 1$ de pesos. O resultado é um vetor \mathbf{c} $I \times 1$.

9. Covariância

Dado dois vetores centrados na média \mathbf{x}_j e \mathbf{x}_k , a covariância da amostra é expressa como

$$\text{cov}(\mathbf{x}_j, \mathbf{x}_k) = \mathbf{x}_j^t \mathbf{x}_k$$

Em que o escalar $I - 1$ é implicado para uma amostra.

10. Matriz identidade

A matriz $J \times J$ com 1's na diagonal principal e 0's em todas as outras posições, *i.e.*

$$\mathbf{I}_J = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

É chamada de matriz identidade de ordem J . Para toda matriz \mathbf{X} $I \times J$, $\mathbf{X}\mathbf{I}_J = \mathbf{X}$ e $\mathbf{I}_I \mathbf{X} = \mathbf{X}$.

O subscrito em \mathbf{I}_J é negligenciado quando a dimensão é óbvio a partir do contexto.

11. Inversão matricial

Para uma dada matriz quadrada $\mathbf{X}_{J \times J}$, a matriz $\mathbf{V}_{J \times J}$ que satisfaz as condições $\mathbf{XV} = \mathbf{I}_J$ e $\mathbf{VX} = \mathbf{I}_J$ é chamada de inversa de \mathbf{X} e é denominada por $\mathbf{V} = \mathbf{X}^{-1}$. Nem todas as matrizes quadradas são invertíveis – a matriz zero é um exemplo trivial, mas há também muitas matrizes diferentes da matriz zero que não são invertíveis. Uma matriz invertível é definida como não-singular e uma matriz quadrada que não possui uma inversa é chamada de matriz singular.

Embora, computacionalmente, a inversão de matrizes não deva ser usada (e não é) para resolver o sistema $\mathbf{Xb} = \mathbf{y}$, há algumas poucas aplicações onde o conhecimento explícito de \mathbf{X}^{-1} é desejável. Para isto, os dois mais eficientes algoritmos (não descritos aqui) que podem ser usados se baseiam na eliminação de Gauss-Jordan e na decomposição LU.

O uso da eliminação Gauss-Jordan para inverter uma matriz consiste do seguinte princípio: $[\mathbf{X} | \mathbf{I}] \xrightarrow{\text{Gauss-Jordan}} [\mathbf{I} | \mathbf{X}^{-1}]$. Este cálculo exige J^3 multiplicações/divisões e $J^3 - 2J^2 + J$ adições/subtrações, portanto o custo computacional de uma inversão é muito alto e há maneiras para resolver o sistema $\mathbf{Xb} = \mathbf{y}$ com custo extremamente inferiores ao procedimento de inversão explícita.

12. Sistema linear mal condicionado

Uma matriz não-singular \mathbf{X} é dita ser mal-condicionada se uma pequena mudança relativa em \mathbf{X} causa uma grande mudança relativa em \mathbf{X}^{-1} . De outro modo, o sistema é dito ser bem condicionado. O grau de mal-condicionamento é medido por um número de condição $\kappa = \|\mathbf{X}\| \cdot \|\mathbf{X}^{-1}\|$, em que $\|\cdot\|$ é a norma da matriz.

Em outras palavras, quando \mathbf{X} é bem condicionada (*i.e.* quando κ é pequeno), isto garante que pequenas incertezas relativas em \mathbf{y} não afetam significativamente a solução, mas quando \mathbf{X} é mal condicionada (*i.e.*, quando κ é grande), uma incerteza relativamente pequena em \mathbf{y} pode resultar em uma incerteza relativamente grande em \mathbf{b} .

13. Independência linear

Conjuntos linearmente independentes são aqueles que não contêm relação de dependência e conjuntos linearmente dependentes são aqueles em que no mínimo um vetor é a combinação dos outros.

Deixe \mathbf{X} ser uma matriz $I \times J$. Se $\text{posto}(\mathbf{X}) = J$, as colunas de \mathbf{X} formam um conjunto linearmente independente e se o $\text{posto}(\mathbf{X}^t) = I$, as linhas de \mathbf{X} formam um conjunto linearmente independente.

Quando \mathbf{X} é quadrada, cada uma das seguintes afirmações é equivalente a dizer que \mathbf{X} é não-singular.

- As colunas de \mathbf{X} formam um conjunto linearmente independente.
- As linhas de \mathbf{X} formam um conjunto linearmente independente.

14. Sistema linear determinado

Cada uma das seguintes afirmações é equivalente a dizer que o sistema linear $[\mathbf{X}|\mathbf{y}]$ é determinado.

1. Em uma redução de linhas de $[\mathbf{X}|\mathbf{y}]$ (eliminação de Gauss), uma linha da seguinte forma nunca aparece: $(0 \ 0 \ \dots \ 0|\alpha)$, em que $\alpha \neq 0$.
2. \mathbf{y} é uma coluna que não é uma base de $[\mathbf{X}|\mathbf{y}]$.
3. $\text{posto}[\mathbf{X}|\mathbf{y}] = \text{posto}(\mathbf{X})$
4. \mathbf{y} é uma combinação das colunas linearmente independentes em \mathbf{X} .

15. Equações normais

Para um sistema $(I \times J) \mathbf{X}\mathbf{b} = \mathbf{y}$, o sistema associado de equações normais é definido ser o sistema $(J \times J) \mathbf{X}^t\mathbf{X}\mathbf{b} = \mathbf{X}^t\mathbf{y}$.

$\mathbf{X}^t\mathbf{X}\mathbf{b} = \mathbf{X}^t\mathbf{y}$ é sempre determinado, mesmo quando $\mathbf{X}\mathbf{b} = \mathbf{y}$ não é determinado.

Quando $\mathbf{X}\mathbf{b} = \mathbf{y}$ é determinado, sua solução concorda com aquela de $\mathbf{X}^t\mathbf{X}\mathbf{b} = \mathbf{X}^t\mathbf{y}$. As equações normais fornecem soluções de quadrados mínimos para $\mathbf{X}\mathbf{b} = \mathbf{y}$ quando $\mathbf{X}\mathbf{b} = \mathbf{y}$ é indeterminado.

$\mathbf{X}^t\mathbf{X}\mathbf{b} = \mathbf{X}^t\mathbf{y}$ tem uma solução única se e somente se o $\text{posto}(\mathbf{X}) = J$. Neste caso a solução única é $\mathbf{b} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y}$.

Quando $\mathbf{X}\mathbf{b} = \mathbf{y}$ é determinado e tem uma solução única, então o mesmo é verdadeiro para $\mathbf{X}^t\mathbf{X}\mathbf{b} = \mathbf{X}^t\mathbf{y}$, e a solução única para ambos os sistemas é dados por $\mathbf{b} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y}$.

16. Problema geral de quadrados mínimos

Muitos problemas, quando modelados, levam a sistemas lineares $\mathbf{X}\mathbf{b} = \mathbf{y}$, que são indeterminados (*i.e.*, não possuem solução), apesar dos problemas que os originaram necessitarem de uma solução. A não existência de solução vem com frequência devido a erros experimentais no vetor \mathbf{y} ou devido ao mal-condicionamento na matriz \mathbf{X} . Uma forma de resolver o sistema indeterminado é através das projeções por quadrados mínimos, ou seja,

Para $\mathbf{X} \in \mathfrak{R}^{J \times J}$ e $\mathbf{y} \in \mathfrak{R}^J$, deixe $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}(\mathbf{b}) = \mathbf{X}\mathbf{b} - \mathbf{y}$. O problema geral de quadrados mínimos é encontrar um vetor \mathbf{b} que minimiza a quantidade:

$\boldsymbol{\varepsilon}^t\boldsymbol{\varepsilon} = (\mathbf{X}\mathbf{b} - \mathbf{y})^t (\mathbf{X}\mathbf{b} - \mathbf{y}) = \|\mathbf{X}\mathbf{b} - \mathbf{y}\|^2$, que nada mais é que minimizar a soma quadrática dos erros.

Qualquer vetor que fornece um valor mínimo para esta expressão é chamado de solução de quadrados mínimos.

- O conjunto de todas as soluções de quadrados mínimos é precisamente o conjunto de soluções das equações normais.

- Há uma única solução de quadrados mínimos se e somente se o posto(\mathbf{X}) = J . Neste caso a única solução é $\mathbf{b} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y}$.

- Se $\mathbf{X}\mathbf{b} = \mathbf{y}$ é determinado, então a solução para $\mathbf{X}\mathbf{b} = \mathbf{y}$ é a mesma fornecida pela solução de quadrados mínimos.

17. Norma vetorial Euclidiana

Para um vetor $\mathbf{x}_{J \times 1}$, a norma euclidiana de \mathbf{x} é definida por

$$\|\mathbf{x}\| = \left(\sum_{i=1}^J \mathbf{x}_i^2 \right)^{1/2} = \sqrt{\mathbf{x}^t\mathbf{x}}, \text{ em que } \mathbf{x} \in \mathfrak{R}^{J \times 1}$$

18. Ortogonalidade

Em um produto interno do espaço \mathcal{V} , dois vetores $\mathbf{x}, \mathbf{y} \in \mathcal{V}$ são ortogonais (um em relação ao outro) para \mathfrak{R}^J , quando $\mathbf{x}^t \mathbf{y} = \langle \mathbf{x} | \mathbf{y} \rangle = 0$, e isto é indicado escrevendo $\mathbf{x} \perp \mathbf{y}$.

19. Ortonormalidade

Considere um conjunto de vetores $\beta = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_J\}$. Este conjunto é chamado de ortonormal quando $\|\mathbf{u}_i\| = 1$ para cada i e $\mathbf{u}_i \perp \mathbf{u}_j$ para todo $i \neq j$. Em outras palavras,

$$\langle \mathbf{u}_i | \mathbf{u}_j \rangle = \begin{cases} 1 & \text{quando } i = j, \\ 0 & \text{quando } i \neq j. \end{cases}$$

- todo conjunto ortonormal é linearmente independente.
- todo conjunto ortonormal de J vetores de um espaço J -dimensional \mathcal{V} é uma base ortonormal para \mathcal{V} .

20. Matrizes ortogonais

Uma matriz ortogonal é definida como uma matriz real $\mathbf{U}_{J \times J}$ cujas colunas (ou linhas) constituem uma base ortonormal para \mathfrak{R}^J .

Matrizes ortogonais têm algumas características excelentes, uma delas é o fato que elas são fáceis de inverter. Note que as colunas de $\mathbf{U}_{n \times n} = (\mathbf{u}_1 | \mathbf{u}_2 | \dots | \mathbf{u}_n)$ são um conjunto ortonormal se e somente se,

$$[\mathbf{U}^t \mathbf{U}]_{ij} = \mathbf{u}_i^t \mathbf{u}_j = \begin{cases} 1 & \text{quando } i = j, \\ 0 & \text{quando } i \neq j. \end{cases} \Leftrightarrow \mathbf{U}^t \mathbf{U} = \mathbf{I} \Leftrightarrow \mathbf{U}^{-1} = \mathbf{U}^t.$$

Note que devido $\mathbf{U}^t \mathbf{U} = \mathbf{I} \Leftrightarrow \mathbf{U} \mathbf{U}^t = \mathbf{I}$, as colunas de \mathbf{U} são ortonormais se e somente se as linhas de \mathbf{U} sejam ortonormais.

Outra característica destas matrizes é que a multiplicação por um vetor não muda o comprimento do vetor. Apenas a direção pode ser alterada, assim:

$$\|\mathbf{U} \mathbf{x}\|^2 = \mathbf{x}^t \mathbf{U}^t \mathbf{U} \mathbf{x} = \mathbf{x}^t \mathbf{x} = \|\mathbf{x}\|^2 \text{ para todo } \mathbf{x} \in \mathfrak{R}^{J \times 1}.$$

21. Decomposição dos Valores Singulares[‡]

Para todo $\mathbf{X} \in \mathfrak{R}^{I \times J}$ de posto h , existem matrizes ortogonais $\mathbf{A}_{I \times I}$, $\mathbf{P}_{J \times J}$ e uma matriz diagonal $\mathbf{D}_{h \times h} = \text{diag}(\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_h)$ tais que

$$\mathbf{X} = \mathbf{A} \begin{pmatrix} \mathbf{D} & 0 \\ 0 & 0 \end{pmatrix}_{I \times J} \mathbf{P}^t \text{ com } \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_h > 0$$

Os σ_i 's são chamados de *valores singulares* diferentes de zero de \mathbf{X} . Quando $h < (p = \min\{I, J\})$, \mathbf{X} é dito ter $p - h$ adicionais valores zeros. Esta fatoração é chamada de decomposição dos valores singulares (SVD) de \mathbf{X} e as colunas em \mathbf{A} e \mathbf{P} são chamadas *vetores singulares* do lado esquerdo e direito de \mathbf{X} , respectivamente.

- Os quadrados dos valores singulares (σ_i^2) são os autovalores de $\mathbf{X}^t\mathbf{X}$ (e $\mathbf{X}\mathbf{X}^t$).
- Os vetores \mathbf{a}_i em \mathbf{A} , são os autovetores de $\mathbf{X}\mathbf{X}^t$ associados a σ_i^2 .
- Os vetores \mathbf{p}_i em \mathbf{P} , são os autovetores de $\mathbf{X}^t\mathbf{X}$ associados a σ_i^2 .

22. Moore-Penrose Pseudoinversa[§]

Em termos da decomposição SVD, pode se escrever da seguinte maneira a Moore-Penrose Pseudoinversa de \mathbf{X} :

Sabendo que $\mathbf{X}_{I \times J} = \mathbf{A} \begin{pmatrix} \mathbf{D}_{h \times h} & 0 \\ 0 & 0 \end{pmatrix}_{I \times J} \mathbf{P}^t$, a pseudo inversa de \mathbf{X} é:

$$\mathbf{X}_{I \times J}^+ = \mathbf{P} \begin{pmatrix} \mathbf{D}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \mathbf{A}^t.$$

- Quando $\mathbf{X}\mathbf{b} = \mathbf{y}$ é determinado, $\mathbf{b} = \mathbf{X}^+\mathbf{y}$ é a solução de norma euclidiana mínima.

[‡] A SVD foi descoberta e redescoberta várias vezes. O crédito para os primeiros desenvolvedores vai para Eugenio Beltrami (1835 – 1899) em 1873, M. E. Camille Jordan (1838 – 1922) em 1875, James J. Sysvester (1814 – 1897) em 1889, L. Autonne em 1913 e C. Eckart e G. Young em 1936.

[§] Em honra a Eliakim H. Moore (1862–1932) e Roger Penrose (um famoso e contemporâneo físico matemático inglês). Cada um formulou um conceito da inversão matricial generalizada — O trabalho de Moore foi publicado em 1922, e o trabalho de Penrose apareceu em 1955. E. H. Moore é considerado por muitos, como o primeiro grande matemático americano.

- Quando $\mathbf{Xb} = \mathbf{y}$ é indeterminado, $\mathbf{b} = \mathbf{X}^+\mathbf{y}$ é a solução de quadrados mínimos de norma euclidiana mínima.

Referências

1. Meyer, C. D. *Matrix analysis and linear algebra*. SIAM Publications: New York, 2000; p 718.
2. Strang, G. *Linear algebra and its applications*. 3 ed.; Thomson Learning, Inc.: London, 1988; p 505.
3. Smilde, A.; Bro, R.; Geladi, P. *Multi-way analysis with applications in the chemical science*. John Wiley & Sons Inc: Chichester, 2004; p 379.
4. Naes, T.; Isaksson, T.; Fearn, T.; Davies, T. *Multivariate calibration and classification*. Nir publications: Chichester, 2002; p 344.
5. Mason, R. L.; Gunst, R. F.; Hess, J. L. *Statistical design and analysis of experiments*. 2 ed.; John Wiley & Sons Inc: Hoboken, 2003; p 730.
6. Winefordner, J. D. *Chemometrics: From basics to wavelet transform*. John Wiley & Sons Inc: New Jersey, 2004; Vol. 164, p 292.

